Genetic dissection of *Anopheles gambiae* epithelial responses to gut commensal bacteria

Stavros Stathopoulos

Division of Cell and Molecular Biology

Department of Life Sciences

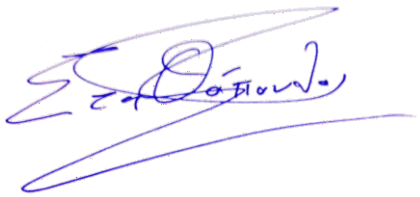
Imperial College London

A thesis submitted in accordance with the requirements of Imperial College London for the degree of Doctor of Philosophy, November 2013

Declaration of originality

I hereby certify that this thesis and the research upon which it is based is the product of my own work. Any ideas, quotations as well as published or unpublished work from other authors is clearly attributed and properly referenced. All scientific collaborations are clearly listed, attributed and acknowledged.

Heraklion, 12/11/2013



Stavros Stathopoulos

Copyright declaration

Acknowledgements

After more than four years since commencing this PhD project, it is time to conclude this effort with the submission of the current thesis, hopefully wiser. I would like to thank my supervisor, Professor George Christophides, for help and advice during the course of this PhD project and a productive collaboration that drove this project forward, towards the right direction. I would also like to thank all members of the Kafatos/Christophides lab along with visiting students and researchers during the period I was member of the group that rendered work in the lab a pleasant experience.

As a vital part of working in such group is maintenance of the mosquito colony, on which the success of each individual project, and thus the progress of the whole group, relies upon. I would especially like to thank Kasia Sala for the tedious work of mosquito maintenance and all members of the group that also contributed to mosquito maintenance during the course of my project, especially during Kasia's maternity leave.

During this period I had the pleasure to work in the vibrant Imperial College South Kensington campus and the modern SAF building that allowed a fruitful co-habitation with members of other research groups that was also vital to a pleasant work experience during this period. The sometimes tedious mosquito work was made much more pleasant with the countless coffee breaks with Aristotelis Antonopoulos but also with George.

This PhD project could not have been implemented without the generous financial support of the Wellcome Trust that not only contributed to my comfortable stay in London but also allowed the successful development of this project.

Last but not least, a unique experience during the course of this PhD project included the, brief but fruitful, encounters with Fotis Kafatos, whose immense contribution to the vector biology field underlies the development of the current project.

List of scientific collaborators and contributors

- **Guido Favia**, University of Camerino, Italy

Shared the *Asaia SF2.1 (GFP)* stock

- **Dominique Ferrandon**, IBMC, France

Shared the *Serratia marcescens Db11-GFP* stock

- **Tibebu Habtewold**, Imperial College London, UK

Mosquito colony rearing and maintenance

- **Amanda Jackson**, Imperial College London, UK

DNA microarrays: Sample preparation, hybridization, scanning, data analysis

- **Mara Lawniczak**, Imperial College London, UK

SNP genotyping arrays: Shared reagents, statistical analysis of derived data

- **Marc Muskavitch**, Broad Institute, USA, Harvard School of Public Health, USA & Boston College, USA

SNP genotyping arrays: Shared reagents, sample preparation, array hybridization & scanning, data analysis

- **Daniel Neafsey**, Broad Institute, USA

SNP genotyping arrays: Shared reagents, sample preparation, array hybridization & scanning, data analysis

- **Katarzyna Sala**, Imperial College London, UK

Mosquito colony rearing and maintenance

- **Robert Sinden**, Imperial College London, UK

Allowed use of *Zeiss Axiophot* fluourescence microscope

- **Dina Vlachou**, Imperial College London, UK

DNA microarrays: Statistical analysis of derived data

Funding

Wellcome Trust PhD studentship (086723/Z/08/A) awarded to SS as part of the Wellcome Trust 4-year PhD training programme “Molecular and Cellular Basis of Infection”.

List of abbreviations

400k	400,000
aa	amino acid
ACT	artemisinin combination therapies
Ae. aegypti	*Aedes aegypti*
AMPK	adenosine monophosphate protein kinase
An.	*Anopheles*
An. gambiae s.s.	*Anopheles gambiae sensu stricto*
An. stephensi	*Anopheles stephensi*
ANCOVA	analysis of co-variance
AT7IP	activating transcription factor 7-interacting protein
ATP	adenosine triphosphate
BC	before Christ
BMW	Bayerische Motoren Werke AG
bp	base pair
C. elegans	*Caenorhabditis elegans*
CAFE	capillary feeder
cDNA	complementary DNA
C_T	threshold cycle
CTL	C-type lectin
Cyto	cytoplasmic
D. melanogaster	*Drosophila melanogaster*
DAP	diaminopimelic acid
DDT	dichlorodiphenyltrichloraethane
dH_2O	distilled water
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
Dscam	down syndrome cell adhesion molecule
dsDNA	double stranded DNA
dsLacZ	double stranded *LacZ*
dsRNA	double stranded RNA
DUOX	Dual oxidase
E. coli	*Escherichia coli*
e.g.	*exempli gratia* (for example)
EST	Expressed sequence tag
FBN	fibrinogen-related immunolectin
FN3D	type III fibronectin domain
G6PD	glucose-6-phosphate dehydrogenase
GABA	gamma-aminobutyric acid
gDNA	genomic DNA
GFP	Green Fluourescent Protein
GNBP	Gram-negative binding protein
GO	gene ontology
GPCR	G protein-coupled receptor
GPS	gentamicin-penicillin-streptomycin
GPSC	gentamicin-penicillin-streptomycin-carbenicillin
Gr	gustatory receptor
gr	gram
GWAS	genome-wide association study
i.e.	*id est* (that is)
Ig	immunoglobulin
IMD	immune deficiency
IR	insecticide resistance
IRS	indoor residual spraying
ITN	insecticide treated nets
kg	kilo gram
l	litre
LB	Luria-Bertani
log	logarithmic
LPS	lipopolysaccharide
LRIM	leucine-rich immune protein
LRR	leucine-rich repeat
MAF	minor allele frequency
MAF Diff	minor allele frequency difference
MAPK	mitogen-activated protein kinase
mg	milli gram
ml	milli litre
MM	mismatch
mM	milli molar
NCAM	neural cell adhesion molecule
NCBI	National Center for Biotechnology Information
ng	nano gram
nm	nano metre
NOS	nitric oxide synthase
NPF	neuropeptide F
NS	non-significant
^oC	Celsius
OD600	optical density at 600 nm
ONNV	*O'nyong-nyong* virus
P. aeruginosa	*Pseudomonas aeruginosa*
P. berghei	*Plasmodium berghei*
P. cynomology	*Plasmodium cynomology*
P. entomophila	*Pseudomonas entomophila*
P. falciparum	*Plasmodium falciparum*
P. gallinaceum	*Plasmodium gallinaceum*
P. knowlesi	*Plasmodium knowlesi*
P. malariae	*Plasmodium malariae*
P. ovale	*Plasmodium ovale*
P. vivax	*Plasmodium vivax*
PBS	phosphate buffered saline
PCA	principal component analysis
PCR	polymerase chain reaction
PM	perfect match
PPO	prophenoloxidase
PQQ-ADH	pyrroloquine quinone-dependent alcohol dehydrogenase
qPCR	quantitative real-time PCR
qRT-PCR	quantitative reverse transcription real-time PCR
R^2	R-squared
Repl	replicate
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNAi	RNA interference
rpm	rotations per minute
rRNA	ribosomal RNA
RT-PCR	reverse transcription PCR
S. marcescens	*Serratia marcescens*
SEM	standard error of means
SNP	single nucleotide polymorphism
SOC	super optimal catabolite repressor
SSRI	selective serotonin reuptake inhibitor
TEP	thioester-containing protein
TM	transmembrane
US	United States
v/v	volume to volume
WHO	World Health Organization
WHO TDR	World Health Organization Tropical Diseases training group
µg	micro gram
µl	micro litre
µM	micro molar

List of research paper publications based on work presented in this thesis

Genetic Dissection of *Anopheles gambiae* Gut Epithelial Responses to *Serratia marcescens.* Stavros Stathopoulos, Daniel E. Neafsey, Mara K. N. Lawniczak, Marc A. T. Muskavitch and George K. Christophides. *PLoS Pathogens* 10(3): e1003897. doi:10.1371/journal.ppat.1003897.

The complete manuscript can be found in the appendix and is also available in the following web address: http://www.plospathogens.org/article/info:doi/10.1371/journal.ppat.1003897

Abstract

Commensal bacteria inhabiting the mosquito gut have been shown to profoundly influence the outcome of infection with malaria *Plasmodium* parasites. Genetic variation within mosquito populations has also been associated with the outcome of *Plasmodium* infections, especially with regard to immune factors, many of them targeting both bacteria and *Plasmodium* parasites. Although bacteria are considered a major evolutionary force that may affect survival, no link has yet to emerge between mosquito genetic variation and the outcome of bacterial infections. To test this possibility, a model of oral infections was established in *Anopheles gambiae*, the major malaria vector, with two common members of the mosquito gut microbiota, *Asaia* and *Serratia marcescens*. Using genome-wide SNP genotyping and expression analyses, mosquito genes associated with the outcome of *Serratia* infection and differentially expressed following *Asaia* or *Serratia* infection were identified. Three genes encoding type III fibronectin domains (*FN3D1-3*), associated with the outcome of *Serratia* infection, were shown to limit *Serratia* abundance and shift the mosquito gut bacterial population structure by limiting mainly *Enterobacteriaceae*, suggesting that mosquito responses shape the gut bacterial community with much higher specificity than previously thought. Furthermore, silencing of the gene encoding the gustatory receptor Gr9, also associated with the *Serratia* infection outcome, was shown to radically increase *Serratia* levels following oral infection, in a behavioural response than mostly relied on changes in neuropeptide F expression. Both the expression and SNP genotyping analysis indicated the existence of an epithelial and a behavioural mode of immunity following oral bacterial infection and revealed the complexity and specificity of this intricate biological system that is expected to shape malaria transmission dynamics.

Table of contents

List of Figures

List of Tables

Chapter 1

Introduction

Introduction

1.1 Mosquitoes and public health

This study focuses on the elucidation of mosquito immune responses against bacteria in the mosquito gut. The identification of mosquito epithelial responses against bacteria presents a dual significance: a biological significance of harnessing the mosquito as a model system to identify principles in immunity likely to be relevant in other organisms, including humans, but also an immense public health significance as mosquitoes are vectors for a number of disease-causing agents, including dengue virus (Racloz et al., 2012), West Nile virus (Mackenzie et al., 2004), filarial worms (Erickson et al., 2009), and, most importantly, malaria *Plasmodium* parasites (Sinden, 2002). The mosquitoes used in this study belong to the *Anopheles gambiae* species, the main malaria vector in sub-Saharan Africa.

1.2 The *Plasmodium* life cycle and malaria transmission

Malaria is caused by protozoan *Plasmodium* parasites. Five *Plasmodium* species can infect humans: *P. falciparum,* which is responsible for most deaths in Africa (Miller et al., 2013), but also *P. vivax* (Guerra et al., 2006), *P. ovale*, *P. malariae* and *P. knowlesi* (Maguire and Baird, 2010). Transmission to humans requires a bite by an infected female anopheline mosquito. Mosquitoes ingest *Plasmodium* gametocytes through a blood meal from an infected human. Gametocytes differentiate into gametes in the mosquito midgut, fuse to form zygotes which transform into motile banana-shaped ookinetes that, about 24 hours after blood feeding, cross the peritrophic matrix and the mosquito midgut epithelium and form oocysts in the extracellular space between the midgut epithelium and the basal lamina. Several days later, oocysts release thousands of sporozoites that traverse through the haemolymph and reach the mosquito salivary glands, ready to be transmitted to an uninfected human through another mosquito bite and transmit malaria (Ghosh et al., 2000) (Figure 1.1).

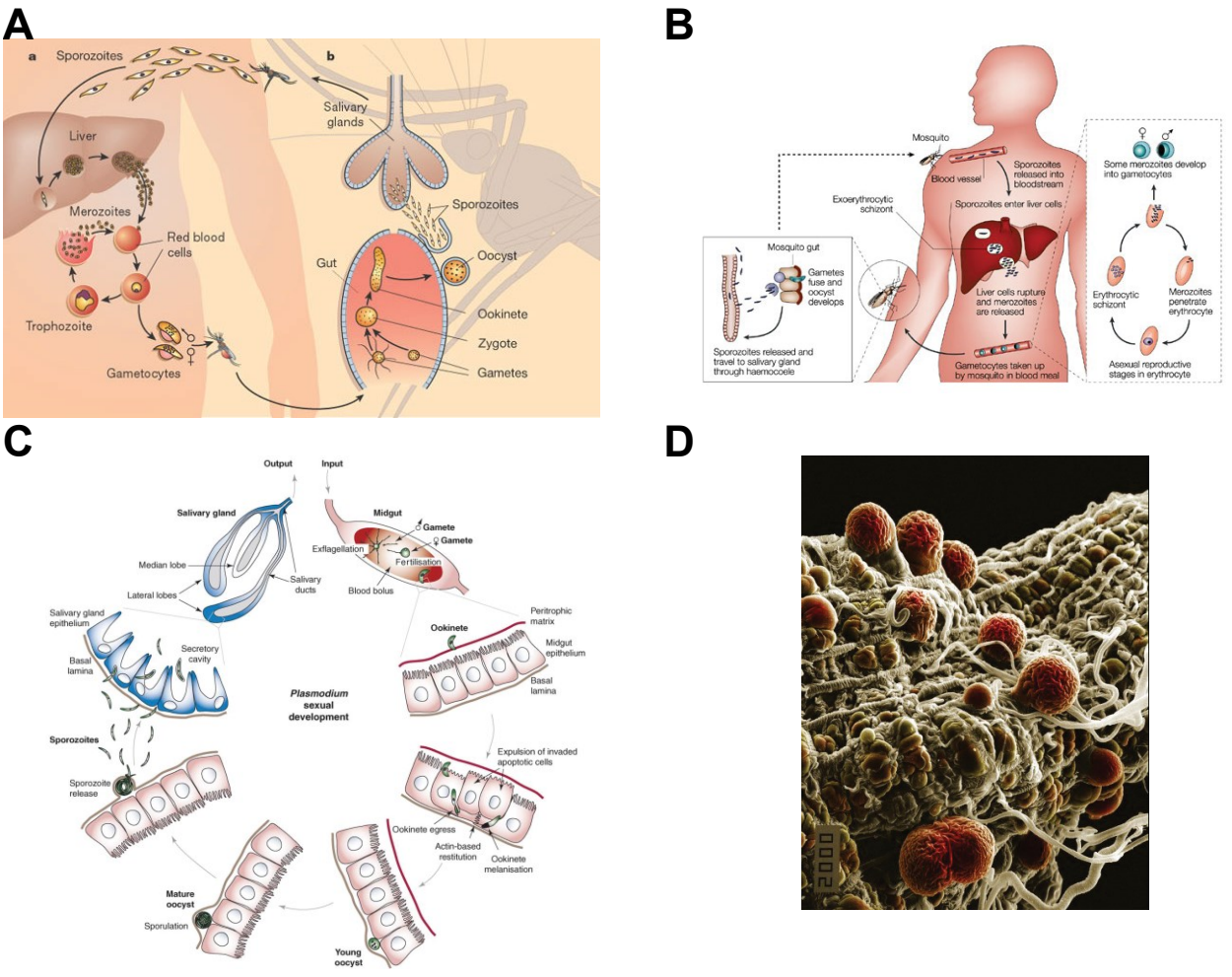


Figure 1.1: The life cycle of malaria *Plasmodium* parasites. A: The *Plasmodium* life cycle spans across a vertebrate and an invertebrate mosquito host stage. A *Plasmodium*-infected mosquito transmits malaria by biting an uninfected human host and thus transmitting parasites in their sporozoite form. Adapted from: (Wirth, 2002). **B:** Sporozoites end up in the liver where they develop into merozoites which invade red blood cells and multiply in their trophozoite stage. In some red blood cells, merozoites develop into gametocytes. Infected red blood cells burst, causing most malarial clinical features. Adapted from: (Stevenson and Riley, 2004) **C:** When a mosquito bites an infected human, it takes up gametocytes present in the blood meal. Sexual development of the *Plasmodium* parasites in the mosquito midgut includes development into male and female gametes, which fuse to form a zygote. Zygotes develop into motile, banana-shaped ookinetes which invade the mosquito midgut about 24 hours after entering the mosquito gut and develop into oocysts in the space between the midgut epithelium and the basal lamina. Oocysts subsequently burst and release thousands of sporozoites that reach the mosquito salivary glands, ready to be transmitted to an uninfected human through a mosquito bite. Adapted from: (Vlachou et al., 2006) **D:** Oocysts of *P. yoelii* parasites developing in

the midgut of *An. stephensi.* Copyright: Hillary Hurd, Wellcome Images. Used under the Creative Commons Attribution-Non-Commercial version 2.0 license for England & Wales.

Plasmodium sporozoites that reach humans through a mosquito bite invade hepatocytes and develop into merozoites, which are released into the blood and invade red blood cells. The rupture of infected red blood cells causes fever and rigours, indicative of symptomatic malaria (Miller et al., 2013). Some merozoites further develop into gametocytes which complete the *Plasmodium* life cycle when transmitted to a mosquito that bites the infected individual (Nakazawa et al., 2011). In most cases, especially if treated early with antimalarial drugs that target the parasite blood stage, malaria is not life threatening and, following mild symptoms such as intermittent fever episodes, nausea and headache, it is resolved through host innate and adaptive immune mechanisms (Stevenson et al., 2011; Stevenson and Riley, 2004). Such immune responses do not eliminate but rather reduce parasite levels (Ramasamy, 1998), while repeated exposure to infection can lead to development of immunity that further limits clinical symptoms but can also participate in immune pathology (Artavanis-Tsakonas et al., 2003).

1.3 Progression to fatal forms of severe malaria

Malaria caused mainly by P. *falciparum*, when untreated and especially in vulnerable groups such as children under the age of 5, pregnant women and non-immune individuals can lead to its severe form that is the cause of malaria-related deaths. Typically, 10% of *P. falciparum* malaria can be classified as severe while 10% of severe malaria leads to death (Pasvol, 2005). The severe form of malaria manifests itself through cerebral malaria that includes neurological symptoms and can lead to acute coma, severe anaemia and metabolic acidosis that may lead to respiratory distress and is usually indicative of whether the incident is fatal or not (Miller et al., 2002). Development of cerebral malaria largely depends on the ability of infected red blood cells to adhere to the endothelium of microvessels, causing occlusion but also inducing an immune response that contributes to the manifested pathology (Artavanis-Tsakonas et al., 2003; Mazier et al., 2000). Aberrant regulation of immune responses also contributes to the pathogenesis of severe anaemia (Perkins et al., 2011).

A key issue that remains poorly understood is the identification of host factors that facilitate progression to severe malaria. Susceptibility of non-immune individuals suggests that polymorphisms of immune factors involved could influence progression to severe malaria, with several such polymorphisms being associated with protection against severe malaria (Mazier et al., 2000; Munde et al., 2012; Ong'echa et al., 2011; Schuldt et al., 2011; Velavan et al., 2012). Furthermore, deficiency in glucose-6-phosphate

dehydrogenase (G6PD) has been associated with protection against severe malaria (Ruwende et al., 1995). The prevalence of *G6PD* variants that, despite causing haemopathologies, protect against severe malaria, comprises one of the best studied examples of natural selection in humans (Sabeti et al., 2002; Tishkoff et al., 2001). Several red blood cell polymorphisms (Williams, 2006), with most strikingly the sickle cell anaemia trait (Williams et al., 2005a; Williams et al., 2005b), also protect against severe malaria. The genetic basis of resistance to severe malaria is also highlighted by the association of loci related to erythrocytes and endothelial cells with severe malaria pathogenesis (Timmann et al., 2012), suggesting that future interventions against malaria could focus on enhancing the ability shown in human populations to disrupt progression to fatal forms of malaria.

1.4 Malaria through history

The origin of *Plasmodium* parasite lineages greatly precedes human history and is estimated at more than 150 million years ago, about at the time of divergence of birds from reptiles (Escalante and Ayala, 1994). The three main agents of human malaria, *P. falciparum*, *P. vivax* and *P. malariae,* are also thought to have diverged before the advent of hominids, with phylogenetically independent events resulting in their parasitic association with humans (Escalante and Ayala, 1994). The virulent *P. falciparum* strains are considered to have an ancient origin from gorilla parasites (Liu et al., 2010b), with the last common ancestor originating up to 180,000 years ago, coinciding with human population expansion (Mu et al., 2002), while a more recent expansion is considered to have taken place at about 10,000 years ago (Joy et al., 2003). Severe *P. falciparum* malaria is considered to have become endemic for the first time in tropical Africa, 5,000 to 10,000 years ago, coinciding with the advent of agriculture and the diversification of anthropophilic mosquitoes (Hartl et al., 2002). *P. falciparum*, as well as the other human malaria parasites, is thought to have expanded from pre-agricultural Africa to the rest of the world (Carter and Mendis, 2002).

The incidence of malaria has been documented in historical times, through its characteristic clinical symptoms of intermittent fever episodes and enlarged spleen (Neghina et al., 2010). Malarial epidemics have been tracked in the Chinese Canon of Medicine (Nei Ching) dated at 2700 BC (Hoeppli, 1956) while *P. falciparum* has been traced in ancient Egyptian mummies dated at about 4,000 years ago (Nerlich et al., 2008). Malarial paroxysm was also described by Hippocrates at around 500 BC (Cunha and Cunha, 2008). Malaria-affected parts of the Roman Empire, at least since 200 BC, had been negatively correlated with agricultural activity and economic prosperity, while malaria became widespread in the

Mediterranean shores of Southern Europe in the Christian era (Carter and Mendis, 2002). Malaria, under the name of ague, severely plagued England, causing illness and death from the 16th until the 19th century, when it started to severely decline (Reiter, 2000).

1.5 Malaria causal agents and transmission

For centuries, the cause of malaria transmission was attributed to "miasmata". Indeed, the word malaria stems from the Italian words *mal'aria*, meaning bad/evil/corrupted air (Neghina et al., 2010). Despite this prevailing theory, reports linking malaria to the environment, the presence of marshes or mosquitoes, have been noted before the elucidation of the cause of malaria (Cook and Webb, 2000). The establishment of bacteria as agents of disease by Pasteur and Koch in the 1870s (Karamanou et al., 2012) incentivized the search for a similar explanation for the cause of malaria (Figure 1.2A). In 1880, Alphonse Laveran, a French army officer, discovered protozoan parasites in the blood of malaria patients while in 1897, the work of Ronald Ross and Patrick Manson implicated culicine mosquitoes in the transmission cycle of avian *Plasmodium* parasites (Manson, 2002) (Figure 1.2B). At about the same time, Giovanni Batista Grassi, Amico Bignami and Giuseppe Bastianelli demonstrated transmission of human malaria in anopheline mosquitoes (Cox, 2010) (Figure 1.2B).

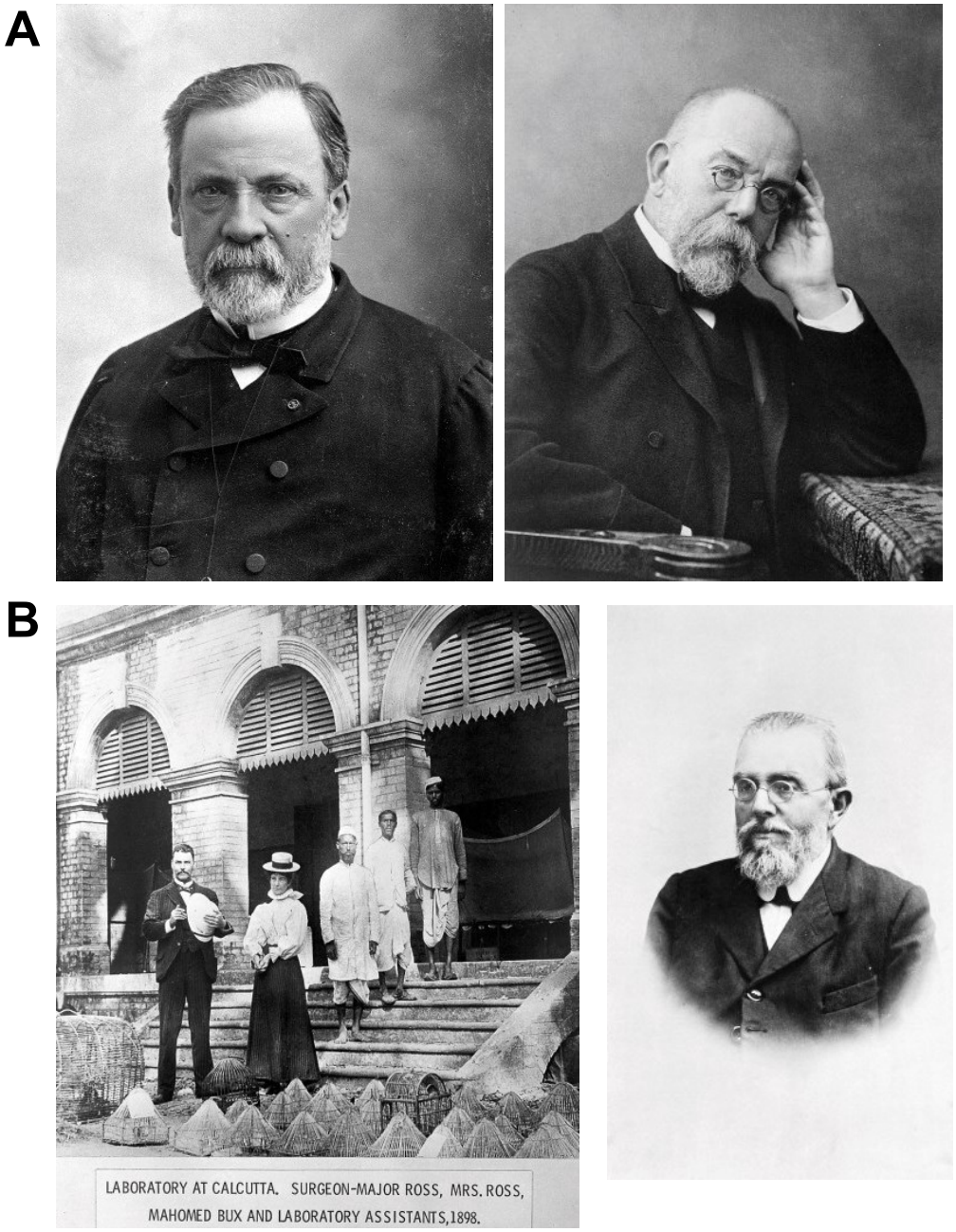


Figure 1.2: Malaria is caused by *Plasmodium* parasites and transmitted through mosquitoes. A: Louis Pasteur (left) and Robert Koch (right) laid the foundation of the hypothesis that disease is caused by microbial agents and led to the initiation of the search for a pathogenic agent that causes malaria. Copyright: Wellcome Library, London. **B:** Ronald Ross, a British army officer, contributed to the implication of mosquitoes in the transmission of malaria (left) and Giovanni Batista Grassi, contributed to the demonstration that humans can contract malaria through the bite of anopheline mosquitoes (right). Copyright: Wellcome Library, London. Used under the Creative Commons Attribution-Non-Commercial version 2.0 license for England & Wales.

1.6 Malaria eradication campaigns

Soon after mosquitoes were recognized as malaria vectors, the first campaigns to contain or eradicate malaria commenced (Figure 1.3A-B). In 1905, a campaign to contain malaria and yellow fever epidemics in the Panama Canal construction workers was highly successful (Le Prince, 1916). The advent of chloroquine as a potent antimalarial (Krafts et al., 2012) and DDT as a highly effective insecticide (Raju, 1999) rendered malaria eradication feasible. In 1955, the Global Malaria Eradication Programme was initiated by the World Health Organization (WHO) (WHO, 1973) (Figure 1.3C-D). This programme relied heavily on vector control using indoor residual spraying with DDT and indeed assisted in malaria elimination, especially in countries that already had functional control programmes (Najera et al., 2011). Although malaria was eliminated from 49 countries, including most of Europe and North America, from 1950 to 1978, little attention was paid to the instability and socioeconomic issues of particular countries, with Africa not included in the initial campaign (Anders and Hay, 2012). Further complications that arose included resistance to chloroquine (Mita and Tanabe, 2012) and DDT (Livadas and Georgopoulos, 1953), environmental concerns regarding the widespread use of DDT (Rosner and Markowitz, 2013), resurgence of malaria after disruption of control programmes (Cohen et al., 2012), which, along with diversion of funding especially from the United States of America that comprised the main contributor, resulted in a change of policy by the WHO in 1969 from malaria eradication to minimization of health damage, with a shift from vector control to antimalarial treatment (WHO, 1973).

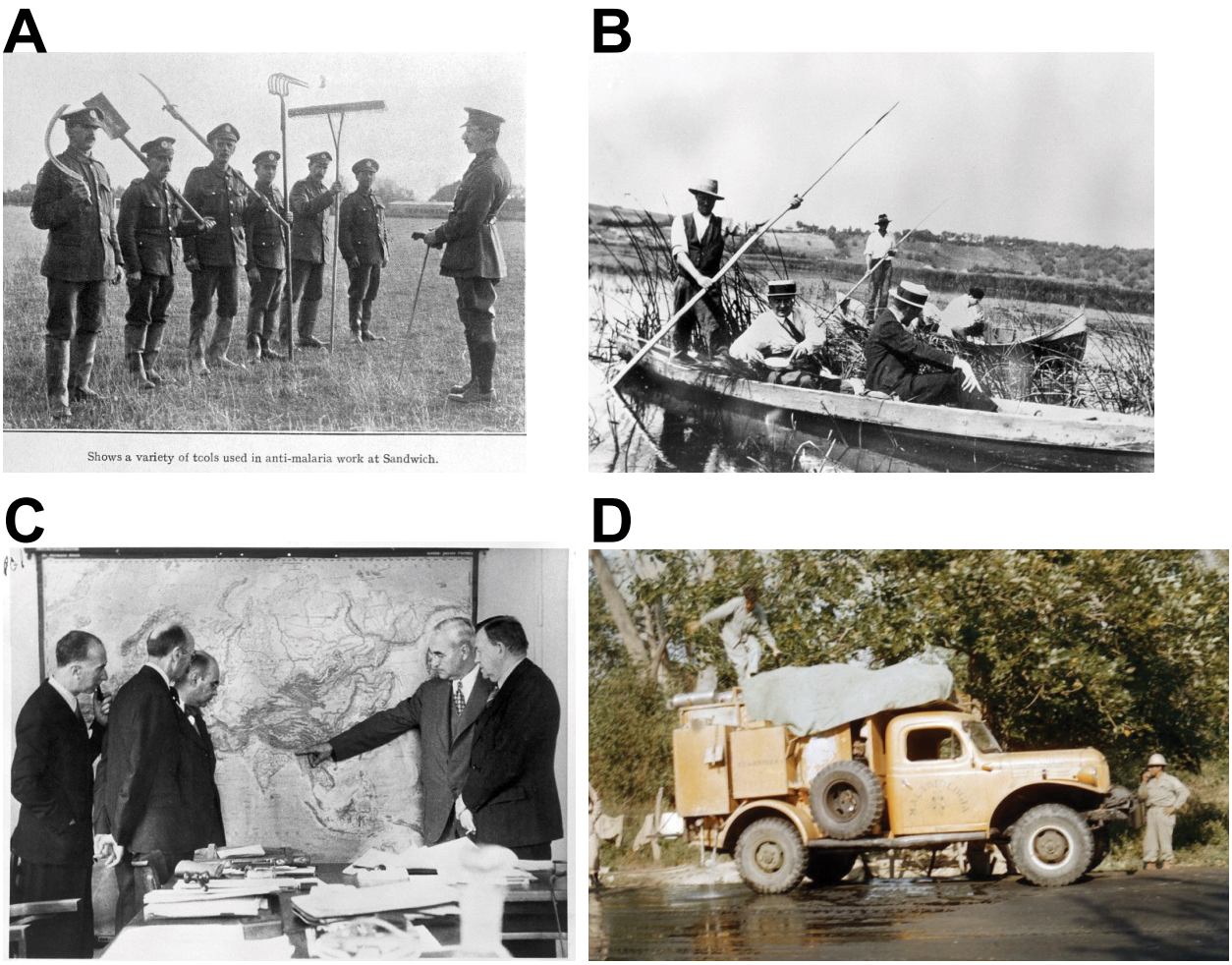


Figure 1.3: Malaria eradication campaigns. A: Anti-malaria brigade, Great Britain War Office, 1919. Copyright: Wellcome Library, London. **B:** Members of the Malaria Commission of the League of Nations collecting larvae on the Danube Delta, 1929. Copyright: Wellcome Library, London. **C:** World Health Organization interim committee on malaria, 1947. Copyright: Wellcome Library, London. **D:** Malaria eradication programme in Venezuela, mid 1950s. Copyright: Wellcome Library, London. Used under the Creative Commons Attribution-Non-Commercial version 2.0 license for England & Wales.

1.7 The burden of malaria today

Public perception in developed countries still regards malaria as a scourge of the past: a preventable, treatable disease that has been all but eliminated. Some sporadic incidents that occur are easily tackled by proper medication (Kousoulis et al., 2012). Reality cannot be further from this perception: malaria remains one of the major global public health emergencies[1] (WHO, 2012) (Figure 1.4A-B). In 2010, malaria was responsible for an estimated 660,000 deaths resulting from 219 million malaria incidents. Although 104 countries are considered endemic, with 99 countries with ongoing malaria transmission and 3.3 billion people at risk, over 90% of mortality occurs in sub-Saharan Africa, especially targeting children under 5 years old. This concentration of the burden in few under-developed countries is likely responsible for the limited sense of urgency seen in the developed world in fighting malaria. Indeed, 14 countries in 2010 accounted for about 80% of total malaria-related deaths, with the Republic of Congo and Nigeria alone sustaining 40% of malaria-related mortality.

[1] The source for all subsequent facts and figures in this section is the WHO World Malaria Report 2012. WHO (2012). World Malaria Report 2012 (Geneva: WHO Press).

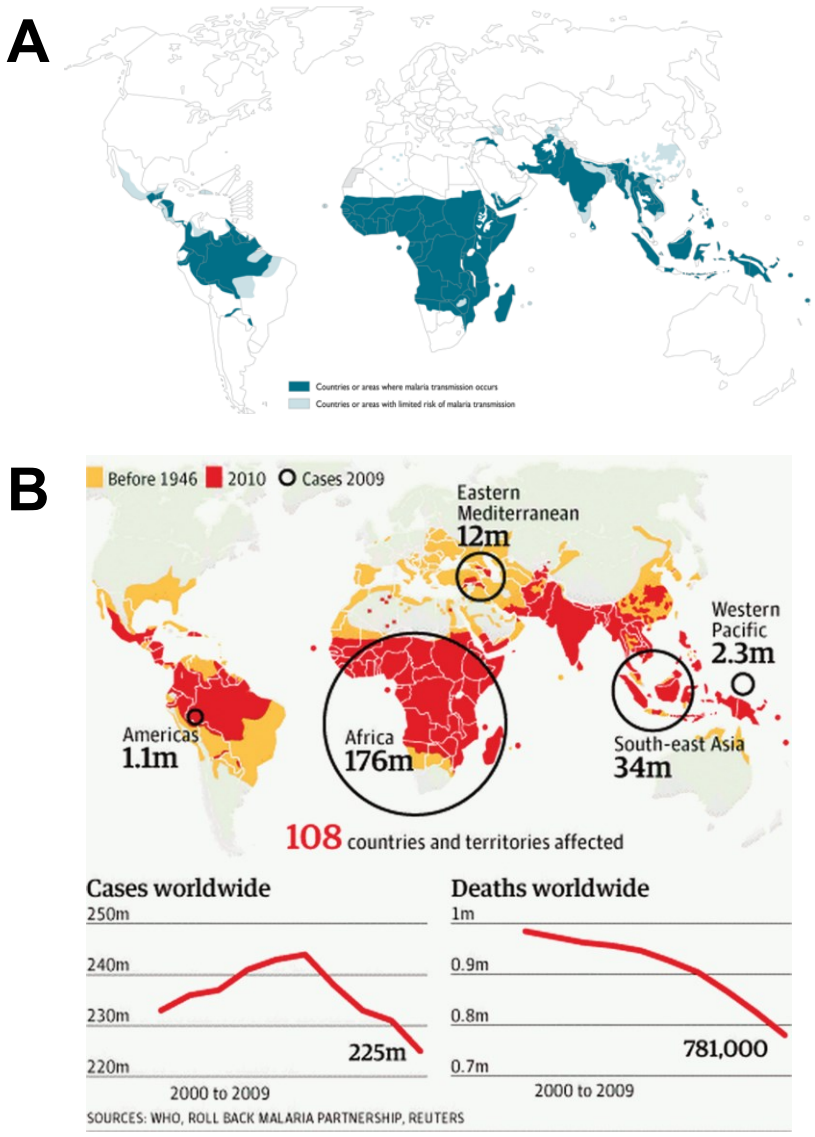


Figure 1.4: The burden of malaria today. A: Countries or areas where malaria transmission occurred in 2010. Source: WHO. **B:** Affected countries, malaria cases and deaths in 2009. Adapted from: (Boseley, 2011).

It has to be noted, though, that in the last decade or so, a surge in funding both by organizations disbursing public funds, such as the Global Fund (Lancet, 2013), but also by initiatives through privately funded philanthropies such as the Bill and Melinda Gates foundation (Greenwood et al., 2012; McCoy et al., 2009), has been effective in considerably decreasing the burden of malaria especially in poverty-stricken countries. International disbursement on fighting malaria increased from 100 million US dollars in 2000 to 1.84 billion US dollars in 2012, for a total of 2.3 billion US dollars if domestic government spending on malaria is also accounted for. Funded interventions focused on vector control, preventive treatments, diagnostic tools and treatment of infected individuals. Indeed, from 2000 to 2010, malaria-related mortality decreased by 25%, with a 33% decrease in Africa alone. A major focus of this campaign was the widespread use of insecticide-treated nets (ITNs). ITN coverage increased from 3% of households in sub-Saharan Africa in 2000 to 53% in 2012, to a total of 66 million ITNs. Another strategy employed for vector control was indoor residual spraying (IRS) whose coverage of protected households in the African region increased from 5% in 2005 to 11% in 2011, covering 153 million protected individuals. The intermittent preventive treatment of vulnerable groups in high transmission areas with antimalarial therapies has been also employed to some degree.

1.8 Current strategies for malaria containment and treatment

As resistance to the once potent antimalarial drug chloroquine is widespread in *P. falciparum* populations (Lehane et al., 2012; Payne, 1987; Sidhu et al., 2002), the current first-line treatment of *P. falciparum* infected individuals is the use of artemisinin-based combination therapies (ACTs). Artemisinin was discovered in 1971 from traditional herbs in China, in an initiative involving more than 500 Chinese scientists aiming to aid the North Vietnamese plagued by chloroquine-resistant *P. falciparum* malaria during the Vietnam war (Miller and Su, 2011) (Figure 1.5). Artemisinin is currently the most potent antimalarial but due to its short-lived potency, its use in combination with other partner drugs is required for malaria clearance. For *P. vivax* malaria, chloroquine is still in use, with ACTs employed if resistance to chloroquine is observed. Indeed, in 2011, 278 million treatment courses of ACTs were disbursed. Most worryingly, as has happened in the past for other treatments, resistance of *P. falciparum* to ACTs has been observed (Dondorp et al., 2009). Artemisinin resistance has been confirmed in Southeast Asia, first in Cambodia in 2008 and subsequently in Myanmar, Thailand and Vietnam (Phyo et al., 2012). Remarkably, resistance to past chloroquine, pyrimethamine and sulfadoxine antimalarial therapies first emerged in the same area in Southeast Asia and subsequently spread to Africa (Mita and Tanabe, 2012). The emergence

of resistance to artemisinin as well as to past antimalarial therapies in this region of Southeast Asia is most likely due to high levels of genetic differentiation of local *P. falciparum* populations (Miotto et al., 2013). The paramount importance of ACTs in controlling malaria-related mortality makes it imperative to slow the spread of such resistance, mainly by eliminating artemisinin monotherapies still employed in some regions and develop novel antimalarial therapies (Miller et al., 2013). Nevertheless, past experience has shown that spread of such resistance is inevitable and could result in a surge of malaria-related deaths if no other interventions have emerged at that point in time. At the same time, resistance has also been documented for second-line antimalarials such as quinine (Okombo et al., 2011) and mefloquine (Cowman et al., 1994), which, along with their cost, recorded efficacy and side-effects, make them unlikely replacements for artemisinin (Schlagenhauf et al., 2010; Yeka et al., 2009; Yeka et al., 2013).

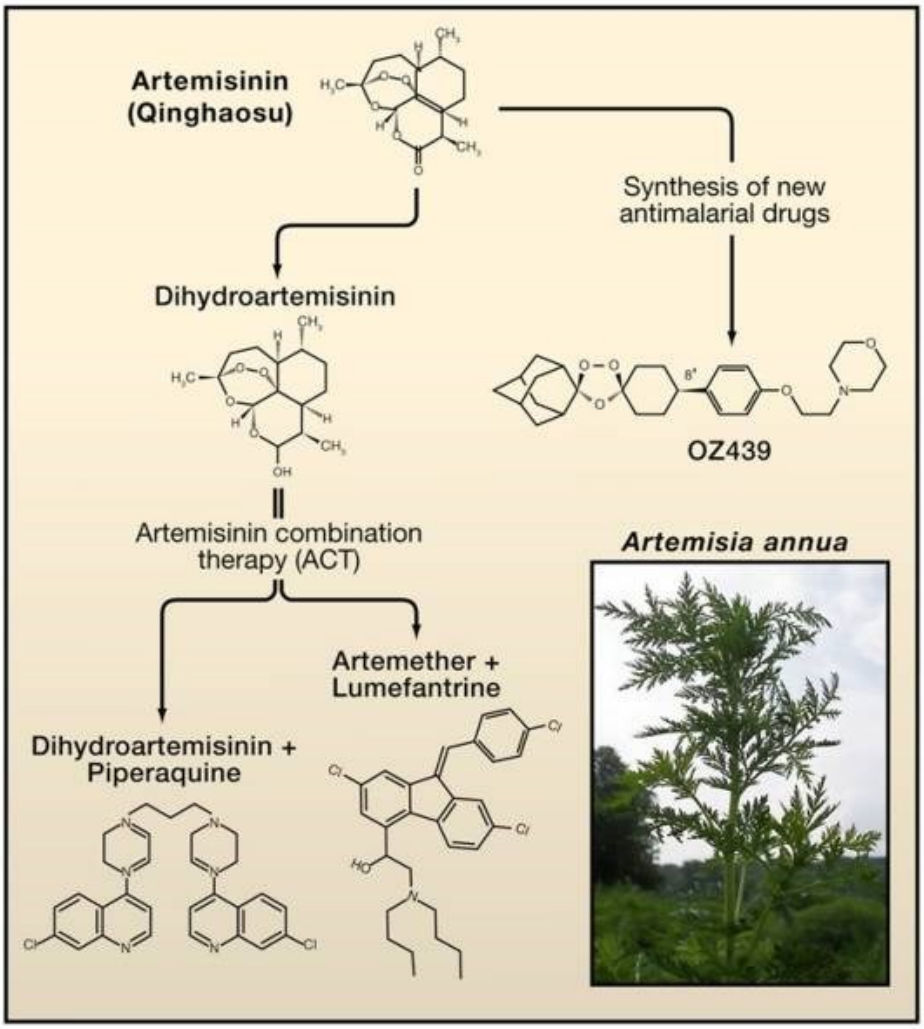


Figure 1.5: Artemisinin-based therapies. The antimalarial drug artemisinin was identified from the plant *Artemisia annua L.,* a plant used in traditional Chinese therapies. Many derivatives have been synthesized based on the initial artemisinin molecule, such as dihydroartemisinin, which is more active than artemisinin. Artemisinin or its derivatives are typically used in combination with other antimalarial drugs so as to effect malaria clearance and avoid development of resistance. Adapted from: (Miller and Su, 2011).

Another worrying issue in the current intervention regime is the spread of insecticide resistance (IR) (Figure 1.6A-B). IR has been documented in 64 countries and involves all *Anopheles* species but, most importantly, all classes of insecticides in use, including pyrethroids, the most common class of insecticides used in ITNs (Ranson et al., 2011) but also organochlorines that include DDT (Livadas and Georgopoulos, 1953; Ndiath et al., 2012; Tikar et al., 2011). Therefore, the identification of new classes of insecticides is also of paramount importance in the fight against malaria, as growing IR can take a substantial toll on the efficacy of current interventions.

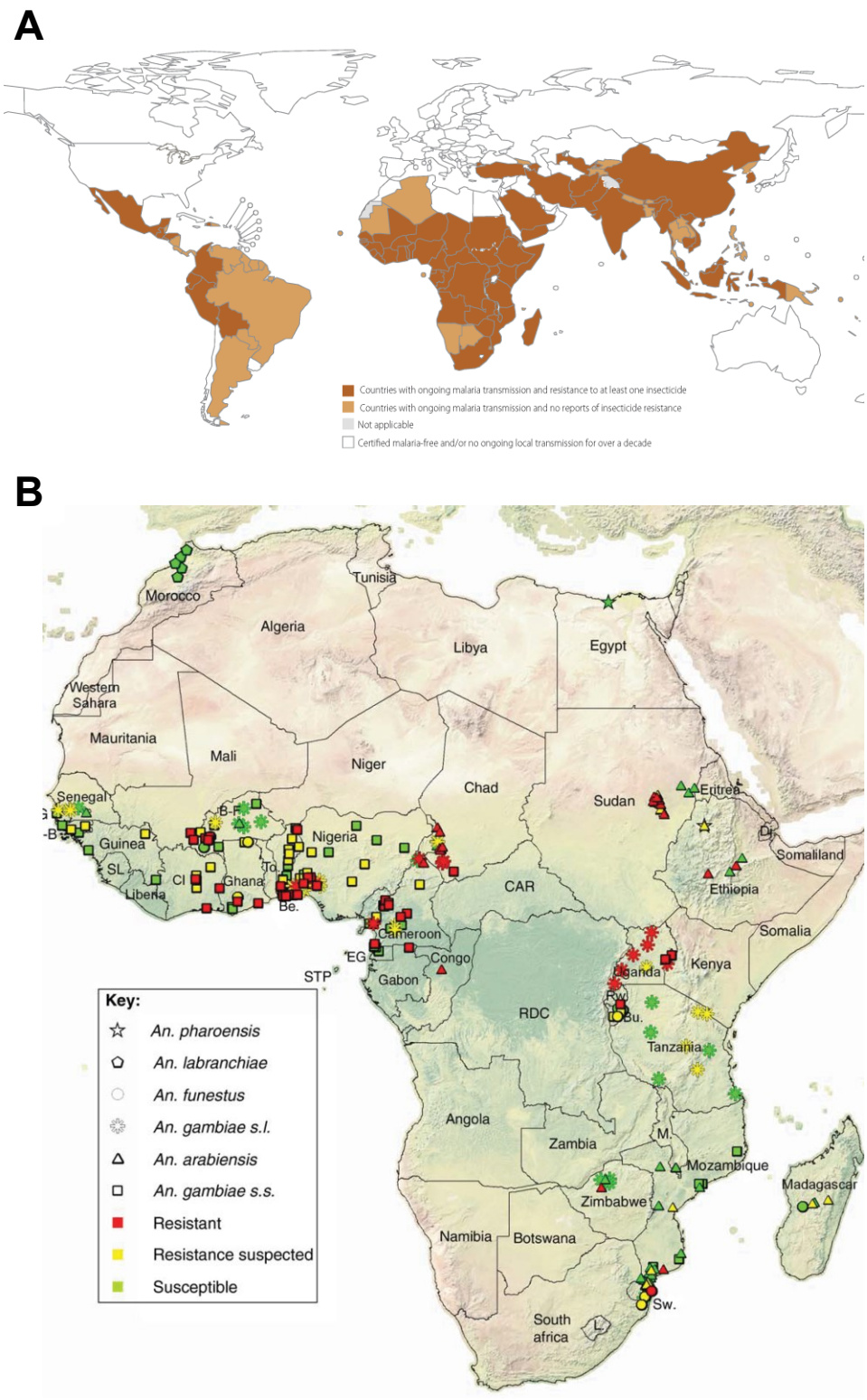


Figure 1.6: Insecticide resistance of the major malaria vector mosquitoes. A: Countries with exhibited insecticide resistance in at least one major malaria vector in 2012. Source: WHO Malaria report 2012. **B:** Distribution of pyrethroid resistance in Africa. Areas with *Anopheles* mosquitoes showing resistance, suspected resistance or susceptibility to pyrethroids. Based on data collected from 2000 to 2010. Adapted from: (Ranson et al., 2011).

As has been made clear, although malaria is treatable and preventable, no magic bullet has yet to emerge in the effort towards malaria elimination. Poor infrastructure, multiple breeding sites and unstable political and socioeconomic conditions in affected countries further hinder interventions against malaria. Furthermore, it is not only the emergence of resistance to ACTs or IR that could lead to a surge in malaria-related mortality; a halt to current interventions due to reduced funding or disruptive events such as civil wars could also lead to a precipitous increase of malaria-related deaths, as malaria incidents in populations that lack immunity acquired by previous exposure (Ramasamy, 1998) are expected to increase incidents of severe malaria.

1.9 Novel approaches needed in face of expensive and volatile malaria intervention strategies

The development of an efficacious vaccine could greatly assist in efforts to fight malaria (Targett and Greenwood, 2008). Currently, no such vaccine has been developed and only one candidate, *RTS,S/AS01* is in phase III clinical trials (White, 2011). *RTS,S/AS01* is expected to go on the market in 2015 providing partial protection against malaria as initial results show a modest 31% reduction in malarial clinical episodes (Vogel, 2012). Therefore, the cost and partial efficacy of such approach make it questionable whether such intervention will achieve a substantial blow in the fight against malaria, although it may supplement ongoing efforts. An alternative approach pursued is the development of transmission-blocking vaccines that, unlike *RTS,S/AS01* that target the parasite liver stages, target sexual or mosquito stages of the parasite (Alonso et al., 2011) or even mosquito antigens (Dinglasan and Jacobs-Lorena, 2008; Lavazec and Bourgouin, 2008) in an effort to disrupt malaria transmission, thus protecting the population rather than just the vaccinated individual (Figure 1.7). Despite various approaches at developing transmission-blocking strategies (Blagborough et al., 2013a; Blagborough et al., 2013b; Chowdhury et al., 2009; Wu et al., 2008), no such vaccine is expected to imminently enter the market.

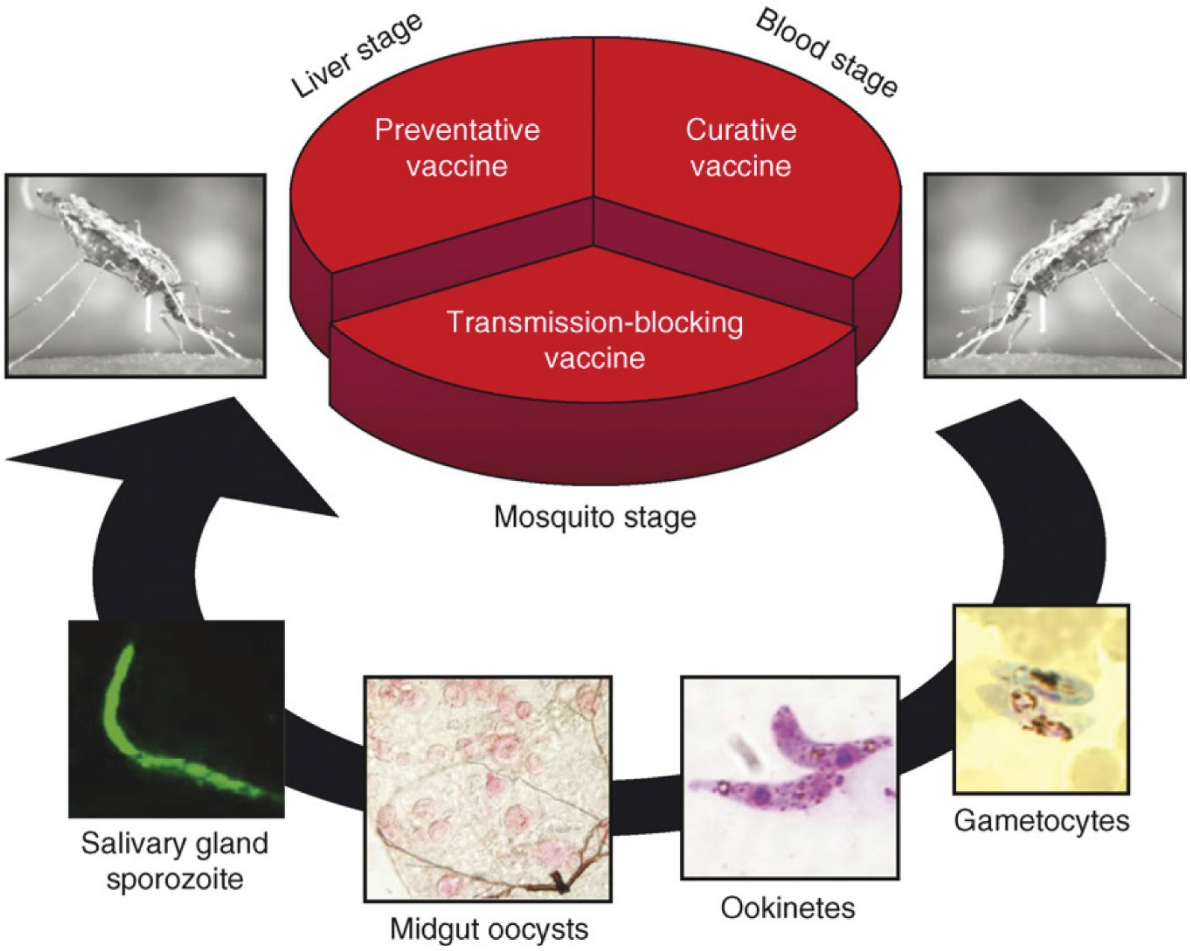


Figure 1.7: Malaria vaccines potentially target different stages of the *Plasmodium* life cycle. Preventative or curative vaccines target parasite liver or blood stages inside the human host. Transmission-blocking vaccines target stages of the parasite in the mosquito host, from ingested gametocytes to ookinetes, oocysts or sporozoites. Adapted from: (Dinglasan and Jacobs-Lorena, 2008).

The conclusion to be drawn is that, despite all efforts, malaria remains a huge burden especially in sub-Saharan Africa, accounting for 660,000 deaths globally. The current interventions have a substantial cost of 2.3 billion US dollars annually, while the estimate for a more sustained effort in malaria interventions is 5.1 billion US dollars annually for the period from 2011 to 2020. Furthermore, anticipated IR or resistance to ACTs can profoundly undermine these efforts. The current malaria intervention regime, therefore, is expensive and partially effective while it is highly uncertain for how long this partial efficacy can be sustained. It becomes clear that the fight against malaria urgently requires innovative approaches on how malaria transmission can be disrupted.

1.10 The mosquito taxonomic status

Mosquitoes belong to the *Insecta* class, *Diptera* order and *Culicidae* family (*Insecta*: *Diptera*: *Culicidae*). Currently, according to the Mosquito Taxonomic Inventory (Harbach, 2013), *Culicidae* includes 3,531 different species classified in the *Anophelinae* and *Culicinae* subfamilies. *Anophelinae* are further subdivided into 3 genera, one of them the *Anopheles* Meigen, 1818 genus that comprises 469 formally recognized species (Harbach, 2013). Mosquitoes exhibit remarkable morphological diversity, especially in tropical regions, and almost global distribution, being “as ubiquitous as water” (Foley et al., 2007; Grimaldi, 2005). Divergence leading to the *Anophelinae* and *Culicinae* subfamilies is dated more than 200 million years ago, with an expected rapid radiation early in the diversification process (Reidenbach et al., 2009).

Remarkably, less than 200 mosquito species are of medical importance as disease vectors (Reidenbach et al., 2009), while about 40-60 anopheline species can transmit malaria (Hay and Snow, 2006; Manguin, 2008) (Figure 1.8). Even within this limited number of anopheline mosquito species susceptible to malaria transmission, vectorial capacity is highly variable, for reasons that remain poorly understood and are still subject of intense research interest (Cohuet et al., 2010; Neafsey et al., 2013).

A

B

Figure 1.8: The *Anopheles* mosquito. A: A blood-fed *An. stephensi* mosquito in flight. Copyright: Hugh Sturrock, Wellcome Images. **B:** An *An. gambiae* mosquito while biting human skin. Copyright: Wellcome Library, London. Used under the Creative Commons Attribution-Non-Commercial version 2.0 license for England & Wales.

1.11 *Anopheles* population structure

Anopheles mosquitoes show a high level of genetic diversity, exhibiting a highly intricate phylogenetic diversification and population structure (Loaiza et al., 2012). Ecological adaptations and anthropophilic feeding behaviour have largely driven this genetic diversification (Ayala et al., 2010; Besansky et al., 1994), which remains a subject of intense and ongoing research interest due to its importance in the elucidation of mosquito-pathogen co-evolution (Ndiath et al., 2011), the differential involvement of isolated populations in malaria transmission (Coluzzi, 1992; Mendes et al., 2008) or the effect exerted on future interventions to fight malaria transmission of a highly diverse and fragmented mosquito population (Alphey et al., 2002).

The main malaria vector in sub-Saharan Africa that was used in this study, *An. gambiae,* is in fact just one of the seven closely related and morphologically indistinguishable members of the *An. gambiae* complex. The species referred to here as *An. gambiae* corresponds to *An. gambiae sensu stricto* (*s.s.*), meaning in the strict sense. The other six members of the *An. gambiae* complex are: *An. arabiensis, An. bwambae, An. melas, An. merus and An. quadriannulatus* (Besansky et al., 1994; Hunt et al., 1998). Diversification within this complex is believed to have been driven by an inversion-based speciation process (Coluzzi, 1982). Paracentric chromosomal inversions are abundantly observed within the *An. gambiae* complex, with >120 inversions detected in natural populations and 10 inversions fixed in different species (Coluzzi et al., 1979). These chromosomal rearrangements can reduce recombination rates for part of the genome, capture and stabilize blocks of co-adapted genes and thus can be employed in natural selection processes for adaptation to different environments, further ecotypic differentiation and exploitation of ecological niches (Coluzzi et al., 2002).

A prominent example of a chromosomal inversion that confers a selective advantage within *An. gambiae s.s.* is the 2La inversion (Sharakhov et al., 2006), which is non-randomly distributed in relation to the degree of humidity, showing fixation in the arid north of West Africa and absence in favor of the 2L+^{a} karyotype in the humid south (Simard et al., 2009; White et al., 2007). The 2La inversion has been implicated in adaptive processes that enable *An. gambiae s.s.* to exploit a broad range of climatic conditions, possible by capturing sets of alleles that confer an advantage in arid conditions such as desiccation resistance and thermal tolerance (Gray et al., 2009; Rocca et al., 2009).

Except for *An. quadriannulatus*, which has a standard chromosomal rearrangement (Coluzzi et al., 2002), a large number of hosts and feeding preferences for animal blood (Habtewold et al., 2008), all other members of the *An. gambiae* complex are vectors of human malaria, with *An. gambiae s.s.* exhibiting the

highest vectorial capacity (Coluzzi et al., 2002). *An. bwambae,* which breeds in mineral water as well as *An. merus* and *An. melas* that breed in saltwater, are minor malaria vectors with narrow geographical distribution (Coetzee et al., 2000; Coluzzi et al., 1979) (Figure 1.9, lower panels). *An. gambiae s.s*. and *An. arabiensis* are highly anthropophilic, they breed in temporary, sunlit freshwater, with *An. gambiae s.s*. mainly found in humid areas as rainforests and *An. arabiensis* in more arid savannas (Kamali et al., 2012). Nevertheless, both *An. gambiae s.s.* and *An. arabiensis* have a broad distribution in Western and Eastern Africa as well as Madagascar and are often found in sympatry (Coetzee et al., 2000) (Figure 1.9, upper panels).

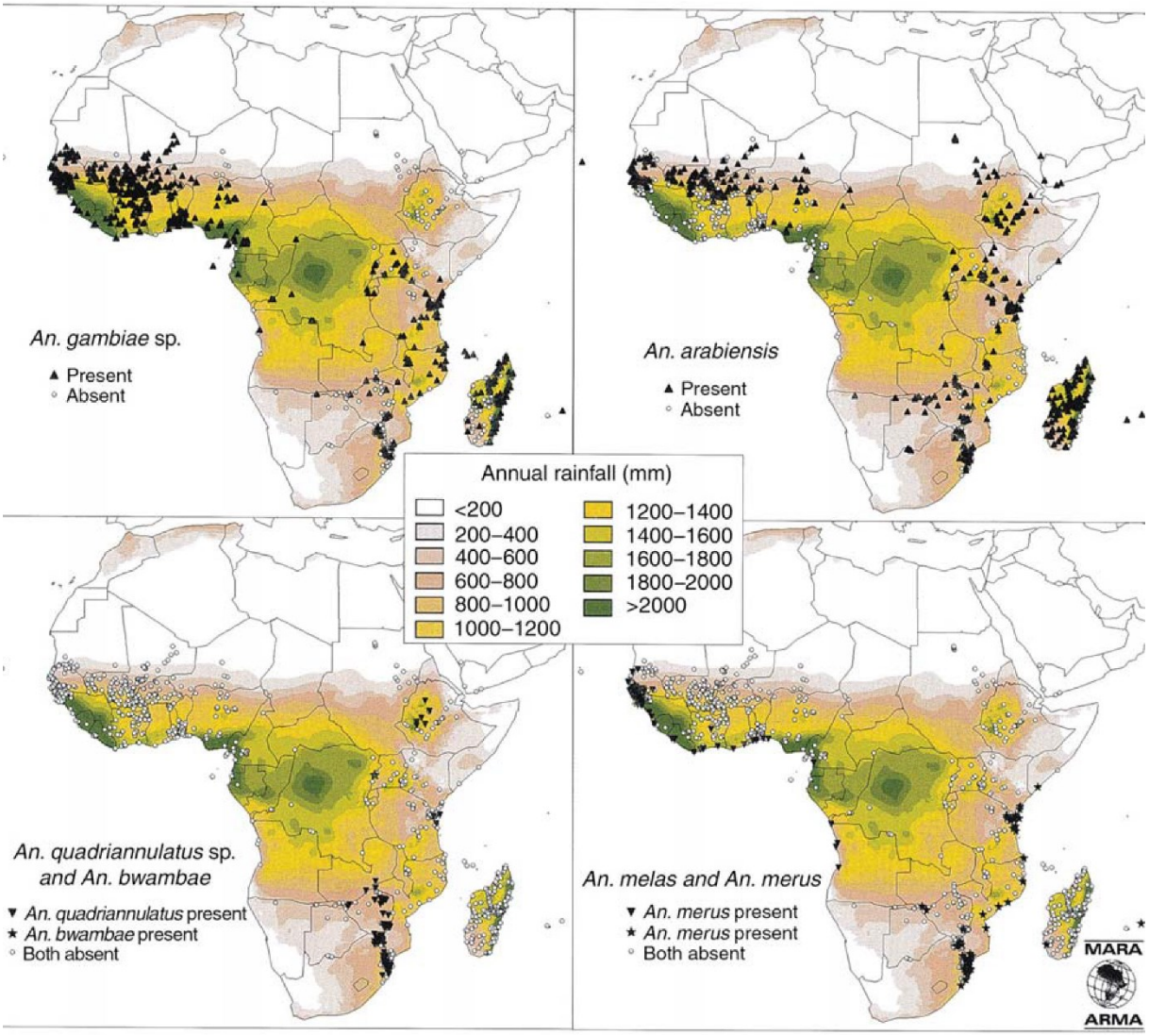


Figure 1.9: Distribution of *An. gambiae* complex mosquitoes. The geographic location of collection sites in which the respective members of the *An. gambiae* complex have been found or were not recorded. The distribution of mean annual rainfall is also shown. Adapted from: (Coetzee et al., 2000).

Vectorial capacity for members of the *An. gambiae* complex is believed to have originated repeatedly through independent events following species diversification (Kamali et al., 2012). The speciation process of the highly anthropophilic *An. gambiae s.s.* was most likely driven by the impact of human agricultural activities on the environment. This process is believed to have commenced more than 3,000 years ago in African tropical rainforests. The combination of the availability of human hosts for blood with larval breeding opportunities due to agricultural activities could have provided strong selection for anthropophilic traits observed in *An. gambiae s.s.* (Coluzzi et al., 2002).

1.12 *Anopheles gambiae s.s.* chromosomal and molecular forms (or the M/S mess)

The occurrence of non-random distributions of inversions within *An. gambiae s.s.* populations, especially in the central part of the 2R chromosomal arm, suggested additional taxonomic splitting within the species, in an incipient speciation process, as well as a putative role for these chromosomal changes in polymorphisms that confer adaptation to ecological niches (Coluzzi et al., 2002; della Torre et al., 2002) (Figure 1.10A). In Mali and Burkina Faso, three sympatric chromosomal forms of *An. gambiae s.s.* were designated as *Mopti*, *Savanna* and *Bamako,* referring to sub-populations with different 2R chromosomal rearrangements and evidence of reproductive isolation between them (Favia et al., 1994). The *Forest* and *Bissau An. gambiae s.s.* chromosomal forms have also been proposed to a total of five different chromosomal forms (Coluzzi, 1985; Gentile et al., 2002). The *Mopti* chromosomal form was further associated with temporal and spatial climatic differences in Mali, indicating a role for 2R inversions that characterize the *Mopti* chromosomal form in adaptive processes leading to speciation events (Toure et al., 1994). *Savanna* and *Bamako* were found to prevail in humid savannas while *Mopti* showed higher ecological flexibility, with presence in various ecological zones but also dominance in flooded or irrigated areas (Toure et al., 1998).

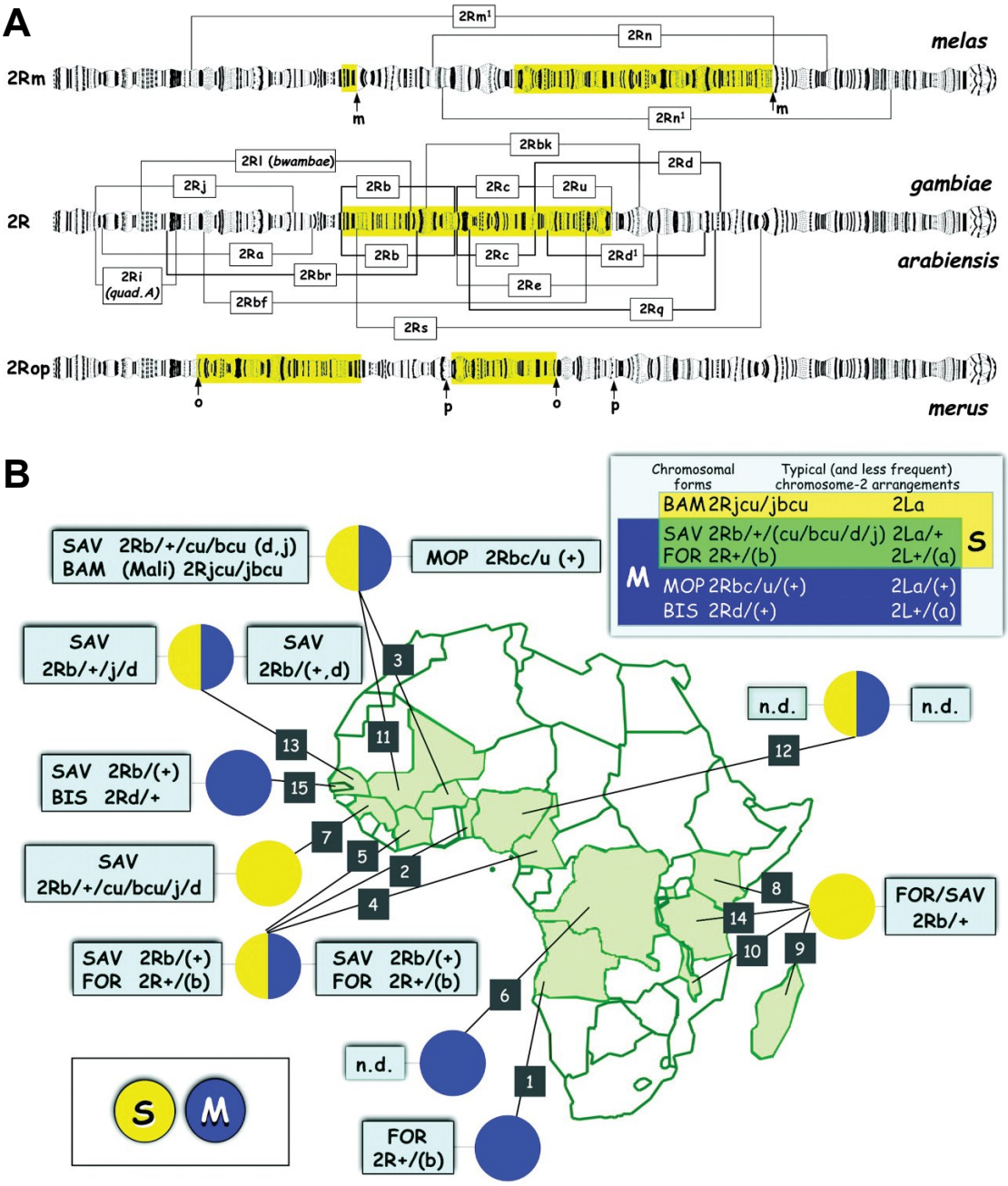


Figure 1.10: Chromosomal inversions in the *An. gambiae* complex and the *An. gambiae s.s.* M/S molecular forms. A: The main fixed and polymorphic 2R paracentric inversions in the sibling species of the *An. gambiae* complex. Adapted from: (Coluzzi et al., 2002). **B:** Geographical distribution of the M/S molecular forms of *An. gambiae s.s.* and their relation to chromosomal forms. Adapted from: (della Torre et al., 2002).

Paradoxically, despite evidence for the presence of adaptive processes and reproductive isolation occurring between chromosomal forms (Toure et al., 1998), with the *Mopti* form showing even complete reproductive isolation to the sympatric *Savanna* and *Bamako* forms in Mali (Favia et al., 1997; Favia et al., 2001), no signs of genetic differentiation were found between chromosomal forms, outside of inversions and some other chromosomal areas such as the intergenic spacer and the transcribed spacer of the ribosomal DNA locus on the X chromosome (della Torre et al., 2001; Gentile et al., 2001; Lee et al., 2013b). One explanation that has been suggested for this discrepancy is the existence of low levels of contemporary gene flow that could account for the observed reproductive isolation, while historical gene flow and co-ancestry could account for the limited genetic differentiation or could bias the gene flow estimates (Tripet et al., 2005). Nevertheless, the lack of connection between chromosomal constitution, as exhibited by the different chromosomal forms, and reproductive isolation, led to the introduction of two non-panmictic taxonomic units based on molecular markers, mainly X-linked ribosomal DNA, the M and S molecular forms of *An. gambiae s.s.* (della Torre et al., 2002; della Torre et al., 2001) (Figure 1.10B).

Based on molecular patterns in the X-linked ribosomal DNA intergenic spacer, which are consistently differentiated between the *Mopti* and *Savanna/Bamako* chromosomal forms in Mali (Favia et al., 1997; Favia et al., 2001), the *An. gambiae s.s.* M/S molecular forms could be discriminated using PCR restriction fragment length polymorphism (RFLP) (della Torre et al., 2001), corresponding to different ribosomal DNA types with fixed nucleotide differences in the intergenic spacer region of the ribosomal DNA cluster (Gentile et al., 2002; Gentile et al., 2001). The M/S molecular forms are essentially non-panmictic, showing almost complete reproductive isolation, with less than 3‰ hybrid patterns (della Torre et al., 2001), comparable to cross mating between *An. gambiae s.s.* and *An. arabiensis* (Toure et al., 1998). Nevertheless, heterogamous insemination by transferring sperm between M/S forms stands at about 1%, suggesting that there is potential for low-level gene flow, restricted through pre-mating barriers (della Torre et al., 2002; Tripet et al., 2001). The S form is the only one found outside West Africa, while the M form is confined to West Africa, found there in sympatry with the S form (Gentile et al., 2001). The relationship between chromosomal and molecular forms is quite intricate; in Mali, Burkina Faso and Guinea, M corresponds to the *Mopti* while S corresponds to the *Savanna/Bamako* chromosomal forms (della Torre et al., 2001; Gentile et al., 2001). In areas outside these regions, such as Senegal, Gambia, Benin or the North Ivory Coast, the association between chromosomal forms and ribosomal DNA types corresponding to different molecular forms breaks down, with M corresponding to a variety of inversion arrangements, from *Mopti* to *Savanna* or *Bissau,* even showing higher frequency for *Forest* or *Savanna* chromosomal arrangements in Ivory Coast and Benin (della Torre et al., 2001; Gentile et al., 2001).

Interestingly, the reproductive isolation between M/S forms seems not to be universal, with several geographical areas in the extreme west of Africa, including Senegal, Gambia and Guinea Bissau, identified with high M/S hybridization rates (Caputo et al., 2008; Ndiath et al., 2008; Oliveira et al., 2008), most likely due to a breakdown in ecological post-mating barriers due to changing environmental conditions (Caputo et al., 2011; Nwakanma et al., 2013; Reidenbach et al., 2012). This geographic mosaic of reproductive isolation, with various hybridization rates between M/S forms in different areas, further complicates the relationship between M/S divergence and genetic differentiation due to gene flow (Marsden et al., 2011; Weetman et al., 2012).

Despite compelling evidence for the presence of two non-panmictic taxonomic units in *An. gambiae s.s.*, no fixed differences were initially found in DNA regions other than the ribosomal DNA regions used for their discrimination, which could be due to recent isolation that prevents significant divergence of non-coding DNA, the faster spreading of ribosomal DNA copy variants that are not necessarily causal to reproductive isolation or limited but ongoing gene flow (Gentile et al., 2001). Furthermore, gene expression differences between M/S forms seem to be limited but not confined to areas of genetic differentiation (Cassone et al., 2008; Turner et al., 2005). Indeed, restricted gene flow between the M/S forms has been documented (Taylor et al., 2001; Wondji et al., 2002), while indications for genetic differentiation between the M/S forms have also emerged in limited genomic regions of genetic divergence referred to as “speciation islands” (Stump et al., 2005; Turner et al., 2005; White et al., 2010; Wondji et al., 2002).

A recent whole genome sequencing effort of the M/S forms of *An. gambiae s.s.*, employing Sanger sequencing in laboratory colonies established from Mali, has suggested that M/S genetic divergence is widespread in the *An. gambiae s.s.* genome, indicative of an advanced state in the speciation process between the M/S forms (Lawniczak et al., 2010). Such widespread genetic divergence contradicts the model of ongoing speciation with gene flow, suggesting that genetic barriers between the M/S *An. gambiae s.s.* forms are much more advanced, thus severely restricting gene flow. Based on the identification of about 400,000 single nucleotide polymorphisms (SNPs) in the *An. gambiae s.s.* genome derived from the same sequencing effort (Lawniczak et al., 2010), an array-based SNP genotyping approach confirmed the degree of genetic divergence between the M/S *An. gambiae s.s.* forms using field-collected mosquitoes from Mali, thus ensuring that the observed genetic divergence is not a result of inbreeding within the laboratory colonies used for sequencing the M/S forms (Neafsey et al., 2010). Further SNP genotyping of M/S *An. gambiae s.s.* forms collected from West and Central Africa, using the

same array, also confirmed the widespread divergence observed in the previous studies, further suggesting that the M/S *An. gambiae s.s.* forms are collectively evolving over a broad geographic area in Africa, although postmating barriers that account for the M/S genetic isolation may vary, depending on local ecological or environmental conditions (Reidenbach et al., 2012). Indeed, SNP genotyping of M/S *An. gambiae s.s.* forms, using the same array, derived from sites in West Africa with high proportions of M/S hybrid forms, indicated that genetic divergence was limited to differentiation in the pericentromeric region of the X chromosome, used for molecular form discrimination, suggesting asymmetric introgression and backcrossing, most likely due to a breakdown in the postmating genetic barriers between the two molecular forms (Nwakanma et al., 2013).

In any case, the identification of loci contributing to the M/S incipient speciation process remains a subject of ongoing research efforts. Nevertheless, the observed level of genetic divergence between the M/S *An. gambiae s.s.* forms indicates that the identification of such loci contributing to the speciation process would be extremely difficult (Lawniczak et al., 2010; Neafsey et al., 2010). The conclusion to be drawn from the aforementioned intricacies of the population structure of *An. gambiae s.s.* is that the M/S differentiation of the most important malaria vector has created barriers to gene flow, in an incipient speciation process (della Torre et al., 2002). This differentiation has ecological repercussions that may influence malaria transmission dynamics. The M/S forms contrast significantly in the way they exploit limiting resources, with the M form showing closest association with domestic environments created by human activities, such as larval habitats occurring as a result of agricultural activities, while the most likely ancestral S form shows higher dependency on rain-dependent temporary breeding sites (Caputo et al., 2008; Toure et al., 1994). Therefore, the M/S incipient speciation process may be a result of an adaptive process that enables *An. gambiae s.s.* to exploit new opportunities that extend the malaria transmission potential in space and time while this process limits detrimental within-species competition (Costantini et al., 2009; della Torre et al., 2002).

Furthermore, genetic barriers within *An. gambiae s.s.* populations and the possibility of incipient speciation in the main malaria vector raise concerns regarding vector control efforts through the use of insecticides. Differentiation between M/S forms has been observed in the *kdr* locus, encoding a sodium channel important for pyrethroid resistance (Donnelly et al., 2009), with *kdr* genetic variants exhibiting divergence between the *An. gambiae s.s.* M/S forms (Diabate et al., 2003; Weill et al., 2000). Such differentiation suggests the occurrence of a recent selective sweep in M form mosquitoes following introgression from S form populations (Diabate et al., 2003; Weill et al., 2000). What is more, genes

encoding cuticular proteins implicated in insecticide resistance also show differentiation between M/S *An. gambiae s.s.* forms, possibly by exerting dual functionality in desiccation tolerance, in addition to insecticide resistance (Vannini et al., 2014). What these data suggest is that insecticide efficacy may be variable between *An. gambiae s.s.* populations exhibiting some level of genetic isolation, while the M/S form incipient speciation process may also drive the propagation of insecticide resistance phenotypes. As for other forms of vector control, the intricate *An. gambiae s.s.* population structure greatly complicates efforts to tackle insecticide resistance.

It has been recently proposed that the *An. gambiae s.s.* M form should be regarded as a separate *An. gambiae* complex species, termed *An. coluzzii* (Coetzee M, 2013). Although such taxonomy was not followed here, as the *An. gambiae* colony used here belongs to the M form, the *An. gambiae* species used in this study would be considered to be *An. coluzzii*.

The population structure of *Anopheles* mosquitoes and, more specifically, of the *An. gambiae s.s.* species, illustrates the ability of mosquitoes to genetically diversify and exploit this genetic variation in adaptive processes related to the environmental conditions they encounter. This striking proclivity of mosquitoes to rapidly evolve and adapt forms the basis of the hypothesis that mosquito genetic variation can be driven by mosquito gut microbiota in adaptive processes against symbiotic but also potentially pathogenic gut bacteria, as discussed later on.

1.13 Innovative approaches for vector control

As was made clear in previous sections, current strategies to control malaria are not only insufficient but also prone to disruptions such as development of drug or insecticide resistance that could result in a future surge of malaria-related mortality. It was also made clear that past interventions aiming at malaria eradication relied on vector control rather than therapeutic or preventive interventions of affected individuals. Therefore, the development of innovative vector control strategies that employ biological insights of vector/parasite interactions has been the focus of intense research interest.

A major drive to infuse state of the art genetics, molecular biology and technology into vector biology, in an effort to develop innovative tools for vector control, had been the establishment in 1989 of the Vector Biology Network by the MacArthur Foundation (Beaty et al., 2009) (Figure 7.11). This collaborative research network brought together until then disparate disciplines, from tropical medicine and parasitology to medical entomology, ecology, epidemiology or molecular biology of non-vector systems

in an effort to elucidate molecular and behavioural events that lead to disease transmission but also develop proof of concept demonstrations of interventions that could disrupt such transmission. Specific focus areas included the development of molecular and genomic approaches in Vector Biology, the genetic transformation of insect vectors as well as the study of vector immune systems in an effort to elucidate the molecular and biological basis of vector competence. The progress made since then in these areas has been remarkable.

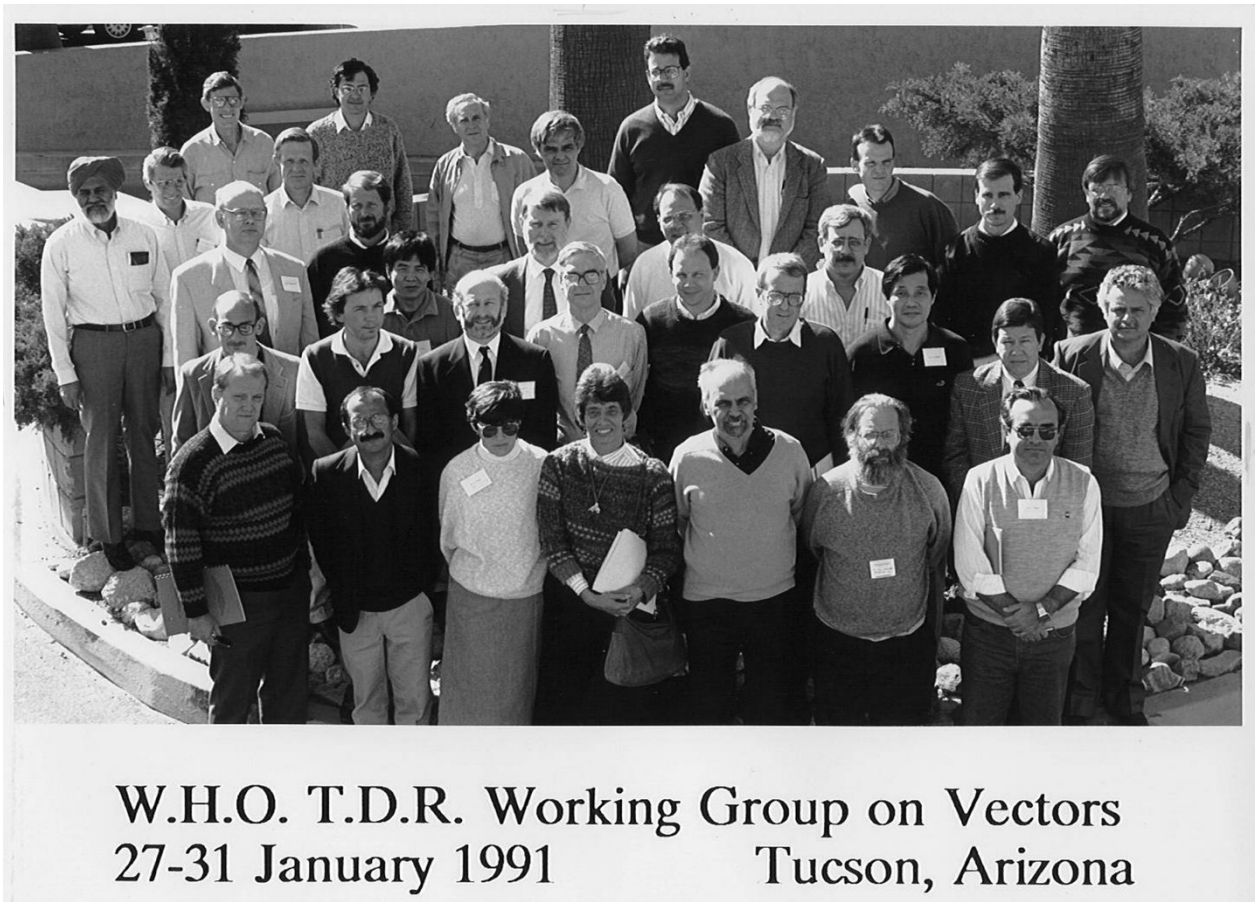


Figure 1.11: The WHO Tropical Diseases Training working group in Tucson, Arizona, USA, January 1991. Participants of the Tucson, Arizona meeting "Prospects for Malaria Control by Genetic Manipulation of its Vectors" held January 27-31, 1991. Adapted from: (Beaty et al., 2009).

1.14 A major bottleneck in the *Plasmodium* life cycle

The part of the *Plasmodium* life cycle that is spent in the mosquito midgut is undoubtedly the most vulnerable stage for the parasite and thus the ideal target for interventions aiming to disrupt malaria transmission. The period of about 24 hours between ingestion of *Plasmodium* gametocytes along with the mosquito blood meal, sexual development in the mosquito gut into motile ookinetes, invasion of the mosquito midgut epithelium by *Plasmodium* ookinetes and oocyst development in the extracellular space between the midgut epithelium and the basal lamina, constitutes a major bottleneck in the parasite life cycle (Ghosh et al., 2000) (Figure 1.12).

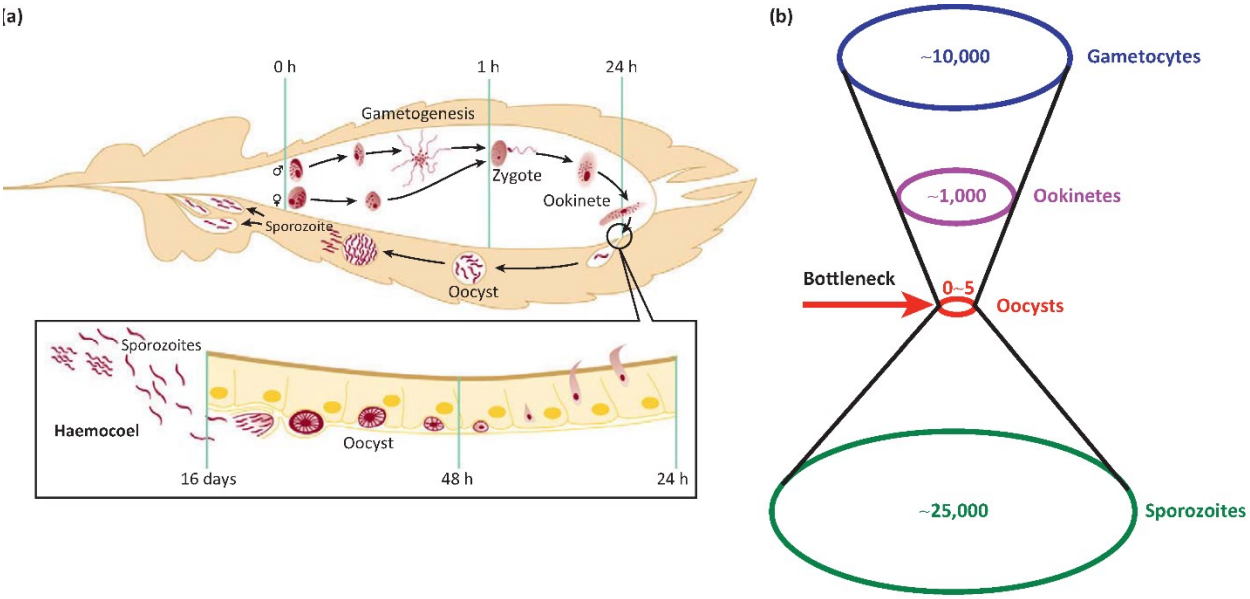


Figure 1.12: Crossing the mosquito midgut epithelium constitutes a major bottleneck in the *Plasmodium* life cycle. (a): The mosquito stage of the *Plasmodium* life cycle includes crossing the mosquito midgut epithelium by banana-shaped motile ookinetes about 24 hours after parasite entry in the mosquito gut through a blood meal. **(b):** The *Plasmodium* parasite sustains heavy losses while crossing the mosquito midgut epithelium, mainly due to mosquito immune responses. From about 10,000 *Plasmodium* gametocytes entering the mosquito through a blood meal, about 1,000 ookinetes develop. Out of these ookinetes, a strikingly low number of oocysts develop in the space between the midgut epithelium and the basal lamina in natural infections, typically up to 5 oocysts. Remarkably, in some mosquitoes sustaining a *Plasmodium* infection, no oocysts develop at all, rendering them refractory to malaria transmission. On the contrary, even a single developed oocyst is sufficient for the mosquito to transmit malaria. When developed oocysts burst several days after the infectious blood meal, typically thousands of sporozoites are released and travel to the mosquito salivary glands, ready to infect a human host. Adapted from: (Wang and Jacobs-Lorena, 2013).

During this stage, the parasite incurs heavy losses, with few or no developed oocysts in natural infections (Alavi et al., 2003). Indeed, the parasite faces a highly hostile environment in the mosquito midgut with regard to the temperature and pH of the midgut, the action of proteolytic enzymes involved in blood digestion and especially the formation of a chitinous peritrophic matrix that separates the blood meal from the midgut epithelium (Beerntsen et al., 2000). Elimination of invading parasites is mainly accomplished through mosquito immune responses (Yassine and Osta, 2010), although induced apoptosis (Han et al., 2000) and actin polymerization of midgut epithelial cells seem to also play such role (Vlachou et al., 2005). The parasite has developed remarkable adaptations for survival and sexual reproduction in such environment that allow, at least in some cases, successful invasion of the midgut epithelium and oocyst development, including exflagellation of the male gametocyte triggered by temperature and mosquito-derived cues, production of chitinases that aid penetration of the peritrophic matrix (Siden-Kiamos and Louis, 2004; Vlachou et al., 2006) as well as mechanisms that allow evasion of the mosquito immune system (Molina-Cruz et al., 2013).

The outcome of *Plasmodium* infections, in the context of the number of developed oocysts, if any, following an infectious blood meal, led to the introduction of the concept of mosquito susceptibility and refractoriness to malaria transmission (Sinden et al., 2004). The existence of both susceptible and refractory mosquitoes is thought to be the result of a dynamic relationship between vector and parasite in an equilibrium that allows a sustainable evolutionary cost of anti-*Plasmodium* responses for the host and prevents fixation of refractoriness in the mosquito population (Sinden et al., 2004). Indeed, refractoriness to malaria transmission, related to immune or non-immune phenotypic traits, seems to be multigenic and quantitative in nature (Beerntsen et al., 2000). Although the genetic basis of vector competence was demonstrated as early as 1929 (Beerntsen et al., 2000; Huff, 1929), it still remains a subject of intense research interest (Blandin et al., 2009; Harris et al., 2010a; Riehle et al., 2006) and will be further addressed later on. The complexity of vector/parasite interactions that lead to susceptibility or refractoriness to malaria transmission is further attested by the effect the combination of different mosquito and parasite genotypes exert on the outcome of infection (Lambrechts et al., 2005), with interactions between mosquito (Riehle et al., 2006) and parasite (Molina-Cruz et al., 2013) genetic variation ultimately deciding the observed susceptible or refractory phenotype (Lambrechts et al., 2006).

1.15 Applied high-tech interventions to disrupt malaria transmission

The existence of a major bottleneck in the mosquito stage of the *Plasmodium* life cycle, with *An. gambiae* mosquitoes exhibiting remarkable variation in their vectorial capacity, raised the possibility that this bottleneck could be exploited in efforts to disrupt malaria transmission. As *An. gambiae* mosquitoes, belonging to populations generally susceptible to malaria transmission, are naturally found to be refractory, it could be possible that isolated refractory strains could subsequently replace susceptible populations. Furthermore, the elucidation of the mechanisms that lead to refractoriness could be utilized either for the genetic manipulation of mosquito populations to render them refractory to malaria transmission and their subsequent release or for the production of anti-malarial substances that in essence render mosquitoes unable to transmit malaria without killing them and thus exerting evolutionary pressure for the development of resistance as conventional insecticides. Another possibility would be the development of transmission blocking vaccines that, through mosquito bites of infected individuals, could deliver such substances to mosquito vectors, rendering them refractory to malaria transmission.

In a seminal study by Collins and colleagues (Collins et al., 1986), the L3-5 strain of *An. gambiae* was isolated, which showed complete refractoriness against the simian malaria parasite *P. cynomolgi* but also some degree of refractoriness against other *Plasmodium* parasites, including *P. falciparum*. Refractoriness of the L3-5 strain was shown to have a genetic basis, related to melanotic encapsulation of parasites due to a chronic state of oxidative stress (Kumar, 2003; Zheng et al., 1997; Zheng et al., 2003). *An. gambiae* mosquitoes were also selected, employing a *P. gallinaceum* ookinete killing mechanism that rendered them refractory to infection (Vernick et al., 1995).

Despite the existence of naturally occurring refractory mosquitoes, an applied intervention cannot rely on such strains, mainly due to reduced fitness, but rather on engineered refractory phenotypes on mosquitoes which, following their release in the field, can spread refractory genes and, by increasing the frequency of parasite resistance, can reduce malaria transmission dynamics (Alphey et al., 2002; James, 2002). The genetic transformation of mosquitoes has been achieved both for the yellow fever vector *Aedes aegypti* (Coates et al., 1998; Jasinskiene et al., 1998) and for anopheline mosquitoes, including *An. stephensi* (Catteruccia et al., 2000) and *An. gambiae* (Grossman et al., 2001), using transposable element approaches. What is more, the expression of antiparasitic genes or single-chain antibodies that target *Plasmodium* in transgenic *An. stephensi* has been shown to impair *Plasmodium* transmission (Isaacs et al., 2012; Isaacs et al., 2011; Ito et al., 2002).

Despite these advances, several challenges hinder the application of such transgenic approaches in efforts to successfully reduce or eliminate malaria transmission in such way (Knols et al., 2007; Riehle et al., 2003). These include the identification of novel effectors that will evade parasite resistance after introduction in wild populations (Moreira et al., 2002). The identification of effector genes whose transgenic expression can sustainably disrupt disease transmission requires an in depth understanding of vector/parasite interactions and especially mosquito immune responses. Effector genes could target parasites, such as single-chain fragments acting as parasite ligands, or parasite-expressed proteins, but could also target mosquito tissues, such as mosquito receptors necessary for parasite invasion (Nirmala and James, 2003). Furthermore, transgenic expression of immune-related genes in *Ae. Aegypti* (Kazura et al., 2011), *An. stephensi* (Corby-Harris et al., 2010; Dong et al., 2012; Dong et al., 2011) but also *An. gambiae* (Kim et al., 2004) has shown that the approach of genetic transformation of mosquitoes with altered immunity can be a viable approach for disease control. Another possibility is the engineering of sterile male mosquitoes that, following their release, can result in the suppression or eradication of field populations (Catteruccia et al., 2009; Thomas et al., 2000).

Even if such refractory or sterile mosquito strains are engineered or selected, a major hurdle is their introduction in wild populations, as the transposable elements used for mosquito transformation are inefficient as drive systems while random integration of the effector elements can lead to reduced fitness (Cirimotich et al., 2011a; Fraser, 2012; O'Brochta et al., 2003). The replacement of populations would require no fitness cost for the released mosquitoes, which seems not to be the case for some transgenic mosquitoes (Catteruccia et al., 2003), although other transgenics show even a fitness advantage (Marrelli et al., 2007). Furthermore, the concomitant reduction of the resident population and extremely complicated logistics due to the number of mosquitoes that would need to be released, especially in vast and under-developed sub-Saharan regions (Moreira et al., 2002; Riehle et al., 2003), render population replacement in a large scale unlikely. Especially for *An. gambiae*, the major malaria vector in sub-Saharan Africa, its intricate population structure and incipient speciation events related to the M/S molecular forms of *An. gambiae s.s*.(Lawniczak et al., 2010; Neafsey et al., 2010), greatly complicate population replacement efforts based on transgenic modification approaches. Therefore, the feasibility of such interventions will most likely require genetic drive systems, in which the effector genes can spread into the population. Several such methods are in development and rely on selfish genetic elements such as homing endonuclease genes that induce chromosome double strand breaks in targeted DNA sequences and repair the chromosome while copying the effector element in the broken chromosome (Windbichler

et al., 2011) or maternal-effect lethal elements that cause lethality in offspring that do not bear the effector element (Akbari et al., 2013; Chen et al., 2007).

It has to be also noted that several avenues of research aim at disruption of malaria transmission through genetic modification of mosquito symbionts instead of mosquitoes themselves. Such efforts include the delivery of anti-*Plasmodium* genes through symbiotic bacteria (Wang and Jacobs-Lorena, 2013). This paratransgenesis approach relies on the expression and secretion of anti-*Plasmodium* effector molecules by symbiotic bacteria that persist through the population (Wang et al., 2012). A promising candidate for such paratransgenic approach is the *Asaia* symbiont, which not only colonizes efficiently *An. gambiae* mosquitoes but can also be vertically transmitted in the mosquito population (Damiani et al., 2010; Damiani et al., 2008; Favia et al., 2008). Transgenic fungi also seem to be very effective in delivering anti-*Plasmodium* effectors (Fang et al., 2011).

Finally, promising efforts to disrupt disease transmission have focused on mosquitoes carrying the intracellular, maternally transmitted, Gram-negative bacterium *Wolbachia*. *Wolbachia* can manipulate the host reproductive system through cytoplasmic incompatibility, as infected males cannot successfully mate with uninfected females and can thus spread readily within the population (Werren et al., 2008). Although not a natural symbiont, certain *Wolbachia* strains have been stably introduced and spread in *Ae. Aegypti* populations, in which they reduced the mosquito lifespan (McMeniman et al., 2009; Xi et al., 2005). A reduced lifespan by itself can reduce disease transmission, but, remarkably, *Wolbachia* can also limit infection of *Ae. aegypti* with various pathogens including *Plasmodium* (Hoffmann et al., 2011; Kambris et al., 2009; Moreira et al., 2009), most likely through induction of oxidative stress (Pan et al., 2011). Indeed, release of *Ae. aegypti* mosquitoes in semi-field conditions has been shown to block transmission of dengue virus, the yellow fever disease agent. Transient infections indicate that *Wolbachia* strains can have similar inhibitory effects against *Plasmodium* in *An. gambiae* (Hughes et al., 2012; Hughes et al., 2011; Kambris et al., 2010). *Wolbachia* is not a natural symbiont of *An. gambiae* either, and an inherited infection has yet to be achieved. The recent stable introduction, though, of *Wolbachia* in *An. stephensi* populations, which renders them refractory to *Plasmodium* infection, raises hopes that a similar introduction in *An. gambiae* can lead to an effective intervention against malaria transmission (Bian et al., 2013).

Despite many promising advances, most efforts to develop innovative vector control applications have yet to bear fruit. Such interventions have the potential to supplement current approaches, especially when current interventions may fail due to e.g. the emergence of insecticide resistance, but, more

importantly, can have a striking impact towards malaria eradication. What is made clear from the aforementioned approaches at vector control is that a deep and detailed understanding of the basic biology and behaviour of the mosquito vector and its interaction with pathogens and symbionts is vital in efforts to successfully develop such interventions.

1.16 The mosquito genomics era

The initiative to elucidate the biology of vector/parasite interactions in a molecular and genetic context required from the outset the adoption of a genomic approach, if any significant breakthrough was to be achieved (Beaty et al., 2009). Such approach required the sequencing of the *An. gambiae* genome, which was accomplished in 2002 (Holt et al., 2002). The significance of the availability and further bioinformatic analysis of the *An. gambiae* genome, combined with the availability of the *Drosophila* genome, accomplished in 2000 (Adams et al., 2000) and other sequencing efforts later on, including the sequencing of the *Ae. aegypti* and *Culex quinquefasciatus* mosquito genomes, the sequencing and comparison of 12 *Drosophila* genomes and the sequencing of the M/S forms of *An. gambiae* (Arensburger et al., 2010; *Drosophila* 12 Genomes et al., 2007; Lawniczak et al., 2010; Nene et al., 2007; Stark et al., 2007) will be evident throughout this study and will be briefly analyzed here.

1.17 From sequence to genes to function

The essence of a genomics approach relies on the premise that during the evolutionary process, core gene/protein features are conserved while others diversify, expand or contract. Therefore, a multi-species comparative analysis of genomic sequences can shed light into the evolutionary history of genes and thus derive genomic, molecular and biological information (Rolff, 2009).

The extraction of biological information from a genome sequence commences with the discovery of functional encoded elements, mainly protein-coding genes but also non-protein elements, e.g. RNA or regulatory elements (Lubec et al., 2005). Such discovery relies on *ab-initio* gene finding algorithms based on sequence homologies (Holt et al., 2002) or the comparison of multiple closely related genomes and the identification of nucleotide conservation or evolutionary signatures associated with functional elements (Stark et al., 2007). This approach is best exemplified in *Drosophila*, which combines the availability of 12 closely related genomes (*Drosophila* 12 Genomes et al., 2007) with extensive experimental information based on decades of research in this genetically tractable model system of

fundamental importance in the immunity and development research fields (Stark et al., 2007). Such combination allows validation of the systematic genomic annotation with manual annotation based on experimental data. The ongoing effort for the sequencing of 17 *Anopheles* genomes could facilitate the improvement of the *An. gambiae* gene annotation in a similar to *Drosophila* manner (Neafsey et al., 2013).

A major challenge following the identification of predicted gene products is the inference of their putative function based on comparative analysis of protein sequences, again assuming that due to a common evolutionary history, similar sequences most likely reflect similar functions. The identification of both protein-coding sequences and homologies with domains of known function relies on sequence similarity searches of a query sequence against a sequence database. One of the most widely used alignment algorithms used for such purposes is the BLAST programme, used for both nucleotide and protein queries (Altschul et al., 1990; Altschul et al., 1997). The inferred function of predicted proteins relies either on sequence similarity with proteins of known function through functional experimental information or on protein signatures, in which alignment of multiple protein sequences has resulted either on a consensus sequence or on conserved domains or residues which might represent conserved active sites or binding sites related to a common function (Mulder et al., 2008). The *Interpro* database integrates and codifies various protein signature databases, in which a query protein sequence can be assigned based on a similarity score to an *Interpro* entry, describing the protein family or domain the sequence shows similarity to, thus predicting the function of the query protein sequence (Mulder et al., 2003; Zdobnov and Apweiler, 2001).

Another way to codify functional information inferred for a gene product is the use of Gene Ontology (GO) terms (Ashburner et al., 2000). GO terms assigned to a gene product functionally describe it in terms of inferred biological processes it participates, cellular components in which it can be expressed and molecular functions it putatively performs. Both GO terms and *Interpro* IDs provide a structured way to assign putative functional information on a hypothetical gene or protein product irrespective of the organism it belongs to or any previous experimental information.

1.18 Homologues, orthologues, paralogues and gene families

The comparison of gene sequences between genomes of different organisms and a subsequent phylogenetic analysis can reveal the evolutionary history of each gene and thus establish homologies between genes in the same or different genomes (Rolff, 2009). Homologies generally refer to genes that share a common origin in an ancestral genome. The basic concept of homology includes orthologues,

referring to genes derived from a single ancestral gene through a speciation process and paralogues, referring to genes derived from a duplication event (Koonin, 2005). Although the concept of homology shows no requirement for a common or similar biological function, it is generally accepted that homologous genes typically perform equivalent functions, although it is entirely possible that a non-homologous gene can be recruited to perform an equivalent function in another organism and thus deviate from typical functions of its own homologues (Koonin, 2005).

Comparison of genomes in closely related organisms can define one-to-one orthologues, with a clear phylogenetic relationship between two genes in different organisms and a high likelihood of conserved function, although it is more typical to define orthologous groups, referring to all descendants of a common ancestral gene (Waterhouse et al., 2013). Orthologous groups can be similar but conceptually distinct from the definition of gene families. Gene families can be based on sequence similarity and thus homology but can also be more narrowly defined, based on the presence of a protein domain or the combination of domains. As in orthologous groups, gene families suggest similarity in function for its members although functional deviation within each family is entirely possible (Rolff, 2009).

The comparison of the genome and proteome between *An. gambiae* and *D. melanogaster*, two dipteran species that diverged 250 million years ago, has shown 47% of *An. gambiae* genes with a one-to-one orthologue in *D. melanogaster* and 56% sequence identity, mainly in genes related to basic biological processes that most likely conserved their function (Zdobnov et al., 2002). Nevertheless, considerable expansions/contractions can be observed in orthologous groups and defined gene families, most likely reflecting adaptation to different ecologies, such as the adoption of haematophagy in the mosquito, with an expansion of fibrinogen-related proteins in *An. gambiae* being the most obvious example (Christophides et al., 2002; Zdobnov et al., 2002). As *Drosophila* remains one the most thoroughly studied organisms, the identification of these orthologous relationships in *An. gambiae* genes with functionally annotated *Drosophila* counterparts, forms the basis of prioritization of candidate genes for experimental testing in various aspects of *An. gambiae* physiology, including immunity (Rolff, 2009).

1.19 The mosquito innate immune system

To sustain attacks from various pathogens it encounters during its life cycle, the mosquito has developed an intricate innate immune system, able to recognize and respond to pathogens but also modulate these responses through homeostatic mechanisms that reach an optimal cost-benefit balance for the mosquito. The availability of the *D. melanogaster* genome (Adams et al., 2000), along with the availability of the *An.*

gambie (Holt et al., 2002) and, later on, the *Ae. aegypti* genomes (Nene et al., 2007), allowed for comparative genomic analyses of these insect immune systems that, based on the prior extensive study of innate immunity in *Drosophila*, led to a detailed mapping of the *An. gambiae* innate immune system (Christophides et al., 2002; Waterhouse et al., 2007).

An initial comparative genomic analysis between the *An. gambiae* and *D. melanogaster* innate immune repertoires led to the identification of 242 *An. gambiae* genes implicated in immune responses, over 18 gene families (Christophides et al., 2002). Further refinement of this analysis based on the availability of the *Ae. aegypti* genome extended the number of *An. gambiae* immune-related genes to 338, over 31 gene families (Waterhouse et al., 2007). A detailed description of these immune-related genes has been previously reviewed (Christophides et al., 2004; Osta et al., 2004b; Yassine and Osta, 2010) and is also available at the ImmunoDB database (http://cegg.unige.ch/Insecta/immunodb).

The *An. gambiae* innate immune system shows considerable diversification compared to the *D. melanogaster* immune system (Christophides et al., 2002). Therefore, there is considerable gene expansion and a deficit of orthologues: there are 91 orthologous trios and 57 pairs between the *An. gambiae*, *Ae. aegypti* and *D. melanogaster* immune repertoires, with some families, including apoptosis inhibitors, oxidative defence enzymes and scavenger receptors, exhibiting mainly orthologue trios, some highly divergent functional groups, including the antimicrobial peptide family, exhibiting no orthologous trios and some families, including the C-type lectins, exhibiting an intermediate level of divergence with large expansions but also conserved orthologous trios or pairs (Waterhouse et al., 2007) (Figure 1.13). Therefore, different modes of evolutionary pressure on different parts of the innate immune system equip it with flexibility to adapt to diverse selective challenges (Waterhouse et al., 2007).

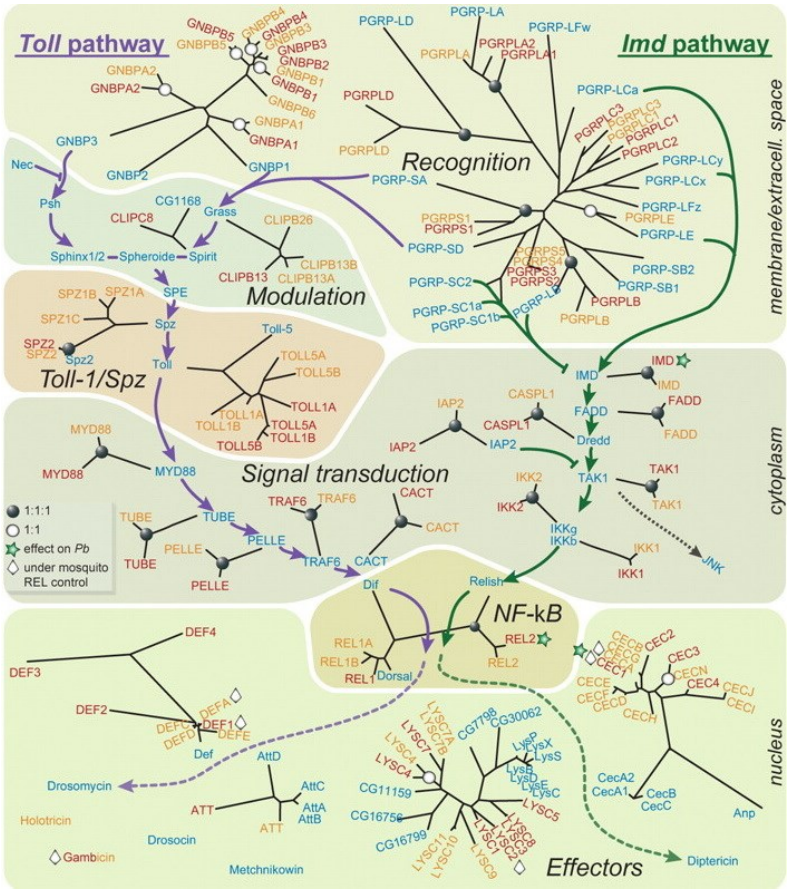


Figure 1.13: Diversification of the *Drosophila* and mosquito innate immune systems. Genes and gene families implicated in the Toll and Imd immune signalling pathways in *Drosophila* (blue colour), *An. gambiae* (red colour) or *Ae. aegypti* (yellow colour). Orthologous trios or mosquito 1:1 orthologues are indicated by solid or open circles, respectively. Adapted from: (Waterhouse et al., 2007).

1.20 Functional groups of *An. gambiae* immune genes

Innate immune genes in *An. gambiae* are generally categorized under three functional groups: pattern recognition, signal transduction and effector genes (Christophides et al., 2002). Pattern recognition genes are related to the recognition of infectious non-self by various receptors and triggering of signalling pathways that lead to mounting an immune response. These include peptidoglycan recognition proteins, with PGRPLC known to recognize DAP-type peptidoglycan and trigger the IMD/REL2 pathway (Meister et al., 2009) but also thioester-containing proteins, including TEP1, a member of a complement-like system that binds on the surface of *Plasmodium* parasites (Blandin et al., 2004; Fraiture et al., 2009; Povelones et al., 2009), Gram-negative binding proteins (GNBPs) (Warr et al., 2008), fibrinogen-like proteins (FREPs) (Dong and Dimopoulos, 2009), carbohydrate-binding C-type lectins (Osta et al., 2004a), galectins (Kamhawi et al., 2004) and MD2-like receptors, which recognize lipids (Osta et al., 2004b).

Signal transduction involves the modulation and amplification of signals originating from a pattern recognition molecule that culminates in the activation of effector mechanisms against pathogens. Such amplification in *An. gambiae* is mainly accomplished by CLIP-type serine proteases and corresponding serine protease inhibitors (serpins) that modulate the transduced signal (Christophides et al., 2004; Volz et al., 2006; Volz et al., 2005). Finally, effector genes include genes encoding anti-microbial peptides but also prophenoloxidases (PPOs), which, following their activation by CLIP-type serine proteases to active phenoloxidases, catalyze the synthesis of melanin, necessary for the melanotic encapsulation of pathogens during the melanization reaction, an insect killing mechanism against various pathogens (Barillas-Mury, 2007; Christophides et al., 2004; Paskewitz et al., 2006).

1.21 Immune pathways in *An. gambiae*

Comparative genomic analysis between mosquitoes and Drosophila has also identified several conserved immune pathways in *An. gambiae* (Christophides et al., 2002; Waterhouse et al., 2007). The Toll pathway is strikingly conserved in *An. gambiae*, probably due to participation in developmental processes (Christophides et al., 2002), although phylogenetic relations between the *An. gambiae* and *D. melanogaster* Toll pathway members are intricate, with some members having clear orthologues and others, including six mosquito Toll receptors, being a result of recent expansions (Christophides et al., 2004). *REL1* is the mosquito homologue of the *Drosophila Dorsal*, a transcription factor activated by the Toll pathway. *REL1* has been implicated in anti-fungal responses in *Ae. aegypti* (Shin et al., 2005) but also responses against dengue virus (Ramirez and Dimopoulos, 2010; Xi et al., 2008) and possibly bacteria, as

Ae. aegypti REL1 overexpression regulates a subset of IMD-controlled genes (Kazura et al., 2011). The REL1 pathway has also been implicated in responses against the rodent malaria parasite *P. berghei* in *An. gambiae* (Frolet et al., 2006; Garver et al., 2009; Riehle et al., 2008).

The IMD pathway is also conserved between *An. gambiae* and *D. melanogaster*, with *REL2* being the orthologue of the *Drosophila Relish* (Christophides et al., 2004). REL2, and thus the IMD pathway, is involved in defence against both Gram-negative and Gram-positive bacteria (Meister et al., 2005). REL2 also plays a central role in defence against the human malaria parasite *P. falciparum* (Dong et al., 2011; Frolet et al., 2006; Garver et al., 2012; Garver et al., 2009). The main receptor for the IMD/REL2 pathway is the peptidoglycan recognition protein PGRPLC, which is also important for defence against bacteria and *P. falciparum* (Christophides et al., 2002; Lin et al., 2007; Meister et al., 2009).

Finally, the JAK/STAT pathway is also present in *An. gambiae*, with two STAT genes involved in responses against bacteria and *Plasmodium* (Christophides et al., 2004; Gupta et al., 2009*)*. The JAK/STAT pathway also plays a central role in anti-*Plasmodium* and anti-dengue viral defence in *Ae. aegypti* (Kazura et al., 2011; Souza-Neto et al., 2009).

1.22 Analysis of *An. gambiae* immunity through RNAi and microarrays

The study of the *An. gambiae* innate immune system was greatly facilitated by two further advances: the RNAi-mediated silencing of *An. gambiae* genes and the expression analysis of the *An. gambiae* transcriptome using DNA microarrays. Although transgenesis has been achieved in *An. gambiae* (Grossman et al., 2001), limitations related both to the difficulty of achieving *An. gambiae* with specific gene knock-outs and to the maintenance of such transgenic lines, preclude the routine use of such screens, as they are used in *Drosophila* (Cronin et al., 2009; Duffy, 2002; Piccin et al., 2001). Therefore, *in vivo* functional analysis of candidate genes in *An. gambiae* routinely utilizes a double stranded RNA (dsRNA) knockdown system, in which dsRNA targeting the gene transcript is intrathoracicly injected in an adult mosquito and, employing the mosquito RNAi machinery, results in silencing of the respective transcript (Blandin et al., 2002) (Figure 1.14). Despite its limitations, including transient, partial silencing without any tissue specificity, which is manually applied to each individual mosquito, for want of a more efficient knock-out approach, RNAi-mediated silencing has proven invaluable for the functional analysis of the mosquito immune system.

Figure 1.14: RNAi-mediated silencing of *An. gambiae* genes. Mosquitoes are microinjected in their abdomen with a fine needle containing dsRNA. Typically, 69 nl of a solution containing dsRNA at a 3 μg/μl concentration are injected in each instance. The injection is performed manually while observing the CO_2 anaesthetized mosquito under a stereoscope. The injected dsRNA triggers the mosquito's RNAi machinery that leads to silencing of the targeted mosquito transcript. Adapted from the Kafatos/Christophides laboratory web page, http://openwetware.org/wiki/Kafatos:Kafatos/Christophides_Lab.

The use of DNA microarrays in the study of *An. gambiae* innate immunity greatly facilitated the identification of genes which, by showing a significant up or down regulation following an immune challenge, were very likely to play pivotal roles in such immune responses. Since the first expression analysis in *An. gambiae*, using microarrays derived from expressed sequence tag collections (Dimopoulos et al., 2000; Dimopoulos et al., 2002), microarrays have been extensively used to identify differential expression following *Plasmodium* infection (Dong et al., 2006a; Mendes et al., 2011) or related to bacterial infections (Baton et al., 2009; Dong et al., 2009), while, currently (release VB-2013-04), 58 microarray experiments in *An. gambiae* or other mosquitoes, have been deposited to Vectorbase, including transcription profiling following pathogen challenge, in different development stages, following blood feeding, in insecticide resistant or susceptible mosquitoes, between male or female mosquitoes or between M/S forms of *An. gambiae.*

The combination of comparative genomic analyses of the *An. gambiae* innate immune system with the availability of expression analyses following a pathogen challenge and RNAi-mediated silencing led to the elucidation of many aspects of *An. gambiae* innate immune responses against *Plasmodium* or bacteria. What emerged as of central importance to parasite killing was a complement-like response orchestrated by the thiosester-containing protein TEP1 and two leucine-rich repeat proteins, LRIM1 and APL1C. Leucine-rich repeat protein genes were initially identified to be potently upregulated following a bacterial or *Plasmodium* challenge (Dimopoulos et al., 2002). Silencing *LRIM1* resulted in a precipitous increase in developed *P. berghei* oocysts (Osta et al., 2004a), and the same effect was observed following *TEP1* silencing (Blandin et al., 2004). Although LRIM1 seemed to have a minor effect on *P. falciparum* infections (Aguilar et al., 2005), which has been shown to rely on the intensity of infection (Garver et al., 2012), TEP1 was shown to be involved in defences also against the human malaria parasite (Dong et al., 2006a). LRIM1 was further shown to form a high molecular weight complex with another leucine-rich repeat protein, APL1C and interact with TEP1, promoting its localization on the surface of *P. berghei* parasites, leading to their killing (Fraiture et al., 2009; Povelones et al., 2009).

1.23 Population genetics and *An. gambiae* immunity

In combination with the aforementioned approaches, another way of identifying candidate genes involved in *An. gambiae* immune responses has been the utilization of population genetics screens. Indeed, resistance to human malaria parasites in *An. gambiae* has been shown to have a genetic basis (Niare et al., 2002). The identification of the *APL1* involvement in anti-*Plasmodium* responses had been

based on its presence within a genomic region shown to regulate resistance to natural *P. falciparum* infections, by interrogating genetic variation between isofemale pedigrees isolated from field-collected mosquitoes (Riehle et al., 2006). Similarly, polymorphisms in the *TEP1* gene, which possesses resistant and susceptible alleles, were shown to explain much of the variation in the outcome of *P. berghei* infections (Blandin et al., 2009). Divergence in *TEP1* alleles between the M/S forms of *An. gambiae*, were also shown to correlate with resistance to *P. falciparum* (White et al., 2011). Therefore, genetic variation associated with the outcome of *Plasmodium* infections has been used to identify genes involved in anti-*Plasmodium* responses but also elucidate allele combinations that can explain phenotypic variation tightly linked with the mosquito ability to transmit malaria.

1.24 Antibacterial and anti-*Plasmodium* immunity

Early on in the study of *An. gambiae* immune responses, it became clear that such responses profoundly influence the outcome of *Plasmodium* infections. Silencing of immune genes such as *LRIM1, APL1* or *TEP1* profusely increased the number of developed *Plasmodium* oocysts while silencing of the C-type lectins *CTL4* and *CTLMA2* decreased the *Plasmodium* infection intensity, thus acting in an agonistic manner relative to *Plasmodium* development (Blandin et al., 2004; Osta et al., 2004a; Riehle et al., 2006). At the same time, though, that *An. gambiae* immune responses target malaria parasites, the mosquito also mounts responses aimed to contain bacteria that are either found in the mosquito gut, or, having crossed the midgut epithelium, are located in the mosquito haemolymph.

It also became clear early on that the antibacterial and anti-*Plasmodium* components of *An. gambiae* immunity are inextricably linked, with immune factors that despite being triggered by bacteria also target *Plasmodium*. Indeed, TEP1 was initially identified due to its involvement in phagocytosis of Gram-negative bacteria (Levashina et al., 2001). Furthermore, CTL4 and CTLMA2 were also shown to be involved in haemolymph defences against Gram-negative bacteria (Schnitger et al., 2009). This link is best exemplified, though, in immune responses related to the IMD/REL2 pathway. Although REL2 was shown to be involved in antibacterial responses against both Gram-negative and Gram-positive bacteria, the full length REL2-F, produced through alternative splicing, was shown to be involved in responses against the rodent malaria parasite *P. berghei* (Meister et al., 2005). What is more, the IMD/REL2 pathway was shown to regulate the expression of *LRIM1*, shown to have anti-*Plasmodium* effects (Meister et al., 2005; Osta et al., 2004a) and also be involved in the regulation of TEP1 levels (Frolet et al., 2006). Despite these initial reports for an effect of REL2 on *P. berghei* infection (Frolet et al., 2006; Meister et al., 2005), later studies

indicated that REL2 mainly affects the development of the human malaria parasite *P. falciparum* (Garver et al., 2012), while silencing of the major IMD/REL2 negative regulator *Caspar* in *An. gambiae* (Garver et al., 2009) or overexpression of REL2 in *An. stephensi* transgenic mosquitoes (Dong et al., 2011) decreased the *P. falciparum* infection intensity.

The only known receptor that triggers the IMD/REL2 pathway is PGRPLC. PGRPLC, both in *Drosophila* (termed as PGRP-LC) (Choe et al., 2002; Gottar et al., 2002) and *An. gambiae* (Meister et al., 2009), is triggered by DAP-type peptidoglycan (Kaneko et al., 2004), common to all Gram-negative bacteria, and leads to activation of the Imd pathway in *Drosophila* (Choe et al., 2005) and the respective IMD/REL2 pathway in *An. gambiae* (Lin et al., 2007). PGRPLC also affects the outcome of both *P. berghei* and *P. falciparum* infections, as *PGRPLC* silencing increases *Plasmodium* infection intensity (Meister et al., 2009). Intriguingly, though, the PGRPLC anti-*Plasmodium* effect is only observed in the presence of midgut bacteria, indicating that the activation of the IMD/REL2 pathway by Gram-negative bacteria through PGRPLC leads to an anti-*Plasmodium* response (Meister et al., 2009).

1.25 The mosquito gut microbiota

The mosquito gut is rife with commensal bacteria that, altogether with fungi and viruses also found in the mosquito gut, constitute what is defined as the mosquito gut microbiota. As with other insects and higher organisms, including humans, immense interest has recently developed into how the symbiotic relationship between host and microbiome has shaped the host and how this relationship affects many aspects of the host's physiology, and, more importantly, what role it plays during disease. The microbiota of insects such as *Drosophila* and *Anopheles* have a much simpler composition and can thus be used as models aiming to elucidate complex interactions also relevant for human physiology (Chandler et al., 2011; Dillon and Dillon, 2004; Douglas, 2011). The studies on co-evolution between the insect host and the resident microbiota have mostly focused on nutrient provision by the microbe or the host, the resistance to colonization by pathogenic to the host microbes due to the presence of beneficial microbiota and the influence the microbiota can exert on the host's behaviour (Dillon and Charnley, 2002; Dillon and Dillon, 2004; Ezenwa et al., 2012). Symbiotic microbiota in *Drosophila* have been shown to influence various homeostatic programmes, including developmental rate, body size, energy metabolism and intestinal stem cell activity (Shin et al., 2011; Storelli et al., 2011), while symbionts of the tsetse fly are required for the maturation and proper development of the host's immune system (Weiss et al., 2011a).

In mosquitoes, midgut bacteria have been shown to be involved in blood digestion and egg production in *Ae. aegypti* (Gaio Ade et al., 2011).

The composition of the mosquito gut microbiota has been extensively investigated, initially with culture-dependent methods (Demaio et al., 1996; Gonzalez-Ceron et al., 2003; Pumpuni et al., 1996), later on with PCR-based methods (Favia et al., 2007; Kampfer et al., 2006a; Kampfer et al., 2010; Kampfer et al., 2006b; Lindh et al., 2005; Rani et al., 2009; Terenius et al., 2008) and, more recently, using metagenomic profiling by high-throughput sequencing (Boissière et al., 2012; Osei-Poku et al., 2012; Wang et al., 2011b). The conclusion to be drawn by these studies is that the mosquito gut is inhabited mainly by Gram-negative enterobacteria, which show great diversity at the individual level at any given population as well as among different mosquito populations. Despite this diversity, though, there is a limited number of bacterial taxa represented at any individual mosquito midgut while the bacterial families with the highest ability to colonize the mosquito gut are also limited (Gendrin and Christophides, 2013).

1.26 The human host-gut microbiota mutualism

In humans, the gut microbiota comprise a subject of intense research interest (Ferreira and Veldhoen, 2012; Gordon, 2012; Mueller et al., 2012; Pennisi, 2010), pointed to many different directions, including the sequencing of the human microbiome (Muegge et al., 2011; Nelson et al., 2010; Proctor, 2011), how the microbiota have evolutionarily shaped the human adaptive immune system (Hooper et al., 2012; Lee and Mazmanian, 2010) as well as what is the microbiota involvement in autoimmune diseases (Mathis and Benoist, 2011; Wu et al., 2010) and many other pathological states that impact human health (Balter, 2012; Clemente et al., 2012; Curtis et al., 2011; Littman and Pamer, 2011; Plottel and Blaser, 2011; Virgin and Todd, 2011).

The symbiotic relationship derived from the interactions of the human host with gut-inhabiting bacteria is known to affect many aspects of human physiology, and, consequently, also influence many manifestations of pathology (Balter, 2012; Clemente et al., 2012; Leser and Molbak, 2009; Proctor, 2011). Mammalian gut microbiota are known to impact the development and function of the immune system (Hooper et al., 2012), with alterations in the symbiotic microbial communities shown to be linked to the manifestation of allergic or autoimmune disorders, such as autoimmune arthritis, type 1 diabetes or inflammatory bowel disease (Chung et al., 2012; Mathis and Benoist, 2011; Wen et al., 2008; Wu et al., 2010).

Human intestinal flora exhibit immense diversity but also inter-subject variability (Eckburg et al., 2005). Nevertheless, clusters of gut bacterial communities detected in human populations, referred to as enterotypes, suggest the existence of a limited number of well-balanced symbiotic states (Arumugam et al., 2011). In many cases, disease pathology is thought to be the result of a disturbance in such mutualistic balance of the gut microbial community, referred to as dysbiosis (Littman and Pamer, 2011). Many factors can contribute to such disease-related pathogenic shifts in gut microbial communities, including external factors such as antibiotic treatment that eliminates commensal bacteria involved in colonization resistance of pathogens (van de Leur et al., 1997; Vollaard et al., 1992), but also influences mediated by the host's inflammatory response (Hand et al., 2012; Stecher et al., 2007).

A major question that remains poorly resolved is how this microbial diversity is shaped. A complex interplay that involves both ecological and evolutionary forces is likely to determine the composition of the human gut microbial ecosystem (Dethlefsen et al., 2007; Ley et al., 2006). Many stochastic environmental factors are likely at play, including the host's diet, shown to severely modulate gut microbial communities (Faith et al., 2011; Muegge et al., 2011). At the same time, though, host genetic factors also shape the observed individual microbiome diversity (Benson et al., 2010; Spor et al., 2011).The human host severely limits the diversity of bacterial and archaeal taxa present in its gut, compared to other ecosystems, either due to the limited metabolic niches in the human gut or through active immune responses (Backhed et al., 2005; Costello et al., 2012). At the same time, though, the host is also evolutionarily shaped by the commensal bacterial population (Ganal et al., 2012; Lathrop et al., 2011). Resident bacteria are thought to have co-evolved with the human host into shaping a balance between regulatory and pro-inflammatory factors that determines the host's immune status (Lee and Mazmanian, 2010), with co-evolved host-specific microbial communities being critical for the proper maturation of the host's immune system (Chung et al., 2012).

1.27 Host-gut microbiota interactions in *Drosophila* and other model systems

The complexity of the mammalian gut microbial ecosystem, recently revealed through the use of high-throughput sequencing technologies (Costello et al., 2009; O'Hara and Shanahan, 2006; Spor et al., 2011), greatly complicates research efforts aiming for the elucidation of underlying mechanisms that shape the host-microbe mutualism. Therefore, the use of animal models with much simpler microbial communities, combined with available genetic tools for functional characterization of such underlying mechanisms, can greatly enhance understanding of host-gut microbiota interactions that could be translated to

interventions in human pathological conditions influenced by the gut flora. *Drosophila* is a model system extensively used in the study of antibacterial immunity (Lemaitre and Hoffmann, 2007) and its simple gut microbial ecosystem (Wong et al., 2013) facilitates gut microbiota studies (Erkosar et al., 2013).

Indeed, studies employing the *Drosophila* genetic model have revealed many insights regarding the interaction of gut microbial communities with the *Drosophila* host (Lee and Brey, 2013). Imd regulation through the homeobox transcription factor Caudal shapes the fly's gut bacterial population structure (Ryu et al., 2008). A specific *Acetobacter* strain in the fly's gut community has been shown to encode a pyrroloquinoline quinone-dependent alcohol dehydrogenase (PQQ-ADH) which, most likely through the production of bacterial-derived metabolites, is essential for many aspects of the host's development, including developmental rate, body size, energy metabolism and intestinal stem cell activity (Shin et al., 2011). The effect of PQQ-ADH, mediated by the host's insulin/insulin-like growth factor signalling pathway, entails that changes in the host's microbial structure profoundly influence developmental and metabolic homeostasis (Shin et al., 2011). Another member of the *Drosophila* gut microbial community, *Lactobacillus plantarum*, has been shown to promote larval growth upon nutrient scarcity through signalling of the host's TOR-dependent nutrient sensing pathway, most likely through bacterial-derived metabolites (Storelli et al., 2011).

A direct link between bacterial-derived metabolites and DUOX-dependent immune responses has been recently reported in *Drosophila* (Lee et al., 2013a). The DUOX pathway has been shown to be activated by bacterial-derived uracil, most likely by an unidentified G protein-coupled receptor (GPCR). Intriguingly, DUOX-mediated immune responses are triggered by uracil-producing opportunistic pathogens but not commensal bacteria (Lee et al., 2013a; Valanne and Ramet, 2013), suggestive of an evolutionary co-adaptation process, in which either the fly's immune system was adapted so as to recognize uracil produced by pathogenic bacteria that would affect survival, or commensal bacteria adapted to the fly's gut ecosystem and lost their uracil-producing ability, thus limiting the triggering of detrimental for them immune responses. Furthermore, as chronic DUOX activation by bacterial-derived uracil has been shown to cause gut pathology (Lee et al., 2013a), it is also possible that loss of uracil production capacity by commensal bacteria or colonization of the fly's gut by bacteria that do not produce uracil, may be a result of enhanced host fitness by limiting the oxidative burst elicited by the DUOX pathway activation.

The tsetse fly (*Glossina* sp.), the primary vector of *Trypanosoma brucei* parasites, harbours three maternally transmitted endosymbionts, *Wigglesworthia*, *Sodalis* and *Wolbachia* (Wang et al., 2013). Field tsetse fly populations harbour additional strains of commensal bacteria, including *S. marcescens* (Lindh

and Lehane, 2011). The tsetse fly larval gut microbiota have been shown to be essential for the maturation of the host's immune system (Weiss et al., 2011a), while they also modulate the integrity of host immune barriers at the larval stage in a way that influences the establishment of trypanosome parasite infections (Weiss et al., 2013). Indeed, midgut bacteria in sandflies (Schlein et al., 1985) and tsetse flies (Azambuja et al., 2004; Welburn and Maudlin, 1999) have been shown to influence parasite infections with *Leishmania* or trypanosomes, respectively (Azambuja et al., 2005).

1.28 The influence of mosquito gut bacteria on the *Plasmodium* infection outcome

The influence of the mosquito gut microbiota on the outcome of *Plasmodium* infection was known long before any mechanistic link could explain this influence. Co-infection of Gram-negative bacteria with *P. falciparum* in *An. stephensi* was shown to inhibit the parasite sporogonic development (Pumpuni et al., 1993; Pumpuni et al., 1996). The addition of the potent antimicrobial gentamicin was also shown to affect *P. falciparum* development, although it was not made clear whether gentamicin could directly influence *P. falciparum* development or its effect was due to its antibacterial potency (Beier et al., 1994). Co-infection of enterobacteria and *P. vivax* was also shown to block the parasite's sporogonic development in *An. albimanus* (Gonzalez-Ceron et al., 2003).

The growing body of evidence suggesting a direct or indirect anti-*Plasmodium* effect for the mosquito gut microbiota was confirmed by two studies which showed that elimination of the *An. gambiae* gut microbiota increased *Plasmodium* infection intensity while their re-introduction eliminated this effect (Dong et al., 2009; Meister et al., 2009). Furthermore, these studies suggested that the gut microbiota influence on the outcome of *Plasmodium* infections is indirect and is mediated by mosquito immune responses, with bacteria-induced immune factors exhibiting anti-*Plasmodium* effects (Dong et al., 2009; Meister et al., 2009).

The mosquito gut microbiota were also shown to exert an indirect anti-*Plasmodium* effect upon reinfection, in what has been termed immune priming (Rodrigues et al., 2010). A peroxidase/dual oxidase dityrosine barrier formed after blood feeding limits contact of gut bacteria with epithelial cells and thus prevents elicitation of antibacterial responses (Kumar et al., 2010). Barrier disruption by *Plasmodium* ookinetes facilitates the contact of midgut bacteria with epithelial cells, which then triggers differentiation of haemocytes. Haemocyte differentiation has been shown to severely limit *Plasmodium* infection intensity following a second infectious blood meal. This effect relies on the presence of midgut bacteria

and is related to an enhanced ability of haemocytes to express immune factors including TEP1 and LRIM1, known to limit malaria parasites (Rodrigues et al., 2010).

Finally, the possibility that midgut bacteria can exert a direct anti-*Plasmodium* effect without the mediation of mosquito immune responses (Azambuja et al., 2005; Pumpuni et al., 1993; Pumpuni et al., 1996) was confirmed as the Enterobacterium *Esp_Z*, isolated from field mosquito gut microbiota, was shown to eliminate *P. falciparum* parasites and render the mosquito refractory to infection, through bacterial-derived production of reactive oxygen species (Cirimotich et al., 2011b).

The dual targeting of bacteria and *Plasmodium* by common components of the mosquito immune system suggests that midgut bacteria can shape immune responses relevant to susceptibility to malaria transmission but, more importantly, can also drive the evolution of the immune system, thus indirectly affecting the mosquito's ability to sustain malaria parasite infections. This evolutionary aspect of the tripartite interactions between midgut bacteria, *Plasmodium* parasites and the mosquito immune system formed the basis of the project's aim to identify genetic variation associated with the outcome of bacterial infections.

1.29 Mosquito gut microbiota composition and malaria transmission dynamics

Similarly to the pathogenic effects of shifts in the composition of the human gut microbiota during dysbiosis (Littman and Pamer, 2011) or the pathogenic effects of the disruption of the fly innate immune homeostasis that also shifts the fly's gut bacterial population structure (Ryu et al., 2008), the load and composition of the mosquito gut microbiota are expected to influence malaria transmission dynamics. Therefore, the direct influence of some bacterial strains inhabiting the mosquito gut on the outcome of *Plasmodium* infections suggests that the population structure of the mosquito gut microbiota can directly influence malaria transmission dynamics. Different bacterial strains inhabiting the mosquito gut have been shown to differentially influence the *Plasmodium* infection outcome. *Serratia marcescens*, a common member of the mosquito gut microbiota, has been shown to influence the outcome of *Plasmodium* infection (Bando et al., 2013; Gonzalez-Ceron et al., 2003), with *S. marcescens* intraspecific variation modulating the anti-*Plasmodium* ability of different *S. marcescens* strains, related to a locus that modulates the flagellum biosynthetic pathway (Bando et al., 2013). Interestingly, S. *marcescens* intraspecific variation related to loci involved in lipopolysaccharide biosynthesis, iron uptake and haemolysin production has been shown to influence *S. marcescens* pathogenicity following infection in the nematode *Caenorhabditis elegans* or *Drosophila* (Kurz et al., 2003). Furthermore, *Serratia*

entomophila pathogenicity in the grass grub *Costelytra zealandica* is also dependent on the presence of a plasmid containing three genes encoding insecticidal toxin analogues (Hurst et al., 2000).

What is more, the ability of the Enterobacterium *Esp_Z* to limit *Plasmodium* parasites relies on the generation of reactive oxygen species, with different *Enterobacteriaceae* members exhibiting differential reactive oxygen species generation ability and thus variable anti-*Plasmodium* potency (Cirimotich et al., 2011b). Therefore, the composition of the mosquito gut microbiota can profoundly influence the, direct or indirect, anti-*Plasmodium* effect exerted by some of its members. As the bacterial population structure in the gut of field-collected or laboratory reared mosquitoes shows immense variation, both at the individual and population level (Boissière et al., 2012; Osei-Poku et al., 2012; Wang et al., 2011b), factors that determine or contribute to shifts of the mosquito gut bacterial population structure can profoundly influence malaria transmission dynamics.

Although environmental variables, including the mosquito food source, bacterial availability or the structure of larval habitats are also expected to influence the mosquito gut microbiota composition, host genetic factors, most likely related to immune factor loci, could also shape the mosquito gut microbiota dynamics in ways that can determine susceptibility or refractoriness to malaria transmission. Therefore, an intriguing ecological aspect of malaria transmission ability emerges, in which genetic variation within the mosquito population influences the ecology of the resident gut flora, which, directly or indirectly, shape the ability of each individual mosquito to transmit malaria.

To further shed light into interactions between the mosquito gut microbiota and the host immune system, a more detailed analysis of mosquito antibacterial responses is required, especially with regard to responses against specific bacterial strains and homeostatic mechanisms related to such responses that shape the mosquito gut bacterial population structure. The experimental design of the present study aimed to address these poorly understood aspects of interactions between the mosquito host and the resident gut bacterial population, as will be made clear in the following chapter of this thesis.

Chapter 2

Project Aims and Experimental Design

Project Aims and Experimental Design

2.1 The tripartite interactions between mosquito immunity, midgut bacteria and *Plasmodium*

At about the same time with the commencement of this PhD project, in August 2009, two papers came out in *PLoS Pathogens*, linking the native bacteria of the mosquito gut with induced immune responses and the outcome of infection with rodent and human malaria parasites (Dong et al., 2009; Meister et al., 2009). The Dong study established that clearance of midgut bacteria through antibiotic treatment results in increased intensity of *P. falciparum* infection, while rechallenge with bacteria eliminates this effect. The same study further established, using DNA microarrays, that immune transcripts are significantly regulated between septic and aseptic mosquitoes, i.e. mosquitoes that retain their gut microbiota and mosquitoes in which midgut bacteria are eliminated through antibiotic treatment, before and after a blood meal, suggesting that the anti-*Plasmodium* effect exerted by the presence of midgut bacteria is mediated through immune responses triggered by bacteria. The Meister study further suggested that midgut bacteria trigger PGRPLC-mediated immune responses that target not only bacteria but also rodent and human malaria parasites. Taken together, these studies established that the outcome of *P. falciparum* infection is profoundly influenced by the presence of midgut bacteria and the immune responses they trigger, before or after blood feeding.

The present PhD project was designed to further refine the tripartite interactions between the mosquito immune system, commensal bacteria residing in the mosquito gut and *Plasmodium* parasites and establish how these interactions might affect susceptibility and refractoriness to malaria transmission. These tripartite interactions, which include direct and indirect influences of midgut bacteria on the outcome of *Plasmodium* infections, are summarized in Figure 2.1 and include insights from studies published after the design of this project.

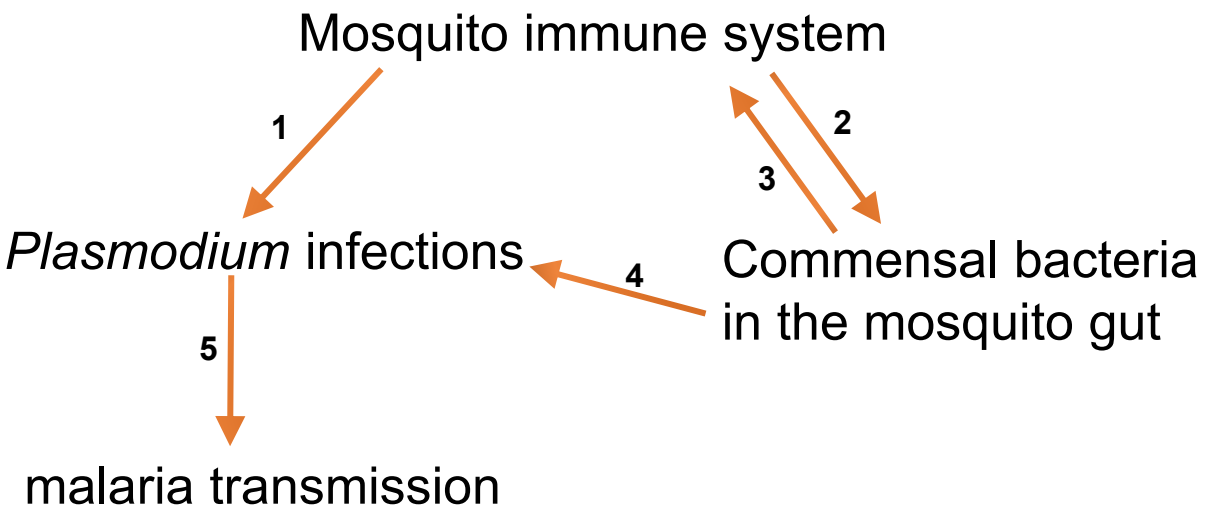


Figure 2.1: The tripartite interactions between mosquito gut microbiota, mosquito immunity and *Plasmodium* infections. 1: Mosquito immunity targets malaria parasites especially at the stage of midgut invasion by *Plasmodium* ookinetes. This is a major bottleneck of the *Plasmodium* life cycle with only few developing oocysts in natural infections (Yassine and Osta, 2010). **2:** Mosquito immunity also targets bacteria inhabiting the mosquito gut both before and after a blood meal. Silencing of the gene encoding the major IMD/REL2 pathway receptor PGRPLC leads to increased bacterial load before and after blood feeding (Meister et al., 2009). **3:** Mosquito gut microbiota can modulate immune responses thus indirectly influencing the outcome of *Plasmodium* infections. Contact of commensal bacteria with the mosquito gut epithelium triggers immune priming, which leads to enhanced immunity against *Plasmodium* parasites upon reinfection (Rodrigues et al., 2010). **4:** Commensal bacteria in the mosquito gut can directly target *Plasmodium* parasites. The enterobacterium *Esp_Z* has been shown to produce reactive oxygen species that kill *Plasmodium* parasites thus rendering the mosquito refractory to malaria transmission (Cirimotich et al., 2011b). **5:** Modulation of the outcome of *Plasmodium* infections can lead to mosquitoes being refractory to malaria transmission, as naturally happens in mosquito sub-populations (Sinden et al., 2004).

2.2 Homeostatic interactions and population structure of the gut microbiota

Despite the importance of the mosquito gut microbiota in influencing the outcome of *Plasmodium* infections and thus malaria transmission, the factors that determine the load and composition of the mosquito gut microbiota and lead to the observed natural variation remain poorly understood. Some bacterial strains are efficient colonizers of the mosquito gut under variable environmental settings while others are rarely or never found to inhabit the mosquito gut. Most likely this is the result of an interplay between the mosquito genetic makeup, especially with regard to immune factors, environmental conditions, including the availability of bacteria vertically transmitted or acquired at the larval stage or ingested at the adult stage and the ability of bacteria to exploit available nutrients, evade mosquito immune responses and sustain the evolutionary pressure exerted by other bacterial populations or mosquito gut conditions. The various components of these interactions remain poorly resolved. Such homeostatic interactions that largely shape the gut microbiota are a subject of intense research interest in insects (Dillon and Dillon, 2004; Douglas, 2011) but also higher organisms, including humans (Costello et al., 2009; Goodman et al., 2009).

2.3 A model of oral bacterial infections in *An. gambiae*

The influence of mosquito gut bacteria on the *Plasmodium* infection outcome made clear that a deeper understanding of the tripartite interactions between mosquito immunity, bacteria inhabiting the mosquito gut and *Plasmodium* infections could prove invaluable in further interventions aiming to disrupt malaria transmission. Further elucidation of these tripartite interactions would require dissection of mosquito antibacterial responses, especially with regard to the establishment of the load and composition of bacteria found in the mosquito gut. *Drosophila* has been extensively used as a model system in the elucidation of antibacterial immune responses (Lemaitre and Hoffmann, 2007). Basic concepts of *Drosophila* antibacterial immunity were essentially transferred to the mosquito, in parallel to comparative genomic analyses between *Drosophila* and *An. gambiae* (Christophides et al., 2002; Waterhouse et al., 2007). This transfer of knowledge between *Drosophila* and *Anopheles* included the involvement of antimicrobial peptides in *Drosophila* (Lemaitre et al., 1997; Tzou et al., 2000) and *Anopheles* (Richman et al., 1997; Vizioli et al., 2000; Vizioli et al., 2001) immunity, the involvement of Dscam in *Drosophila* (Watson et al., 2005) and *Anopheles* (Dong et al., 2006b) immunity and the implication of PGRPLC in the *Drosophila* (Gottar et al., 2002; Kaneko et al., 2004) and *Anopheles* (Christophides et al., 2002; Meister et al., 2009) antibacterial responses. Essentially, the idea upon which this project was based, was to establish a model system of bacterial infections that could be used in a *Drosophila*-like immunity project.

Indeed, models of oral bacterial infections have been extensively used in *Drosophila* in the study of antibacterial epithelial immunity (Buchon et al., 2010; Buchon et al., 2009; Chakrabarti et al., 2012; Cronin et al., 2009; Flyg et al., 1980; Nehme et al., 2007). Such *in vivo* models of single host-microbe associations have also been used in mice and constitute a simplified yet robust approach in studying such complex interactions (Faith et al., 2011; Falk et al., 1998; Rey et al., 2010). Therefore, a similar approach was adopted here for the study of mosquito antibacterial gut epithelial responses, along the lines of similar bacterial infections through the oral route previously conducted in mosquitoes (Damiani et al., 2010; Favia et al., 2007; Lindh et al., 2006).

In all cases, female *An. gambiae* mosquitoes were orally infected with bacteria. As the experimental design in the adopted model of oral bacterial infections does not require haematophagy, restricted to female mosquitoes, male mosquitoes could have conceivably been used in such infection model. Such approach was not adopted here, with female mosquitoes exclusively used for oral infections, so that the implemented model of oral infections establishes relevance with malaria transmission dynamics, effected through bites of female mosquitoes. Furthermore, as all previous efforts to characterize antibacterial responses in *An. gambiae* invariably used female mosquitoes (Cirimotich et al., 2011b; Dimopoulos et al., 2002; Kumar et al., 2010; Meister et al., 2009; Schnitger et al., 2009), transfer of knowledge could be achieved with the approach adopted here. Finally, due to the male gut size, midgut dissection and downstream use of dissected guts of male mosquitoes would have been technically challenging, considerably hindering the practicality of such approach.

2.4 Oral infections with *Serratia marcescens* and *Asaia*

Bacterial infections through the oral route constitute a much closer approximation of natural infections compared to injection of bacteria in the haemolymph and the elicitation of systemic immune responses. Therefore, a similar system of oral infections was adopted, as it was considered that it could be very useful in the elucidation of epithelial immune responses in the mosquito gut, which would be much more relevant in establishing the load and composition of the mosquito's gut microbiota compared to systemic immune responses elicited following crossing of bacteria through the midgut epithelium. A relevant study in *Drosophila* (Nehme et al., 2007) established that *S. marcescens* is highly pathogenic for the fly: if injected in the haemolymph, *S. marcescens* evades systemic immune responses and kills the host while oral infections trigger intestinal immune responses. Therefore oral infections with *S. marcescens* in the mosquito could reveal similar epithelial immune responses essential for fending off this entomopathogen.

The choice of *S. marcescens* was based both on its widespread use as a model system in *Drosophila* but also on its prevalence in lab-reared and field-collected mosquitoes, as has been shown in several studies (Boissière et al., 2012; Dong et al., 2009; Gonzalez-Ceron et al., 2003; Osei-Poku et al., 2012). Two genetic screens in *Drosophila* that utilized *S. marcescens* further supported the adoption of *S. marcescens* in a model system aimed to identify mosquito antibacterial responses: a study in which flies were injected with *S. marcescens* with the aim to identify genetic polymorphism associated with the outcome of infection (Lazzaro et al., 2004) and another study in which a genome-wide RNAi screen was conducted using flies orally infected with *S. marcescens*, aiming to identify genes implicated in antibacterial responses (Cronin et al., 2009).

A second bacterial strain that was adopted belonged to the genus *Asaia*. *Asaia,* an acetic acid bacterium that belongs to alpha-proteobacteria, is also a common symbiont of *Anopheles* as well as most insects, it is also prevalent in lab-reared and field-collected mosquitoes, and it is generally considered less pathogenic than *Serratia* (Crotti et al., 2009; Damiani et al., 2010; Favia et al., 2007).

The adopted model system of mosquito oral infections with *Asaia* or *S. marcescens* is summarized in Figure 2.2. *An. gambiae* mosquitoes are antibiotic treated from the day of emergence to clear their native gut microbiota. Subsequently, they are fed on a sugar meal that includes the fluourescent strain *Db11-GFP* of *S. marcescens*, antibiotics for the selection of this strain and a dye used to establish whether mosquitoes have been fed with this solution. Two days post infection, bacteria-fed mosquitoes are separated, ensuring they have ingested *S. marcescens*-containing sugar, and are further kept in a sugar solution containing antibiotics that select *S. marcescens* bacteria but also further limit any other remaining bacterial strains. Oral infections with *Asaia* employ a similar strategy, with the use of suitable antibiotics for its selection.

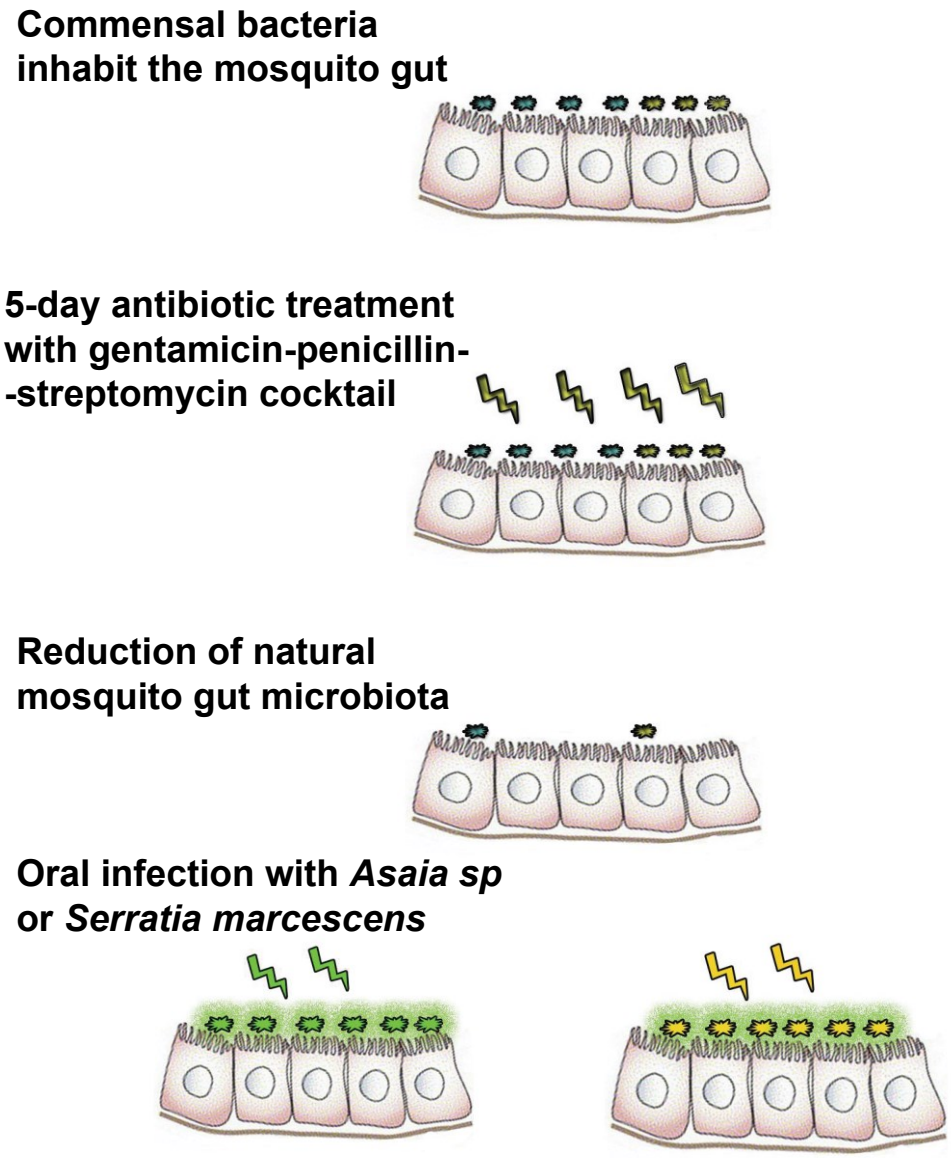


Figure 2.2: A model of *An. gambiae* oral bacterial infections. The gut of adult *An. gambiae* mosquitoes is rife with commensal bacteria. A 5-day antibiotic treatment, starting upon their emergence, with a cocktail of gentamicin, penicillin and streptomycin is used to reduce the natural gut microbiota of adult mosquitoes. Subsequently, mosquitoes are fed with a sugar solution that contains GFP fluorescent strains of *Asaia* or *S. marcescens* and the antibiotics for the selection of the respective strain: kanamycin for *Asaia* or tetracycline and carbenicillin for *S. marcescens*. This antibiotic treatment regime not only selects the respective fluorescent bacterial strains but also eliminates *Asaia* or *S. marcescens* that have lost the plasmid conferring GFP fluorescence or other bacteria that evaded the initial antibiotic treatment. Therefore, the gut microbiota of adult *An. gambiae* mosquitoes is largely replaced by a single fluorescent strain, either of *Asaia* or *S. marcescens*.

2.5 Genome-wide studies for the identification of factors involved in responses to oral bacterial infections

Following confirmation of efficient colonization of the mosquito gut with *S. marcescens* or *Asaia*, as would be expected based on their prevalence in lab-reared and field-collected mosquitoes, the next aim of the project was the implementation of genome-wide expression analysis for the identification of mosquito genes involved in immune responses against the invading bacteria but also gut homeostasis and tolerance which allow the successful colonization of the mosquito gut by these bacterial strains. Further comparisons of induced transcripts by the pathogenic enterobacterium *S. marcescens* and *Asaia* could also identify specific responses against a fraction of the microbiota along with more generalized responses against all Gram-negative bacteria.

The next step following such expression analysis involved phenotypic characterization of identified genes involved in responses against invading bacteria using RNAi-mediated silencing. The phenotypic effects of such genes could encompass involvement in bacterial proliferation, mosquito survival following a bacterial challenge but also effects on the outcome of *Plasmodium* infections.

The commencement of this PhD project also coincided with an effort to genotype single nucleotide polymorphisms (SNPs) in the *An. gambiae* genome. Sequencing of the M and S forms of *An. gambiae* led to the identification of about 400,000 SNPs in the *An. gambiae* genome which were used to construct a SNP genotyping array able to interrogate genetic variation in these variable positions of the *An. gambiae* genome (Lawniczak et al., 2010). Indeed, this array, combined with high-throughput sequencing, was shown to provide useful quantitative information regarding SNP divergence between assayed samples and was used to identify genetic divergence between the M/S forms of *An. gambiae* (Neafsey et al., 2010). The availability of this SNP genotyping array opened the possibility to investigate the genetic basis of various phenotypic traits showing variation within a mosquito population, including the outcome of *Plasmodium* or bacterial infections.

As preliminary results of oral *S. marcescens* or *Asaia* infections in *An. gambiae* indicated great variation regarding the outcome of infection, it became possible to use this SNP genotyping array to investigate SNP divergence associated with the outcome of bacterial infections and, as for gene candidates coming out of the DNA microarray study, phenotypically characterize resulting genes associated with the outcome of infection.

To sum up, the aim of this PhD project was to establish a model system of oral bacterial infections in *An. gambiae*, along the lines of previous studies in *Drosophila* (Cronin et al., 2009; Nehme et al., 2007), and

through the use of genome-wide array-based methods, to identify mosquito genes involved in responses to bacteria, either immune antibacterial responses or homeostatic responses to achieve tolerance. A subsequent aim included phenotypic characterization of candidate genes coming out of these array-based studies in the context of their involvement in bacterial infections but also of their putative contribution in the resulting susceptibility and refractoriness to malaria transmission, largely based on a reverse genetics approach.

Chapter 3

Materials and Methods

Materials and Methods

3.1 Mosquito rearing and maintenance

The *N'gousso* strain of *An. gambiae* was used. This is an M form strain colonized in 2006 (Harris et al., 2010a) and kept in large numbers to retain genetic variation. Rearing and maintenance of the strain was performed as described previously (Crampton et al., 1997). Briefly, mosquito eggs were washed through filter paper containing 1% *Virkon* (Dupont), rinsed with dH_2O and left on 0.001% salted water for 3 days at 27°C. Subsequently, developed L1 larvae were expanded in 500 ml of dH_2O containing approximately 0.3 gr of standard fish food. A similar amount of fish food was added daily. Larvae were further expanded 5 days later. Developed pupae were collected after 4 days by filtering, placed in 100-200 ml of dH_2O and left inside a cage for mosquitoes to emerge. Mosquitoes were further kept at 27°C and a 12 hour light/dark cycle.

Female mosquitoes were collected after emergence and placed in net-covered pots, including approximately 75 to 100 female mosquitoes. A cotton pad was placed on top of each pot dampened with a cocktail of 25 μg/ml gentamicin (Sigma), 10 μg/ml penicillin (Sigma) and 10 units/ml streptomycin (Sigma), diluted in previously autoclaved solution of 10% D-(-)-Fructose (Sigma). This antibiotic treatment regime was used for 5 days, with the antibiotic solution refreshed every 24 hours. At day 5 post emergence, the antibiotic solution was replaced by dH_2O and mosquitoes were starved overnight prior to oral bacterial infection. This antibiotic treatment regime was used in all assays, unless otherwise stated. Untreated mosquitoes were kept on a 10% D-(-)-Fructose solution, also refreshed daily. In some cases, as stated, ciprofloxacin (Sigma) was added to the antibiotics cocktail at a 25 μg/ml concentration.

3.2 Mosquito oral infection with *S. marcescens* or *Asaia*

The *S. marcescens Db11-GFP* strain was used, modified to be GFP-fluourescent and resistant to tetracycline and carbenicillin (Nehme et al., 2007). In all cases, cultures were initiated from *S. marcescens* glycerol stock kept at -80°C, kindly provided by Dominique Ferrandon, which was grown in 5 ml LB cultures containing 50 μg/ml tetracycline and carbenicillin (Sigma) at 37°C. Following overnight incubation, the cultures were expanded to 100 ml, again containing 50 μg/ml tetracycline and carbenicillin and further incubated overnight at 37°C.

Subsequently, OD_{600} and GFP fluourescence (excitation/emission at 485/520 nm) were measured to ensure cultures maintained GFP fluourescence, using the *Fluostar Omega* spectrophotometer (BMG

Labtech). 100 µl of the grown *S. marcescens* culture and 1:2 to 1:16 dilutions of the culture, also at a 100 µl volume, were added in wells of a 96-well plate, with each dilution in triplicate, along with wells only containing LB. For OD_{600} measurements, clear non-coated 96-well plates were used (Fisher) while for GFP fluourescence measurements, black non-coated 96-well plates were used (Fisher). Assessment of GFP fluourescence in the assayed culture was based on comparison of the GFP fluourescence readings of wells corresponding to diluted cultures to the corresponding readings from blank LB-containing wells but also from the gradual decrease of GFP fluourescence readings of wells corresponding to the concentration gradient of the *S. marcescens* culture from 1:1 to 1:16.

Bacterial pellets following centrifugation at 2500 rpm for 5 minutes and removal of supernatant, were washed twice with PBS, again by centrifugation at 2500 rpm and removal of supernatant, and resuspended in such volume of 10% D-(-)-Fructose, so that 1 ml of the bacteria-containing sugar solution corresponded to OD_{600}=0.1 of the initial 100 ml culture, based on the OD_{600} measurement described above. The sugar solution was further diluted 1:12 in a 10% D-(-)-Fructose solution that contained tetracycline and carbenicillin at 50 µg/ml and 5% v/v of scarlet dye (Langdale) and 24 ml of this solution were used to dampen cotton pads placed on top of the corresponding mosquito pots.

Oral infections with *Asaia* were conducted in a similar manner. The *Asaia SF2.1 (GFP)* strain was used, kindly provided by Guido Favia, grown as previously described (Favia et al., 2007). In all cases, *Asaia* cultures were initiated from *Asaia* glycerol stock kept at -80^{o}C, and were grown on a 5 ml volume of GLY medium, containing 1.26 kg/l glycerol (Sigma) and 5 gr yeast extract (Sigma), with an adjusted pH of 5, containing kanamycin at 50 µg/ml, at 30^{o}C. *Asaia* cultures were expanded the following day to 100 ml, also containing kanamycin at 50 µg/ml. GFP fluourescence and OD_{600} were then measured and the *Asaia* culture was washed and prepared for oral infection in the same way as *S. marcescens*.

Mosquitoes were fed with this solution for 2 days. Mosquitoes fed with bacteria-containing sugar were separated based on the presence of the dye in their gut. Mosquitoes were anaesthetized on ice and observed under a stereoscope for signs of red dye in their gut. Mosquitoes with visible dye in their gut were further kept while mosquitoes with no sign of bacteria-containing solution in their gut were discarded at this phase. Subsequently, bacteria-containing cotton pads were discarded and mosquitoes were kept on 10% D-(-)-Fructose containing the respective antibiotics as described above.

3.3 Surface sterilization and gut dissection of mosquitoes

Mosquitoes were collected in 2 ml tubes (Eppendorf) and were treated with 70% v/v ethanol (Sigma) added to each tube. Each tube was shaken on a *Vortex* shaker for 30 seconds and subsequently mosquitoes were removed through a previously sterilized pair of forceps and placed on a tube containing PBS. Mosquitoes were shaken for 30 seconds and washed in the same way 2 additional times in PBS. Subsequently, mosquitoes were placed inside a tube containing *RNA later* (Invitrogen) and kept on ice.

Mosquitoes were individually transferred to a drop of *RNA later* and their midgut was dissected under a stereoscope with 2 pairs of previously sterilized forceps. Dissected guts were placed either on tubes containing *RNA later* or mounted on *Vectashield* medium (Vecta).

3.4 Culture-based growth of mosquito gut microbiota

Dissected guts of antibiotic treated or untreated mosquitoes were homogenized with a pestle motor equipped with sterile pestles for about 3 minutes in SOC medium (Sigma). Subsequently, the homogenate was diluted 1:100 in SOC medium and used to plate LB plates, prepared from standard LB medium (Sigma). Plates were kept at 27^{o}C for 2 days.

3.5 Fluourescence microscopy

The levels of *S. marcescens* infection were determined by microscopic observation of dissected mosquito guts immersed in *Vectashield* mounting medium, immediately after dissection. The *Zeiss Axiophot* fluourescence microscope was used, equipped with light and GFP filters while photos of observed midguts were taken with the *Axiocam HRc* and *Axiovision* software (Zeiss).

3.6 Survival assays

Mosquitoes were kept inside a 27^{o}C room or an incubator at the same temperature and 70% humidity, as stated in each case. Daily, dead mosquitoes were enumerated and removed from the mosquito pot. Generation of survival curves and statistical analysis using the Mantel-Cox logrank or the Gehan-Breslow-Wilcoxon tests were carried out using *Prism* 5.0-7.0 (Graphpad).

3.7 SNP genotyping arrays

All carcasses corresponding to midguts of *S. marcescens* infected mosquitoes were kept numbered in clear 96-well plates (Fisher) immersed in 75% ethanol at -80^{o}C. Carcasses from selected midguts were used for genomic DNA (gDNA) extraction using the *QIAquick Blood and Tissue* kit (QIAGEN). gDNA concentrations were determined using the *Picogreen dsDNA* kit (Invitrogen) and equimolar DNA quantities from each mosquito were pooled. The design and validation of the SNP genotyping array used along with the treatment of gDNA pools, hybridization, calling of SNP genotypes and measurement of differentiation in each pooled hybridization between allele A and B have been described previously (Neafsey et al., 2010; Reidenbach et al., 2012) and were carried out by collaborators at the Broad Institute (Cambridge, MA,USA). The frequency of designated allele A was considered as the minor allele frequency and was used to measure the difference between pooled hybridizations. The permutation analysis used has been described previously (Neafsey et al., 2010), with a modified length of non-overlapping 10-SNP windows, carried out by collaborators at Imperial College London (London, UK). Determination of genes residing in identified genomic areas and homology analysis was performed using *Biomart* 0.7 (http://biomart.vectorbase.org) and the *AgamP3.7 An. gambiae* gene annotation (Lawson et al., 2009). The SNP genotyping array datasets have been deposited to ArrayExpress under the experiment name *Serratia_SNP1* and accession number E-MEXP-3951.

3.8 DNA Microarrays

Total RNA was extracted from midguts using the *Trizol* reagent (Invitrogen), and treated with *Turbo DNAse I* (Ambion). Samples were further purified using the *RNeasy* kit (QIAGEN). Quantification was performed using the *Nanodrop 1000* spectrophotometer (Thermo Scientific) and RNA integrity was assessed using the *RNA 6000 Pico Chip* kit and the corresponding software (Agilent). Total RNA samples with minimal peaks indicating contamination and clear ribosomal RNA peaks were considered to pass quality control and were further used.

Labelling and hybridization were performed using the *Low Input Quick Amp Labeling* kit for two-colour microarray based expression analysis (Agilent), according to the manufacturer's instructions. Agilent custom 4x44k gene expression microarrays were used. The microarray design *Pfalcip_Agamb200*9 (A-MEXP-2324) comprises oligonucleotide probes encompassing all *An. gambiae* annotated genes of the *AgamP3.6* release along with *P. falciparum* probes, with each probe represented in duplicate. Slides were scanned using the *Genepix 4000B* scanner equipped with the *Genepix Pro 6.1* software (Axon

instruments). Manual grid alignment and flagging of features (probes) with evident fluourescence artefacts was performed.

All derived dataset files were normalized using the *Genespring 11.0 GX* software (Agilent). The Lowess normalization method was used while the threshold of raw signals was set to 5, which was sufficient to eliminate background regulation of *P. falciparum* probes. Further analysis of transcriptionally regulated genes and GO analysis was performed using the *Genespring* 11.0 GX software. For GO analysis, GO accession numbers for all *An. gambiae* transcripts were obtained using *Biomart* 0.7 and a hypergeometric test with Benjamini-Hochberg correction was performed on the set of more than 1.75-fold regulated genes. The corrected p-value for testing multiple GO accession numbers for their significance was set to 0.1. The log2-transformed transcriptional regulation for each transcript was extracted from the normalized datasets for each of the two probes corresponding to each transcript and the obtained values from all three independent infections were used in a t-test against zero, with a p-value cut-off of 0.05. The DNA microarray datasets corresponding to the *S. marcescens* infections have been deposited to ArrayExpress under the experiment name *Serratia_infections* and accession number E-MEXP-3952.

Principal component analysis was performed with *Genespring* 11.0 GX and Lowess-normalized datasets. For cluster analysis, Lowess-normalized, log2-transformed fold-change values for all transcripts were extracted using *Genespring* 11.0 GX and only values for *An. gambiae* transcripts were used, filtered in *Excel 2010* (Microsoft). Cluster analysis was performed in *Gene Cluster* 3.0 (Stanford University) and *Java Treeview* 1.1 (Alok).

Functional classification of significantly regulated transcripts was performed with *Biomart* 0.7 based on assigned GO terms, *Interpro* domains or *Drosophila* homologues, based on the *AgamP3.7* annotation of *An. gambiae*.

3.9 Generation of dsRNA used for RNAi-mediated silencing

Primers were designed to target an approximately 250 bp region of the transcript of interest or a 500 bp fragment of the *E. coli LacZ* gene, used as a control. Primer design was performed in *Geneious* 6 (Biomatters) based on exon sequences derived from Vectorbase and corresponding to the *AgamP3.6* annotation of *An. gambiae*, using standard settings. T7 promoter flanking sequences were also included to each primer pair to allow reverse transcription from the cDNA template. Standard desalt purification

of primer pairs was used. The primer sequences used for dsRNA synthesis can be found in Table 3.1 (Appendix).

The corresponding T7 primers were used in a PCR using as template cDNA derived from *N'gousso* sugar fed whole mosquito total RNA. The *GO Taq* DNA polymerase (Promega) was used in a 200 µl reaction containing 40 µl of the respective 5x buffer, according to the manufacturer's instructions, 4 µl of 10 mM dNTPs (Invitrogen) and 10 µl of each primer at a 10 µM concentration, while 2 µl of the cDNA template were used, at an 1:100 dilution. The PCR cycle used comprised a 5 minute step at 94^{o}C, 40 repetitions of a 45 second step at 94^{o}C, a 30 second step at 58^{o}C and a 45 second step at 72^{o}C and one repetition of a final 10 minute step at 72^{o}C. The resulting PCR product was purified using the *PCR Purification* kit (QIAGEN) and DNA concentration was measured using the *Nanodrop* 1000 spectrophotometer. If needed, a second PCR reaction was performed, using as template the purified PCR product and volume was adjusted to achieve a 125 ng/µl concentration. The amplification of the PCR product with the expected length was confirmed by agarose gel electrophoresis of the derived PCR product. For the *dsLacZ* control, the template used comprised previously amplified PCR product from *E. coli*.

For each targeted transcript or the *dsLacZ* control, dsRNA was synthesized using the *T7 Megascript* kit (Invitrogen). For each *T7 Megascript* reaction, 1 µg of purified PCR product was added in a 40 µl reaction containing 16 µl of purified PCR product, 4 µl of each dNTP, 4 µl of pre-warmed at 37^{o}C buffer and 4 µl of T7 enzyme mix, according to the manufacturer's instructions. The T7 reaction was kept at 37^{o}C for 16 hours, subsequently *TURBO DNase I* was added, the reaction mix was further incubated for 15 to 30 minutes at 37^{o}C and further purified using the *RNeasy* kit (QIAGEN) to a concentration of 3 µg/µl. The purification of dsRNA of the expected size was confirmed by agarose gel electrophoresis.

3.10 RNAi-mediated silencing

Mosquitoes were treated with the respective dsRNA at the day of emergence, unless otherwise stated, as described previously (Blandin et al., 2002). Briefly, mosquitoes were anaesthetized with CO_2 for, typically, 15 to 20 minutes, and treated with 69 nl of dsRNA using the *Nanoject II* microinjector (Drummond). The microinjector needle was frequently checked for discharging dsRNA and blunt needles were replaced.

3.11 Total RNA extraction and cDNA synthesis

Total RNA from mosquito midguts was extracted after homogenization with a pestle motor in *RNA later* using the *RNeasy* kit (QIAGEN), according to the manufacturer's instructions. cDNA was synthesized from total RNA using the *QuantiTect Reverse Transcription* kit (QIAGEN). Total RNA was diluted to 1 µg per 12 µl, if exceeding that concentration, and used for the cDNA synthesis reaction, according to the manufacturer's instructions. The resulting cDNA was diluted 1:10 by adding 180 µl of dH_2O to the 20 µl cDNA synthesis reaction.

3.12 Quantitative Real Time PCR

Primers used to target specific transcripts or genes can be seen in Table 3.1, designated in the respective column as qRT-PCR. Template used was cDNA, unless otherwise stated. In all cases, two-step RT-PCR was used, with cDNA synthesized before the reaction, as described above.

The *Fast SYBR Green Master Mix* (Applied Biosystems) was used for the amplification reaction, according to the manufacturer's instructions. The qRT-PCR was carried out using the 7500 Real-Time PCR System (Applied Biosystems) with its respective software to perform the reaction and any further analysis. The relative abundance of each sample was determined using the standard curve method as described in User Bulletin #2 for the ABI Prism 7700 Sequence Detection system (Applied Biosystems).

Primers targeting the housekeeping reference transcript *AgS7* were used as an endogenous control. This is a gene whose expression levels are expected to be constant between samples. In a 20 µl reaction, 1 µl of cDNA template, at a 1:10 dilution, as described above, were added along with 2 µl of each respective primer at a 0.5 to 9 µM concentration, optimized for each primer set. In some cases, gDNA was used as template, as stated, with the qPCR performed in the same way.

The reaction cycle used comprised one repetition of 20 seconds at 95°C, 40 repetitions of 3 seconds at 95°C and a final 30 second repetition at 60°C. Optimization of primer concentrations involved reactions with a cDNA mix used as template and different primer concentrations, 0.5 µM, 3 µM and 9 µM for the forward and reverse primer, resulting in 9 different primer concentration combinations. The reaction cycle included a primer dissociation step, in addition to the steps described above, with one repetition of 15 seconds at 95°C, 1 minute at 60°C, 15 seconds at 95°C and 15 seconds at 60°C. Assessment of primer combinations for amplification efficiency included the lowest threshold cycle (C_T) value combined with a dissociation curve with a single, high peak. PCR products were also run in an agarose gel electrophoresis to confirm the presence of a single band corresponding to the amplified PCR product.

Following the qRT-PCR cycle, an amplification plot was generated for each reaction included in the plate, corresponding to the emitted fluourescence by SYBR Green when binding the respective template, thus capturing the template amplification, against the cycle number in which this fluourescence emission was recorded. The C_T value was determined for each reaction, corresponding to the reaction cycle in which the log-linear amplification phase commenced. The C_T value expresses the relative abundance of template, since a more abundant template would result in linear amplification in fewer cycles compared to a reaction in which a less abundant template would require a higher number of cycles to reach the log-linear amplification phase.

Based on the C_T values of the reactions corresponding to the standard curve gradient, by using either the *AgS7* primers or primers corresponding to the transcript of interest, a standard curve was generated, correlating the C_T value with the relative template abundance. Outlier values of the standard curve gradient resulting in considerable deviation of the calculated R-squared (R^2) value from 1 in a goodness-of-fit linear regression test, were discarded, with an acceptable R^2 value >0.95 in all cases.

Based on the slope and y-intercept of the obtained standard curve for reactions using *AgS7* primers, the relative template abundance for each reaction was calculated. Subsequently, relative template abundance was calculated in the same way, using the slope and y-intercept of the standard curve corresponding to primers targeting the transcript of interest, and normalized to the obtained value of the same reaction using the *AgS7* primers. The obtained relative template abundance for each sample was further normalized to one of the samples included in the reaction plate, with all other samples' relative abundance expressed in fold-change differences compared to the control sample.

3.13 454 pyrosequencing

Obtained cDNA pools representing the transcripts present in the gut of the assayed mosquitoes were used as template in a PCR for the amplification of the bacterial 16S V4-V6 regions corresponding to reverse transcribed bacterial transcripts. PCR was performed as above using the 16S V4-V6 primers shown in Table 3.1 that also included suitable barcode sequences. PCR products were purified using *PCR Purification* and *Gel Extraction* kits (QIAGEN) and the amplification of a single PCR product of a ~560 bp length was confirmed by agarose gel electrophoresis.

PCR products were sequenced by Beckman Genomics (Grenoble, France) using the *Roche 454 GS FLX+* and standard procedures. The resulting FASTA files were filtered to a minimum read length of 250 bp using *Galaxy* (Goecks et al., 2010) and blasted against the NCBI *16SMicrobial* database using *BLAST+* and

prfectBLAST (Santiago-Sotelo and Ramirez-Prado, 2012) with standard *blastn* algorithm settings and 10 maximum target sequences. Further phylogenetic analysis was performed using *MEGAN4* (Huson et al., 2011).

3.14 Meal size and two-choice preference assays

Sugar meal size was determined through a modified capillary feeder assay (Ja et al., 2007). Mosquitoes treated with *LacZ* or *Gr9* dsRNA were antibiotic treated for 5 days, starved overnight and, subsequently, individual mosquitoes placed inside a pot were fed on a 5 µl glass capillary (VWR) containing 10% D-(-)-Fructose and 5% v/v scarlet dye. For alive mosquitoes, sugar consumption was determined 16 hours later through the reduction of sugar solution in each capillary. The two-choice preference assay was also conducted based on a previously described capillary feeder assay (Ja et al., 2007). Mosquitoes treated with *LacZ* or *Gr9* dsRNA were antibiotic treated for 5 days, starved overnight and placed in pools of 8-11 mosquitoes. Mosquitoes were offered to feed from two capillaries, one containing a sugar solution, as above, and one also containing *S. marcescens*, prepared as described above for oral infection. Water-containing cotton pads were also used and pools with mosquito mortality were disregarded. 16 hours later, consumption for each capillary was determined based on the reduction of the sugar solution in each capillary.

Chapter 4

Dynamics of *Anopheles gambiae* oral infection with *Asaia* or *Serratia marcescens*

Introduction

4.1 *Serratia marcescens* and mosquitoes

To investigate gut epithelial responses in *An. gambiae*, a model of oral infections was established with single strains of bacteria that commonly inhabit the mosquito gut under natural settings. Two bacterial strains were selected to be used in such model of oral infections, *S. marcescens* and bacteria of the genus *Asaia*. *S. marcescens* is a common entomopathogen that has been extensively used in studies of *Drosophila* epithelial (Cronin et al., 2009; Flyg et al., 1980; Nehme et al., 2007) but also systemic immunity (Lazzaro et al., 2006; Lazzaro et al., 2004; Nehme et al., 2007), therefore its adoption was expected to facilitate transfer of knowledge between the two model organisms. *Serratia* is commonly found in plants, nematodes and many other invertebrate and vertebrate hosts (Grimont and Grimont, 1978; Grimont et al., 1977) including gypsy moths (Broderick et al., 2004), lab-reared and field-collected *Drosophila* (Cox and Gilmore, 2007), while it is also an opportunistic pathogen in human nosocomial infections (Kurz et al., 2003).

S. marcescens is considered highly pathogenic and thus can affect the host's survival if not targeted by immune responses. The use of *S. marcescens* intestinal infection in the nematode model system *Caenorhabditis elegans* has been shown to trigger innate immune responses without which *S. marcescens* kills the host (Mallo et al., 2002). In locusts, *S. marcescens* colonization rapidly kills germ-free hosts while the presence of the normal microbiota controls the *S. marcescens* population and its detrimental effects (Dillon and Charnley, 2002). The *S. marcescens* pathogenicity in *Drosophila* was demonstrated decades ago (Flyg et al., 1980). *S. marcescens* evades systemic immune responses when infected in the fly's haemolymph and kills the host, while oral infection triggers epithelial immune responses that limit *S. marcescens* after crossing the epithelial barriers and reaching the haemolymph, although the host's survival is still compromised due to infection (Nehme et al., 2007).

Bacteria of the genus *Serratia* are commonly found both in lab-reared and field-collected mosquitoes under different environmental settings (Gendrin and Christophides, 2013). They have been traced in *Ae. aegypti* lab-reared or field-collected mosquitoes (Apte-Deshpande et al., 2012; Gaio Ade et al., 2011; Ramirez et al., 2012; Terenius et al., 2012), *An. gambiae* mosquitoes collected in Cameroon (Boissière et al., 2012) or Kenya (Wang et al., 2011b), lab-reared *An. gambiae* (Dong et al., 2009) or *An. stephensi* (Seitz et al., 1987; Sharma et al., 2013) mosquitoes, *An. stephensi* field-collected mosquitoes where *Serratia* was found to be dominant both in larvae and adults (Rani et al., 2009), *An. arabiensis* field mosquitoes from

Kenya (Lindh et al., 2005) and *An. albimanus* mosquitoes collected in Mexico (Gonzalez-Ceron et al., 2003).

Serratia has been shown to modulate the outcome of dengue virus infection in *Ae. aegypti* (Apte-Deshpande et al., 2012), *Trypanosma cruzi* infection in *Rhodnius prolixus* (Azambuja et al., 2004) but also *Plasmodium* infection. A concomitant infection of *P. berghei* and *S. marcescens* was shown to enhance *An. stephensi* mortality (Seitz et al., 1987). *S. marcescens* was also shown to block *P. vivax* infection in *An. albimanus* when offered along with parasite-infected blood (Gonzalez-Ceron et al., 2003). More recently, *S. marcescens* isolates from guts of lab-reared or field-collected mosquitoes were shown to inhibit *P. berghei* development in *An. stephensi* (Bando et al., 2013). Interestingly, the ability to inhibit *Plasmodium* development was correlated with intra-specific variation within *S. marcescens*; non-motile, non-flagellular strains did not affect the outcome of *Plasmodium* infection (Bando et al., 2013).

4.2 *Asaia* and mosquitoes

Bacteria of the genus *Asaia* are also common in lab-reared and field-collected mosquitoes (Boissière et al., 2012; Gendrin and Christophides, 2013; Osei-Poku et al., 2012). Oral infection of fluourescent *Asaia* strains both in *An. stephensi* (Favia et al., 2007) and *An. gambiae* (Damiani et al., 2010) results in a stable association in the midgut and salivary glands of female adults, in the adult male reproductive system and the larval gut. The presence of *Asaia* has been shown to accelerate larval development in *An. gambiae*, suggestive of an *Asaia* role during mosquito development (Mitraka et al., 2013). Furthermore, *Asaia* can be vertically transmitted between generations (Damiani et al., 2008; Mitraka et al., 2013). *Asaia* also influences the outcome of *Plasmodium* infections possibly by triggering immune responses (Capone et al., 2013). All these characteristics make *Asaia* a suitable candidate for paratransgenic efforts in which *Asaia* strains are used for the expression of anti-*Plasmodium* molecules so that their delivery in mosquitoes through *Asaia* infection can render mosquitoes refractory to malaria transmission (Favia et al., 2008; Wang et al., 2012).

Asaia was initially identified in plants (Katsura et al., 2001; Yamada et al., 2000; Yukphan et al., 2004) and is also found in phylogenetically distant insects (Crotti et al., 2009). *Acetobacteriaceae*, to which *Asaia* belongs, show exceptional cross-colonization ability as they form symbiotic relationships with a variety of insects including mosquitoes, fruit flies, bees, butterflies and sugarcane mealybugs (Crotti et al., 2010). Therefore, the colonization mechanisms *Acetobacteriaceae* employ seem to be largely independent of host-specific characteristics, at least for insects, and able to overcome host-specific colonization barriers.

Therefore, *Asaia*, as a commonly found member of *Acetobacteriaceae*, is an ideal candidate for the study of host-microbe interactions that lead to gut homeostasis.

Acetobacteriaceae have also been found to be dominant in gut bacterial populations of *Drosophila* (Corby-Harris et al., 2007; Ryu et al., 2008). Silencing of the modulator of the Imd pathway *Caudal* was shown to result in altered antimicrobial peptide expression in the *Drosophila* gut that shifted the bacterial population structure so that the presence of an *Acetobacteriaceae* strain was diminished; this strain was shown to control the load of another *Acetobacteriaceae* strain belonging to the *Gluconobacter* genus that was highly pathogenic for the fly (Ryu et al., 2008). In *An. gambiae*, silencing of the hypervariable pattern recognition receptor *AgDscam* resulted in increased bacterial load in the mosquito haemolymph, including the increase of *Asaia bogorensis* (Dong et al., 2006b). Therefore, *Acetobacteriaceae* seem to play important roles in interactions that lead to gut homeostasis as they constitute targets of tightly regulated immune responses that shape the gut microbial population. Such interplay between immunity, bacterial interactions and gut homeostasis can thus be studied through oral infections with *Asaia*.

4.3 A model of oral bacterial infection in *An. gambiae*

Adult mosquitoes were collected at the day of emergence and treated with a sugar solution containing a cocktail of gentamicin, penicillin and streptomycin. After 5 days of antibiotic treatment aimed to reduce the mosquito's natural gut bacteria, mosquitoes were starved overnight as the antibiotic-containing sugar solution was replaced by water. In this way, the antibiotic treatment regime was allowed to clear from the mosquito and thus a possible interference of the antibiotic treatment with a subsequent bacterial infection was avoided. At the same time, mosquito starvation contributed to enhanced and, conceivably, more uniform between mosquitoes, bacterial intake. Subsequently, mosquitoes were fed with a sugar solution that contained fluourescent strains of bacteria of the genus *Asaia* or *S. marcescens*, antibiotics for the selection of these strains, kanamycin for *Asaia* or tetracycline and carbenicillin for *S. marcescens* and a dye used for the selection of mosquitoes that ingested this bacteria-containing solution. Such feeding with bacteria-containing sugar will henceforth be referred to as infection.

Two days following infection, mosquitoes with visible signs in their gut of the dye included in the bacteria-containing sugar solution were separated while the remaining mosquitoes were discarded, thus ensuring that mosquitoes had ingested at least some bacteria-containing sugar solution. The two-day period from infection ensures mosquitoes were allowed to feed with the bacteria-containing sugar solution and for the sugar solution to reach the midgut from the crop it was originally stored (Favia et al., 2007), thus making the dye visible in the mosquito gut upon stereoscopic inspection. Subsequently, bacteria-fed mosquitoes were kept on a sugar solution containing the respective antibiotics for the selection of the bacterial strain used for infection. This antibiotic treatment ensures that the bacteria used for infection retain their fluourescent properties, as bacteria without the respective plasmid that confers fluourescence are susceptible to the respective antibiotic treatment. At the same time, other bacteria that inhabit the mosquito gut and possibly evaded the initial antibiotic treatment or subsequently infected the mosquito were further reduced or eliminated.

4.4 Efficacy of antibiotic treatment

The antibiotic treatment regime used here, with minor modifications, has been previously reported to eliminate the resident midgut bacteria in *An. gambiae* mosquitoes (Dong et al., 2009; Meister et al., 2009). As differences in the composition of the mosquito gut microbiota and acquired resistance in some

bacterial strains to the antibiotic regime used can lead to different outcomes, the efficacy of antibiotic treatment was assayed (Figure 4.1). Initially, the efficacy of the antibiotic treatment regime reported in the Meister study (Meister et al., 2009) was assayed using a culture-based method. Mosquitoes were treated only with gentamicin added in a water solution for 3 days while sugar was provided with an added cube. Subsequently, whole mosquitoes were surface-sterilized, plated on LB plates at a 1:100 dilution and were allowed to grow at 27^{o}C, the temperature used for mosquito maintenance. Indeed, this antibiotic treatment was efficient in reducing the resident mosquito gut microbiota, as considerably less colony-forming units grew in the plates corresponding to antibiotic treated mosquitoes, compared to plates corresponding to mosquitoes kept untreated (Figure 4.1A). Nevertheless, the colony-forming units in the antibiotic treated mosquito samples raised the possibility that this antibiotic treatment regime did not eliminate, as reported, the resident mosquito gut microbiota. Another possibility was that the observed colony-forming units in the antibiotic treated mosquito samples were a result of contamination or insufficient surface sterilization rather than remaining bacteria in the mosquito gut.

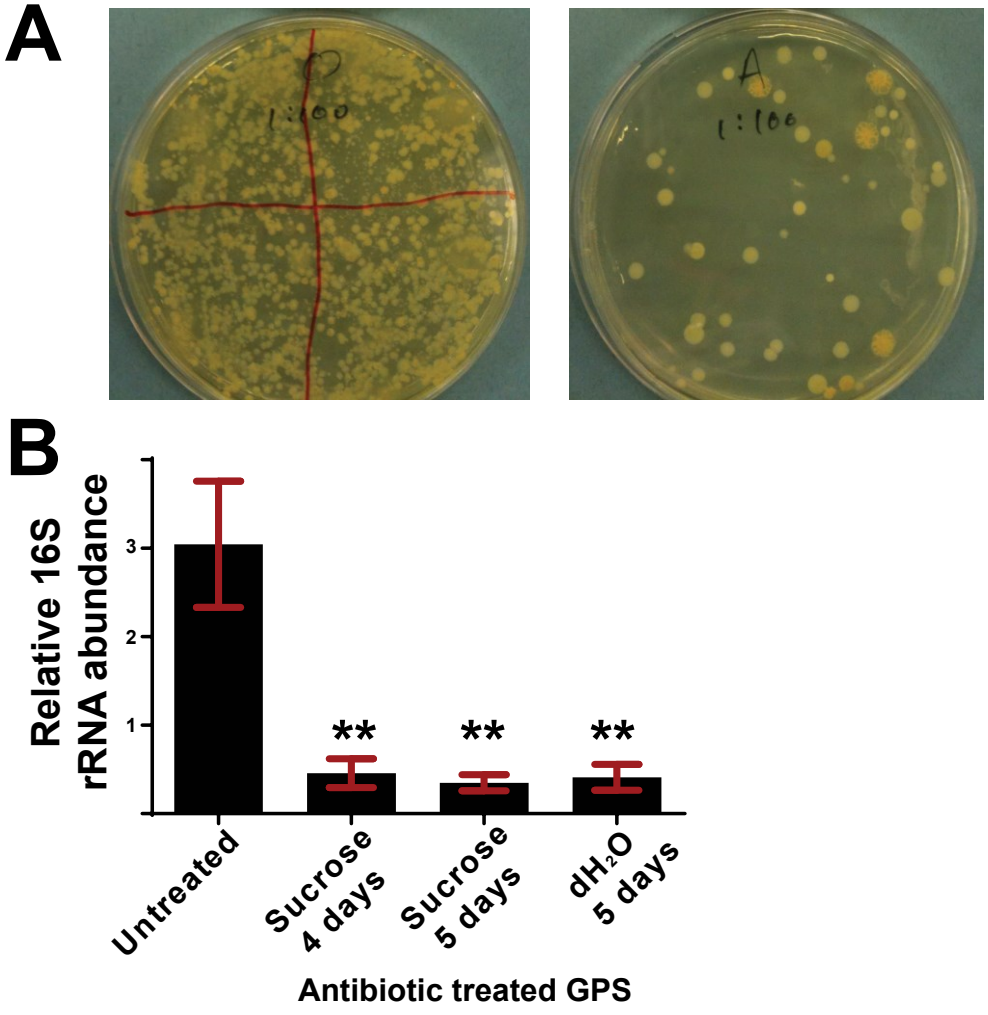

A
P
1:100
A
1:100
B
Relative 16S rRNA abundance
3
2
1
**
**
**
Untreated
Sucrose 4 days
Sucrose 5 days
dH2O 5 days
Antibiotic treated GPS
C
***
Relative 16S rRNA abundance
3
2
1
Untreated
Antibiotic treated GPS
D
***
Relative 16S rRNA abundance
300
200
100
Untreated
Antibiotic treated GPSC

Figure 4.1: Efficacy of antibiotic treatment in reducing the *An. gambiae* natural gut microbiota. A: *An. gambiae* mosquitoes were treated with gentamicin for 3 days and, subsequently, 10-15 whole mosquitoes were surface sterilized, homogenized in SOC medium and plated at a 1:100 dilution in LB agar plates. Plates were incubated at 27^{o}C for 2 days. The same procedure was applied for mosquitoes without any antibiotic treatment. In the left panel, colony-forming units grown after plating untreated mosquitoes can be seen while in the right panel, colony-forming units grown after plating antibiotic treated mosquitoes can be seen. **B:** Mosquitoes antibiotic treated with a cocktail of gentamicin, penicillin and streptomycin (GPS), diluted either in sugar or water, along with mosquitoes remaining untreated, were kept for 5 days. Antibiotic treated mosquitoes with GPS diluted in sugar were also kept for only 4 days. Subsequently, 10-15 whole mosquitoes corresponding to each treatment were surface sterilized and used for gDNA extraction, which was used as template in a qPCR with 16S broad-range bacterial primers. The housekeeping *AgS7* gene was used as control. Bacterial load ±SEM in 4 independent treatments for untreated mosquitoes or mosquitoes antibiotic treated with GPS in sucrose for 4 or 5 days or in water for 5 days can be seen, with the qPCR reaction performed in duplicate. Asterisks over each bar corresponding to a different antibiotic treatment regime indicate significant differences, with a p-value <0.005, in a Mann-Whitney non-parametric test comparing bacterial load in pools of the untreated sample to the respective antibiotic treatment sample. **C:** *An. gambiae* mosquitoes were antibiotic treated for 5 days with a GPS cocktail. Subsequently, 10-15 mosquitoes were surface sterilized and their guts were dissected, homogenized and total RNA was extracted and further used for cDNA synthesis. Mosquitoes kept untreated were processed in the same way. Subsequently, cDNA from antibiotic treated or untreated mosquito pools was used in a qRT-PCR using broad-range bacterial 16S primers while *AgS7* primers were used as controls. Bacterial load ±SEM in 3 independent assays, with the qRT-PCR reaction performed at least in duplicate for each assay, can be seen. Asterisks indicate significant differences, with a p-value <0.0005, in a Mann-Whitney non-parametric test between the bacterial load of untreated and antibiotic treated mosquitoes. **D:** Mosquitoes treated with a GPS antibiotic cocktail that also included ciprofloxacin (GPSC), along with mosquitoes kept untreated, were processed as GPS mosquitoes above. Bacterial load ±SEM over 3 independent assays, with the qRT-PCR performed at least in duplicate, can be seen for untreated or antibiotic treated mosquitoes. Asterisks indicate significant differences, with a p-value <0.0005, in a Mann-Whitney non-parametric test between the bacterial load of untreated and antibiotic treated mosquitoes.

The antibiotic treatment regime, further modified as reported in the Dong study (Dong et al., 2009), was also assayed for its efficacy, employing PCR-based methodology. Mosquitoes were antibiotic treated with a cocktail containing gentamicin, penicillin and streptomycin diluted in a sugar solution or a water solution, as described above. The efficacy of antibiotic treatment was assayed either at day 4 or 5 post mosquito emergence, for the antibiotic cocktail diluted in sugar solution, or at day 5 for the antibiotic cocktail diluted in water. Again, whole mosquitoes were surface sterilized and gDNA was extracted and used as template for qPCR using broad-range 16S primers, targeting the bacterial 16S ribosomal DNA. As shown in Figure 4.1B, antibiotic treatment was again efficient in reducing gut bacteria in antibiotic treated mosquitoes by 85% to 89%, compared to untreated controls kept on a sugar solution. The presence of residual bacteria in antibiotic treated mosquitoes was detected through the identified bacterial load in the antibiotic treated mosquito samples. Furthermore, no difference was observed between antibiotic dilution in sugar or water, a methodological difference between the Dong and Meister studies (Dong et al., 2009; Meister et al., 2009) or between 4 or 5 day treatment for antibiotics diluted in a sugar solution. Taken together, these results confirm the reduction but not the elimination of the resident bacteria in the mosquito gut using previously reported (Dong et al., 2009; Meister et al., 2009) methodologies.

Another possibility, though, is that the observed bacterial load shown in antibiotic treated mosquitoes using PCR-based methods was a result of dead bacteria remaining in the gDNA pools. To further clarify whether antibiotic treatment reduces or eliminates resident gut bacteria, mosquitoes were treated with a gentamicin, penicillin and streptomycin cocktail diluted in sugar and after 5 days mosquitoes were surface sterilized, their guts were dissected and total RNA was extracted for cDNA synthesis used to assay bacterial load in a qRT-PCR using 16S primers. The acquisition of cDNA compared to gDNA as above excludes the possibility that ribosomal DNA from dead bacteria can be traced while the extraction of guts also excludes the possibility that bacteria traced are located in gonads or salivary glands that antibiotics may not reach. As shown in Figure 4.1C, antibiotic treatment significantly reduced the resident gut bacteria by about 82% but did not eliminate them from the mosquito gut, thus strengthening the possibility that the antibiotic treatment regime used is not able to completely eliminate the mosquito gut flora.

To confirm the partial efficacy of the antibiotic treatment used, the antibiotic ciprofloxacin, which is typically used to eliminate mycoplasma from cell cultures (Fleckenstein, 1996), was added in the antibiotic regime used. Again efficacy of antibiotic treatment was assayed as in Figure 4.1C. This time the combination of gentamicin, penicillin and streptomycin with ciprofloxacin was able to almost eliminate

bacteria from the guts of antibiotic treated mosquitoes (Figure 4.1D). The higher efficacy of the antibiotic treatment regime that includes ciprofloxacin suggests that the previously reported gentamicin, penicillin and streptomycin combination is not able to eliminate mosquito gut bacteria, at least in the laboratory mosquito colony used in this study.

Despite the higher efficacy of ciprofloxacin addition in reducing the resident gut bacteria in mosquitoes used in this study, the antibiotic treatment regime further used in this study employed only gentamicin, penicillin and streptomycin, as reported previously (Dong et al., 2009; Meister et al., 2009), with minor modifications, using a cocktail of gentamicin, penicillin and streptomycin diluted in a sugar solution and provided to mosquitoes for 5 days.

4.5 Dynamics of *An. gambiae* oral infection with *Asaia*

Antibiotic treated *An. gambiae* mosquitoes were orally infected with fluourescent strains of bacteria of the genus *Asaia*. Bacteria-fed mosquitoes were separated 2 days post infection and mosquitoes were further kept on a sugar solution containing kanamycin for the selection of the fluourescent *Asaia* strain. From day 2 to day 6 post infection, *Asaia*-infected mosquitoes were dissected and their guts were observed under a fluourescent microscope (Figure 4.2A). Considerable variation was observed regarding the achieved level of infection. The guts of some mosquitoes were highly infected with *Asaia*, with intense fluourescence spanning most areas of the gut. Other mosquitoes displayed an intermediate level of infection; fluourescence was evident but restricted to a part or some parts of the gut. Mosquitoes exhibiting this intermediate level of infection will henceforth be referred to as lowly infected, even if fluourescence was evident in several parts of the gut but was distinguishably lower than highly infected mosquitoes. Finally, some mosquitoes showed no sign of fluourescence in their gut despite ingesting *Asaia*-containing sugar. These mosquitoes will be further referred to as non-infected.

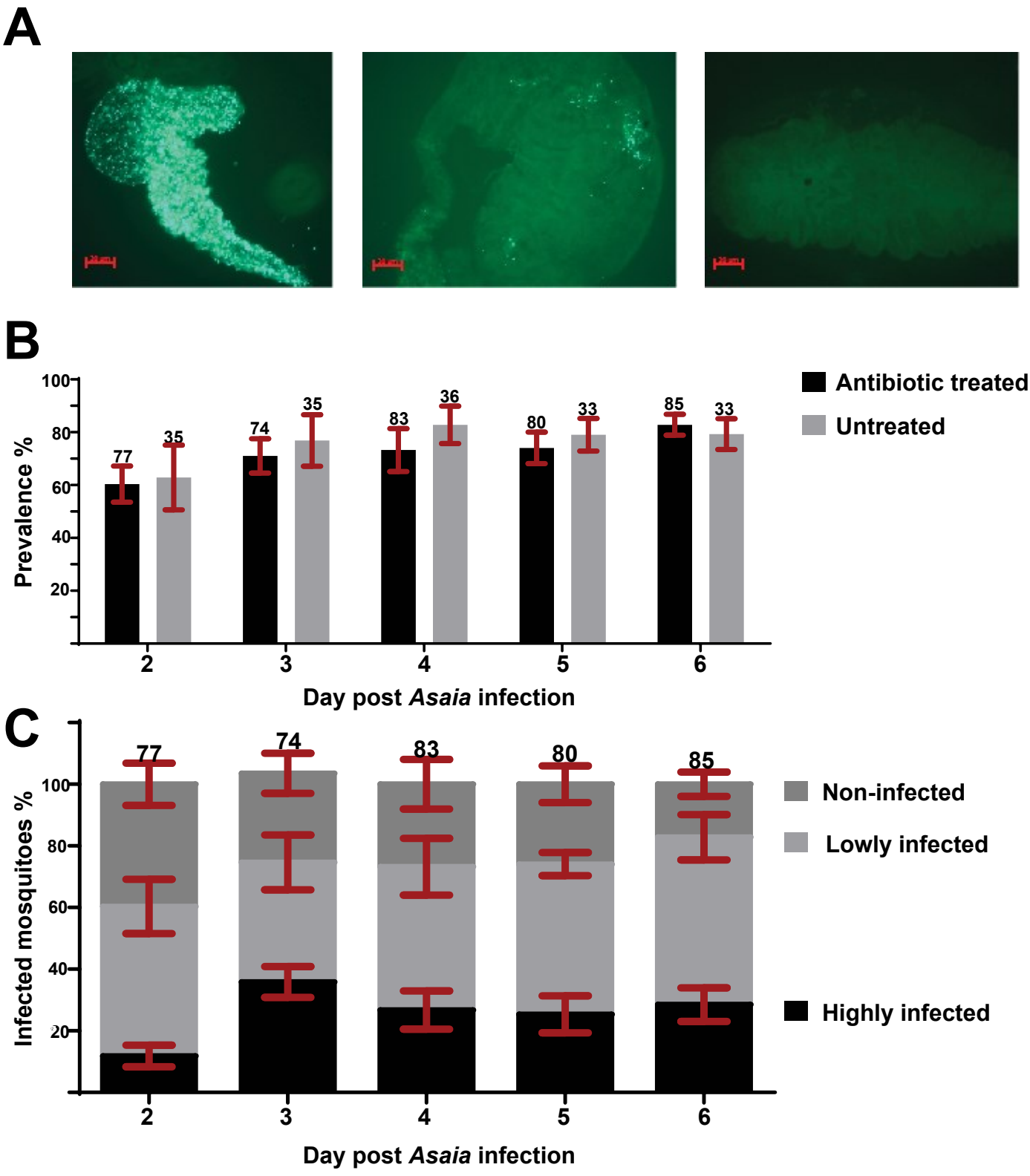

A
B
Prevalence %
100
80
60
40
20
77
35
74
35
83
36
80
33
85
33
2
3
4
5
6
Day post Asaia infection
Antibiotic treated
Untreated
C
Infected mosquitoes %
100
80
60
40
20
77
74
83
80
85
2
3
4
5
6
Day post Asaia infection
Non-infected
Lowly infected
Highly infected

Figure 4.2: Dynamics of *An. gambiae* oral infection with *Asaia*. *An. gambiae* antibiotic treated mosquitoes were orally infected with the *SF2.1 (GFP) Asaia* strain. Bacteria-fed mosquitoes were separated at day 2 post infection and, subsequently, the level of infection was determined from day 2 to day 6 post infection following midgut dissection and microscopic observation under a GFP fluourescence microscope. **A:** Mosquitoes showed great variation regarding the outcome of infection with highly infected mosquitoes (left panel), in which intense fluourescence was evident throughout the gut, lowly infected mosquitoes (middle panel), in which the level of fluourescence was distinguishably lower than highly infected mosquitoes, and non-infected mosquitoes (right panel), in which no sign of fluourescence was traced despite ingestion of bacteria-containing sugar. **B:** The prevalence of infection was determined for *Asaia* infected mosquitoes, which were either antibiotic treated with gentamycin, penicillin and streptomycin for 5 days or were kept untreated, as they were fed with a sugar solution for 5 days prior to infection. The percentage of *Asaia*-infected mosquitoes showing some sign of fluourescence in their mosquito gut when dissected, from day 2 to day 6 post *Asaia i*nfection, was determined in 4 independent assays and the mean prevalence ±SEM can be seen for each time point, either for initially antibiotic treated or untreated mosquitoes. The total number of mosquitoes dissected at each time point can be seen over each respective bar. **C:** The proportions of highly, lowly and non-infected mosquitoes were determined for *Asaia* infected mosquitoes that were initially antibiotic treated and whose prevalence over 4 independent assays was shown above. At each time point, from day 2 to day 6 post *Asaia* infection, the mean percentage ±SEM of highly, lowly or non-infected mosquitoes can be seen.

The proportions of highly, lowly and non-infected mosquitoes following *Asaia* infection were quantified from day 2 to day 6 post infection (Figure 4.2B). Each day, mosquitoes were dissected and the level of fluourescence in their gut was determined through microscopic observation. The proportions determined over 4 independent infections were used to assess the prevalence of infection, i.e. the proportion of mosquitoes with evident GFP fluourescence suggesting the presence of *Asaia* in their gut, being either highly or lowly infected. The prevalence of infection showed some variation from day 2 to day 6 post infection, ranging from about 60% at day 2 to about 80% at day 6. Mean prevalence showed an increase from 60.3% at day 2 to 71% at day 3, subsequently showed minor variations from day 3 to day 5, from 71% at day 3 to 74% at day 5, while at day 6 mean prevalence increased to 82.8%.

To assess whether the presence of the natural gut microbiota affects the outcome of *Asaia* infection, untreated mosquitoes retaining their natural gut microbiota were orally infected with *Asaia*. Although kanamycin was added for the selection of the *Asaia* fluourescent strain following infection, these mosquitoes were thought to retain a higher number of midgut bacteria despite kanamycin treatment compared to mosquitoes initially antibiotic treated with a gentamicin, penicillin and streptomycin cocktail. The prevalence of infection was determined as above from day 2 to day 6 post infection and the results were highly similar to mosquitoes that were initially antibiotic treated, despite the smaller sample size (Figure 4.2B). The mean prevalence of infection over 4 independent assays showed an increase from 62.8% at day 2 to 76.8% at day 3 and stayed relatively stable until day 6 post infection, ranging from 79% to 82%. These results suggest that the initial antibiotic treatment seems not to significantly affect the dynamics of *Asaia* infection, from day 2 to day 6 post infection.

The proportions of highly, lowly and non-infected mosquitoes for *Asaia*-infected mosquitoes that were initially antibiotic treated can be seen in Figure 4.2C, from day 2 to day 6 post infection. The proportion of highly infected mosquitoes increased substantially from a mean of 11.8% at day 2 to 35.8% at day 3 and stayed relatively stable from day 4 to day 6, ranging from 25% to 28%. The mean proportion of non-infected mosquitoes inversely followed the prevalence reported above, as prevalence includes both highly and lowly infected mosquitoes, by remaining relatively stable from day 3 to day 5 post infection, within the range of 26% to 29%.

These results indicate that bacteria of the genus *Asaia* are able, as expected, to colonize the midgut of *An. gambiae*. The outcome of infection is highly variable, though, with highly, lowly and non-infected mosquitoes. This variability was quantified as proportions of highly, lowly and non-infected mosquitoes

and was shown to be relatively stable, at least from day 3 to day 5 post infection, while the presence of resident gut microbiota, at least to some extent, did not seem to influence the dynamics of *Asaia* infection.

4.6 Oral *Asaia* infection reduces mosquito fitness

To examine whether oral infection with *Asaia* affects mosquito fitness, the survival of *Asaia*-infected mosquitoes was compared with uninfected mosquitoes that remained antibiotic treated. So as to track survival from the day of infection, separation of *Asaia*-fed mosquitoes was not performed at day 2 post infection as previously. The number of dead mosquitoes was tracked daily and dead mosquitoes were removed, until all mosquitoes were dead. Survival was tracked for a total of 272 antibiotic treated and 404 *Asaia* infected mosquitoes, over 4 independent assays. The results were used to create Kaplan-Meier survival curves for antibiotic treated and *Asaia*-infected mosquitoes, which are shown in Figure 4.3. Indeed, the survival curves of antibiotic treated and *Asaia*-infected mosquitoes were significantly different, as determined by the Mantel-Cox logrank as well as the Gehan-Breslow-Wilcoxon test, with the latter conferring extra statistical weight in earlier time points (Jung and Jeong, 2003). Therefore, *Asaia* infection was shown to result in reduced longevity of infected mosquitoes, compared to uninfected mosquitoes, which were also antibiotic treated. Interestingly, even antibiotic treated mosquitoes that remained uninfected showed considerable mortality, albeit lower than *Asaia* infected mosquitoes, with 40% survival at day 10 post infection (day 15 post emergence) and less than 10% survival at day 15 post infection (day 20 post emergence), probably due to the conditions of the assay.

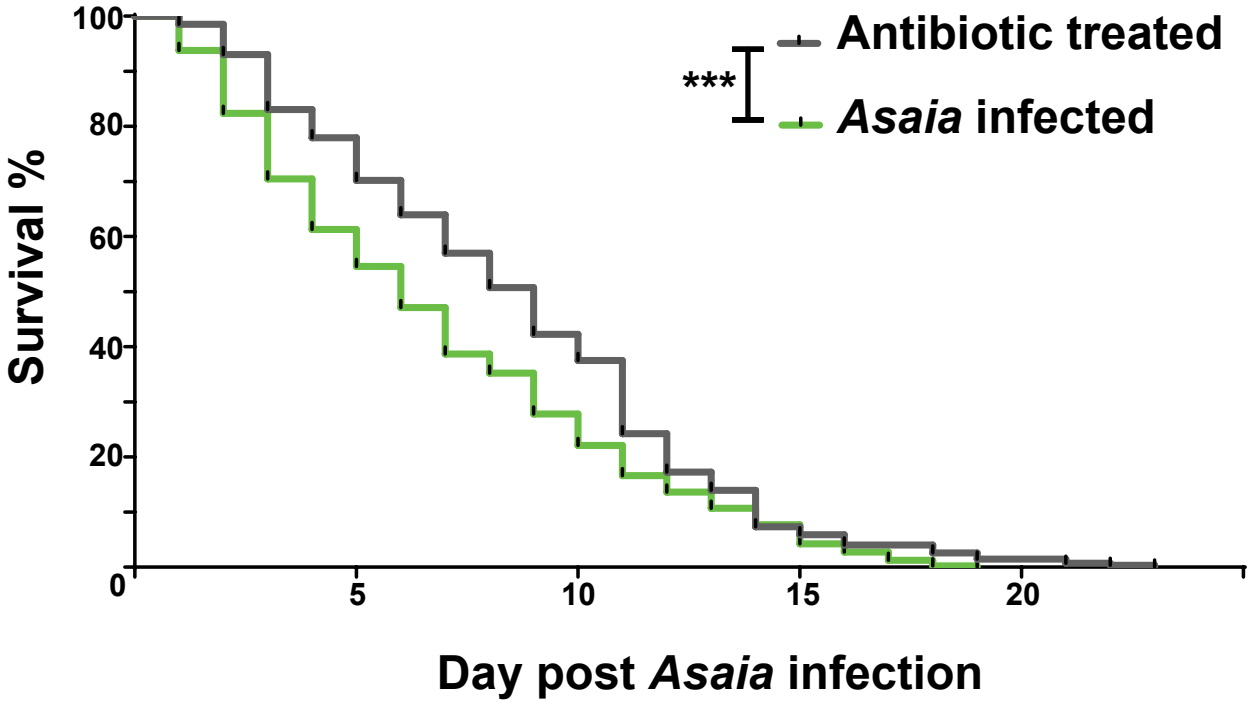


Figure 4.3: Survival curves of *Asaia* infected and uninfected mosquitoes. Antibiotic treated mosquitoes were infected with *Asaia* or kept uninfected and their survival was tracked daily with dead mosquitoes removed until all mosquitoes were dead. Bacteria-fed mosquitoes were not separated so that survival was tracked from day 1 post infection. *Asaia* infected and uninfected mosquitoes were kept in an incubation room which showed fluctuations in temperature and humidity that might have affected survival for both treatments. Kaplan-Meier curves were generated after tracking 404 *Asaia* infected and 272 antibiotic treated mosquitoes, over 4 independent assays. Significant differences between the two curves were assessed using the Mantel-Cox logrank and the Gehan-Breslow-Wilcoxon tests and in both cases the computed p-value was <0.0005, indicated by 3 asterisks.

4.7 Dynamics of *An. gambiae* oral infection with *S. marcescens*

Antibiotic treated mosquitoes were orally infected with the *Db11-GFP* fluourescent strain of *S. marcescens*, in the same way described above for *Asaia*. Tetracycline and carbenicillin were added for the selection of the *S. marcescens* fluourescent strain, instead of the kanamycin used for *Asaia*. Considerable variation was again observed regarding the outcome of *S. marcescens* infection, with highly, lowly and non-infected mosquitoes (Figure 4.4A).

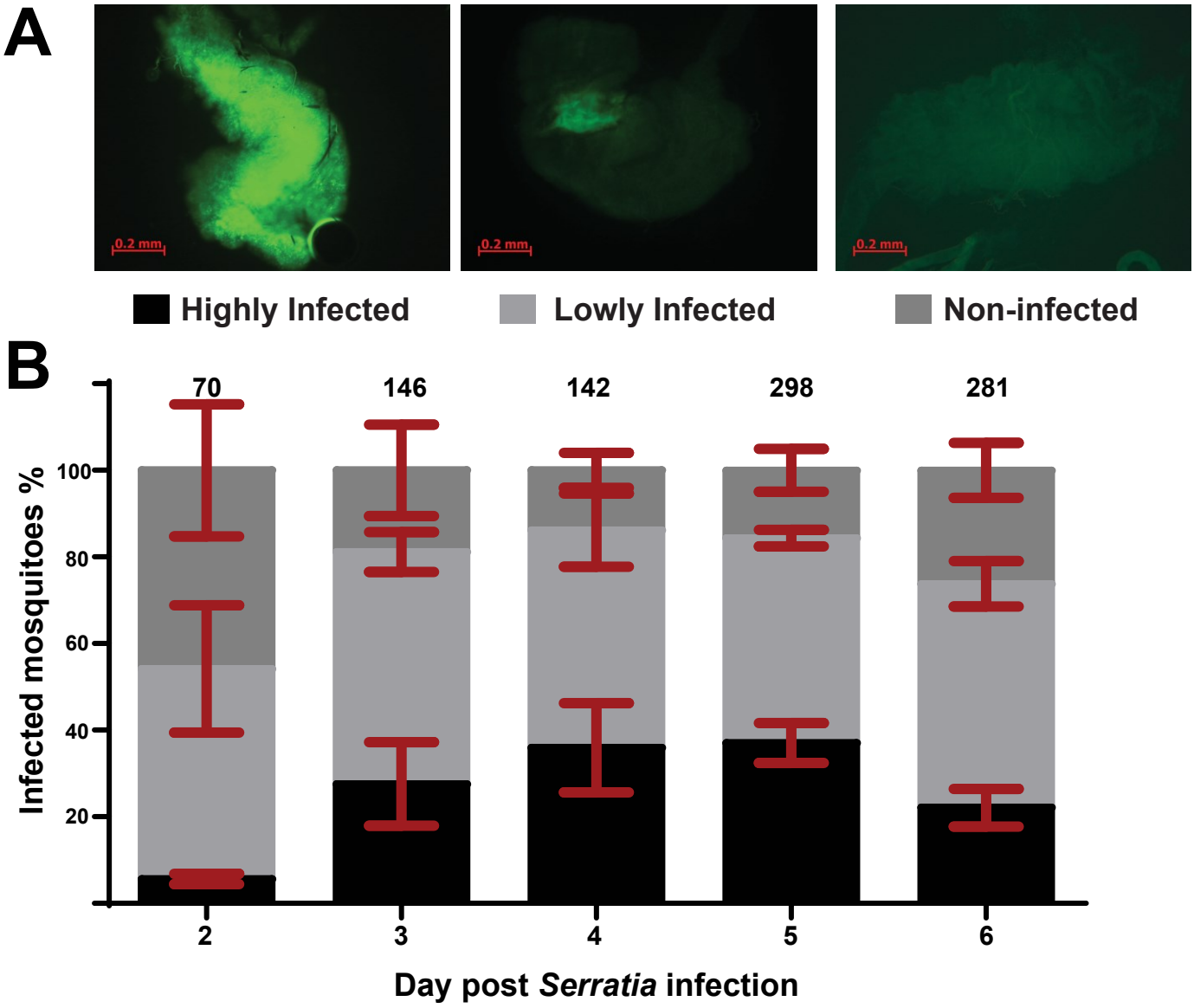


Figure 4.4: Dynamics of *An. gambiae* oral infection with *S. marcescens*. Antibiotic treated mosquitoes were orally infected with the fluourescent *Db11-GFP* strain of *S. marcescens* and the level of infection was tracked from day 2 to day 6 post infection through midgut dissection of mosquitoes and microscopic observation under a GFP fluourescence microscope. **A:** Great variation was observed regarding the outcome of *S. marcescens* infection with highly infected mosquitoes (left panel), lowly infected mosquitoes (middle panel) or non-infected mosquitoes (right panel). **B:** The proportions of highly, lowly and non-infected mosquitoes following *S. marcescens* infection were tracked from day 2 to day 6 post infection, in the same way described above for *Asaia.* The mean percentage ±SEM of highly, lowly and non-infected mosquitoes can be seen for each time point, as determined over 4 independent infections. The total number of mosquitoes dissected at each time point so as to determine the level of *S. marcescens* gut infection can be seen over each respective bar.

The proportions of highly, lowly and non-infected mosquitoes following *S. marcescens* infection were tracked from day 2 to day 6 post infection. The sample size used in this case was considerably larger for time points from day 3 to 6 post infection, compared to the respective sample sizes used to track the dynamics of *Asaia* infection (Figure 4.2B-C). Over 4 independent infections, 146 and 142 mosquitoes were dissected at day 3 and 4 post infection, respectively, while for day 5 and 6 post infection, the sample size was even higher with 298 and 281 dissected mosquitoes, respectively. The mean proportion of highly infected mosquitoes considerably increased from 5.6% at day 2 to 27.5% at day 3 post infection, stayed relatively stable until day 5 post infection, with gradual increases to 35.9% at day 4 and 37% at day 5 post infection and decreased to 22% at day 6 post infection (Figure 4.4B). Non-infected mosquitoes showed inverse dynamics compared to highly infected mosquitoes; they decreased considerably from 45.8% at day 2 to 19.9% at day 3 post infection, stayed relatively stable at about 15% until day 5 post infection and subsequently increased to 26.2% at day 6 post infection.

Although the differences in sample size can complicate comparisons between time points or between infection dynamics of *Asaia* and *S. marcescens* infected mosquitoes, it is notable that the dynamics between day 4 and day 5 post *S. marcescens* infection are remarkably similar despite the sample size difference while between day 5 and day 6 post *S. marcescens* infection there are notable differences despite similar sample sizes (Figure 4.4B). Therefore, although increased sample size enhances statistical power, with lower margins of error, even lower sample sizes were thought to provide relatively robust estimates for the dynamics of infection at the given time point.

4.8 Oral *S. marcescens* infection reduces mosquito fitness

The effect of *S. marcescens* oral infection on mosquito fitness was also examined (Figure 4.5). As antibiotic treated mosquitoes showed previously relatively high mortality rates (Figure 4.3), both untreated and antibiotic treated mosquitoes were used this time to examine whether antibiotic treatment affects survival. Furthermore, the assay was performed in a dedicated incubator with much more limited temperature and humidity fluctuations that might affect survival. Therefore, survival of uninfected mosquitoes was much improved compared to the *Asaia* survival assay (Figure 4.3) due either to the use of an incubator or due to differences in the fitness of the mosquito colony used which could also fluctuate in different time periods. The survival for a total of 101 antibiotic treated, 381 untreated and 259 *S. marcescens* infected mosquitoes was assayed over 2 independent assays. As shown in Figure 4.5, no significant differences were observed between antibiotic treated mosquitoes and mosquitoes retaining

their natural gut microbiota, while overall mortality was considerably reduced, with about 80% of antibiotic treated or untreated mosquitoes alive at day 15 post emergence and about 60% alive at day 20 post emergence. Oral infection with *S. marcescens* was shown to significantly reduce longevity, compared to antibiotic treated or untreated mosquitoes, as determined by both the Mantel-Cox logrank and the Gehan-Breslow-Wilcoxon tests (Figure 4.5). *S. marcescens* infected mosquitoes showed an about 40% survival at day 10 post infection, compared to about 80% for uninfected mosquitoes while at day 15 post infection survival was down to about 20% for *S. marcescens* infected mosquitoes compared to about 60% for uninfected mosquitoes. The methodological differences between the two survival assays preclude any comparison between the effects of *Asaia* and *S. marcescens* oral infection on mosquito fitness.

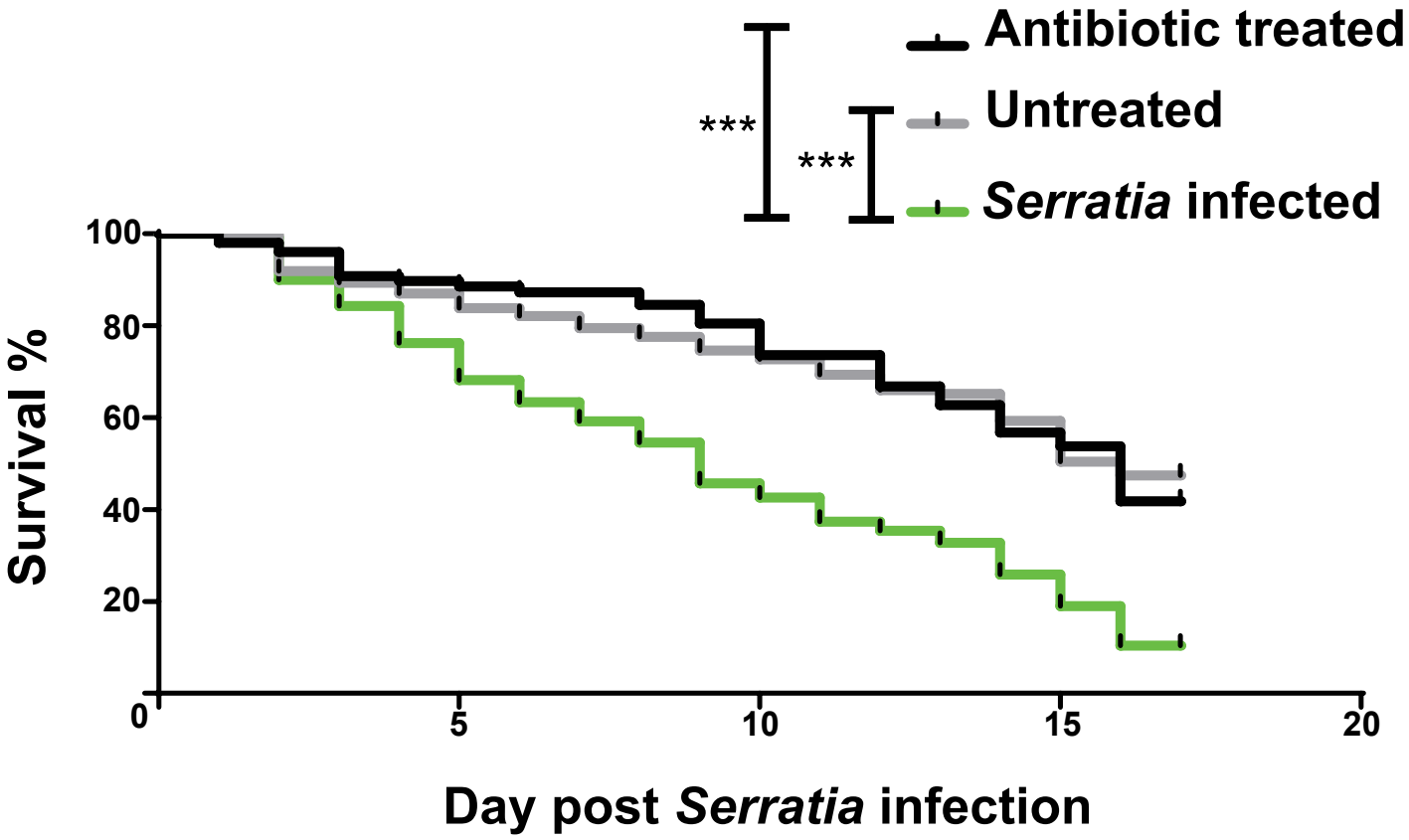


Figure 4.5: Survival curves of *S. marcescens* infected and uninfected mosquitoes. Antibiotic treated mosquitoes were orally infected with *S. marcescens* and their survival was tracked along with antibiotic treated and untreated mosquitoes that were kept uninfected up to 17 days post infection. The assay was performed in a dedicated incubator. Bacteria-fed mosquitoes were not separated and survival was tracked starting at day 1 post infection. Kaplan-Meier curves were generated after tracking survival for 259 *S. marcescens* infected, 101 uninfected and antibiotic treated and 381 uninfected and untreated mosquitoes, over 2 independent assays. Significant differences between survival of *S. marcescens* infected mosquitoes and either antibiotic treated or untreated mosquitoes were determined using both the Mantel-Cox logrank and the Gehan-Breslow-Wilcoxon tests. In both cases and for both tests, a p-value <0.0005 was computed, indicating significant differences between the assayed survival curves, denoted by three asterisks.

Discussion

4.9 Oral bacterial infections and immunity

The objective of this part of the project was to establish a model of oral bacterial infections in *An. gambiae* that could be further used to investigate mosquito antibacterial responses, with a particular interest in homeostatic interactions that determine the load and composition of the mosquito gut microbiota and the factors underlying the broad variation observed in mosquito gut flora at the individual and population levels. The dissection of such complex interactions between mosquito immunity and diverse populations of commensal bacteria led to the adoption of the concept of gnotobiotic model systems used in mice and elsewhere, in which the natural microbiota are replaced by a known (gnoto: from the Greek γνώσις, meaning knowledge) single bacterial strain (Falk et al., 1998).

Similar approaches have been adopted in the study of epithelial immune responses in *Drosophila*, in which flies are orally infected with single strains of Gram-negative bacteria (Buchon et al., 2010; Buchon et al., 2009; Chakrabarti et al., 2012; Cronin et al., 2009; Nehme et al., 2007). These studies were able to provide a detailed knowledge of intestinal responses to Gram-negative bacteria by adopting oral infection models with *S. marcescens* (Cronin et al., 2009; Nehme et al., 2007), *Erwinia carotovora* (Buchon et al., 2010; Buchon et al., 2009) or *Pseudomonas entomophila* (Chakrabarti et al., 2012). Importantly, though, these oral infection studies in *Drosophila* were considerably different from a gnotobiotic approach. These studies in *Drosophila* did not aim to clear the resident gut flora and replace it with a single bacterial strain but relied on massive natural infections of wild-type flies with bacteria. As flies were inundated with a high concentration of a bacterial strain added to their food source, the aim was not to track the variation in the outcome of infection or the colonization ability of these strains but rather track transcriptional responses to such natural infection or the effect of silencing a given gene in the fly's survival after infection due to its involvement in immunity or gut homeostasis.

A different approach was adopted in the study of the effect of *Caudal* on the fly's gut homeostasis (Ryu et al., 2008), in which germ-free flies were used for the introduction of a single *Acetobacteriaceae* strain either at the embryonic stage or through oral infection in adult flies. This approach enabled the determination of the pathogenicity of different *Acetobacteriaceae* strain when dominant in the fly's gut and the mechanism through which tight *Caudal*-mediated regulation of antimicrobial peptide production, via the Imd pathway, controls their prevalence. This study provided a proof-of-concept validation of how a gnotobiotic approach can provide insights on how gut homeostasis is determined through an interplay

of immunity, interactions between members of the gut microbiota and the importance of the achieved balance in the microbial population structure for the fly's survival.

Importantly, though, all above studies in *Drosophila* employed fly lines with limited genetic diversity. Such genetic homogeneity undoubtedly enhances reproducibility of the observed experimental outcomes but misses out on investigating gut microbiota variation present in field fly populations, which is similar to mosquitoes (Chandler et al., 2011; Cox and Gilmore, 2007; Wong et al., 2013). This ecological aspect of host-microbe interactions is completely missing from these oral infection models in *Drosophila*.

4.10 Oral bacterial infections and colonization ability

Models of oral bacterial infection in several organisms have been employed to study the colonization efficiency of certain bacterial strains and its effects on the host rather than focus on immunity or gut homeostatic interactions. In *Drosophila*, a model of oral infections following antibiotic treatment of flies with bacteria of the genus *Enterococcus* focused on the colonization ability of these native members of the fly's gut microbiota and its pathogenicity upon the host (Cox and Gilmore, 2007). Similar studies have been performed in the desert locust *Schistocerca gregaria* using *S. marcescens* (Dillon and Charnley, 2002). In mosquitoes, studies of oral infection with bacterial strains also focused on their colonization ability, which would make such strains suitable in the implementation of paratransgenic approaches (Damiani et al., 2010; Favia et al., 2007; Lindh et al., 2006; Terenius et al., 2012). As bacteria used for paratransgenesis should be efficient colonizers under natural settings, competing with a diverse population of resident gut bacteria in field mosquitoes, these studies did not aim at clearing the resident gut microbiota and replacing them with a single strain but rather investigated the colonization ability of these bacteria in the presence of the natural gut microbiota. Favia and colleagues (Favia et al., 2007) employed fluourescent strains of bacteria of the genus *Asaia* to examine its colonization efficiency in *An. stephensi*, identifying stable *Asaia* colonization in the midgut and salivary glands up to 25 days post oral infection. A similar study showed the ability of *Asaia* to colonize *An. gambiae* for up to 20 days post oral infection with *Asaia* fluourescent strains (Damiani et al., 2010). A similar study in *Ae. aegypti*, utilizing oral infection with various bacterial strains, also traced the dynamics of infection and showed that host-microbe co-adaptation is correlated with colonization efficiency (Terenius et al., 2012).

4.11 An integrative model of oral bacterial infections in *An. gambiae*

The approach of oral bacterial infections in *An. gambiae* adopted here combines elements of the studies mentioned above and thus innovates in several ways. On the one hand, oral infections of two common members of the mosquito gut microbiota, *Asaia* and *S. marcescens*, are utilized to dissect mosquito antibacterial responses and factors that lead to gut homeostasis, along the lines of similar studies in *Drosophila*. At the same time, the utilization of fluourescent bacterial strains for oral infections in *An. gambiae* mosquitoes, in which the natural gut microbiota have been previously reduced through antibiotic treatment, allows tracking of the dynamics of infection and correlation of these dynamics with underlying immune and homeostatic responses. Finally, the use of a recently established laboratory mosquito colony, so as to retain much of the genetic variation of field mosquito populations, allows the study of ecological aspects of gut microbiota variation and correlation of these aspects with infection dynamics and underlying immune and homeostatic forces. Therefore, the model system of oral bacterial infections utilized here allows the study of mosquito epithelial responses against bacteria while integrating aspects of infection dynamics and ecological variation which remain poorly understood.

The approach adopted here provided an efficient and robust way to study infection dynamics and epithelial responses through the use of fluourescent strains of *Asaia* and *S. marcescens*. The use of an oral infection model provides a close approximation of bacterial infections and corresponding responses under natural settings and constitutes the only way to study epithelial responses compared to bacterial injections in the mosquito haemolymph that bypass epithelial immunity and directly trigger systemic immune responses (Schnitger et al., 2009). Although a control for mosquitoes that did not feed with bacteria-containing sugar was implemented, the sugar meal size of bacteria-fed mosquitoes can be variable, as happens in natural infections and thus the number of bacteria taken up cannot be precisely quantified, as in bacterial injections. This inevitable level of variation can only be controlled, as in studies employing blood feeding of mosquitoes with *Plasmodium* parasites, as a mosquito blood meal can also be of variable size (Cator et al., 2013), by removing non-fed mosquitoes and through averaging out differences by a large sample size.

Other conceivable approaches to tackle the same questions in *An. gambiae* would be impractical or even impossible to implement compared to the approach adopted here. One possibility would be to eliminate bacteria in the mosquito eggs or larvae and infect with a single bacterial strain so as to generate monoassociated mosquitoes, as has been previously done in *Drosophila* (Ryu et al., 2008). This approach cannot be used for the study of triggered epithelial responses due to oral bacterial infection in adults

while it is impractical to implement in mosquito colonies kept in large numbers. Differences in mosquito physiology can also complicate such approach as the larval microbiome is much different compared to the adult microbiome in mosquitoes and also plays important roles in larval development (Chouaia et al., 2012; Rani et al., 2009).

Another possibility would be to track bacterial dynamics in single mosquitoes throughout the course of infection. Indeed, that would be a more straightforward way to track infection dynamics rather than dissection of different mosquitoes for each time point. In such assay, statistical power would be enhanced as all mosquitoes would log an infection load at each time point and not only the fraction of mosquitoes dissected each day, as in the approach followed here. Furthermore, it would be easier to track changes in the level of infection throughout the course of the assay, instead of comparing time points with different dissected mosquitoes, thus reducing the statistical power of such comparison and also requiring large sample sizes. This could not have been possible in the implemented model as it was required for the mosquito to be killed and its gut dissected so as to determine the level of infection. One possibility would have been to construct bioluminescent bacteria and track the level of infection in a way similar to murine infection models (Johnson et al., 2013; Leevy et al., 2007). Regrettably, such model of infection could not have been implemented in mosquitoes; beside the construction of bioluminescent bacteria and the difficulties of background luminescence due to oral infection either in other tissues except the gut or outside the mosquito, the main obstacle was that the difference in size between mice and mosquitoes precluded the use of existing equipment designed to track such infections in mice. Therefore, it would have been infeasible to implement such assay in a meaningful way.

4.12 The implementation of antibiotic treatment prior to oral infection

The approach adopted here included the clearance of the natural gut microbiota through antibiotic treatment prior to oral infection with *Asaia* or *S. marcescens*. As mentioned previously, most studies implementing oral bacterial infections either to study epithelial immune responses or infection dynamics did not make use of clearance of gut bacteria prior to oral infection (Buchon et al., 2010; Buchon et al., 2009; Chakrabarti et al., 2012; Cronin et al., 2009; Damiani et al., 2010; Favia et al., 2007; Nehme et al., 2007). The main motivation to clear the mosquito's natural gut bacteria here was to eliminate interactions between bacteria that would detract from the reproducibility of experiments due to variation in the load and composition of the gut microbiota. Furthermore, as antibiotic treatment of mosquitoes increases the intensity of *Plasmodium* infections (Dong et al., 2009; Meister et al., 2009), oral bacterial infection of

previously antibiotic treated mosquitoes could have been used to discriminate whether the strains used for infection or strains inhabiting the mosquito gut prior to antibiotic treatment are responsible for triggering this anti-*Plasmodium* effect.

As shown in Figure 4.1, antibiotic treatment regimes previously reported to eliminate mosquito gut bacteria (Dong et al., 2009; Meister et al., 2009), had partial efficacy as they reduced the natural gut bacteria in the mosquito colony used here by about 80% to 85%. Differences in the load and composition of the mosquito gut microbiota and acquisition of antibiotic resistance by bacterial strains can account for this difference. High-throughput sequencing of the gut microbiota inhabiting the mosquito colony used here in different time periods, reported later on in this study (chapter 7, Figure 7.8), suggests that the most likely candidates to exhibit antibiotic resistance are bacteria belonging to the *Flavobacteriaceae* family, most likely of the genus *Elizabethkingia*, known to display multiple resistance to several antibiotics including ampicillin, kanamycin and tetracycline (Kampfer et al., 2010).

The confirmation of the presence of bacteria evading the antibiotic treatment regime used here was based on the addition of ciprofloxacin to this regime that enabled clearance of the mosquito gut microbiota in antibiotic treated mosquitoes. Nevertheless, the antibiotic treatment regime implemented here did not make use of ciprofloxacin. The reason for that was that several experiments have already been performed using the gentamicin, penicillin and streptomycin antibiotic cocktail, therefore, for reasons of consistency, the antibiotic regime could not have been modified later on. Furthermore, as the use of ciprofloxacin has not been previously reported in mosquitoes, side-effects on mosquito physiology or delayed clearance that could interfere with subsequent bacterial infections would have been much more detrimental to the study, compared to the benefit of total clearance of the mosquito gut microbiota. In any case, this initial step of antibiotic treatment was not shown to be essential, with the dynamics of, at least, *Asaia* infection were shown not to be influenced by residual bacteria evading the initial antibiotic treatment. Therefore, this initial antibiotic treatment step was employed as an additional control to limit the gut microbiota and possible background interactions that would reduce reproducibility, it was not implemented in several previous studies and the partial efficacy of the antibiotic treatment did not seem to influence the experimental outcomes reported here.

4.13 Great variation in the outcome of oral *Asaia* infection

Oral infection with *Asaia* confirmed that *Asaia* can indeed efficiently colonize *An. gambiae* mosquitoes. From day 3 to day 5 post infection, a mean of 71% to 74% of infected mosquitoes exhibited signs of

fluourescent *Asaia* in their gut while at day 6 mosquitoes with *Asaia* in their gut increased to a mean of 82.8%. When mosquitoes retained their natural gut microbiota, *Asaia* colonization efficiency was remarkably similar, up to 6 days post infection, with a mean of 79% to 82% of orally infected mosquitoes bearing fluourescent *Asaia* in their gut. These results are consistent with a study tracking *Asaia* colonization in *An. gambiae* (Damiani et al., 2010). Using the same *SF2.1 (GFP) Asaia* strain as here, 50 out 72 initially untreated mosquitoes that were infected with *Asaia* showed fluourescent *Asaia* in their gut, tracked over a period of 20 days post infection, thus reaching a prevalence of *Asaia* colonization of 69.4%. The lower observed prevalence by about 10 percentage points compared to the prevalence reported here can be explained by the fact that in the Damiani study bacteria-fed mosquitoes were not separated after infection, so mosquitoes that did not feed with *Asaia*-containing sugar were included. Furthermore, the different time points used between the two studies, over 20 days in the Damiani study instead of 6 days used here could also have influenced the total observed prevalence.

The similar *Asaia* infection dynamics observed here between mosquitoes that were initially antibiotic treated and mosquitoes that retained their natural gut microbiota indicate that *Asaia* colonization efficiency is not influenced by the presence of other bacterial strains, at least for the gut microbiota composition of the mosquito colony used here. That would be expected for an efficient colonizer as *Asaia*, broadly found in lab-reared and field-collected mosquitoes (Chouaia et al., 2010). This observation also lends more credence to the possibility that the partial efficacy of antibiotic treatment hardly affected the dynamics of *Asaia*, and most likely also *S. marcescens*, infection.

One of the most interesting findings in this part of the study was that, during the course of tracking the dynamics of oral *Asaia* infection, great variation was observed regarding the outcome of infection. Some mosquito guts showed intense fluourescence covering most of the gut, indicating that fluourescent *Asaia* had essentially taken over the gut, in some cases fluourescence was distinguishably lower and in some cases there was no sign of fluourescent *Asaia*, despite the separation of sugar-fed mosquitoes which ensured that these mosquitoes had taken a sugar meal containing *Asaia*.

One possible explanation for the presence of non-infected mosquitoes could have been that *Asaia* were present but had lost their fluourescence ability. Although this possibility cannot be ruled out, it is highly unlikely. To ensure that the *Asaia* strains used for oral infection retained their fluourescence, *Asaia* cultures used were invariably initiated from *Asaia* glycerol stock, so that the possibility of the loss of the plasmid conferring fluourescence during prolonged culturing was ruled out. Furthermore, prior to infection, GFP fluourescence of the *Asaia* culture used was measured using spectrophotometry, to

confirm the fluourescence of the culture used in relation to its OD_{600} density. The same methodology was also used for *S. marcescens* infections. What is more, during infection with bacteria-containing sugar and afterwards, until 6 days post infection, mosquitoes were treated with suitable antibiotics for the selection of the fluourescent *Asaia* or *S. marcescens* strains. Thus, bacteria that had lost the plasmid conferring GFP fluourescence would be susceptible to the antibiotic treatment the mosquitoes were undertaking and, thus, would be expected to be cleared from the mosquito gut. Therefore, the combination of the use of *Asaia* or *S. marcescens* cultures initiated from glycerol stock with confirmation of GFP fluourescence for the resulting cultures prior to infection, ensured that antibiotic resistance without the presence of the respective plasmid conferring GFP fluourescence had not been generated.

The sample size used in this assay was rather limited, as 74 to 85 mosquitoes were interrogated regarding the outcome of infection at each time point, over 4 independent infections. The proportions of highly, lowly and non-infected mosquitoes following *Asaia* infection, though, were highly similar from day 3 to day 6 post infection. The consistency of the dynamics of *Asaia* infection in different time points suggests that the employed methodology constitutes a robust way to track the proportions of highly, lowly or non-infected mosquitoes following oral *Asaia* infection, despite the limited sample size. Furthermore, although the margins of error are considerable, they allow the formation of a relatively clear picture regarding the dynamics of *Asaia* infection. A larger sample size, restricted to day 5 post infection, was used in another oral *Asaia* infection assay intended for SNP genotyping, which will be presented later on (Figure 5.1B). Compared with the dynamics presented here, this assay shows a lower percentage of non-infected mosquitoes combined with a higher percentage of highly infected mosquitoes. Despite these differences which are related to the sample size used in each assay, the main conclusions stemming from this assay remain that oral infection with *Asaia* shows great variation in the outcome of infection whereas the infection dynamics show relative stability over a time course from 3 to 5 days post infection.

4.14 Great variation also in the outcome of oral *S. marcescens* infection

The dynamics of oral infection with *S. marcescens* were broadly similar with the corresponding *Asaia* dynamics. Great variation regarding the outcome of infection was observed, with highly, lowly and non-infected mosquitoes. Although the sample size used for *S. marcescens* infection was higher compared to *Asaia*, especially at days 5 and 6 post infection, the proportions of highly, lowly and non-infected mosquitoes following *S. marcescens* infection were also highly similar to the corresponding proportions following *Asaia* infection, at least from day 3 to day 5 post infection. One observed difference between

the dynamics of *Asaia* and *S. marcescens* oral infection was the higher percentage of highly infected mosquitoes with a concomitant reduction of non-infected mosquitoes, especially at days 5 and 6 post infection. As mentioned above, this difference was most likely due to differences in sample size as indicated when the dynamics of *Asaia* and *S. marcescens* infected mosquitoes restricted to day 5 post infection were compared using similar sample sizes (Figure 5.1B).

Furthermore, the dynamics following oral *S. marcescens* infection showed a considerable increase of highly infected mosquitoes from day 2 to day 3 post infection with a parallel decrease of non-infected mosquitoes, and also a distinguishable decrease in highly infected mosquitoes from day 5 to day 6 post infection, again with a parallel increase in non-infected mosquitoes. Therefore, during the infection time course tracked here, there is a window of 3 days, from day 3 to day 5 post infection, in which the proportions of highly, lowly and non-infected mosquitoes remain relatively constant, following the establishment of infection. This window of relatively constant infection dynamics will be the basis for the establishment of the time points used for the SNP genotyping and DNA microarray assays that will be presented in following chapters.

The dynamics of oral *Asaia* or *S. marcescens* infection are generally consistent with a previous study of oral bacterial infections in *Ae. aegypti* (Terenius et al., 2012). Oral infections with bacteria of the genus *Pantoea,* which, as *S. marcescens*, belong to *Enterobacteriaceae*, showed a precipitous increase in bacterial load in the first 2 to 3 days, as bacteria containing sugar move from the crop to the midgut, and bacterial load peaks, at least for some of the strains used, prior to day 5 post infection (Terenius et al., 2012). The combination of these data with the reduction in highly infected mosquitoes following *S. marcescens* infection from day 5 to day 6 post infection strengthen the choice of the time course followed here to track oral bacterial infection dynamics, up to 6 days post infection.

4.15 Oral *Asaia* or *S. marcescens* infection and reduced survival

Tracking of the survival of *Asaia* or *S. marcescens* infected mosquitoes, compared with uninfected mosquitoes, suggests that oral infection with any of the two bacterial strains reduces the fitness of infected mosquitoes. As bacteria-fed mosquitoes were not separated in both cases, the survival difference shown can be even more pronounced, taking into account that the survival of non-fed mosquitoes was tracked in the infected mosquito pools. Differences in the survival level of uninfected mosquitoes between the *Asaia* and *S. marcescens* infection assays preclude a comparison between the effects of the two bacterial strains, although it is unlikely that an assay combining *Asaia* and *S. marcescens* infections would

show any statistically significant differences in survival. In both cases, the detrimental effect following *Asaia or S. marcescens* infection that leads to mortality seems to take place in the initial stages of the infection, up to 5 days post infection, especially following *Asaia* infection. One explanation would be that induced responses to infection that occur during this initial stage may affect mortality while, after a homeostatic balance has been achieved, the presence of *Asaia* or *S. marcescens* no longer affects mosquito mortality, as would be expected for two common members of the mosquito gut microbiota.

The mortality observed following *Asaia* or *S. marcescens* infection, about 20% up to day 5 post *S. marcescens* infection and about 40% up to day 5 post *Asaia* infection, imply that during the course of infection used to track the dynamics of *Asaia* or *S. marcescens* infection, a portion of mosquitoes died and was excluded from tracking of infection dynamics. As infection experiments presented subsequently in this study, both for *Asaia* and *S. marcescens* infections, were conducted in an incubator with smaller temperature and humidity fluctuations that detract from mosquito survival, mortality during *Asaia* or S*. marcescens* infection most likely followed the *S. marcescens* mortality dynamics. For mosquitoes that died during the course of infection, it can be postulated that they might have been highly infected with *Asaia* or *S. marcescens*, with the high infection load contributing to the observed mortality. Alternatively, a highly efficacious immune response that completely eliminated the invading bacteria might account for the mortality observed following *Asaia* or *S. marcescens* oral infection, due to the activation of mosquito immune responses that detract from mosquito fitness. Therefore, infection dynamics for mosquitoes that died during the course of infection need to be further investigated so as to reach any firm conclusions.

The observed effect on survival is somewhat surprising for *Asaia*, which is considered less pathogenic than *S. marcescens*. Nevertheless, certain *Acetobacteriaceae* strains that naturally inhabit the *Drosophila* gut have been shown to be pathogenic (Ryu et al., 2008), so it is conceivable that *Asaia* can elicit some pathogenic effects following a massive infection. Essentially, both *Asaia* and *S. marcescens*, as common members of the mosquito gut microbiota, although they can be pathogenic, they are not lethal to the mosquito as symbiosis has probably resulted in adaptation from both ways, i.e. loss of pathogenic characteristics by the commensal bacteria and development of immune defences by the host. On the other hand, a massive infection of even innocuous and beneficial bacterial strains is most likely to elicit pathogenic effects, either in a direct manner, through secreted metabolites or indirectly due to aberrant activation of host immunity, thus rendering beneficial bacteria into pathogens. Therefore, a massive *Asaia* or *S. marcescens* infection might result in aberrant regulation of immune responses, with the end result being that the activation of such responses limits mosquito longevity due to energy consumption or

autoimmune reactions. It has been shown in *Drosophila* (Maillet et al., 2008), and also suggested in the mosquito (Meister et al., 2009), that regulation of peptidoglycan fragment availability is important for the regulation of the Imd (IMD/REL2) pathway. Therefore, a massive oral infection could result in a surge of available peptidoglycan which triggers antibacterial responses that also limit mosquito longevity.

As the mosquito gut microbiota consist of, in addition to commensal bacteria, fungi and viruses, the effect of non-bacterial members of the mosquito gut microbiota in influencing mosquito mortality following antibiotic treatment or oral bacterial infection remains to be determined. Fungal pathogens are known to influence mosquito longevity (Blanford et al., 2005; Fang et al., 2011; Knols et al., 2010) while mosquito immune responses are known to target infections with entomopathogenic fungi such as *Beauveria bassiana* (Yassine et al., 2012). Furthermore, interactions between fungi and commensal bacteria have been observed during symbiotic associations in other organisms (Kaltenpoth et al., 2005) and cannot be ruled out in mosquitoes. Therefore, disruption of the mosquito gut microbiota homeostasis, through antibiotic treatment or oral bacterial infection, along with induced mosquito immune responses could conceivably influence the status of entomopathogenic fungi resident in the mosquito gut. The repercussions of such interactions are beyond the scope of this study, but any observed results should also be interpreted in the context of a complex ecosystem in the mosquito gut that includes fungi and viruses.

4.16 Conclusions

The results presented in this part of the project establish a model of oral bacterial infections with bacteria of the genus *Asaia* or *S. marcescens*. Both strains are able to colonize the guts of *An. gambiae*, as expected for two common members of the mosquito gut microbiota, and limit mosquito longevity, suggesting that such oral infections trigger antibacterial responses that could be further investigated in subsequent parts of the project. The most important finding in this part of the project, though, is the variability of the infection outcome with *Asaia* or *S. marcescens*, with highly, lowly and non-infected mosquitoes following oral infection. This phenotypic variation could be correlated with the variable load and composition of the natural gut microbiota at the individual and population level, as determined in recent metagenomic analyses in lab-reared or field-collected mosquitoes (Boissière et al., 2012; Osei-Poku et al., 2012; Wang et al., 2011b). The relatively constant proportions of highly, lowly and non-infected mosquitoes following *Asaia* or *S. marcescens* infection, at least from day 3 to day 5 post infection, suggest that these proportions

might be genetically determined. This suggestion led to the investigation of the genetic basis of the outcome of *S. marcescens* infection that will be presented in the following chapter.

Chapter 5

Identification of *Anopheles gambiae* SNP divergence associated with the outcome of *Serratia marcescens* infection

Introduction

5.1 Ecological aspects of vectorial capacity and genetic variation in *An. gambiae*

An. gambiae mosquitoes show great variation regarding the outcome of *Plasmodium* infections, with mosquitoes refractory or susceptible to malaria transmission naturally found in field populations (Niare et al., 2002; Riehle et al., 2006; Sinden et al., 2004). Field *An. gambiae* mosquitoes show an intricate population structure with a high degree of nucleotide diversity (della Torre et al., 2002; Loaiza et al., 2012; Wilding et al., 2009). Furthermore, *An. gambiae* mosquitoes, as all insects, constitute a rapidly evolving and adapting species, showing a 2 to 3-fold higher evolutionary rate compared to vertebrates (Wyder et al., 2007). What is more, most likely due to haematophagy and adaptation to variable environmental settings in which mosquitoes are faced with diverse pathogen pressures, *An. gambiae* immunity genes show a much more pronounced degree of diversification, compared to other gene families, in an "arms race" to efficiently target diverse and also rapidly evolving pathogens (Christophides et al., 2002; Zdobnov et al., 2002).

Several lines of evidence suggest that genetic variation within *An. gambiae* populations influences vectorial capacity and thus malaria transmission, especially with regard to genes encoding immune factors (Mitri and Vernick, 2012). This correlation between the observed ecological variation regarding pathogen susceptibility and genetic variation, especially within immunity genes, suggests that the naturally observed refractoriness phenotype in a subset of the mosquito population has an underlying genetic basis. The identification of such genetic basis can be further exploited in the identification of immune factors and their corresponding genotypes responsible for refractoriness to malaria transmission. Furthermore, the underlying refractoriness genetic factors can form the basis of interventions aiming to disrupt malaria transmission, including the genetic modification of mosquitoes so that they are refractory to malaria transmission, but they also possess no fitness deficit that would compromise their release in the field and low possibility for adaptation of the mosquito or the parasite, which would result in a relapse to susceptibility to malaria transmission.

5.2 Association of *An. gambiae* genetic variation with the outcome of *Plasmodium* infections

The isolation of *An. gambiae* strains naturally refractory to malaria transmission had long suggested a genetic basis underlying the outcome of *Plasmodium* infections (Collins et al., 1986; Vernick et al., 1995).

Indeed, genetic mapping of *An. gambiae* isogenic families resulting from crosses between refractory and susceptible strains, using identification of quantitative trait loci through microsatellite marker genotyping (Zheng et al., 1996), suggested a genetic basis for the melanotic encapsulation trait, the major killing mechanism for the identified refractory strains (Collins et al., 1986), following infection with *P. cynomolgi* (Zheng et al., 1997; Zheng et al., 2003) or *P. berghei* (Gorman et al., 1997*).* Although these studies did not pinpoint to causative loci conferring refractoriness to *Plasmodium* infection, they introduced the concept of genomic regions controlling resistance to *Plasmodium* infection, whose heritability from generation to generation suggested the existence of one or few alleles that control the various killing mechanisms such as melanotic encapsulation, which play a major role in rendering the mosquito refractory to *Plasmodium* infection.

Indeed, *An. gambiae* field-collected isofemale families with diverse vectorial capacity regarding the outcome of *Plasmodium* infections allowed a field-based genetic analysis that identified a large genetic component responsible for parasite resistance, with segregating resistance alleles largely explaining the susceptibility to natural *P. falciparum* infections (Niare et al., 2002). A similar study also identified quantitative trait loci controlling refractoriness to *P. falciparum* (Menge et al., 2006), while another study employing isofemale pedigrees of field-collected mosquitoes and genetic mapping through microsatellite markers allowed the identification of *Plasmodium*-resistance islands (Riehle et al., 2006). Despite the considerable size of these genomic regions, extending to several megabases for each of them, the gene encoding the leucine-rich immune gene *APL1* was detected in one of these regions and was shown to explain a large component of *An. gambiae* resistance to *P. falciparum* infection (Riehle et al., 2006). Indeed, further phenotypic analysis of its APL1C paralogue confirmed its involvement in forming a high-molecular weight complex with LRIM1 that directs the complement factor TEP1 on the surface of *P. berghei* parasites, leading to their killing (Fraiture et al., 2009; Povelones et al., 2009).

Another study employing microsatellite linkage mapping following *P. berghei* infection of *An. gambiae* isogenic families stemming from the cross of refractory and susceptible strains, identified a large genomic region that conferred resistance to *P. berghei* infection and included the gene encoding the complement factor TEP1 (Blandin et al., 2009). Polymorphisms in the *TEP1* gene were further shown to differentially affect the outcome of *Plasmodium* infections, with *TEP1* alleles conferring resistance or susceptibility to *Plasmodium* infection, thus explaining most of the variability regarding the outcome of *Plasmodium* infection (Blandin et al., 2009). Interestingly, divergence of *TEP1* alleles has also been identified between

M/S forms of *An. gambiae*, with the diverged alleles showing differential susceptibility to *P. falciparum* infection (White et al., 2011).

The association of *An. gambiae* genetic variation with the outcome of natural *P. falciparum* infections has also been shown using targeted SNP genotyping of immune genes. The interrogation of 157 SNPs in 67 *An. gambiae* immune genes identified 5 SNPs significantly associated with the outcome of *P. falciparum* infection, suggesting that polymorphisms within immunity genes can influence the mosquito's ability to transmit malaria (Harris et al., 2010a). A similar approach has implicated *An. gambiae* SNPs linked to the Toll-like receptor TOLL5B locus with the *P. falciparum* infection outcome (Horton et al., 2010).

5.3 *Plasmodium* an unlikely candidate for driving *An. gambiae* genetic variation

Although previously proposed to be the case (Niare et al., 2002), the effect of *Plasmodium* infections on mosquito fitness, despite intense research efforts, still remains a matter of debate (Sangare et al., 2013). On the other hand, bacteria are considered a major evolutionary force that can directly affect survival. It is generally accepted in mosquitoes or *Drosophila* that deficiency in immune components targeting bacteria results in a shorter lifespan following bacterial infections (Lemaitre and Hoffmann, 2007; Lemaitre et al., 1995; Meister et al., 2009; Schnitger et al., 2009). Therefore, bacterial infections, if not sufficiently sustained by the immune system, increase mortality, implying a strong selection pressure on such immune responses and a strong drive for adaptation either to novel pathogens, introduced to the host's environmental setting, or derived from continuous evolutionary processes of the pathogens themselves. Although this possibility seems exceptionally plausible, the identification of signatures of natural selection that would formally demonstrate the contribution of bacteria in driving genetic variation remains extremely difficult. A recent study in humans, harnessing large-scale data from a human genome variation mapping project using 1000 genomes (1000 Genomes Project Consortium et al., 2010), identified causal variants in candidate signals of natural selection, including a non-synonymous variant in the Toll-like receptor 5 (TLR5) locus that alters NF-κB signalling in response to bacterial flagellin (Grossman et al., 2013). As in humans, it is very likely that also in mosquitoes, adaptations to bacterial infections can also drive the selection of variants, especially in immune receptor genes involved in mounting or regulating antibacterial responses.

What is more, adaptation to *Plasmodium* infections that would drive genetic variation implies that mosquito immune components recognize and subsequently target malaria parasites. This area also remains highly controversial. So far no *Plasmodium* molecular patterns have been identified as been

recognized by mosquito immune factors that would subsequently trigger an anti-*Plasmodium* response. An indication of specificity in *Plasmodium* recognition can be inferred by the differential ability of the leucine rich immune protein LRIM1 (Cohuet et al., 2006; Garver et al., 2012) and the *APL1* paralogues APL1A and APL1C (Garver et al., 2012; Holm et al., 2012; Mitri et al., 2009) to kill human and rodent malaria parasites. Importantly, though, neither LRIM1, APL1A nor APL1C have been shown to bind *Plasmodium* parasites (Fraiture et al., 2009; Mitri et al., 2009; Povelones et al., 2009). On the other hand, a recent study has indicated that epithelial nitration of *P. berghei* parasites as they traverse through the mosquito midgut is responsible for parasite targeting by TEP1-mediated complement defences (Oliveira et al., 2012). What is more, the *P. falciparum* gene *Pfs47* has been shown to suppress mosquito midgut nitration responses, which results in evasion of the mosquito's complement killing mechanism (Molina-Cruz et al., 2013). What these studies suggest is that as epithelial nitration renders the parasites visible to the mosquito complement system and leads to their killing, there is no specific recognition of *Plasmodium* molecular patterns by the complement system and complement killing efficiency relies on the efficiency of the protein nitration mechanism of *Plasmodium* parasites as they traverse the midgut epithelium. It would be difficult to reconcile a mechanism of selective pressure exerted by *Plasmodium* parasites on *TEP1* alleles with no specific recognition of *Plasmodium* parasites and thus no effect of *Plasmodium* genetic adaptations on the ability for complement-mediated killing of these parasites, while at the same time at least a subset of *P. falciparum* parasites even evade this response.

Furthermore, the adaptive divergence of *TEP1* alleles differentially affecting the outcome of *P. falciparum* infections between the M/S forms of *An. gambiae* (White et al., 2011) has been a particularly surprising finding as no divergence has been detected between *An. gambiae* M/S forms regarding susceptibility or refractoriness to *P. falciparum* infection (Gneme et al., 2013; White et al., 2011). One suggested possibility which could explain this finding is that the adaptive divergence of *TEP1* alleles between the *An. gambiae* M/S forms could be due to selective pressures exerted by bacteria present in the mosquito larval habitats, perhaps related to TEP1 antibacterial activities (Levashina et al., 2001; White et al., 2011). Therefore, these recent findings (Oliveira et al., 2012; White et al., 2011), render the possibility of infection by *Plasmodium* being the driving factor of mosquito genetic variation underlying susceptibility and refractoriness to malaria transmission highly questionable.

5.4 Gut microbiota as a major evolutionary force driving *An. gambiae* genetic variation

The studies mentioned above investigating the genetic basis of the outcome of *Plasmodium* infections (Blandin et al., 2009; Harris et al., 2010a; Niare et al., 2002; Riehle et al., 2006) disregarded the direct or indirect influence on the outcome of *Plasmodium* infections exerted by the mosquito gut microbiota, as revealed by more recent studies (Bando et al., 2013; Cirimotich et al., 2011b; Dong et al., 2009; Meister et al., 2009; Rodrigues et al., 2010). As a result, these studies did not take into account the possibility that genetic divergence associated with the outcome of *Plasmodium* infections might have a causal basis in influencing the load and composition of the mosquito gut microbiota, whose variation in turn can also directly or indirectly determine susceptibility or refractoriness to *Plasmodium* infection.

As genetic components might directly shape the mosquito gut microbiota and thus indirectly influence susceptibility and refractoriness to malaria transmission, it would be expected that some correlation should exist between the load and composition of the mosquito gut microbiota and the outcome of *Plasmodium* infections. Although this possibility is strongly suggested by the anti-*Plasmodium* effect of certain bacterial strains such as the enterobacteria *S. marcescens* (Bando et al., 2013) and *Esp_Z* (Cirimotich et al., 2011b), no direct demonstration of such possibility has been achieved. Interestingly, though, a study in which *An. gambiae* field mosquitoes were infected with *P. falciparum* and the microbiota composition inhabiting their gut was identified using 454 pyrosequencing, a significant correlation was detected between susceptibility and refractoriness to *P. falciparum* infection and the presence of *Enterobacteriaceae* in the mosquito gut (Boissière et al., 2012). This correlation is consistent with the possibility that genetic components targeting *Enterobacteriaceae*, broadly variable within the mosquito population, result in broad ecological variation of *Enterobacteriaceae* load and composition in the mosquito gut, which in turn influences the outcome of *Plasmodium* infections.

This possibility entails that genetic variation related to the outcome of bacterial infections is retained in the mosquito population, thus underlying ecological variation observed through variation in the load and composition of the mosquito gut microbiota. Genetic variation driving adaptations to bacterial infections is very likely to follow such pattern of balancing selection, rather than resulting in allele fixation within the mosquito population. Genetic components evolved to optimally sustain the mosquito gut microbiota would be expected to reach a balance between an antibacterial response selected due to the effect of bacteria on survival if left unrestrained, the detrimental effect of unrestrained immune responses and the achievement of gut homeostasis that allows the colonization of the mosquito gut by beneficial bacteria. This balancing selection requirement is much more likely for antibacterial rather than anti-*Plasmodium*

components. Commensal bacteria inhabiting the mosquito gut can form mutually beneficial symbiotic relationships with the host, by contributing either to the suppression of pathogenic strains as shown in *Drosophila* (Ryu et al., 2008) or to the nutritional requirements of the host as in other insects (Akman Gunduz and Douglas, 2009; MacDonald et al., 2011). On the other hand, there is no conceivable reason for the mosquito to avoid fixation of alleles that confer refractoriness to *Plasmodium* infection, especially if such immune efficiency would not drain additional energy resources from the host.

The conclusion to be drawn from the above is that the tripartite interactions between the mosquito immune system, commensal bacteria inhabiting the mosquito gut and *Plasmodium* infections dramatically complicate the efforts of genetic mapping alleles responsible for refractoriness to malaria transmission. What is more, the expected influence of midgut bacteria in driving genetic variation throughout the evolutionary process, suggests that mosquito adaptations resulting in broadly variable genetic components within the population are much more likely to reflect adaptations to bacterial rather than *Plasmodium* infections.

5.5 A high-resolution SNP genotyping array to interrogate genetic variation in *An. gambiae*

The novelty of this part of the project lies not only in tackling the previously unexplored aspect of the genetic basis of the outcome of bacterial infections in the main malaria vector *An. gambiae*, but also in the implementation of high-resolution genetic mapping through a 400k SNP genotyping array. The genetic tools employed by previous studies investigating the outcome of *Plasmodium* infections (Blandin et al., 2009; Niare et al., 2002; Riehle et al., 2006), involved quantitative trait loci mapping using microsatellite markers (Zheng et al., 1996), and allowed the identification of large genomic areas, typically including hundreds of loci, as associated with the outcome of infection. Such microsatellite markers are too infrequent in the genome to allow a comprehensive analysis, in comparison to the abundance of SNPs throughout the genome, while their size precludes fine-scale mapping for the identification of single causal loci (Wilding et al., 2009), as the identification of *TEP1* (Blandin et al., 2009) or *APL1 (*Riehle et al., 2006*)* in *Plasmodium* resistance islands, comprising several megabases, relied on their previously reported (Blandin et al., 2004) or highly suspected (Osta et al., 2004a) involvement in anti-*Plasmodium* responses.

SNPs are widespread both in humans (Sachidanandam et al., 2001) and insects (Berger et al., 2001; Cohuet et al., 2008) and have been extensively used as genetic markers in genome-wide association studies (Baker, 2010a). The main limitation in implementing a SNP-based genetic mapping approach in *An.*

gambiae is the discovery and validation of a large number of such polymorphisms. Studies implementing targeted SNP genotyping limited to genomic areas in the proximity of a subset of genes (Morlais et al., 2004; Wilding et al., 2009), or, in some cases, immune genes (Cohuet et al., 2008; Harris et al., 2010a; Horton et al., 2010) or insecticide resistance related genes (Weetman et al., 2010), lacked the resolution of a genome-wide approach. SNPs identified through the sequencing of *An. gambiae* (Holt et al., 2002) showed low frequency and uneven distribution and, as this initial sequencing relied on the PEST strain, which was a mix of M/S *An. gambiae* forms, the identified SNPs showed a bias towards divergence between these two forms rather than representing polymorphisms in natural populations (Morlais et al., 2004; Wilding et al., 2009).

The SNP genotyping array utilized in this study adopted a previously reported and validated approach in high-resolution SNP genotyping in the *An. gambiae* genome (Neafsey et al., 2010), which was based on complete genome sequencing pools of laboratory-reared mosquitoes belonging either to M or S *An. gambiae* forms, using Sanger sequencing (Lawniczak et al., 2010). Sequencing of the M/S *An. gambiae* forms yielded about 2 million SNPs per form, compared to 450,000 identified SNPs in the PEST strain (Holt et al., 2002). About 400,000 of these SNPs were further utilized in constructing and validating a high-density customized Affymetrix 400k SNP genotyping array, as described previously (Neafsey et al., 2010). This subset of identified SNPs was selected based on proximity to annotated *An. gambiae* coding regions and heterozygosity in both the M/S sequences or fixed differences between them. As a result, this subset of SNPs provided a comprehensive genome-wide coverage with a median distance of 300 bp between SNPs and 98% of annotated *An. gambiae* genes containing at least one SNP. Hybridization of this SNP genotyping array with gDNA from 20 field-collected *An. gambiae* mosquitoes as well as with pooled gDNA from the same individuals, indicated that this array can provide useful quantitative information regarding allele frequency divergence between pooled samples. Furthermore, genotype calls derived from Gdna hybridization of two *An. gambiae* mosquitoes which had been also sequenced using Illumina sequencing were also used for array validation, yielding about 66,000 SNPs with validated calls. Therefore, measured allele differentiation between hybridizations utilizing this SNP genotyping array was considered to provide useful information regarding genetic divergence between the hybridized samples assayed (Neafsey et al., 2010).

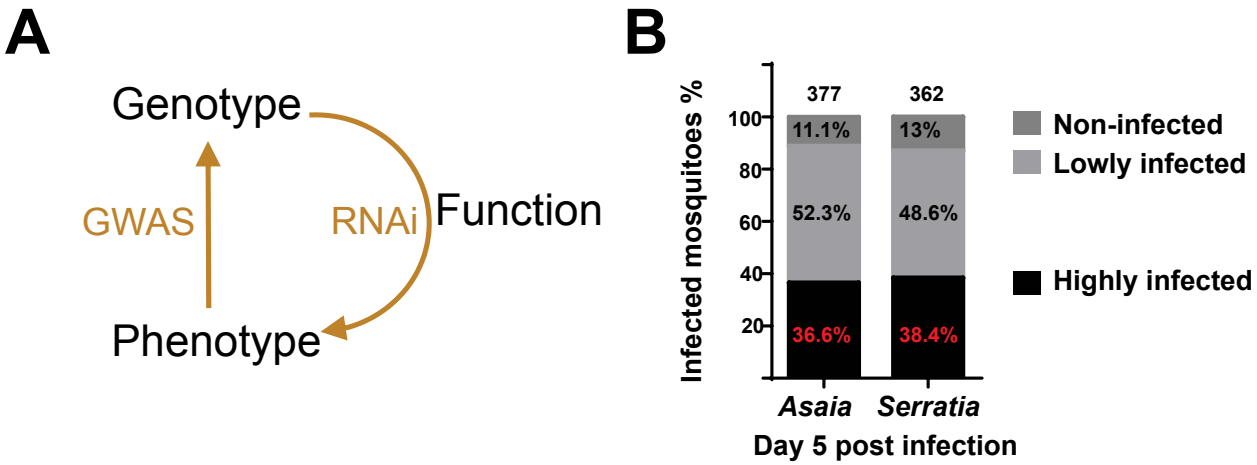
The implementation of this SNP genotyping array has been previously reported in identifying SNP divergence between sympatric or allopatric *An. gambiae* M/S forms as well as between *An. gambiae* and *An. arabiensis* field mosquitoes (Neafsey et al., 2010; Nwakanma et al., 2013; Reidenbach et al., 2012).

Here, this array was used to identify SNP divergence between gDNA pools from mosquitoes that were either highly infected or non-infected following oral *S. marcescens* infection.

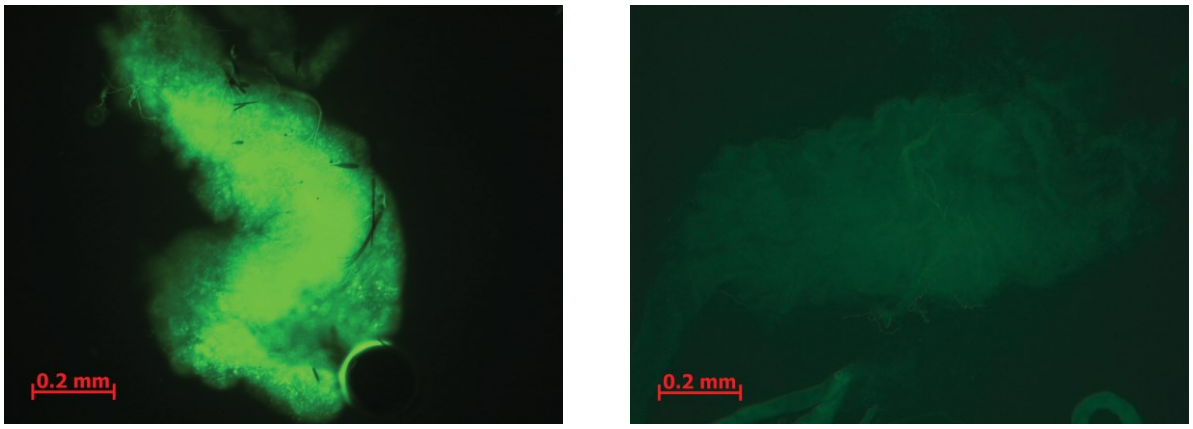
5.6 A forward genetics approach for identification of a genetic basis in the outcome of bacterial infections

Oral infection of antibiotic treated *An. gambiae* mosquitoes with *S. marcescens* (Figure 4.4) or bacteria of the genus *Asaia* (Figure 4.2) indicated great variation regarding the outcome of infection, with highly, lowly or non-infected mosquitoes, as determined by microscopic observation of fluourescent bacteria inhabiting the mosquito gut. The proportions of highly, lowly and non-infected mosquitoes showed relative stability, at least from day 3 to day 5 post infection, both after *Asaia* or *S. marcescens* infection. Such stability suggested that the observed phenotypic variation regarding the abundance of fluourescent bacteria in the mosquito gut might have a genetic basis, which could, at least partly, underlie the outcome of infection, perhaps along with stochastic processes that may also shape the observed phenotypic variation.

The implementation of a forward genetics approach provides a non-biased way to establish whether an observed phenotypic trait has a genetic basis. Genome-wide association studies (GWAS) can identify significant SNP segregation between two phenotypic pools, which may serve as genetic markers of causal polymorphisms underlying the interrogated phenotypic variation (Figure 5.1A). This correlation between the observed phenotypic trait and genetic differences between the pooled individuals in such genome-wide studies does not rely on any *a priori* assumptions as to whether there is a genetic basis or which genetic factors contribute to it. Associations arising from a forward genetics approach can be validated in two ways: through identification of loci known or highly suspected to be involved in relevant functions based on existing homologies, or implementation of a reverse genetics approach, in which silencing of an associated gene results in a phenotypic alteration that confirms the genetic basis of the observed phenotypic variation and the causal involvement of the respective gene (Figure 5.1A).



C



Affymetrix 400k SNP genotyping array

Interrogation of genetic variation in ~400,000 variable positions in the *Anopheles gambiae* genome

Figure 5.1: A forward genetics approach for the identification of a genetic basis in the outcome of bacterial infections. A: Graphical representation of forward and reverse genetics integration for the functional characterization of genes underlying the outcome of bacterial infections. Association studies (GWAS) correlate phenotypic variation, in this case the outcome of bacterial infections, with genetic variation, in this case SNP segregation between highly and non-infected phenotypic pools. Genetic markers can be linked to candidate loci, putatively possessing causal relationships with the outcome of bacterial infection. Through silencing of these genes using RNAi-mediated silencing (RNAi), the functional characterization of these genes can reveal the effect of these genes on the outcome of bacterial infection, thus validating their initial association with the outcome of infection (Function). **B:** Antibiotic treated mosquitoes were orally infected with fluourescent bacteria of the genus *Asaia* or *S. marcescens*, bacteria-fed mosquitoes were separated at day 2 post infection and, at day 5 post infection, the level of infection was determined through microscopic observation of dissected mosquito midguts under a fluourescence microscope. Based on the observed level of infection, mosquitoes were categorized as highly, lowly or non-infected. The level of infection was determined for 377 *Asaia* infected mosquitoes and for 362 *S. marcescens* infected mosquitoes, in two independent infections for both cases. Each bar represents the proportions of highly, lowly or non-infected mosquitoes following *Asaia* (left bar) *or S. marcescens* (right bar) infection. The percent representation for each level of infection is indicated within the part of the bar representing the respective proportion, as is indicated in the colour legend, while the total number of assayed mosquitoes is indicated above the respective bar. **C:** Graphical representation of the adopted experimental design in identifying SNP divergence associated with the outcome of *S. marcescens* infection. gDNA from pools of highly infected mosquitoes (left) or non-infected mosquitoes (right) following oral *S. marcescens* infection were individually hybridized to customized Affymetrix SNP genotyping arrays, which can interrogate genetic variation at ~400,000 variable positions in the *An. gambiae* genome, based on a previous M/S *An. gambiae* resequencing project (Lawniczak et al., 2010; Neafsey et al., 2010).

5.7 Establishing highly and non-infected phenotypic pools for *Asaia* and *S. marcescens*

To implement such a genome-wide approach in interrogating the genetic basis of the phenotypic variation regarding the bacterial infection outcome, *An. gambiae* mosquitoes were orally infected either with *Asaia* or *S. marcescens* and the proportions of highly, lowly or non-infected mosquitoes were determined at day 5 post infection, a time point in which the dynamics of infection have been established and have remained stable for at least 3 days (Figure 4.2 for *Asaia* and Figure 4.4 for *S. marcescens*). The sample size used for both *Asaia* and *S. marcescens* infection comprised a total of 377 and 362 individuals interrogated regarding the outcome of infection in *Asaia* or *S. marcescens* infected mosquitoes, respectively, over two independent assays in both cases. For the *S. marcescens* infection, in which the sample sizes were relatively comparable, with 362 assayed mosquitoes compared to 298 previously assayed at day 5 post infection (Figure 4.4B), the observed proportions were remarkably similar, with a mean of 38.4% of mosquitoes characterized as highly infected, compared to 37% previously, 48.6% lowly infected, compared to 47.3% previously and 13% non-infected, compared to 15.7% previously (Figure 5.1B). Such consistency regarding the dynamics of oral *S. marcescens* infection, as determined overall in 6 independent infections, attests to the robustness of the implemented assay in determination of the dynamics of oral *S. marcescens* infection and suggests that the sample size used is sufficient to identify these dynamics.

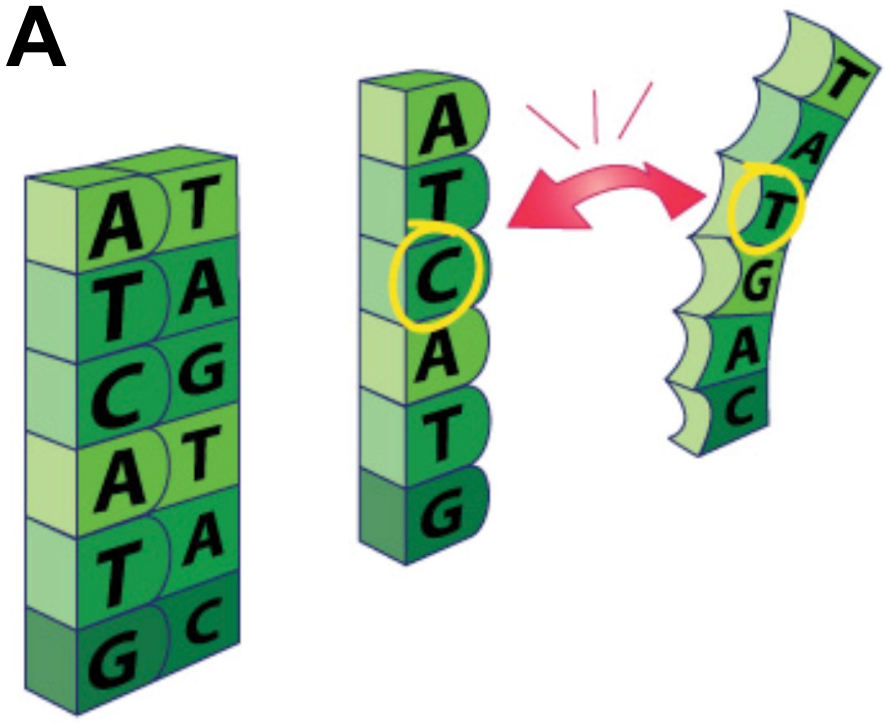
The proportions of highly, lowly and non-infected mosquitoes following *Asaia* infection showed the same general trend as previously, although they exhibited a lower proportion of non-infected mosquitoes, most likely due to the higher sample size used, with 377 assayed mosquitoes compared to 80 mosquitoes assayed previously (Figure 4.2C). A mean of 36.6% of mosquitoes were characterized as highly infected, compared to 25.3% previously, 52.3% were characterized as lowly infected, compared to 48.7% previously and 11.1% were characterized as non-infected, compared to 26% previously (Figure 5.1B). The larger sample size used for oral *Asaia* infection, comparable to the respective sample size used for the *S. marcescens* infection dynamics at day 5 post infection, suggests that the dynamics between oral infection with *Asaia* and *S. marcescens* are remarkably similar and previously observed differences were due to the different sample sizes used (Figure 4.2C and 4.4B).

Extracted gDNA from mosquitoes that were either highly infected or non-infected following *Asaia* infection was used for whole-genome sequencing that is not presented in this study.

5.8 Identification of genetic divergence associated with the outcome of *S. marcescens* infection by utilizing a 400k SNP genotyping array

The genetic basis of the outcome of *S. marcescens* infection was investigated by employing an array-based approach. Equal quantities of gDNA from 15 highly infected mosquitoes following *S. marcescens* infection were pooled, and in the same way, gDNA from 15 non-infected mosquitoes was also extracted and equal quantities were pooled. These highly and non-infected gDNA pools, from representative mosquitoes of the highly and non-infected phenotypic pools, were hybridized to individual customized Affymetrix SNP genotyping arrays (Figure 5.1C). These arrays can interrogate genetic variation at ~400,000 variable positions in the *An. gambiae* genome and have been previously shown to provide useful quantitative information regarding SNP divergence between pooled samples (Neafsey et al., 2010).

The Affymetrix 400k *Genechip* used here has been customized to explore sequence variability in ~400,000 variable positions of the *An. gambiae* genome, as determined by whole-genome sequencing of the M/S forms of *An. gambiae* (Lawniczak et al., 2010). This array contains 25 bp oligonucleotide probe sequences. For each of the ~400,000 SNPs, a SNP tiling strategy is employed. The two main variants for this polymorphism are designated as allele A and B, as previously determined by whole-genome sequencing. For each of these alleles, a probe based on the 25 bp sequence encompassing the SNP is included in the array, with the particular SNP in a central position as a perfect match that is expected to hybridize to gDNA corresponding to the same genotype as the allele and, in another oligonucleotide, as a mismatch, with the SNP central base inverted so it is not expected to hybridize if gDNA is of the same genotype as the allele (Figure 5.2A). Therefore, for both alleles A/B of the particular SNP, there is a probe quartet consisting of 4 probes: a perfect match and mismatch probe for allele A and similarly for allele B. What is more, offset probes are also included in the array, with the same probe quartet but in slightly different positions within the 25 bp sequence (Figure 5.2B) (Dalma-Weiszhausz et al., 2006).


A
T
A
T
C
G
A
T
T
A
G
C
A
T
C
A
T
G
T
A
T
G
A
C

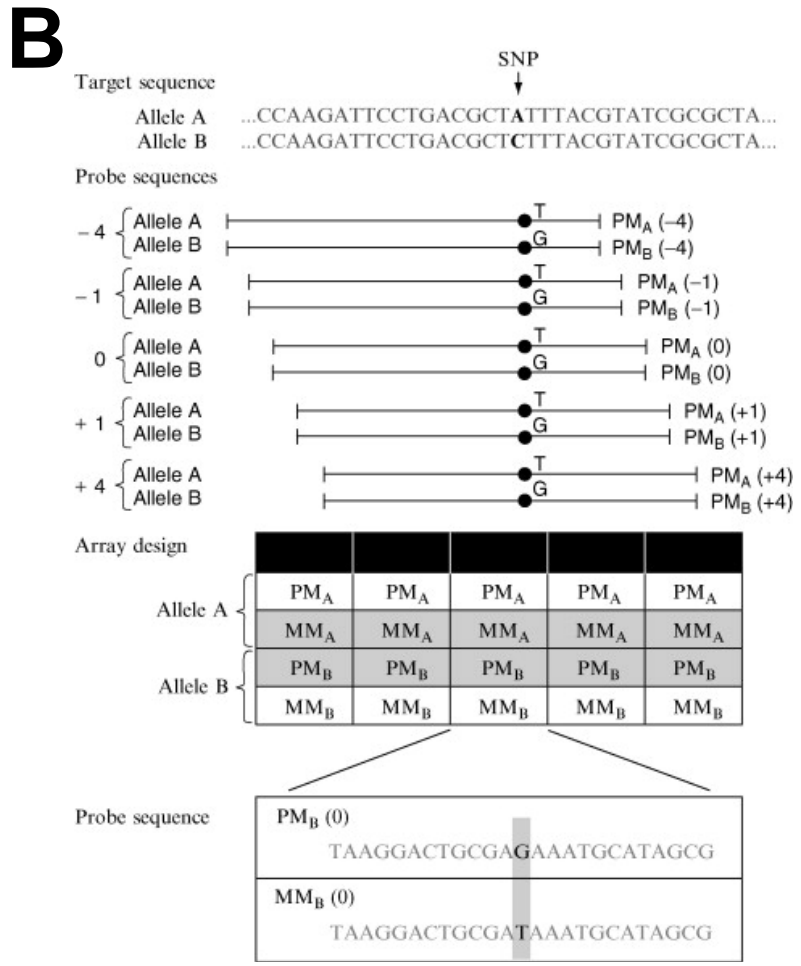

B
SNP
Target sequence
Allele A ...CCAAGATTCCTGACGCTATTTACGTATCGCGCTA...
Allele B ...CCAAGATTCCTGACGCTCTTTACGTATCGCGCTA...
Probe sequences
−4 Allele A T PMA (−4)
Allele B G PMB (−4)
−1 Allele A T PMA (−1)
Allele B G PMB (−1)
0 Allele A T PMA (0)
Allele B G PMB (0)
+1 Allele A T PMA (+1)
Allele B G PMB (+1)
+4 Allele A T PMA (+4)
Allele B G PMB (+4)
Array design
Allele A
Allele B
PMA MMA PMB MMB
Probe sequence
PMB (0)
TAAGGACTGCGAGAAATGCATAGCG
MMB (0)
TAAGGACTGCGATAAATGCATAGCG

Figure 5.2: The Affymetrix *Genechip* platform. A: A good match sticks, a bad does not. 25 bp oligonucleotides on the array include the SNP allele A in a central or offset position. In perfect match oligonucleotides, gDNA of the same genotype is expected to hybridize to the complementary sequence on the oligonucleotide (left). In mismatch probes, the SNP base is inverted so gDNA of the same genotype with allele A is not expected to hybridize to the oligonucleotide (right). Adapted from: How Affymetrix *Genechip* microarrays work, Affymetrix. **B:** Affymetrix *Genechip* array design. Target sequence: 25 bp gDNA sequences that include the SNP in a central position, denoted by an arrow. The SNP can be found in 2 allele genotypes, allele A (nucleotide A) or allele B (nucleotide C). These main SNP genotypes have been determined by previous sequencing of natural populations. Probe sequences: 25 bp oligonucleotides included in the *Genechip* array designed to be complementary to the target sequence so that they hybridize in all positions around the SNP. For the SNP, perfect match probes for allele A or allele B (PM_A or PM_B) include a complementary nucleotide base so that they hybridize to allele A or allele B gDNA, respectively. For allele A/B perfect match probes, in addition to the probe containing the SNP in central position (0), 4 additional offset probes are included in the array in which the SNP is designed in offset positions relative to the central position (-1, -4, +1, +4 bases). Array design: For the SNP in question, a total of 20 oligonucleotide probes are included in the *Genechip* array. For allele A, 5 perfect match probes (PM_A) that are expected to hybridize to target gDNA sequences of the same genotype, and 5 mismatch probes (MM_A) that are not expected to hybridize if the target gDNA sequence is of the same genotype. Another 10 oligonucleotide probes are included for allele B in a similar manner. Probe sequence: Perfect match and mismatch probes are identical except for the SNP position in which the inverted nucleotide base is included. Therefore, if these probes are hybridized to gDNA with genotype AA, target sequences that include A in the SNP position, are going to hybridize to PM_A (T) and not MM_A (A), PM_B (G) nor MM_B (C). Similarly, hybridization to gDNA with genotype BB, target sequences that include C in the SNP position are going to hybridize to PM_B (G) and not MM_B (C), PM_A (T) nor MM_A (A). Adapted from: (Dalma-Weiszhausz et al., 2006).

The mapping assay involves 250 ng of gDNA that is restriction digested to a range of sizes, ligated with a common set of adaptors and PCR amplified for fragments in the 250 bp to 1000 bp range. These fragments are labelled with biotin so that hybridization efficiency is detected through emitted fluourescence. Following a normalization step that ensures that the overall median chip intensity is the same across all samples assayed, sufficient specific hybridization is determined for each probe through emitted fluourescence of gDNA fragments that may have hybridized to the respective probe.

The hybridization efficiency of the assayed gDNA on this set of oligonucleotide probes is used to determine the genotype call for this particular SNP: AA, AB or BB. Measured differentiation between alleles A and B is determined by the ratio in hybridization signal intensity between the two alleles, based on signal intensity of the corresponding oligonucleotide probes. As the assayed gDNA is a pool from different individual mosquitoes that may vary for this particular genotype, the magnitude of each genotype is determined through its hybridization efficiency to identify the dominant genotype for this particular SNP, expressed through the frequency of allele A, which is designated as the minor allele frequency (MAF).

The quantification, pooling, digestion, labelling and hybridization of gDNA samples corresponding to highly infected or non-infected pools in individual SNPchips, along with downstream analysis and genotype calling, was performed by collaborators at the Broad Institute (Boston, USA) and has been described previously (Neafsey et al., 2010; Reidenbach et al., 2012). The resulting analysis determined for each phenotypic pool the MAF corresponding to each of ~400,000 SNPs assayed, and these readings were used to measure the MAF difference between the highly infected and non-infected pools for each SNP. This MAF difference provides a measure of genetic divergence between the two pools for each SNP. The determined MAF differences for each SNP can be found in the relevant table deposited in ArrayExpress (Experiment name *Serratia_SNP1*; E-MEXP-3951).

To further assess statistically significant genetic divergence between the highly infected and non-infected pools, a permutation analysis was performed by collaborators at Imperial College London (London, UK) as previously described (Neafsey et al., 2010). Measured differentiation in non-overlapping windows of 10 adjacent SNPs was determined based on the average absolute values of MAF difference for each SNP between the highly infected and non-infected pools. In this way, the average MAF difference between the two pools was determined for ~40,000 sliding 10-SNP windows of variable size due to the non-uniform physical spacing of adjacent SNPs. Statistical significance of the difference between the average MAF difference for each 10-SNP window and the average MAF difference of 10 random SNPs was determined by randomly permuting the assays 100 million times and assaying the frequency by which the average

MAF difference of the 10-SNP window was observed at equal or greater magnitude for the same comparison of pools. This permutation analysis resulted in the assignment of a p-value for each 10-SNP window expressing the statistical confidence that the average MAF difference of this 10-SNP window is significantly higher than the average MAF difference of 10 random SNPs. The threshold of statistical significance, following Bonferroni correction for the number of tests conducted, was set to 10^{-5}. Therefore, genomic areas within 10-SNP windows with p-value $<10^{-5}$ were considered to be genetically divergent between the highly and non-infected phenotypic pools.

5.9 Genomic areas and corresponding loci associated with the outcome of *S. marcescens* infection

Two approaches were employed to assess genotypic association with the outcome of *S. marcescens* infection: individual SNP loci with MAF difference >0.5, indicating a preponderance of different genotypes between the two pools for the respective locus, and genomic areas within 10-SNP windows for which the permutation analysis indicated a statistically significant genetic divergence between the two interrogated phenotypic pools. The two approaches detected 140 out of 400,071 SNPs with MAF difference >0.5 (Table 5.1) and 44 out of 40,007 10-SNP windows with p-value $<10^{-5}$ in the permutation analysis conducted (Table 5.2). These highlighted individual SNPs and genomic regions encompassing these significant 10-SNP windows formed distinctive clusters throughout the *An. gambiae* genome that were designated as peaks and can be seen in Figure 5.3. These peaks were numbered so that they are discerned from one another and encompass highlighted SNPs or 10-SNP windows with significant p-values or both, although assessed association was limited to genes within a 5 kb radius of highlighted SNPs or within 10-SNP windows.

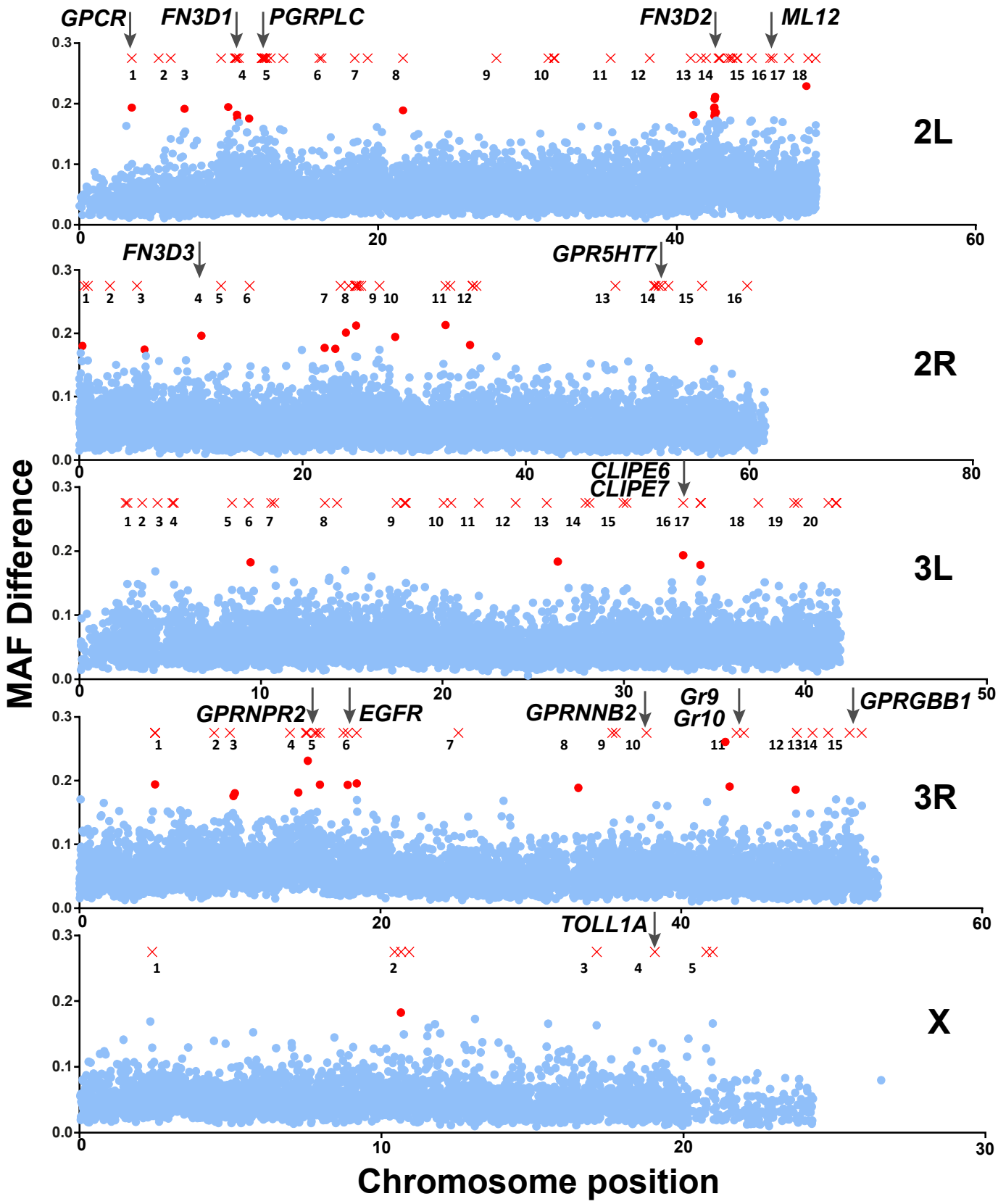

GPCR
FN3D1
PGRPLC
FN3D2
ML12
2L
FN3D3
GPR5HT7
2R
CLIPE6
CLIPE7
3L
GPRNPR2
EGFR
GPRNNB2
Gr9
Gr10
GPRGBB1
3R
TOLL1A
X
MAF Difference
Chromosome position

Figure 5.3: Mapping of SNP divergence associated with the outcome of *S. marcescens* infection. Extracted gDNA from highly infected or non-infected mosquito pools following oral *S. marcescens* infection was hybridized to individual 400k SNP genotyping arrays interrogating genetic divergence at ~400,000 SNPs in the *An. gambiae* genome. For each of these SNPs, the MAF difference between the highly and non-infected phenotypic pool was determined. Individual SNPs with MAF difference >0.5 are shown in their respective genomic position as red Xs. A permutation analysis was also carried out in which the average MAF difference of 10 adjacent SNPs was compared to the average MAF difference of 10 random SNPs. Significant differences were assessed for ~40,000 sliding 10-SNP windows and 10-SNP windows with a Bonferroni-corrected p-value $<10^{-5}$ were considered as associated with the outcome of *S. marcescens* infection. 10-SNP windows with significant p-values are plotted in their respective genomic position in relation to their average MAF difference, shown as red dots, while 10-SNP windows with non-significant p-values are also plotted in the same way, indicated by a blue dot. Genomic areas with clusters of neighbouring highlighted SNPs and/or significant 10-SNP windows are numbered so that these areas are discerned from one another. Loci within a 5 kb radius of highlighted SNPs or within genomic areas delineated by significant 10-SNP windows were considered associated with the outcome of *S. marcescens* infection. The locus position for a subset of genes of interest associated with the outcome of *S. marcescens* infection is indicated, with arrows pointing to their genomic position.

Based on the *AgamP3.7* annotation of the *An. gambiae* genome, as presented in Vectorbase (Lawson et al., 2009), 118 annotated *An. gambiae* genes were found to reside within a 5 kb radius of the 140 highlighted SNPs while 27 genes fell within genomic areas delineated by 10-SNP windows with significant p-values in the permutation analysis conducted. The two approaches combined detected 138 genes, as there was an overlap of 7 genes between the two sets. One of the genes that was detected by both the individual SNP and permutation analysis, in peak 3R-6 (chromosomal arm 3R, peak 6 as will be used from now on to denote the respective peaks) was the gene encoding the epidermal growth factor receptor EGFR, an orthologue of the *Drosophila* EGFR, which has been shown to be involved in stem cell proliferation and gut epithelial remodeling following oral bacterial infection (Buchon et al., 2010).

In peak 2L-5, the gene encoding the peptidoglycan recognition receptor PGRPLC was found within a 5 kb radius of a highlighted SNP. PGRPLC is known to be involved in immune responses against Gram-negative bacteria such as *S. marcescens*, as is the orthologue of the homonymous fruit fly receptor of DAP-type Gram-negative bacterial peptidoglycan that activates the IMD/REL2 NF-κB signalling pathway thus eliciting antibacterial responses (Choe et al., 2005; Choe et al., 2002; Gottar et al., 2002; Meister et al., 2009).

The identification of *PGRPLC* within a 5 kb radius of an associated SNP suggests, along with the relevance, as discussed later on, of other genes associated with the outcome of infection based on such proximity with highlighted SNPs, that the nucleotide distance used here to correlate proximal loci with SNP genetic markers associated with the outcome of *S. marcescens* infection captured relevant information regarding genetic divergence between the two interrogated phenotypic pools. The physical distance of up to 5 kb between a locus and a SNP used here to consider the SNP as a linked genetic marker for the respective locus, is in line with the average gene size of 5.7 kb, including introns, in *An. gambiae* (Lawniczak et al., 2010) and the higher linkage disequilibrium that is expected in laboratory colonies, such as the *N'gousso* colony used here, compared to field mosquito populations (Harris et al., 2010b).

Furthermore, the relevance of genes associated with the outcome of *S. marcescens* infection based on the individual SNP or the permutation analysis, along with the genes that overlap between the two approaches, including *EGFR*, suggest that both the identification of individual SNPs with striking divergence between the two phenotypic pools, one of them in the proximity of the *PGRPLC* locus, and 10-SNP windows with more subtle but statistically significant divergence, were considered to provide useful information regarding genetic variation associated with the outcome of *S. marcescens* infection, with both approaches pointing to relevant genes.

The list of the 138 *An. gambiae* genes associated with the outcome of *S. marcescens* infection can be found in Table 5.3.

5.10 The locus of PGRPLC, the main receptor of the IMD/REL2 pathway, is associated with the outcome of *S. marcescens* infection

As mentioned above, *PGRPLC*, the gene encoding the main receptor for the IMD/REL2 pathway, was found to be associated with the outcome of *S. marcescens* infection. PGRPLC is known to be involved in mosquito antibacterial responses both before and after a blood meal, while the IMD/REL2 pathway has also been shown to be constitutively triggered by mosquito gut bacteria thus maintaining an elevated level of antimicrobial peptide production (Lin et al., 2007; Meister et al., 2009). Therefore, the identification of genetic divergence associated with the outcome of *S. marcescens* infection related to PGRPLC, a receptor known to be involved in antibacterial responses, confirms that the approach adopted here can provide useful information regarding the genetic basis of the outcome of *S. marcescens* infection. Furthermore, the association of *PGRPLC* with genetic divergence that influences the outcome of *S. marcescens* infection suggests that genetic variation within the mosquito population can influence the ability to mount an antibacterial response via the IMD/REL2 pathway thus contributing to the observed variation in the abundance of the mosquito gut microbiota.

Adjacent to the *PGRPLC* locus, the gene encoding another peptidoglycan recognition protein, *PGRPLA*, was found to be associated with the outcome of *S. marcescens* infection (peak 2L-5). For associated loci found in clusters, the current approach cannot discriminate whether association is due to causal polymorphisms in one or all cluster loci or regulatory sequences affecting one or all loci. In such cases, association could be a result of a selective sweep in the area affecting loci proximal to the causal polymorphism. Therefore, further sequencing-based analysis is required to elucidate the putative contribution of associated genes found in clusters in proximity to associated genetic markers. Nevertheless, PGRPLA has been implicated in antibacterial responses in *Drosophila*, regulating Imd-mediated responses in barrier epithelia (Gendrin et al., 2013) and is highly suspected to play similar roles also in *An. gambiae*.

5.11 EGFR, a known player in antibacterial responses, is associated with the outcome of *S. marcescens* infection

Several other genes associated with the outcome of *S. marcescens* infection also showed homologies that suggested involvement in antibacterial responses, including the association of *EGFR*, detected in the prominent peak 3R-6 both by the individual SNP and the permutation analysis. EGFR is expressed in various developmental stages in *An. gambiae*, showing variation in its protein size and localization (Lycett et al., 2001). The orthologous EGFR pathway in *Drosophila* has been shown to be involved in gut homeostasis following oral bacterial infection (Buchon et al., 2010). Epithelial damage following bacterial infection requires gut repair through elimination of damaged cells by apoptosis, stem cell proliferation and differentiation as well as integration of the newly formed cells in the gut epithelium. These epithelial renewal processes are tightly regulated by both the JAK/STAT (Beebe et al., 2009; Jiang et al., 2009) and EGFR pathways, with several members of the EGFR pathway transcriptionally induced following oral bacterial infection (Buchon et al., 2009) and the EGFR ligand Vein also controlled by the JAK/STAT pathway (Buchon et al., 2010). Disruption of the EGFR pathway in *Drosophila* impairs stem cell proliferation in response to damage induced by bacterial infection, and the failure to repair the gut epithelium following bacterial infection results in increased mortality (Buchon et al., 2010). The association of *EGFR* with the outcome of *S. marcescens* infection suggests that the EGFR pathway may play similar roles in *Anopheles*, thus influencing the outcome of bacterial infection through synergistic roles in gut homeostasis.

5.12 Two associated CLIP domain serine proteases, members of a gene family with broad involvement in immune responses

The genes *CLIPE6* and *CLIPE7* were also associated with the outcome of *S. marcescens* infection by both the individual SNP and permutation analyses. They were found adjacent to each other in peak 3L-16 and they encode CLIP domain serine proteases of the E subtype. CLIP domain serine proteases are known to be involved in immune effector cascades in *Drosophila* but also in *Aedes* and *Anopheles* mosquitoes (Christophides et al., 2002; Jiang and Kanost, 2000; Zou et al., 2010). CLIP domain serine proteases were initially implicated in activating the Toll pathway during *Drosophila* embryogenesis leading to formation of the dorso-ventral axis (Chasan and Anderson, 1989; Jang et al., 2008; Stein and Nusslein-Volhard, 1992). Several CLIP domain serine proteases have also been shown to activate the Toll pathway in *Drosophila* in response to infection with fungi or Gram-positive bacteria (Jang et al., 2006; Kambris et al., 2006; Ligoxygakis et al., 2002).

In a similar manner to the mammalian coagulation (Shrivastava et al., 2007) and complement (Whaley and Schwaeble, 1997) systems, proteinase cascades are also used in arthropods for rapid, local and focused activation of immune reactions utilizing CLIP domain serine proteases (Jiang and Kanost, 2000). Several such examples have been elucidated including the haemolymph clotting system in horseshoe crabs (Iwanaga et al., 1998) but also the prophenoloxidase system in *Manduca sexta* (Jiang et al., 2003; Wang et al., 2006; Zhang et al., 2004*)*. Based on these biochemical studies that implicated CLIP domain serine proteases in the prophenoloxidase activation, it was hypothesized that *An. gambiae* CLIPs may also play similar roles in the melanization response against *Plasmodium* parasites. Indeed, several CLIP domain serine proteases have been shown either to participate or regulate the melanization response in *An. gambiae*, which kills *Plasmodium* parasites in susceptible strains and clears lysed parasites in refractory strains (Barillas-Mury, 2007; Volz et al., 2006). CLIP domain serine proteases have also been implicated in defences against fungi in mosquitoes, both in *Aedes,* in a response mediated by the Toll pathway (Zou et al., 2010) but also in *Anopheles* in a response requiring CLIPA8 but also the complement factor TEP1 (Yassine et al., 2012). Furthermore, the *An. gambiae* CLIP domain serine proteases CLIPB14 and CLIPB15 have been shown to be involved not only in the melanization response against *Plasmodium* but also in defences against Gram-negative bacteria, with both *CLIPB14* and *CLIPB15* showing a marked transcriptional induction following bacterial infection (Christophides et al., 2002; Volz et al., 2005).

CLIP domain serine proteases belonging to the B, C or D subfamilies largely retain their catalytic proteinase activity which is regulated by the CLIP domain, while members of the A and E subfamily are characterized as serine protease homologues as they have lost their proteinase catalytic activity and their functions are related to regulation of immune effector cascades in which they act as cofactors (Kambris et al., 2006; Volz et al., 2006; Waterhouse et al., 2007). The E subfamily encompasses previously unannotated CLIPs in the other four subfamilies, thus comprising a diverse subfamily (Waterhouse et al., 2007). A member of the E subfamily, SPCLIP1, has been shown to be upregulated following infection with *P. falciparum* or the Gram-negative bacterium *E. coli* and has been implicated in defences against both the rodent and human malaria parasites (Dong et al., 2006a). This SPCLIP1 anti-*Plasmodium* activity has been recently shown to be related to the complement response (Povelones et al., 2013).

The association of the yet to be functionally characterized *CLIPE6* and *CLIPE7* with the outcome of *S. marcescens* infection suggests that these serine protease homologues may be involved in regulatory activities related to antibacterial responses.

5.13 Three previously uncharacterized genes encoding FN3 domains are associated with the outcome of *S. marcescens* infection

The permutation analysis also revealed 3 genes, out of a total of 27, encoding proteins with type III fibronectin domains (FN3D) in different peaks: *FN3D1* in peak 2L-4, which was also detected by the individual SNP analysis, *FN3D2* in peak 2L-14 and *FN3D3* in peak 2R-4. The identification of a group of 3 genes in the set of 138 genes associated with the outcome of *S. marcescens* infection with FN3 domains was quite remarkable, taking into account that only 65 genes annotated in the *An. gambiae* genome contain FN3 domains. Therefore, this overrepresentation of associated FN3Ds, not in one cluster but in different peaks, strongly suggested a functional importance for this previously uncharacterized group of genes.

Several of the 65 *An. gambiae* genes possessing FN3 domains encode proteins that function as receptors for diverse signalling pathways. These include AgDscam, a hypervariable pattern recognition receptor (Dong et al., 2006b), the insulin receptor INR (Luckhart and Riehle, 2007) and the JAK/STAT receptor DOME (Bahia et al., 2011). While *FN3D1* only has an FN3 domain, *FN3D2* and *FN3D3* also possess putative transmembrane domains and immunoglobulin domains. A total of 36 *An. gambiae* genes combine FN3 and immunoglobulin domains, including *AgDscam* but also the immunoglobulin superfamily members *IRID5* and *IRID6*. Both IRID5 and IRID6 have been shown to be involved in antibacterial responses, as their silencing significantly reduced mosquito lifespan following a bacterial challenge with *E. coli*, but also, for IRID6, in responses that limit the intensity of *P. berghei* or *P. falciparum* infection (Garver et al., 2008).

5.14 FN3D2 shows homologies with Dscam, a hypervariable pattern recognition receptor

FN3D2 is an orthologue of *Drosophila Dscam4. Drosophila* Dscam was initially characterized as an axon guidance receptor in the *Drosophila* embryo, recognizing guidance signals and recruiting components required for connecting neurons, with *Dscam* silencing resulting in severe defects in axonal organization of the central nervous system (Schmucker et al., 2000). Further studies have established multiple roles for Dscam in the *Drosophila* central nervous system (Hummel et al., 2003; Schmucker, 2007; Wang et al., 2002). Its human homologue, Down syndrome cell adhesion molecule (DSCAM) is also involved in neuronal differentiation in the central nervous system and DSCAM defects are believed to contribute to the development of Down syndrome, as it maps in a region of the human chromosome 21, found in trisomy in affected individuals (Yamakawa et al., 1998).

An intriguing aspect of *Drosophila Dscam* is its extraordinary molecular diversity, encoding a vast repertoire of isoforms generated through alternative splicing. The extracellular region of Dscam consists of both immunoglobulin and FN3 domains; alternative splicing of exons 4, 6, 9 and 17 can modify 3 out of 10 immunoglobulin domains as well as the transmembrane domain, potentially generating up to 38,016 different isoforms (Schmucker et al., 2000) (Figure 5.4A). The *An. gambiae* orthologue, *AgDscam*, can also potentially generate up to 31,920 different isoforms through alternative splicing of exons 4, 6, and 10 (Dong et al., 2006b; Graveley et al., 2004) (Figure 5.4B). Interestingly, considerable divergence is observed between the *Drosophila* and *Anopheles Dscam*, with limited orthologous relationships through sequence similarity between exon variants (Graveley et al., 2004). Such divergence is consistent with the possibility of diverse selective pressures that shape the sequence and thus the function of the respective gene. The molecular diversity of *Drosophila* Dscam has been shown to be critical for some aspects of its functionality in neuron connectivity, suggesting that the generated diversity of Dscam isoforms is critical for the determination of neuronal wiring specificity (Chen et al., 2006). One possibility that could explain the Dscam mode of action, at least in some cases, is that the Dscam binding specificity is related to homophilic interactions between Dscam molecules expressed in different cells, as identical isoforms have been shown to bind to each other through their variable immunoglobulin domains (Hattori et al., 2009; Hughes et al., 2007; Wojtowicz et al., 2004; Wojtowicz et al., 2007).

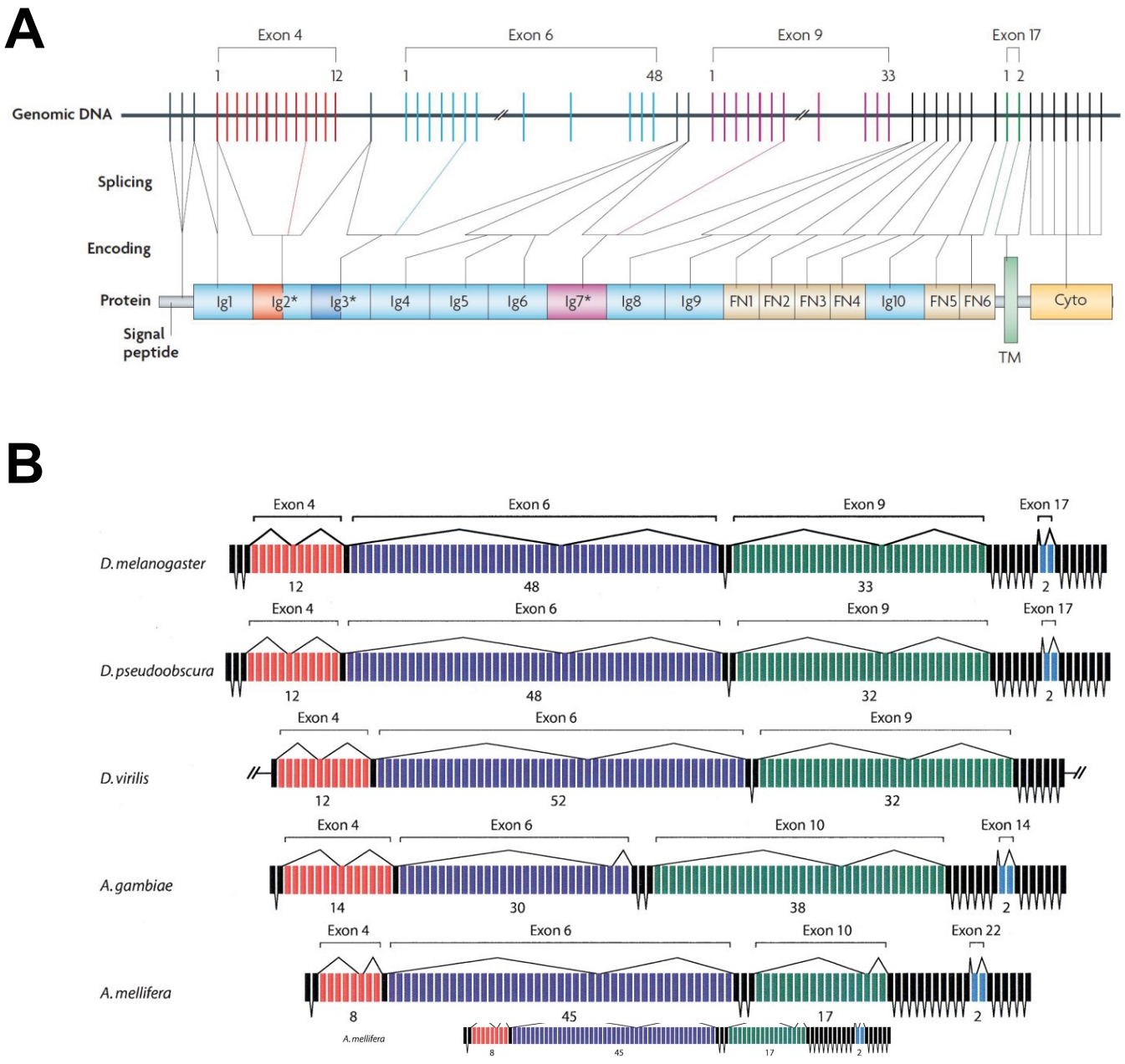

A
Exon 4
Exon 6
Exon 9
Exon 17
1
12
1
48
1
33
1 2
Genomic DNA
Splicing
Encoding
Protein
Signal peptide
Ig1
Ig2*
Ig3*
Ig4
Ig5
Ig6
Ig7*
Ig8
Ig9
FN1
FN2
FN3
FN4
Ig10
FN5
FN6
Cyto
TM
B
D. melanogaster
12
48
33
2
D. pseudoobscura
12
48
32
2
D. virilis
12
52
32
A. gambiae
Exon 10
Exon 14
14
30
38
2
A. mellifera
Exon 22
8
45
17
2
A. mellifera
8
45
17
2

Figure 5.4: *Drosophila* and *Anopheles* Dscam gene structure and protein architecture. A: The *Drosophila* Dscam gene structure and protein architecture. Graphical representation of the Dscam gene comprising 24 exons. Exons 4, 6, 9, and 17 can be alternatively spliced with 12, 48, 33, and 2 alternative exons, respectively, which participate in mutually exclusive splicing for each respective exon. Therefore, *Dscam* splicing can theoretically produce 38,016 splice variants. These splice variants encode a protein product that comprises 10 immunoglobulin domains (Ig1-Ig10), 6 FN3 domains (FN1-FN6), a transmembrane helix (TM) and a cytoplasmic segment (Cyto). As indicated, splice variants differ in the Ig2, Ig3 and Ig7 domains as well as in the transmembrane domain. Adapted from: (Schmucker, 2007). **B:** *Dscam* gene structure in 3 *Drosophila* and 2 *Anopheles* species, as indicated adjacent to each graphical representation. Constitutively spliced exons are in black while alternatively spliced exons are coloured. For each exon, the mutually exclusive alternative exon number is indicated below each exon cluster. For *An. gambiae*, exons 4, 6, 10 and 14 are alternatively spliced, with 14, 30, 38 and 2 alternative exons, respectively. Exons 4, 6 and 10 correspond to immunoglobulin domains while exon 14 corresponds to the transmembrane domain. Adapted from: (Graveley et al., 2004).

Remarkably, *Drosophila* Dscam was shown to have a dual functionality, not only in neuronal wiring but also in the *Drosophila* innate immune system (Watson et al., 2005). A vast repertoire of more than 18,000 *Dscam* isoforms was shown to be expressed in *Drosophila* haemocytes and secreted in the haemolymph. Dscam was further shown to be involved in phagocytosis and directly bind *E. coli* bacteria, suggesting a role for Dscam as a co-receptor or signalling receptor during phagocytosis (Watson et al., 2005). The structure of Dscam isoforms support this dual role of Dscam both in the *Drosophila* central nervous and innate immune system (Meijers et al., 2007); two distinct epitopes are formed on the Dscam surface by variable amino acid residues of the immunoglobulin domains, one involved in homophilic interactions between identical Dscam isoforms and one for non-Dscam ligands possibly involved in innate immune responses but also in additional functionality in the central nervous system (Matthews and Grueber, 2011). Dscam in the crustacean crayfish *Pacifastus leniusculus* was also shown to be involved in bacterial clearance and phagocytosis (Watthanasurorot et al., 2011).

The *An. gambiae Dscam* orthologue, *AgDscam*, also plays a role in immune responses (Dong et al., 2012; Dong et al., 2006b*)*. Silencing *AgDscam* has been shown to decrease mosquito survival following injection of Gram-positive or Gram-negative bacteria in the mosquito haemolymph, combined with an increase in the respective bacterial abundance. The AgDscam antibacterial effect was correlated with an involvement in phagocytosis of both Gram-positive and Gram-negative bacteria (Dong et al., 2006b). Furthermore, AgDscam was shown to influence the outcome of *P. berghei* (Dong et al., 2006b) as well as *P. falciparum* (Dong et al., 2012) infections. Interestingly, *AgDscam* showed no changes regarding its transcript abundance but a shift in isoform representation following a pathogen challenge, with different bacterial or *Plasmodium* elicitors inducing diverse *AgDscam* splice form repertoires, with the induced variants showing higher affinity to their eliciting pathogen (Dong et al., 2006b). Although different *AgDscam* isoform repertoires are generated following a challenge with *P. berghei* or *P. falciparum*, a mix of diverse *P. falciparum* isolates did not show further specificity in the induction of *AgDscam* splice forms (Smith et al., 2011).

The generation of *AgDscam* isoforms that target *P. falciparum* with high affinity has been shown to be controlled by the IMD/REL2 pathway. IMD/REL2 responses to *P. falciparum* infection include the transcriptional regulation of splicing factors that lead to the generation of *AgDscam* isoforms that specifically target *P. falciparum (*Dong et al., 2012*)*. The Toll pathway was also implicated in influencing the AgDscam isoform repertoire, generating variants that show higher affinity for *P. berghei* (Dong et al., 2012). AgDscam variants showing affinity for either *P. berghei* or *P. falciparum* were also shown to be

involved in controlling bacterial load in the mosquito gut (Dong et al., 2012). These results suggest that AgDscam functions as an effector molecule; upon induction of specific isoform variants through the IMD/REL2 or Toll pathway, it targets gut bacteria, *P. falciparum* or *P. berghei*.

The orthologous relationship between *FN3D2* and *Drosophila Dscam4* suggests that the *FN3D2* function might be discrete from the functionally characterized *Dscam*. The *Drosophila* genome contains a family of 4 *Dscam* genes. Except *Dscam, Dscam2-4* remain poorly characterized, although *Dscam2* has been shown to be involved in the neural development of the *Drosophila* visual system (Millard et al., 2007). Furthermore, alternative splicing or generation of a diverse isoform repertoire for *Dscam2-4* has not been reported. A more detailed study of alternative splicing features in *FN3D2* is needed to identify possible expressed isoforms, as in *Dscam*, although no splice variants have been bioinformatically annotated in the current *AgamP3.7 An. gambiae* genome annotation.

5.15 The phylogenetically unrelated FN3D3 shares the same domain architecture with FN3D2 and Dscam

Drosophila Dscam has been shown to play multiple roles in neural circuit assembly (Zipursky and Sanes, 2010). The Dscam mode of action includes self-recognition among neurites that leads to repulsion. This self-avoidance mechanism is thought to depend on recognition of molecular identity based on the expression of different Dscam isoforms (Schmucker, 2007). Homophilic interactions between identical Dscam molecules due to increased binding affinity can lead to repulsion and thus self-avoidance (Hughes et al., 2007; Wojtowicz et al., 2004; Zipursky and Sanes, 2010). Another Dscam mode of action is thought to be the recognition of heterophilic ligands triggering a guidance receptor signalling cascade through the recruitment of additional adaptor proteins, eventually leading to attraction (Evans and Bashaw, 2010; Schmucker et al., 2000; Zhan et al., 2004). This involvement of *Drosophila* Dscam in axon guidance is mediated by the recognition of the chemoattractant Netrin, as is the case for its human counterpart DSCAM (Andrews et al., 2008; Liu et al., 2009a; Ly et al., 2008; Matthews and Grueber, 2011).

Axon guidance in flies is a complex process that relies on molecules that act as attraction signals, including Netrin or as repellants, such as Slit (Evans and Bashaw, 2010). These guidance responses involve several receptors which co-ordinate a balance between attraction and repulsion. Dscam and Frazzled recognize Netrin and act as attractive receptors whereas Roundabout (Robo) mediates repulsion in response to Slit (Andrews et al., 2008; Dickson and Gilestro, 2006; Garbe and Bashaw, 2007). The *Drosophila* axon guidance receptors Dscam (Wojtowicz et al., 2004), Frazzled (Kolodziej et al., 1996) and Robo (Kidd et al.,

1998) constitute phylogenetically unrelated receptors, independently identified in having discrete but similar functions in neural wiring. Remarkably, though, these receptors share a similar domain architecture in their extracellular part that includes immunoglobulin and FN3 domains. The mammalian transmembrane receptor Boc and several cell adhesion molecules, including NCAM, also involved in axon guidance, share a similar architecture, possessing immunoglobulin and FN3 domains in their extracellular region (Sanchez-Arrones et al., 2012; Walsh and Doherty, 1997). One explanation for the observed similarity in the constituent domains of these receptors is that independent selection processes resulted in the recruitment of receptors with immunoglobulin and FN3 domains in discrete axon guidance functions in *Drosophila* neural wiring.

FN3D3, also associated with the outcome of *S. marcescens* infection, shows no phylogenetic similarity with *FN3D2* or *Dscam*, yet it shares a similar domain architecture, consisting of immunoglobulin and FN3 domains as well as a putative transmembrane helix (Figure 5.5). *Drosophila* or *Ae. aegypti* orthologues of *FN3D3* have not been functionally characterized. An attractive hypothesis is that *FN3D2* and *FN3D3*, although they show no phylogenetic similarity, might serve discrete yet similar functions in antibacterial immune responses, based on their common domain architecture consisting of immunoglobulin and FN3 domains, along the lines of the observed similarities between domain architecture and function of axon guidance receptors, including Dscam.

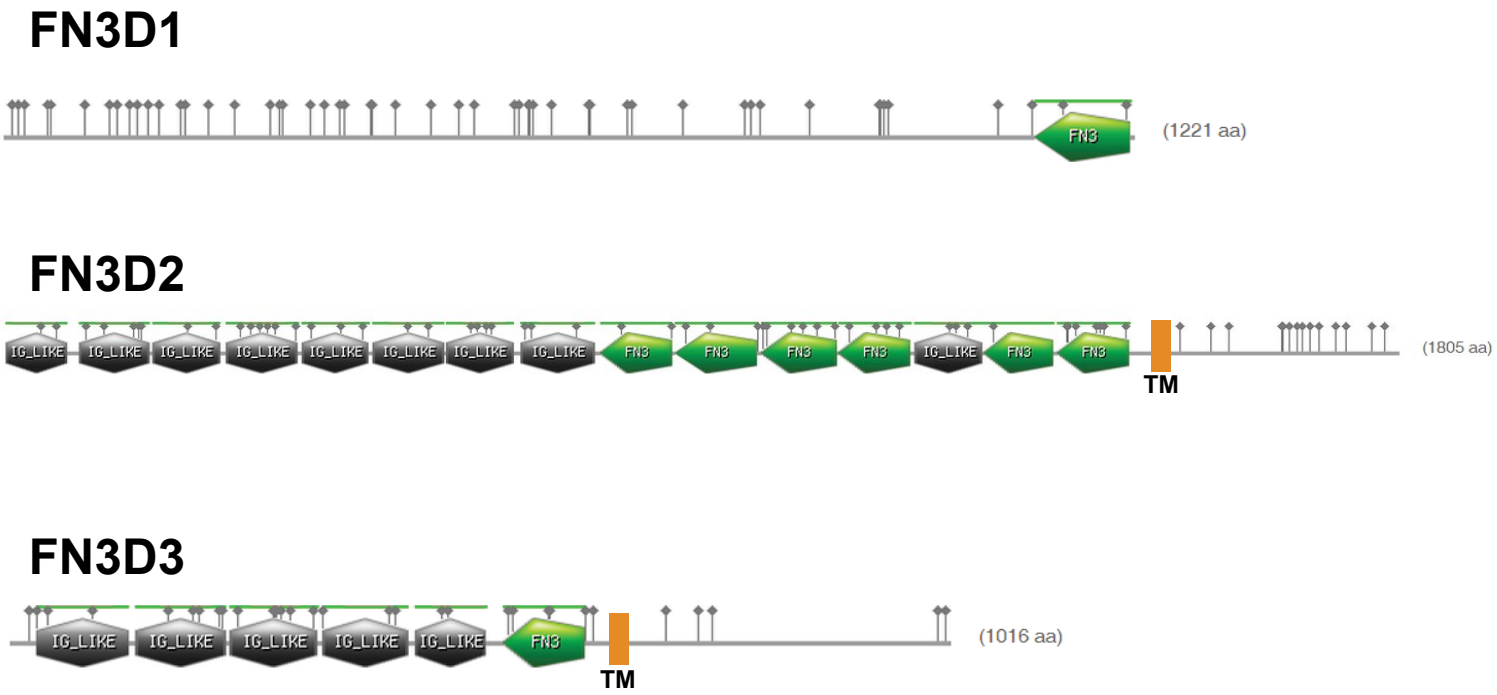


Figure 5.5: Domain architecture of the FN3D1-3 predicted proteins. The gene sequence of *FN3D1*, *FN3D2* and *FN3D3* was used to generate the domain architecture of the respective predicted encoding protein, using the PROSITE collection of motifs (Sigrist et al., 2013). The predicted protein size in amino acids (aa) can be found adjacent each graphical representation. IG_LIKE corresponds to immunoglobulin predicted domains, FN3 corresponds to FN3 domains and TM denotes a predicted transmembrane domain.

5.16 FN3D1, a putative regulator of gene expression, also shares an FN3 domain

FN3D1 shows a distinct domain architecture compared to FN3D2 and FN3D3. FN3D1 does not seem to possess a putative transmembrane domain, so it is unlikely for FN3D1 to play a role as a transmembrane receptor. Except for the FN3 domain, no other domains are predicted based on the FN3D1 protein sequence (Figure 5.5). As Dscam interactions in *Anopheles* and *Drosophila*, including binding to bacteria or *Plasmodium*, occur through the immunoglobulin domains, which are also the domains showing sequence variability among isoforms (Dong et al., 2012; Meijers et al., 2007), FN3D1 is unlikely to show any functional similarity to Dscam. Interestingly, though, the FN3D1 protein sequence shows similarity to the human activating transcription factor 7-interacting protein (ATF7IP). ATF7IP is a nuclear protein acting as a transcriptional co-activator or co-repressor by interacting with components of the basal transcription machinery (De Graeve et al., 2000; Liu et al., 2009b). Therefore, FN3D1 can play a similar role in the regulation of gene expression following bacterial infection in *An. gambiae*. *FN3D1* shows a one-to-one orthologous relationship with *Drosophila windei,* encoding a transcriptional cofactor that acts during germ line development for the nuclear localization of the methyltransferase Egg, while it also localizes in the euchromatin (Koch et al., 2009). This orthologous relationship suggests that FN3D1 may also function in a similar manner in altering gene expression through epigenetic regulation following bacterial infection.

5.17 A possible correlation of FN3 domains and antibacterial functionality

As was made clear above, no functional relevance can be ascribed to the FN3 domains present in FN3D1-3 based on prior functional analysis of Dscam or other axon guidance receptors, which also combine immunoglobulin and FN3 domains. In Dscam, variability between isoforms and homophilic or heterophilic interactions involve the immunoglobulin domains (Dong et al., 2012; Meijers et al., 2007; Schmucker, 2007). FN3 domains are similar to C2 immunoglobulin domains and are typically found in extracellular matrix proteins, contributing to cell adhesion through an arginine-glycine-asparagine motif (Leahy et al., 1996; Main et al., 1992; Ruoslahti, 1984). FN3 domains are also found, though, in addition to the axon guidance receptors already mentioned, in several other cell adhesion molecules involved in neural development, including the *Drosophila* neuroglian and the human neural cell adhesion molecule NCAM, again in conjunction with immunoglobulin domains (Huber et al., 1994; Walsh and Doherty, 1997). The FN3 domain in NCAM has a direct functional relevance to NCAM involvement in axonal outgrowth, as it is the FN3 domain that interacts with the fibroblast growth factor receptor FGFR (Anderson et al., 2005; Carafoli et al., 2008; Kiselyov et al., 2003). Furthermore, an FN3 domain in NCAM is also essential for N-glycan polysialylation of adjacent immunoglobulin domains (Mendiratta et al., 2006). Therefore, the FN3

domains found in FN3D1-3 can have a direct functional role, possibly in complex formation or regulatory processes, which remains to be determined.

Interestingly, though, FN3 domains of the human fibronectin, found in the extracellular matrix, have been shown to bind to adhesion proteins derived from the surface of bacteria of the genus *Moraxella*, in interactions important for bacterial colonization (Agnew et al., 2011). It is also possible, therefore, that, as immunoglobulin domains in Dscam, FN3 domains can directly interact with bacteria.

5.18 FN3D1-3 antibacterial activities underlying their association with the outcome of *S. marcescens* infection?

The association with the outcome of *S. marcescens* infection of three genes with FN3 domains in different peaks, with two of them sharing a similar domain architecture with immunoglobulin and FN3 domains despite no phylogenetic relation, strongly suggests that the observed association may be a result of a causal function of these genes, with this novel family of FN3Ds playing a role in antibacterial responses. Furthermore, the homology of *FN3D2* with *Dscam*, a hypervariable pattern recognition receptor known to bind bacteria and mediate antibacterial responses (Dong et al., 2012; Watson et al., 2005), further strengthens this possibility. The experimental design of this association study cannot provide further insights regarding the function of FN3D1-3 or allelic variation that results in the observed association. Further functional characterization can shed more light not only in any putative involvement of FN3D1-3 in antibacterial responses but also in any epistatic interactions that led to their association with the outcome of *S. marcescens* infection.

5.19 Five GPCRs are associated with the outcome of *S. marcescens* infection

Five genes encoding G protein-coupled receptors (GPCRs) were found in different peaks: the serotonin receptor GPR5HT7 in peak 2R-14, a putative GPCR with a 7-transmembrane rhodopsin-like domain (AGAP004783) in peak 2L-1, the neuropeptide receptor GPRNPR2 in peak 3R-5, the orphan class B GPCR GPRNNB2 in peak 3R-10 and the GABA-B family GPCR GPRGBB1 in peak 3R-15. The association of five GPCRs in different peaks, as for the FN3Ds, suggests a functional significance for this receptor family in influencing the outcome of *S. marcescens* infection.

GPCRs, both in humans (Fredriksson et al., 2003) and *An. gambiae (*Hill et al., 2002*)*, play pivotal roles in signal transduction of most fundamental biological processes. There are more than 800 GPCRs in the

human genome (Fredriksson et al., 2003), while 276 GPCRs were initially annotated in *An. gambiae* (Hill et al., 2002), with more genes including GPCR signatures having been further discovered (Nowling et al., 2013). GPCRs are characterized by a central domain with 7 transmembrane regions (Baker, 2010b) and recognize a wide array of ligands, including chemosensory cues in olfactory (Spehr and Munger, 2009) or taste receptors (Kinnamon, 2012), small molecules functioning as hormones or neurotransmitters such as adenosine (Hino et al., 2012; Lebon et al., 2011), noradrenaline (Lebon et al., 2011), opioids such as morphine or heroin (Granier et al., 2012; Thompson et al., 2012; Wu et al., 2012), neurotensin (White et al., 2012) or acetylholine (Haga et al., 2012; Kruse et al., 2012) or function as photoreceptors such as rhodopsin, whose activation by light leads to vision (Palczewski et al., 2000). A GPCR conformational change following ligand binding relays a signal to G proteins, further leading to activation of various downstream processes (Flower, 1999; Venkatakrishnan et al., 2013). Here, the classification of GPCRs includes only non-sensory GPCRs. Odour or taste receptors possessing 7-transmembrane chemoreceptor domains are considered a separate class due to differences in topology from canonical GPCRs that relate to their function as ion channels (Benton et al., 2006; Sato et al., 2008). Taste receptors controlling gustation (gustatory receptors) and odourant receptors in both mosquito and *Drosophila* have been a subject of intense research interest (Carey et al., 2010; Hallem et al., 2006; Weiss et al., 2011b) and will be addressed separately further on.

GPCRs constitute the most common of drug targets, with more than 50% of modern drugs exerting their action by altering GPCR activities (Flower, 1999). The potential of targeting GPCRs in *An. gambiae* in efforts to disrupt malaria transmission, for example through insecticide development that target GPCRs, remains under-exploited (Hill et al., 2002; Nowling et al., 2013; Van Hiel et al., 2010). One explored possibility is the screening of GPCRs which are phylogenetically distant from mammalian receptors for chemical compounds that may act as antagonists to these GPCRs, thus exerting significant toxicity to larval or adult mosquitoes, with minimal side effects to humans or other organisms (Meyer et al., 2012).

5.20 Neurotransmitter-triggered GPCRs and innate immunity

The contribution of GPCRs to mosquito immunity remains poorly understood. A study investigating transcriptional regulation following *P. berghei* or *P. falciparum* infection in *An. gambiae,* identified the GPCR family as significantly overrepresented in the set of differentially expressed genes following *P. falciparum* infection (Mendes et al., 2011). Several of these GPCRs, including the neuropeptide receptor GPRNPY3, were shown to influence the outcome of *Plasmodium* infections, with *GPRNPY3* silencing

increasing the *P. falciparum* infection intensity (Mendes et al., 2011). The mechanism through which these GPCRs exert their agonistic or antagonistic effect on the outcome of *Plasmodium* infection remains unclear.

One possibility is that GPCRs expressed in the nervous system might regulate innate immune responses. In the widely used model system of the nematode *Caenorhabditis elegans,* the GPCR NPR-1, related to mammalian neuropeptide Y receptors, is required for defence to infection with the pathogenic bacterium *Pseudomonas aeruginosa* (Styer et al., 2008). NPR-1 is involved in a neural circuit that modulates *C. elegans* feeding behaviour based on external cues such as oxygen and food availability (Coates and de Bono, 2002; Gray et al., 2004; Rogers et al., 2003). The role of NPR-1, though, in defences against *P. aeruginosa* was shown to be independent of a behaviour phenotype and related to regulation of innate immune responses through the PMK-1/p38 mitogen-activated protein kinase (MAPK) signalling pathway (Styer et al., 2008). Another study identified a behavioural component in the NPR-1 antibacterial effect (Reddy et al., 2009). Genetic polymorphisms in the *NPR-1* gene were found to influence *C. elegans* susceptibility to *P. aeruginosa* infection while the NPR-1 involvement was found to be related to oxygen-dependent behavioural avoidance of a *P. aeruginosa*-containing food source (Reddy et al., 2009). These studies indicate a tight link between the nervous and innate immune systems, with GPCRs regulating this interplay that involves both behavioural immune responses but also regulation of innate immune responses (Aballay, 2009).

Three of the GPCRs associated with the outcome of *S. marcescens* infection, *GPRNPR2*, *GPR5HT7* and *GPRGBB1,* are likely to be triggered by neurotransmitters. Therefore, the implication for these GPCRs is that they may modulate the outcome of *S. marcescens* infection either through regulation of innate immune responses in an interplay between the nervous and innate immune systems or through modulation of behavioural immune responses related to aversion to bacteria or feeding suppression. Further phenotypic characterization of these GPCRs can shed more light on their putative mode of action.

The interplay between the nervous system and immunity, with neurotransmitters playing key modulatory roles at the interface of such interactions, has been extensively studied in mammalian adaptive immunity (Pacheco et al., 2010). This is especially true for the role of serotonin (5-HT), a neurotransmitter involved in processes within the central nervous system that lead to the control of mood and thus are responsible for neuropsychiatric disorders (Baganz and Blakely, 2013; Duerschmied et al., 2013; Qu and Chaikof, 2010), which is also a key modulator of immune responses. At least 14 5-HT receptors are expressed on the surface of immune cells (Nichols and Nichols, 2008), while both T cells and dendritic cells are known

to release 5-HT (Katoh et al., 2006). Activation of Toll-like receptors leads to 5-HT intake by dendritic cells that is subsequently released to T cells, while T cells also release 5-HT during inflammatory responses (O'Connell et al., 2006; Pacheco et al., 2010). Several lines of evidence (Baganz and Blakely, 2013) indicate a cross-talk between the immune and nervous systems, including pathological expression of 5-HT receptors in NK cells of Alzheimer's patients (Martins et al., 2012), increased 5-HT blood levels in autism (Veenstra-VanderWeele and Blakely, 2012) or alterations in the activity of lymphocyte populations in individuals receiving anti-depressant treatments based on selective serotonin re-uptake inhibitors (SSRIs) (Evans et al., 2008; Hernandez et al., 2010). 5-HT is also expressed in enteric endocrine cells and influences intestinal inflammation and gut physiology (Cirillo et al., 2011; Khan and Ghia, 2010).

5.21 An NPF link between the *GPR5HT7* and *GPRGBB1 Drosophila* orthologues

The *Drosophila* orthologue of *GPR5HT7*, *5-HT7*, has been implicated in several behavioural processes, including sleep, circadian rhythms, courtship, mating, olfactory learning and memory (Becnel et al., 2011; Johnson et al., 2011). Furthermore, 5-HT in the *Drosophila* brain modulates aggressive behaviour through serotonin receptors, including 5-HT7, in a process that is also modulated by neuropeptide F (NPF) (Dierick and Greenspan, 2007).

As 5-HT, gamma-aminobutyric acid (GABA), another neurotransmitter, is also involved in modulation of adaptive immune responses (Tian et al., 2004). The *Drosophila* orthologue of *GPRGBB1*, *GABA-B-R1*, has been implicated in behavioural responses related to alcohol sensitivity (Dzitoyeva et al., 2003). Interestingly, NPF has also emerged as a major modulator of alcohol sensitivity (Shohat-Ophir et al., 2012; Wen et al., 2005) and could possibly be involved in circuits related to GABA-B-R1.

Despite the compelling evidence presented above for a link between neurotransmitters and immunity in mammals, the involvement of such neural circuits in mosquito immunity remains to be determined. The association of GPCRs with the outcome of *S. marcescens* infection, especially those likely to recognize neurotransmitters such as GPRNPR2, GPR5HT7 or GPRGBB1, is the first indication for the involvement of these GPCRs or their *Drosophila* orthologues in immune responses. An intriguing finding is the putative cross-talk between the *Drosophila* orthologues of *GPR5HT7* and *GPRGBB1* through NPF. The involvement of NPF in mosquito responses against *S. marcescens* will be identified and discussed later on in this study. The possible cross-talk between GPR5HT7, GPRGBB1 or GPRNPR2 with NPF signalling could also be a subject of further research, indicating complex circuits underlying the response to *S. marcescens*.

5.22 Two gustatory receptors are associated with the outcome of *S. marcescens* infection

Two genes encoding the gustatory receptors (Gr) Gr9 and Gr10 were associated with the outcome of *S. marcescens* infection, found in peak 3R-11. These are chemosensory receptors possessing 7-transmembrane chemoreceptor domains. 76 putative Grs have been characterized in *An. gambiae* (Hill et al., 2002). *Gr9* is one of only five *An. gambiae* Grs for which alternative splicing has been hypothesized, with 14 NH_2-terminal exons, each hypothesized to be alternatively spliced to 2 short COOH-terminal exons encoding 7-transmembrane regions (Figure 5.6A) (Hill et al., 2002). Indeed, *Gr9* has 13 predicted splice variants, according to the *AgamP3.7 An. gambiae* genome release, corresponding to transcripts AGAP009805-RA to AGAP009805-RM. Taking into account that *PGRPLC* also has 8 predicted splice variants, shown to have functional importance in regulation of IMD/REL2 antimicrobial responses (Lin et al., 2007; Meister et al., 2009), it is intriguing to speculate whether alternatively spliced genes show particular proclivity in genetic variation that alters their function, resulting in genetic associations. Therefore, the *Gr9* gene structure suggests that *Gr9* is more likely than *Gr10* to influence the outcome of *S. marcescens* infection, leading to SNP divergence in the specific genomic area.

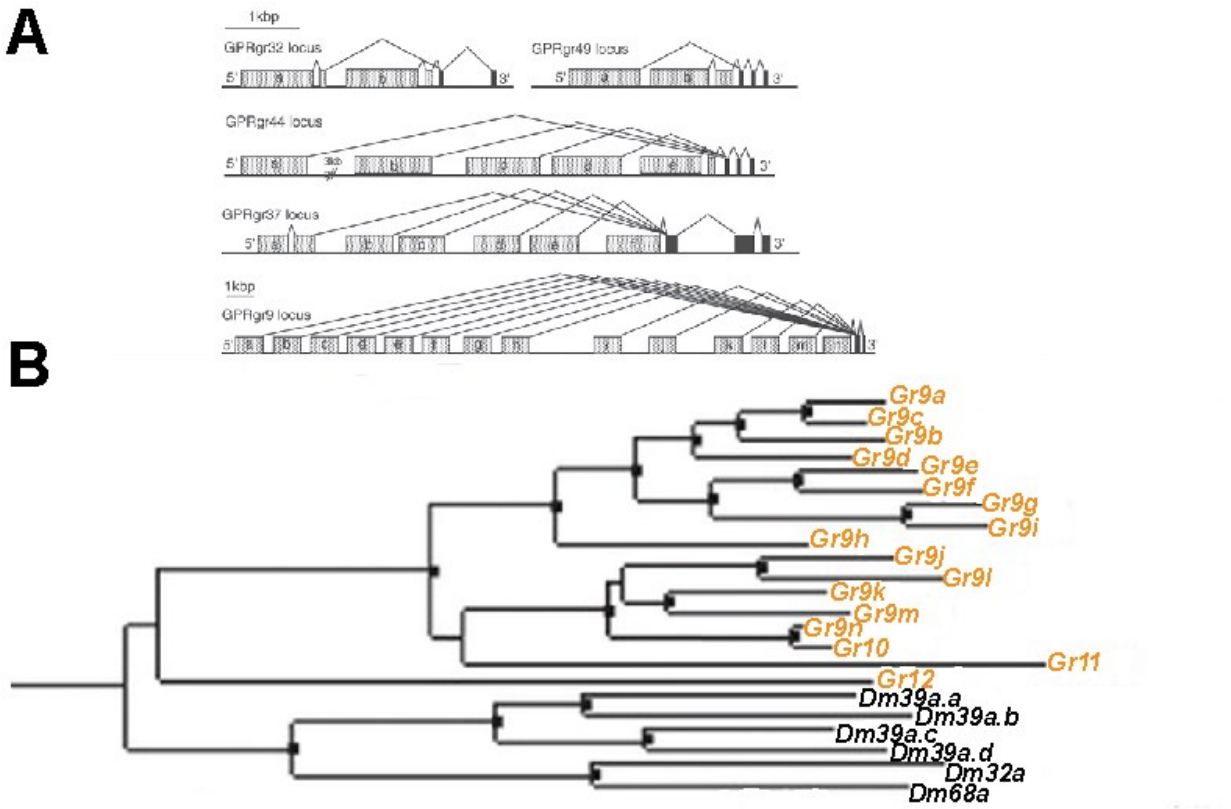


Figure 5.6: *Gr9* alternative splicing and phylogenetic relationships of *Gr9* and *Gr10*. A: Gene structure and proposed alternative splicing for five *An. gambiae* Grs. *Gr9* is indicated as *GPRgr9*. Gray boxes indicate N-terminal exons unique to differentially spliced products and black boxes indicated shared C-terminal exons. **B:** Phylogenetic relationships of *An. gambiae Gr9* and *Gr10* with *Drosophila* Grs. *An. gambiae Gr9* splice variants and *Gr10* are indicated in orange type while neighbouring *Drosophila* Grs characterized as orthologues of *Gr9/Gr10* are indicated in black type. Adapted from: (Hill et al., 2002).

Due to extensive gene duplication in mosquito and *Drosophila* Grs, as is the case for olfactory receptors, orthologous relationships of *An. gambiae* Grs with *Drosophila* counterparts are complicated, typically resulting in many-to-many orthologues (Hill et al., 2002). *Gr9* and *Gr10* form a group of paralogues, along with *Gr11* and *Gr12*, with both *Gr9* and *Gr10* exhibiting an orthologous many-to-many relationship with the *Drosophila Gr32a*, *Gr39a* and *Gr68a* (Figure 5.6B). *Gr68a* has been shown to be expressed in gustatory bristles in the *Drosophila* forelegs and act as a pheromone receptor, involved in male courtship performance (Bray and Amrein, 2003). Gr32a is also involved in modulation of mating behaviour through pheromone recognition, inhibiting male courtship to mated females or other males through recognition of respective inhibitory pheromones (Miyamoto and Amrein, 2008), while Gr39a has been implicated, through 4 splice variants, in sustaining courtship behaviour in males (Watanabe et al., 2011). Furthermore, Gr32a has been shown to recognize small non-volatile hydrocarbons, most likely acting as gustatory pheromones, and regulate behaviours related to mating but also regulation of aggressive behaviour (Wang et al., 2011a). Therefore, all members of the *Drosophila* group of Grs exhibiting an orthologous relationship with Gr9 and Gr10 were shown to mediate the behavioural effects of small hydrocarbons acting as pheromones, mainly on mating behaviour. Although binding of such molecules on Grs was not demonstrated, it is very likely that Grs are the receptors for these molecules, acting individually or through formation of Gr complexes.

In addition to pheromone recognition related to mating behaviour, the expression of Gr32a and Gr39a in bitter taste neurons suggests a possible involvement in discerning bitter taste compounds and initiating aversive behaviour (Weiss et al., 2011b). Furthermore, Gr32a, expressed in the labellum, situated in the *Drosophila* proboscis, has been implicated in feeding suppression following recognition of DEET, one of the most potent insecticides, and other naturally occurring antifeedants (Lee et al., 2010). Therefore, Gr32a, Gr39a and Gr68a, in response to a multitude of chemosensory cues, through their expression in gustatory receptor neurons in external sensory organs situated in the proboscis or the forelegs, are able to initiate aversive behavioural responses. Based on the orthologous relationship of *An. gambiae Gr9* and *Gr10* with these Grs, it would be expected that Gr9 and Gr10 would show some functional conservation with their *Drosophila* counterparts in recognition of chemosensory cues and initiation of aversive behaviour related to mating behaviour or taste.

Functional characterization of other *Drosophila* Grs has implicated them in aversive taste of caffeine (Lee et al., 2009; Moon et al., 2006) as well as various antifeedants including quinine or strychnine (Moon et al., 2009) but also the insecticide L-canavanine (Lee et al., 2012). Two *Drosophila* Grs, and most likely also

their *An. gambiae* orthologues, have been implicated in responses to CO_2 (Jones et al., 2007), which in mosquitoes are important for their host-seeking behaviour. Their *Ae. aegypti* orthologues have also been involved in CO_2 sensitivity, as their silencing resulted in reduced CO_2 perception (Erdelyan et al., 2011). A subfamily of *Drosophila* Grs has been characterized in recognition of sugar, including Gr5a, a trehalose receptor (Chyb et al., 2003; Dahanukar et al., 2001; Ueno et al., 2001), and six Gr64 genes, required for responses to glucose, sucrose, maltose and other sugars (Jiao et al., 2007; Slone et al., 2007). The requirement for two Grs to confer CO_2 sensitivity, the influence of Gr64a on trehalose sensitivity and the partial response of gustatory receptor neurons upon misexpression of caffeine or sugar Grs suggests that the Gr mode of action is likely to include the formation of heteromultimers, heterodimers in the case of CO_2 Grs, upon co-expression in subsets of gustatory receptor neurons (Erdelyan et al., 2011; Jiao et al., 2007; Slone et al., 2007).

The biological significance of sugar sensitivity by the respective *Drosophila* Grs was thought to be the formation of feeding preference in favour of nutritious meals, as the flies search for food. Interestingly, there is broad variation in trehalose sensitivity in natural *Drosophila* populations, largely explained by genetic variation in the *Gr5a* trehalose receptor locus, with variants conferring low trehalose sensitivity, considered the ancestral form in *Drosophila* (Inomata et al., 2004). CO_2 sensitivity in mosquitoes seeking hosts as well as the avoidance of males or mated females in male flies or the avoidance of harmful chemicals or irritants contained in food, conferred by the respective Grs, could also provide selective advantages for these Gr-mediated traits, inferring possible selective pressures exerted on genetic variation related to the respective Gr loci.

Although the function of *Drosophila* Grs has been predominantly studied in external sensory organs possessing gustatory receptor neurons, e.g. the proboscis, antennae or legs, a key regulatory function was discovered for Gr43a through its expression in the brain (Miyamoto et al., 2012; Miyamoto et al., 2013). Gr43a was shown to function as a fructose receptor in taste neurons but also functioned as a nutrient sensor in the *Drosophila* brain. Fructose haemolymph levels were shown to activate Gr43a and regulate feeding behaviour depending on the fly's feeding status. Although fructose sensing by Gr43a promoted feeding in hungry flies, in satiated flies fructose sensing resulted in feeding suppression (Miyamoto et al., 2012).

5.23 *Gr9* is expressed in the mosquito midgut

One poorly understood aspect of Gr function is related to Gr midgut expression. At least 12 *Drosophila* Grs are expressed in the fly's gut (Park and Kwon, 2011). This Gr expression has been demonstrated in *Drosophila* enteroendocrine cells, *i.e.* chemosensory gut cells that can secrete regulatory peptides, where these Grs co-localize with regulatory peptides that include NPF (Park and Kwon, 2011). The Gr role in the fly's gut is thought to involve regulatory responses related to food intake, nutrient absorption or glucose homeostasis (Miyamoto et al., 2013; Park and Kwon, 2011). Several taste receptors show gut expression in mammals, including the T1R2/T1R3 sugar receptors, expressed in enteroendocrine cells on the intestinal luminal epithelium (Miguel-Aliaga, 2012). Activation of these receptors through sugar recognition results in the release of peptide hormones that increase sugar absorption through pancreatic insulin secretion (Behrens and Meyerhof, 2011).

Gr9 has shown significant upregulation, compared to other tissues, in the midgut of blood-fed adult mosquitoes (Marinotti et al., 2006) and also in the midgut of adult mosquito tissues (Baker et al., 2011). *Gr9* is also expressed in external sensory organs, although a comparison of transcription profiles between antennae or maxillary palps and whole body transcriptomes in female mosquitoes, showed a non-significant *Gr9* upregulation of 1.42-fold in antennae and 1.15-fold in maxillary palps (Pitts et al., 2011). The *Gr10* expression profile is quite different; no significant midgut upregulation has been reported in adult mosquitoes (Baker et al., 2011; Marinotti et al., 2006) but *Gr10* shows significant enrichment in antennae, with a 9.5-fold upregulation, and almost no upregulation in maxillary palps (1.01-fold), compared to whole body transcriptomes (Pitts et al., 2011). These data suggest that Gr9 is more likely to play a functional role in the midgut while the Gr10 function might be related to its expression in the mosquito's antennae. In relation to the outcome of *S. marcescens* infection, the *Gr9* midgut upregulation is more plausible to account for influencing the outcome of *S. marcescens* infection, indicating that the *Gr9* association might be more relevant regarding the underlying causality related to the outcome of *S. marcescens* infection.

5.24 Grs and antibacterial responses

No *Drosophila* or mosquito Grs have been implicated in antibacterial responses that could explain the *Gr9/Gr10* association with the outcome of *S. marcescens* infection. Intriguingly, though, the mammalian chemoattractant receptor GPR43 has been shown to recognize short-chain fatty acids of bacterial origin and regulate inflammatory responses (Maslowski et al., 2009). Other chemoattractant receptors, such as

ChemR23, can recognize host-derived lipids and also regulate inflammatory responses (Arita et al., 2007; Serhan et al., 2008). It is thus conceivable that, as their mammalian counterparts, insect Grs might also recognize bacterial or host derived molecules and initiate responses related to immune modulation or feeding behaviour.

Based on known Gr functions, two hypotheses emerge that could explain the *Gr9/Gr10* association with the outcome of *S. marcescens* infection. One possibility, related to the involvement of the *Gr9/G10 Drosophila* orthologues *Gr32a*, *Gr39a* and *Gr68a* in aversive responses following contact chemosensation, is that Gr9 or Gr10 recognize bacterial-derived metabolites or mosquito molecules induced following *S. marcescens* infection and initiate a response that leads to an aversive behaviour. This aversive behaviour can limit or disrupt intake of bacteria-containing sugar and thus influence the outcome of *S. marcescens* infection. The *Gr9* midgut expression could be quite relevant to such aversive response to bacteria, taking into account that the role of midgut expression for *Drosophila* Grs remains poorly resolved but it can be related to NPF-mediated changes in feeding behaviour (Park and Kwon, 2011).

Another possibility is that Gr9 or Gr10 can modulate sugar meal size, possibly through the recognition of nutrients or satiety-related signals. Therefore, by suppressing or promoting feeding, Gr9 or Gr10 may indirectly modulate the abundance of *S. marcescens* taken in by the mosquito. *Gr9* or *Gr10* variants that affect the efficiency of such function can decisively influence the outcome of *S. marcescens* infection, albeit in an indirect way. Although *Gr9/Gr10* show no orthologous relationships with *Drosophila* Grs acting as sugar receptors or nutrient sensors, it is possible that either nutrient recognition has independently evolved in mosquitoes or that Gr9/Gr10 recognize a mosquito-induced cue related to satiation rather than nutrients.

In *Drosophila,* sugar meal size is tightly regulated to achieve a balance between starvation and satiation and several mechanisms have been identified that can modulate feeding behaviour. In adult flies, allatostatin A neurons can suppress while NPF activation can enhance feeding behaviour (Hergarden et al., 2012). The cues that modulate such neural circuits remain elusive; allatostatin A is expressed in gut enteroendocrine cells and proposed mechanisms for its regulation include recognition of gut distention signals or signals related to metabolic changes (Hergarden et al., 2012). Meal size in adult *Drosophila* is also regulated by the leucokinin pathway, in a discrete manner to NPF, while leucokinin neurons are also expressed in the foregut (Al-Anzi et al., 2010). The involvement of midgut-expressed Grs in regulating these neural circuits has not been demonstrated but can be hypothesized. A clear link between neural circuits that regulate feeding behaviour and Grs has been provided by the *hugin* neurons in *Drosophila*

(Melcher and Pankratz, 2005). These neurons regulate the initiation phase of feeding in adult flies based on chemosensory and nutrient signals; in larvae these neurons are connected to external Gr-expressing neurons and internal pharyngeal chemosensory organs (Melcher and Pankratz, 2005).

5.25 Putative modulators of gene expression associated with the outcome of *S. marcescens* infection

In addition to *FN3D1*, encoding a putative transcription cofactor, several genes associated with the outcome of *S. marcescens* infection encoded domains that suggest involvement in regulation of gene expression. AGAP005096 was identified by the permutation analysis, in peak 2L-4, encoding a homeobox domain. AGAP005244, found in peak 2L-5, along with *PGRPLC* and *PGRPLA*, also encoded a homeodomain-like DNA binding domain. The *Drosophila* homeobox gene *Caudal* has been shown to repress the transcription of antimicrobial peptide genes controlled by the Imd pathway, while silencing of *Caudal* resulted in overexpression of Imd-induced antimicrobial peptides that altered the fly's gut microbiota population structure (Ryu et al., 2008). In *Anopheles*, the *Caudal* orthologue also acts as a negative regulator of the IMD/REL2 pathway in the midgut (Clayton et al., 2012). *Caudal* silencing enhances production of immune effectors, including the cecropin and defensin antimicrobial peptides, and thus resistance to bacterial infections, and also reduces the *P. falciparum* infection intensity via the IMD/REL2 pathway (Clayton et al., 2012). Despite its effect in bacterial tolerance in the mosquito gut, *Caudal* was not detected as associated with the outcome of *S. marcescens* infection. This can be attributed to the lack of *Caudal* variants in the gene pool of the assayed mosquito colony, a possible redundancy of its function by other genes or selective pressures on *Caudal* genetic variation due to a possible critical involvement in developmental processes as has been shown in *Drosophila* (Dearolf et al., 1989; Mlodzik and Gehring, 1987). It is possible, though, that other transcription factors may modulate gene expression that could influence the outcome of *S. marcescens* infection. A possible involvement of these putative modulators of gene expression in antibacterial immunity can be further investigated to establish causality related to their association with the outcome of *S. marcescens* infection.

Other genes also putatively related to regulation of gene expression that were associated with the outcome of *S. marcescens* infection included AGAP002492, in peak 2R-7, encoding a Nuclear Respiratory Factor 1 DNA-binding and dimerization domain. Its *Drosophila* orthologue, *Erect wing* (*ewg*), has been shown to be involved in the Wnt/Wingless pathway (Xin et al., 2011). AGAP005156, in peak 2L-4, encoded an ARID/BRIGHT DNA binding domain. Its *Drosophila* orthologue, *retained*, has been implicated in

modulation of male/female sexual behaviour and repression of male courtship (Ditch et al., 2005; Shirangi et al., 2006). The possible involvement of its mosquito counterpart in similar behavioural responses merit follow-up, in relation to similar involvement of Grs in mosquito behaviour, as mentioned above. AGAP005661, in peak 2L-7, encoding a putative ligand-regulated transcription factor, is an orthologue of the *Drosophila* nuclear receptor *FTZ-F1*, involved in juvenile hormone mediated gene expression (Dubrovsky et al., 2011). Juvenile hormone signalling can be related to behavioural responses, since it has been implicated in starvation resistance and trehalose homeostasis in the red flour beetle *Tribolium castaneum* (Xu et al., 2013). Importantly, juvenile hormone signalling in *Ae. aegypti* assesses nutritional information before reproductive maturation, with elevated juvenile hormone levels found in sugar-fed females through allatotropin stimulation and insulin/TOR pathway signalling (Noriega, 2004; Perez-Hedo et al., 2013). Therefore, the involvement in behavioural processes of this putative transcription factor can be further followed up along these lines.

Finally, additional genes associated with the *S. marcescens* infection outcome possibly involved in modulation of gene expression include six genes encoding zinc finger domains, AGAP002664, in peak 2R-8, encoding the transcription cofactor Mediator of RNA Polymerase II transcription subunit 27 MED27, AGAP012397, in peak 3L-20, encoding the DNA-directed RNA polymerase subunit rpb8, AGAP006428, in peak 2L-10, encoding a DNA binding domain that recognizes methylated CpGs and AGAP007387, in peak 2L-16, encoding a DNA methyltransferase associated protein.

5.26 Genes encoding known immune domains associated with the outcome of *S. marcescens* infection

In addition to associated genes already described, including CLIP-domain serine proteases, PGRPs or the immunoglobulin domain containing FN3Ds, several other genes associated with the outcome of *S. marcescens* infection encoded domains that point to an involvement in immune responses. The association of a considerable number of immune genes associated with the outcome of *S. marcescens* infection indicates that the SNP genotyping approach employed here points to relevant genes and the relatively high number of 138 associated genes is the result of an interplay between several genes, and most likely several modes of mosquito immunity, in shaping a complex trait such as the outcome of oral infection with *S. marcescens*.

Several genes associated with the outcome of *S. marcescens* infection encoded leucine-rich repeat (LRR) domains. LRR domains, often in combination with other domains, are believed to be involved in protein-

protein interactions in various processes, but are also common in immune genes involved in pattern recognition, including Toll-like receptors (Deng et al., 2013; Mariuzza et al., 2010; Takken and Goverse, 2012). Based on patterns of LRR domains with cysteine residues and coiled-coil domains, more than 20 *An. gambiae* genes have been identified as leucine-rich immune genes (LRIMs) (Waterhouse et al., 2010). These include *LRIM1* and the *APL1* paralogues *APL1A* and *APL1C*, known to be involved in complement responses against rodent and human malaria parasites (Fraiture et al., 2009; Mitri et al., 2009; Osta et al., 2004b; Povelones et al., 2009; Riehle et al., 2006) but also *LRRD7* (*LRIM17*), involved in responses against *P. berghei* and *P. falciparum* (Dong et al., 2006a).

The gene encoding LRIM15 was found to be associated with the outcome of *S. marcescens* infection, in peak 2L-13, by the individual SNP analysis. *LRIM15* encodes a putative transmembrane domain and thus it is expected to encode a transmembrane receptor protein. Further functional analysis of *LRIM15* could shed more light on a possible LRIM15 involvement in antibacterial immunity. Other LRR-encoding genes associated with the outcome of *S. marcescens* infection included AGAP004405, found in peak 2R-15, encoding a putative glucose-repressible alcohol dehydrogenase transcriptional effector, AGAP006643 and AGAP006644, found adjacent to each other in peak 2L-11 and AGAP010012, encoding a putative LRR and calponin-like protein domain containing protein, a family of cytoskeletal regulators (Foussard et al., 2010).

Furthermore, two Toll-like receptors, *TOLL1A* and a previously unidentified paralogue of *TOLL5B*, were found adjacent to each other, in peak X-4, also associated with the outcome of *S. marcescens* infection by the individual SNP analysis. In *Drosophila,* Toll receptors have been implicated in defences against fungi and Gram-positive bacteria through the Toll pathway (De Gregorio et al., 2002; Lemaitre, 2004; Lemaitre and Hoffmann, 2007; Lemaitre et al., 1996). Little is known, though, for the role of the Toll/REL1 pathway in *Anopheles*, especially with regard to defences against Gram-negative bacteria. Although the IMD/REL2 pathway is considered to be the major pathway for defences against *P. falciparum* and Gram-negative bacteria, while the Toll/REL1 pathway is mainly involved in defences against *P. berghei*, there seems to be an overlap between the two pathways in defences against the human and rodent malaria parasites (Garver et al., 2009). Furthermore, there is also an overlap between REL1 and REL2 mediated transcriptional responses in *Ae. aegypti* (Kazura et al., 2011). The same has also been suggested in *Drosophila*, with several lines of evidence pointing to synergistic interactions between the Toll and Imd pathways (Kounatidis and Ligoxygakis, 2012; Valanne et al., 2011). Therefore, there is reason to believe that defences against Gram-negative bacteria in *An. gambiae* involve an interplay between the IMD/REL2 and Toll/REL1 pathways that remains to be further elucidated.

Three members of *An. gambiae* gene families known to be involved in immune responses were also associated with the outcome of *S. marcescens* infection. *ML12*, in peak 2L-16, belongs to the family of MD2-like receptors, containing a lipid recognition domain. Mammalian MD-2 interacts with lipopolysaccharides (LPS), derived from Gram-negative bacteria, or other lipids and triggers the Toll-like receptor TLR4, leading to downstream NFκB-mediated responses (Maeshima and Fernandez, 2013; Visintin et al., 2006). The *An. gambiae* ML1 has been shown to modulate bacterial load but also the outcome of *P. falciparum* infection (Dong et al., 2006a). In peak 3L-10, the genes encoding the fibrinogen-related protein FREP6 and the galectin GALE4 were found adjacent to each other. FREPs (or FBNs, fibrinogen-domain immunolectins) comprise the largest pattern recognition receptor family in *An. gambiae*, showing considerable expansion compared to *Drosophila*, to a total of 59 members (Zdobnov et al., 2002). FREPs are related to mammalian ficolins, which recognize components of the bacterial cell wall and activate the complement response through the lectin pathway, leading to phagocytosis of bacteria (Endo et al., 2011; Matsushita, 2010, 2013; Matsushita and Fujita, 2002). In *An. gambiae*, several FREPs have shown potent transcriptional regulation following a bacterial or *Plasmodium* challenge while FBN22 and FBN39 have been implicated in defences against gut bacteria, including *Serratia*, while FBN9 has been shown to bind to bacterial surfaces, interact with both Gram-negative and Gram-positive bacteria and also influence the outcome of *P. berghei* and *P. falciparum* infections (Dong and Dimopoulos, 2009). Galectins comprise another family of putative *An. gambiae* pattern recognition receptors. Galectins bind beta-galactoside sugars and have been implicated in humoral immune responses as effector molecules as well as in complement activation (Pace and Baum, 2004; Pace et al., 2002; Thurston et al., 2012). In *An. gambiae*, GALE8 has been implicated in antiviral defences against *O'nyong-nyong* virus (ONNV) infection (Waldock et al., 2012).

In peak 2R-15, a gene encoding a ricin B lectin domain was associated with the outcome of *S. marcescens* infection. The carbohydrate-binding capabilities of lectins raise the possibility of involvement in antibacterial defences through bacterial recognition. The mammalian lectin RegIIIγ has been shown to mediate interactions between the gut microbiota and the intestinal epithelial surface (Vaishnava et al., 2011). The *An. gambiae* C-type lectins (CTLs), CTL4 and CTLMA2, have been implicated in defences against Gram-negative bacteria but also in *P. berghei* melanization, acting as protective agonists during *P. berghei* development (Osta et al., 2004a; Schnitger et al., 2009).

Two genes encoding glycoside hydrolase domains were associated with the outcome of *S. marcescens* infection, AGAP001111, encoding an alpha-glucosidase, in peak 2R-1, and AGAP004032, encoding an

alpha-mannosidase, in peak 2R-13. The association with the outcome of *S. marcescens* infection of two genes with similar domain architecture and thus putative function, in different peaks, indicates a functional significance for these genes in antibacterial effects that influence the outcome of *S. marcescens* infection. Glycoside hydrolase domains hydrolyze glycosilic bonds between carbohydrates and are found in the conserved chitinase gene family, mammalian members of which have been implicated in bacterial clearance and host tolerance in response to bacterial infections (Dela Cruz et al., 2012; Funkhouser and Aronson, 2007).

Finally, AGAP010015, in peak 3R-12, and AGAP010798, in peak 3L-7, encode peptidases that could possibly be involved in antibacterial responses (Chang et al., 2004; Xiong and Jacobs-Lorena, 1995), while AGAP011415 encodes a chitin-binding gene. Chitin-binding genes have been implicated in antibacterial responses both in *Anopheles* and *Drosophila* (Buchon et al., 2009; Danielli et al., 2000; Kuraishi et al., 2011), a matter that will be also discussed later on as several chitin-binding genes showed significant transcriptional regulation following *S. marcescens* infection.

5.27 Non-immune genes with homologies suggesting involvement in responses that shape the outcome of *S. marcescens* infection

Based on their homologies, several of the genes associated with the outcome of *S. marcescens* infection, provide reasons to believe that they might be involved in responses that influence the *S. marcescens* infection outcome. AGAP000595 and AGAP000596, found in peak X-2, encode phospholipid/glycerol acyltransferases involved in phospholipid biosynthesis. Lipid bodies and eicosanoids are part of the *Ae. aegypti* immune response and bioactive lipids are generated following *S. marcescens* infection (Sorgine, MHF; “Molecular and Population Biology of Disease Vectors” meeting, 2011). Therefore, these acyltransferases could be involved in similar immune responses.

AGAP000998, in peak X-4, encodes a mannose 6-phosphate insulin receptor whose *Drosophila* orthologue is involved in lysosomal transport (Dennes et al., 2005). It is thus possible that this receptor could be involved in phagocytosis.

AGAP006405, found in peak 2L-10, encodes a tyrosine protein kinase. Its *Drosophila* orthologue, *derailed2*, is involved in establishing the olfactory circuitry through the Wnt5 signalling pathway (Sakurai et al., 2009). It is thus possible that this *An. gambiae* kinase could modulate olfactory behaviour following *S. marcescens* infection.

In peak 2L-15, the genes encoding the inhibitors of apoptosis IAP4 and IAP5 are found. In *Drosophila*, the inhibitor of apoptosis IAP2 has been shown to regulate Imd signalling (Gesellchen et al., 2005). Therefore, IAP4 or IAP5 could play similar roles in *Anopheles*.

In peak 2L-18, AGAP007657 encodes an orthologue of the *Drosophila syndecan*. Syndecan in mammals is necessary for maintenance of epithelial integrity while in the fruit fly syndecan is involved in Slit signalling during axon guidance (Chanana et al., 2009). It is conceivable that, as Dscam is involved both in axon guidance and innate immunity (Evans and Bashaw, 2010; Watson et al., 2005), the same applies for syndecan.

In peak 2R-8, AGAP013684 encodes a putative miRNA. As miRNAs have emerged as major modulators of gene regulation in insects, affecting many processes, including epithelial immune responses (Lucas and Raikhel, 2013; Lucas et al., 2013; Zhou et al., 2009), it is possible that this putative miRNA is involved in responses against *S. marcescens*.

AGAP012252, found in peak 3L-19, encodes the protein kinase C PKC53E, whose *Drosophila* orthologue has been implicated in NPF-mediated alcohol sensitivity (Chen et al., 2008, 2010). Therefore, the *An. gambiae* counterpart of PKC53E could also be involved in behavioural responses that shape the outcome of *S. marcescens* infection.

In peak 3L-20, AGAP012398 encodes the COP9 signalosome complex subunit 7. In *Drosophila*, the COP9 signalosome regulates Dorsal-dependent transcriptional activities through Dorsal intracellular localization, important for the development of the *Drosophila* immune system (Harari-Steinberg et al., 2007). As Dorsal also plays a role in immune responses, through the Toll pathway (Lemaitre and Hoffmann, 2007), it is conceivable that the COP9 signalosome can also modulate immune-related Dorsal activities in adult flies and possibly play an analogous role in mosquito immunity.

AGAP011363, in peak 3L-11, encodes an orthologue of *Drosophila rab6*, a Rab-type member of the small GTPase superfamily that has been implicated in regulation of phagocytosis (Ye et al., 2012). Furthermore, rab6 has been shown to regulate trafficking of Grk, the EGFR ligand (Coutelis and Ephrussi, 2007; Tian et al., 2013). Therefore, rab6 in *Anopheles* could also be involved in antibacterial immune responses either through regulation of phagocytosis or through regulation of EGFR signalling, important for gut remodeling following oral bacterial infection (Buchon et al., 2010).

AGAP010503, found in peak 3L-4, encoding a putative calcium activated potassium channel, is an orthologue of *Drosophila SK channel*, which has been shown to mediate behavioural plasticity in

Drosophila courtship memory processes (Abou Tayoun et al., 2012). It is thus possible that its *Anopheles* counterpart could also be involved in behavioural responses that shape the outcome of *S. marcescens* infection.

AGAP005216, found in peak 2L-5, encodes a phosphatidylinositol-4-phosphate 5-kinase, an orthologue of *Drosophila* fab1, which co-localizes and, most likely, mediates the lysosomal degradation of necrotic (Rusten et al., 2007; Soukup et al., 2009), a serpin implicated in the Toll pathway (Levashina, 1999; Pelte et al., 2006). It is possible that the association with the outcome of *S. marcescens* infection of its *Anopheles* counterpart has a basis in a similar involvement during lysosomal degradation of mosquito immune regulators.

Finally, AGAP001446, found in peak 2R-3, encodes an AMP-activated protein kinase (AMPK). AMPKs in *Drosophila* have been implicated in starvation-induced hyperactivity, resulting in increased foraging (Johnson et al., 2010). Therefore, AMPKs might modulate feeding behaviour and, in this way, influence the outcome of *S. marcescens* infection.

5.28 Conclusions and future perspectives

The identification of SNP divergence associated with the outcome of *S. marcescens* infection shown here reached an unprecedented level of detail, compared to previous population genetics studies in *An. gambiae* aiming at gene discovery related to an infection outcome. The achieved level of detail was assisted by the implementation of a high-resolution SNP genotyping array, able to interrogate genetic variation at ~400,000 variable positions in the *An. gambiae* genome. The genome-wide SNP genotyping effort in *An. gambiae*, upon which the implemented array is based, has been previously validated and used to identify SNP divergence between pools of M/S forms of *An. gambiae* field mosquitoes (Neafsey et al., 2010; Nwakanma et al., 2013; Reidenbach et al., 2012). Previous association studies in mosquitoes typically employed low-resolution genetic tools, including microsatellite markers or targeted SNP genotyping, in field or laboratory mosquito populations (Blandin et al., 2009; Harris et al., 2010a; Horton et al., 2010; Riehle et al., 2006), allowing identification of associations in large genomic areas or in a limited number of genes. From both perspectives, the resolution of the analysis presented here is improved by orders of magnitude.

Furthermore, the strong evolutionary drive midgut bacteria are expected to exert on mosquito genetic variation can further explain the identification of relevant genes associated with the outcome of *S.*

marcescens infection. In contrast to *Plasmodium* infections, whose effect on mosquito fitness is unclear (Sangare et al., 2013), as gut bacteria affect survival (Schnitger et al., 2009), they can drive genetic variation related to immune genes, whose function limits the bacterial presence or sustains the homeostatic balance of the gut bacterial population.

Finally, the use of a recently established mosquito colony, kept in big numbers, combined the high degree of genetic variation found in field mosquitoes with higher linkage disequilibrium, which is important for gene discovery, especially when array-based SNP genotyping is employed, based on a set number of genetic markers. The use of such laboratory mosquito colony parallels the enhanced gene discovery of human genome-wide association studies employing isolated populations that exhibit higher genetic homogeneity (Heutink and Oostra, 2002). Several such studies employ genome-wide identification of SNP divergence on island populations, in which relative genetic homogeneity facilitates gene discovery that could be further extrapolated on broader populations (Gudmundsson et al., 2012). It is very unlikely that the adoption of field mosquito populations in the study design adopted here would result in any meaningful genetic associations as the linkage disequilibrium of *An. gambiae* populations is considered to be very low (Harris et al., 2010b).

There is no doubt that the implementation of a SNP genotyping platform for the interrogation of genetic variation in *An. gambiae* aimed first and foremost at the identification of SNP divergence associated with the outcome of *P. falciparum* infection. Several ongoing research efforts aim at the identification of a genetic basis of *P. falciparum* infections, using array or sequencing-based approaches that could shed more light on vector/parasite interactions and the genetic variation that underlies these interactions. A detailed comparison of the results presented here with association studies related to the outcome of *P. falciparum* infections or infections with other pathogens using the same SNP genotyping array as here (Lawniczak, ML and Christophides, GK, “Molecular and Population Biology of Mosquitoes and Other Disease Vectors”, July 2013) could be performed following the publication of the latter studies.

Although there is no doubt that the outcome of *Plasmodium* infections is influenced by genetic variation (Niare et al., 2002; Riehle et al., 2006), the influence of gut bacteria on the *Plasmodium* infection outcome (Cirimotich et al., 2011b; Dong et al., 2009; Meister et al., 2009; Rodrigues et al., 2010) opens the possibility that such associations may be indirect, through influences on the mosquito gut microbiota. Furthermore, it is also possible that genetic associations on the *Plasmodium* infection outcome may be driven by evolutionary forces unrelated to *Plasmodium*, including gut bacteria, in the case of immune genes also showing antibacterial activity, or even random drift. The genetic association of traits with

indirect influences on the phenotype and no underlying selective pressure driving allele variation can prove to be elusive. For example, genetic associations of severe malaria in humans are based on the premise of strong frequency increases of the identified variants due to natural selection of the protective trait (Ruwende et al., 1995).

This is the first time that genetic associations with the outcome of bacterial infections are identified in mosquitoes. Although it would be expected that the effect of bacteria on mosquito survival would drive genetic variation, this is the first time such link between bacterial infections and genetic variation is established. No such associations have been established in *Drosophila* either. The most common way to study the genetic basis of various traits in *Drosophila* is the utilization of the *Drosophila* genetic reference panel, a collection of inbred fly lines that have been fully sequenced (Mackay et al., 2012). Phenotypic divergence between such fly lines can be associated with genetic divergence, based on their previously sequenced genetic makeup. Indeed, such genome-wide association studies have been successful in identification of genetic variation associated with various traits, including the outcome of viral infections, yielding some gene candidates that have been verified through reverse genetics (Grubbs et al., 2013; Magwire et al., 2012; Weber et al., 2012). It remains to be determined whether such approach can also identify genetic associations related to the outcome of bacterial infections.

A dual implication can be inferred for the genes associated with the outcome of *S. marcescens* infection: they are putatively involved in responses that influence the infection outcome but also genetic variation within the mosquito population shapes their ability to do so. Further sequencing-based analysis could shed more light on the identification of causal variants that underlie the outcome of infection. Allele divergence that shapes the infection outcome may comprise simple gain-of-function or loss-of-function mutations in coding or regulatory sequences, resulting in gene products with altered functions. Indeed, genetic variation in regulatory sequences can be related to transcriptional or translational efficiency or genetic changes that influence gene expression through epigenetic modifications. Furthermore, the outcome of infection can be shaped by allele combinations found in two or more loci, whose interplay related to epistatic interactions, synergism or combined loss of redundant functions can underlie the observed phenotype. In such cases, reverse genetics approaches, especially genetic screens through silencing of individual genes, cannot capture such involvement.

Several limitations of the association approach, however, should be taken into account. An association study is limited by the variation of the gene pool interrogated that is present in the assayed population, especially with regard to the limited ability to detect rare variants. Furthermore, genes with critical roles,

either in immunity or other processes such as development, may not be detected as their variants are removed from the mosquito colony gene pool due to their lethal phenotype. Finally, detected associations might be false positives, due to array-related artefacts, selective sweeps in genomic areas in the proximity of causal variants or genetic drift within the population. Array-based methods have also inherent limitations due to the detection of probe hybridizations based on a set number of genetic markers. Therefore, phenotypic characterization of candidate genes is required to establish a causal link with the *S. marcescens* infection outcome.

More recently, the adoption of whole-genome sequencing approaches in the identification of genetic associations related to the infection outcome, has been prevalent in mosquitoes and elsewhere. Whole-genome sequencing has been performed for gDNA pools corresponding to both the *Asaia* and *S. marcescens* infection assays but data analysis was ongoing at the time of completion of the current study. Indisputably, a whole-genome sequencing approach would further enhance gene discovery resolution and confirm the SNP divergence associations presented here, but also further identify causal polymorphisms and allele variants that underlie the outcome of infection.

Chapter 6

Identification of *Anopheles gambiae* transcriptional responses following oral infection with *Asaia* or *Serratia marcescens*

Introduction

6.1 DNA microarrays and mosquito antibacterial immunity

The identification of mosquito immune factors involved in antibacterial or anti-*Plasmodium* responses has been largely based to date on expression analysis of transcripts significantly up or down regulated following an immune challenge and further bioinformatic analysis of candidate gene homologies to known immune factors, mainly in *Drosophila*, leading to prioritization of candidate genes for functional characterization (Christophides et al., 2002; Dimopoulos et al., 2002).

Transcriptional responses following infection with human or rodent malaria parasites have been extensively studied using DNA microarrays and have led to the functional characterization of several genes whose silencing modulates the *Plasmodium* infection prevalence and intensity (Dong et al., 2006a; Mendes et al., 2011; Vlachou et al., 2005). Furthermore, these studies, along with others focusing on tissue-specific transcriptional regulation or developmental co-expression (Baker et al., 2011; MacCallum et al., 2011), have provided an invaluable resource for identification or evaluation of the role of immune components in a broad array of studies. Therefore, the use of DNA microarrays has provided a platform for genome-wide expression analysis in *An. gambiae* that has led to the identification of key players in the mosquito anti-*Plasmodium* defence. Examples in which *An. gambiae* expression studies led to functional characterization of immune components include the identification of the LRIM1 role in modulation of *P. berghei* infection (Osta et al., 2004b), the SPCLIP1 role in the complement response based on developmental co-expression with relevant factors (Povelones et al., 2013) or *AgDscam* splicing factors controlled by the IMD/REL2 pathway (Dong et al., 2012).

Although one of first microarray studies in *An. gambiae* included the identification of differential expression following a bacterial challenge (Dimopoulos et al., 2002), little follow-up has been undertaken since then. This study employed first-generation cDNA microarrays from an EST clone collection and investigated transcriptional responses following a systemic bacterial infection by injecting *E. coli* or *Micrococcus luteus* to the mosquito haemolymph or by employing an *in vitro* bacterial challenge in mosquito cell lines (Dimopoulos et al., 2002). Since then, the *An. gambiae* complete genome sequence (Holt et al., 2002) has dramatically improved gene annotation, and oligonucleotide microarrays have also progressed by leaps and bounds in determination of differential expression following bacterial infection.

Few microarray studies have since focused on mosquito transcriptional responses following bacterial infection, especially with regard to epithelial immune responses following ingestion of bacteria. Similar to

the initial microarray study on transcriptional responses following a bacterial challenge, which investigated systemic immunity (Dimopoulos et al., 2002), another microarray study has focused on haemocyte responses to bacteria (Baton et al., 2009). Additional microarray studies investigated the transcriptional response to bacterial infection indirectly, by modulating the activation of the IMD/REL2 pathway in *An. gambiae* (Garver et al., 2009) or *Ae. aegypti* (Kazura et al., 2011*)*.Transcriptional responses following silencing of the negative regulator of the IMD/REL2 pathway *Caspar*, in whole *An. gambiae* mosquitoes, indicated differential expression of several immunity genes but also genes involved in redox metabolism, stress, transcription and translation (Garver et al., 2009). Overexpression of *REL2* in the *Ae. aegypti* fat body indicated that Caspar controls the expression of only a subset of IMD/REL2 mediated genes, with transcriptional responses following *REL2* overexpression again including mainly immunity, stress and metabolism genes (Kazura et al., 2011). Insights into epithelial immune responses against midgut bacteria have been provided by a microarray study between septic and aseptic mosquitoes, i.e. mosquitoes that retain their natural midgut microbiota compared to antibiotic treated mosquitoes in which their gut flora had been considerably reduced (Dong et al., 2009).

Mosquito responses against Gram-negative bacteria are controlled mainly by the IMD/REL2 pathway, which shows similarity but also key differences compared to the *Drosophila* Imd pathway (Figure 6.1) (Garver et al., 2012; Meister et al., 2009; Meister et al., 2005). One such difference involves the orthologue of *Drosophila* Relish (Stoven et al., 2000), REL2, which is alternatively spliced into a long (REL2-F) and short (REL2-S) form, with REL2-S being responsible for responses against Gram-negative bacteria but also *P. falciparum* (Meister et al., 2005; Mitri et al., 2009). A microarray analysis on bacterial-challenged cell lines, in which *REL2* was silenced, identified several immunity genes controlled by the IMD/REL2 pathway including antimicrobial peptides, *CLIPB14* and other CLIP-domain serine proteases as well as *LRIM1* (Meister et al., 2005).

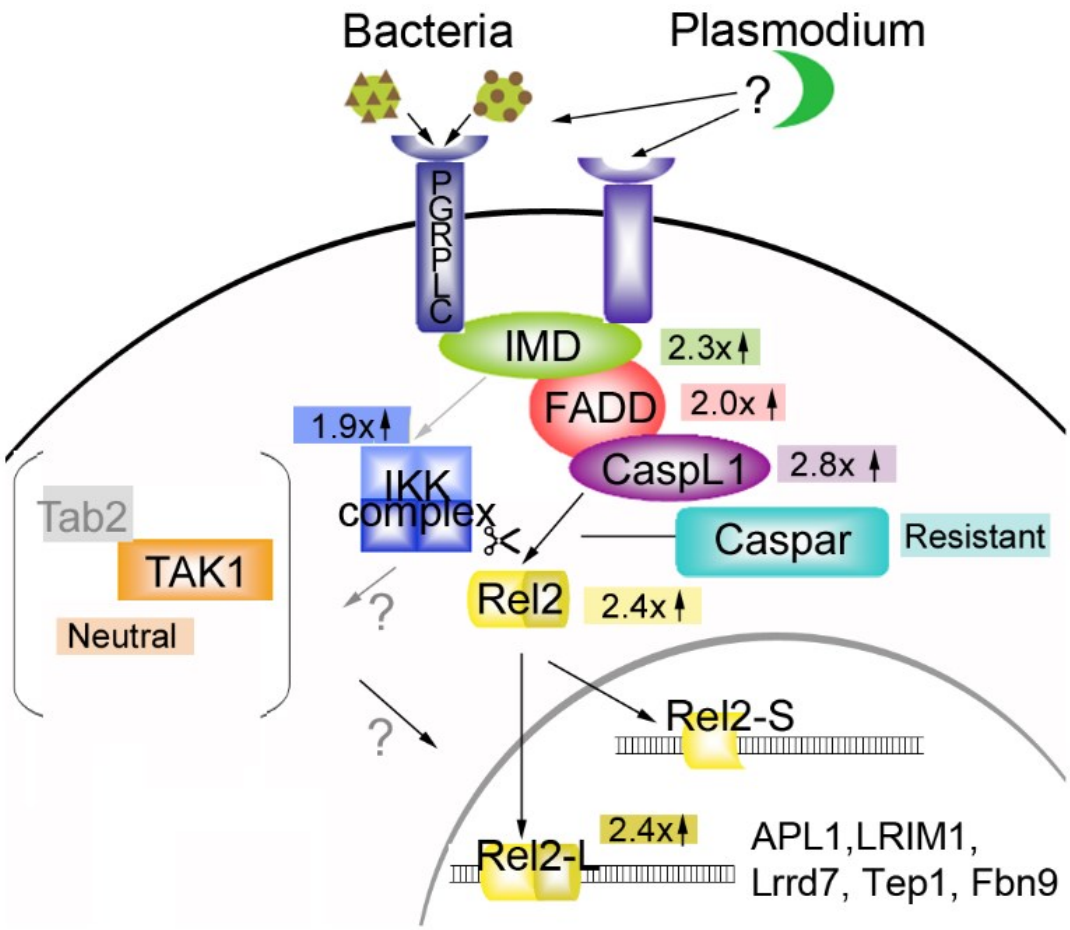
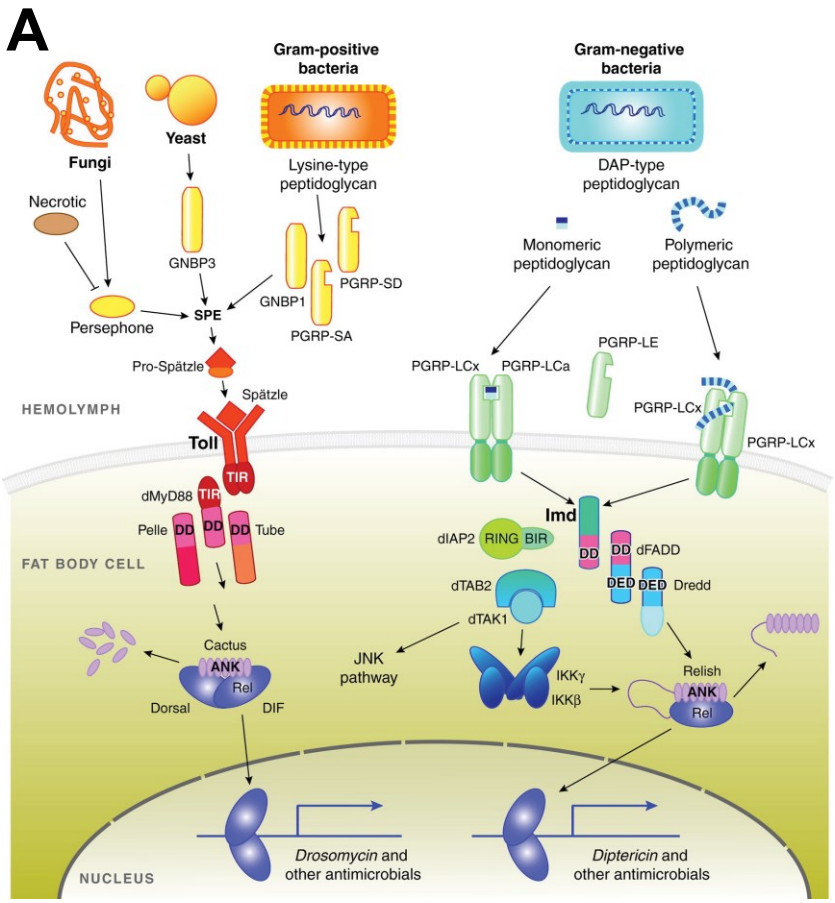


Figure 6.1: The IMD/REL2 pathway in *An. gambiae*. The *An. gambiae* IMD/REL2 pathway is triggered by DAP-type peptidoglycan, found in Gram-negative bacteria, by the peptidoglycan recognition receptor PGRPLC. Other receptors, triggered by Gram-negative bacteria or a subset of bacteria or *Plasmodium*, remain to be discovered. A downstream signalling process initiated by PGRPLC leads to activation of the *Drosophila* Relish orthologue REL2. Through alternative splicing, REL2 is found in 2 isoforms, REL2-L (or REL2-F) and REL2-S. Activation and subsequent nuclear translocation of these transcription factors leads to expression of several immune factors, including the antimicrobial peptides CEC3, GAM1 but also LRIM1 (Meister et al., 2005), but also APL1, LRRD7 (LRIM17), TEP1 and FBN9. Grey arrows indicate possible interactions based on *Drosophila* counterparts. Factors in grey brackets are not involved in *P. falciparum* infection. Fold-change values correspond to transcriptional regulation following *P. falciparum* infection. Adapted from: (Garver et al., 2012).

6.2 *Drosophila* immunity against Gram-negative bacteria

Epithelial responses against Gram-negative bacteria have been extensively studied in *Drosophila* (Lemaitre and Hoffmann, 2007). As shown in Figure 6.2A, while the Toll pathway is mostly responsible for defences against Gram-positive bacteria and fungi, the Imd pathway coordinates the main fly defence mechanisms against Gram-negative bacteria, although several lines of evidence point to synergistic effects of the two pathways in defences against Gram-negative bacteria (Kounatidis and Ligoxygakis, 2012; Valanne et al., 2011). The Imd pathway is important for both systemic and epithelial responses and leads to the production of antimicrobial peptides that limit Gram-negative bacterial populations (Lemaitre et al., 1997). The main receptor for the Imd pathway is PGRP-LC, a peptidoglycan recognition receptor that recognizes monomeric or polymeric DAP-type peptidoglycan, which is found in all Gram-negative bacteria (Choe et al., 2005; Choe et al., 2002; Gottar et al., 2002; Kaneko et al., 2004; Lim et al., 2006). PGRP-LC acts synergistically with another peptidoglycan recognition protein, PGRP-LE, in recognition of monomeric DAP-type peptidoglycan (Kaneko et al., 2006).


A
Gram-positive bacteria
Gram-negative bacteria
Fungi
Yeast
Lysine-type peptidoglycan
DAP-type peptidoglycan
Necrotic
GNBP3
GNBP1
PGRP-SD
PGRP-SA
Monomeric peptidoglycan
Polymeric peptidoglycan
Persephone
SPE
Pro-Spätzle
Spätzle
PGRP-LCx
PGRP-LCa
PGRP-LE
HEMOLYMPH
Toll
TIR
dMyD88
Pelle
DD
Tube
Imd
dIAP2
RING
BIR
dFADD
DED
Dredd
FAT BODY CELL
dTAB2
dTAK1
JNK pathway
Cactus
ANK
Rel
Dorsal
DIF
IKKγ
IKKβ
Relish
Drosomycin and other antimicrobials
Diptericin and other antimicrobials
NUCLEUS

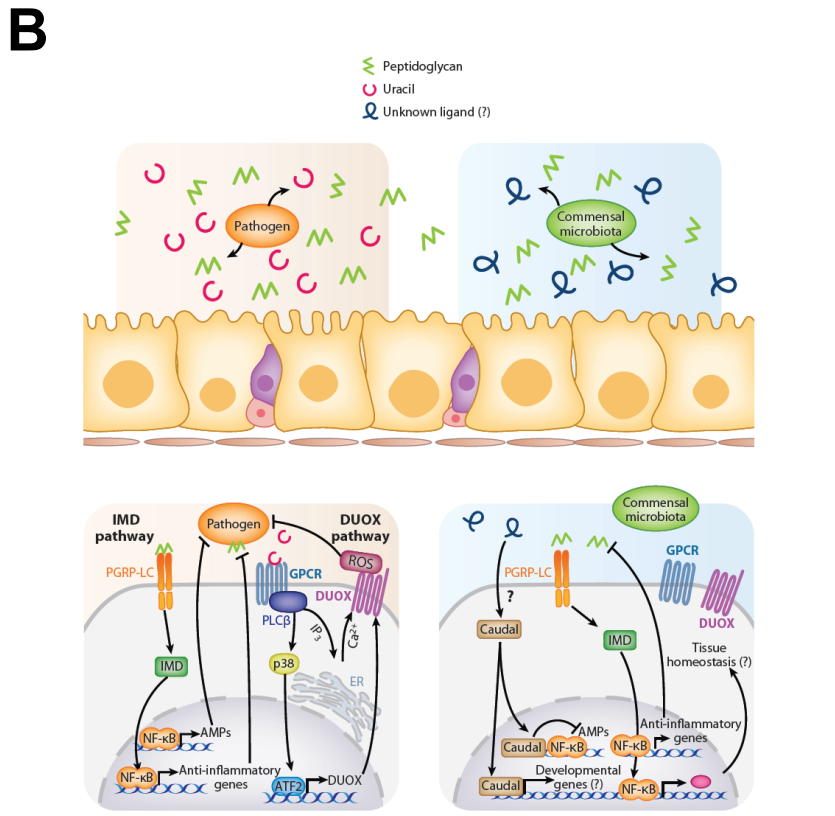

B
Peptidoglycan
Uracil
Unknown ligand (?)
Pathogen
Commensal microbiota
IMD pathway
DUOX pathway
PGRP-LC
GPCR
ROS
DUOX
PLCβ
IP3
Ca2+
IMD
p38
ER
NF-κB
AMPs
Anti-inflammatory genes
ATF2
Caudal
Tissue homeostasis (?)
Developmental genes (?)

Figure 6.2: *Drosophila* antibacterial defences. A: The *Drosophila* Toll and Imd pathways are involved in responses against Gram-positive and Gram-negative bacteria, respectively. The Toll pathway is also activated by fungi. Toll activation involves several PGRP and GNBP pattern recognition factors, leading to proteolytic cascades that activate the Toll receptor through a cleaved form of Spatzle. Downstream signalling processes lead to activation and nuclear translocation of the Rel transcription factors Dif and Dorsal. The Imd pathway is activated by monomeric or polymeric DAP-type peptidoglycan recognized by PGRP-LC. A downstream signalling process involves the cleavage of Relish and nuclear translocation of the Rel domain that leads to activation of immune factors, including antimicrobial peptides such as Diptericin. Adapted from: (Lemaitre and Hoffmann, 2007). **B:** The DUOX pathway is activated through recognition of uracil derived from pathogenic but not commensal bacteria, in contrast to Imd pathway activation by DAP-type peptidoglycan common to all Gram-negative bacteria. Imd regulation through Caudal shapes the gut microbiota population structure through triggers that remain unidentified. Adapted from: (Lee and Brey, 2013).

Although the only elicitor of Imd-mediated responses is DAP-type peptidoglycan, common to all Gram-negative bacteria, the Imd pathway is tightly regulated at several levels and this regulation is considered to shape the composition of the fly's gut microbiota. Regulation of the Imd pathway includes several peptidoglycan recognition proteins including PGRP-LF, which sequesters circulating peptidoglycan (Maillet et al., 2008) and PGRP-LB, which degrades DAP-type peptidoglycan (Zaidman-Remy et al., 2006). Another negative modulator of Imd responses is the PGRP-LC interacting inhibitor of Imd responses, PIMS, whose expression is triggered by commensal bacteria and leads to depletion of PGRP-LC, leading to suppression of Imd-mediated antimicrobial peptide production, thus establishing a balanced immune response that allows tolerance of commensal bacteria (Lhocine et al., 2008). The homeobox transcription factor Caudal also represses the Imd-mediated expression of antimicrobial peptides in ways that shape the population structure of the fly's gut microbiota through regulatory mechanisms that remain poorly understood (Figure 6.2B) (Ryu et al., 2008).

The DUOX pathway also plays a major role in gut antimicrobial activities in response to oral bacterial infections by generating an epithelial oxidative burst against gut bacteria (Figure 6.2B) (Ha et al., 2005). The DUOX pathway is triggered by bacterial-derived uracil through an unknown GPCR that signals the generation of reactive oxygen species (Lee et al., 2013a). Uracil is produced by a subset of Gram-negative bacteria which are predominantly pathogenic, thus providing a level of specificity to the fly's antibacterial responses (Lee and Brey, 2013).

Another important aspect of fly responses against Gram-negative bacteria is the intestinal stem cell proliferation and differentiation required for gut homeostasis following tissue damage. This process is controlled by the JAK/STAT pathway (Beebe et al., 2009; Jiang et al., 2009) in synergism with the EGFR pathway (Buchon et al., 2010) although there are indications that DUOX-mediated signalling is also involved in this process (Lee et al., 2013a).

Furthermore, stress response pathways triggered by highly pathogenic bacteria, such as *Pseudomonas entomophila,* result, mainly due to the host-derived reactive oxygen species production, in a translational blockage that inhibits immune responses and epithelial renewal, which mostly account for the pathogenic nature of these bacteria (Chakrabarti et al., 2012). Finally, barrier formation is also an integral component of the fly's epithelial response. Formation of the peritrophic matrix has been shown to protect against oral infection with pathogenic bacteria such as *P. entomophila* or *S. marcescens* but also regulate Imd-mediated responses, presumably through limiting bacterial sensing (Kuraishi et al., 2011).

The identification of *An. gambiae* genes associated with the outcome of *S. marcescens* infection raised the question of how these associated genes correlate with genes showing differential expression following oral *S. marcescens* infection. Furthermore, transcriptional responses following oral bacterial infection could involve additional components of the mosquito antibacterial response that due to a variety of reasons, such as the lack of variants in the assayed gene pool or redundant function, did not come up in the SNP genotyping analysis. To shed more light on mosquito antibacterial responses and homeostatic interactions that shape the mosquito gut microbial communities and investigate how these responses relate to the identified genes associated with the *S. marcescens* infection outcome but also *Drosophila* antibacterial immunity, DNA microarrays were utilized in identifying transcriptional regulation following oral infection with bacteria of the genus *Asaia* or with *S. marcescens.*

6.3 Identification of mosquito transcriptional responses to oral infection with *Asaia* or *S. marcescens*

To identify *An. gambiae* genes involved in antibacterial responses, a differential expression analysis was carried out (Figure 6.3). Mosquitoes were antibiotic treated for 5 days and were subsequently orally infected with bacteria of the genus *Asaia* or *S. marcescens*. At day 2 post infection, bacteria-fed mosquitoes were selected and the following day total RNA was extracted from the guts of infected mosquitoes. Mosquitoes that were antibiotic treated for 8 days were used as uninfected controls and total RNA was extracted in the same way. The relatively stable infection dynamics, for both *Asaia* and *S. marcescens*, from day 3 to day 5 post infection (Figure 4.2C and 4.4B, respectively), suggest that transcriptional regulation involved in shaping the infection outcome would be already in place at day 3 post infection, the time point selected to assay differential expression.

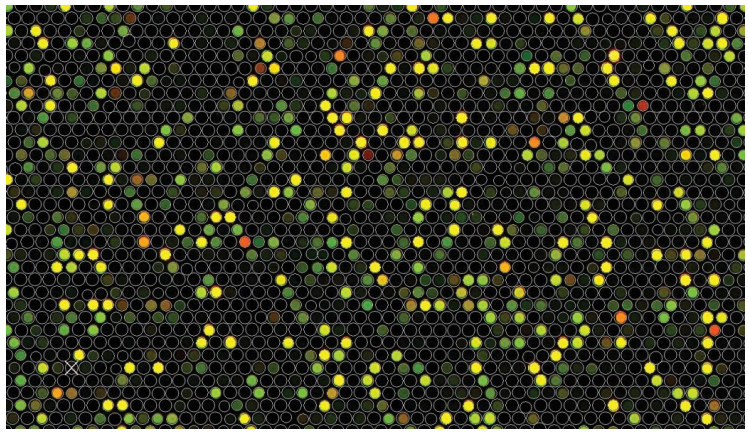
Bacteria naturally inhabiting the mosquito gut

Mosquito 5-day antibiotic treatment commencing at the day of emergence

Reduction of the natural gut microbiota in antibiotic treated mosquitoes

Oral infection with *Asaia* or *Serratia marcescens*

Continue antibiotic treatment for uninfected control mosquitoes



Differential gene expression in the gut of *Asaia* or *Serratia marcescens* infected mosquitoes compared to uninfected control mosquitoes

Figure 6.3: The implemented experimental design in identifying transcriptional responses following oral *Asaia* or *S. marcescens* infection. The same model of *An. gambiae* oral infection with bacteria of the genus *Asaia* or *S. marcescens* presented in chapter 4 is also used here. A 5-day antibiotic treatment with a gentamicin, penicillin and streptomycin cocktail is used, aiming to reduce the natural gut bacteria populations. Subsequently, mosquitoes are orally infected with *Asaia* or *S. marcescens* added to the sugar meal, along with suitable antibiotics for the selection of the respective *Asaia* or *S. marcescens* strain but also to further limit any residual gut bacteria. Control mosquitoes are kept uninfected, with the initial antibiotic treatment further continued. Bacteria-fed mosquitoes are selected 2 days post infection and, at day 3 post infection, infected mosquitoes are dissected and their guts are used for total RNA extraction. Total RNA from guts of uninfected controls is also extracted at the same time point. These total RNA samples are used in a DNA microarray analysis of differential gene expression between the guts of infected and uninfected mosquitoes. The bottom panel shows part of a scanned microarray slide using the suitable fluourescence filters. Each dot represents a separate 60-base oligonucleotide probe. Complementary RNA corresponding to mRNA present in the guts of infected or uninfected mosquitoes has bound to the oligonucleotide probe corresponding to the respective transcript. Complementary RNA from infected or uninfected mosquitoes, labelled with Cy3 or Cy5 cyanine fluourescent dyes, can be visualized in green or red, respectively. Probes in which both Cy3 and Cy5 labelled complementary RNA has hybridized are visualized in yellow.

To identify differentially expressed transcripts between the infected and uninfected mosquito gut pools, a genome-wide two-colour DNA microarray approach was adopted. Total RNA from infected or uninfected guts, following DNAse treatment for removal of gDNA and quality control of the used total RNA samples, was reverse transcribed and labelled with Cy3 or Cy5 cyanine fluourescent dyes, respectively. Equal quantities of Cy3 and Cy5 labelled complementary RNA samples were subsequently hybridized on custom Agilent 4x44k gene expression oligonucleotide arrays. The microarray design used, *Pfalcip_Agamb2009* (ArrayExpress accession ID A-MEXP-2324), comprises 60-base oligonucleotide probes encompassing all *An. gambiae* annotated transcripts based on the *AgamP3.6 An. gambiae* genome release (December 2009), with each transcript represented in the array in duplicate. Additional *P. falciparum* probes were also present in the array but were not relevant to the study design. Cy3 or Cy5 labelled complementary RNA, representing transcripts present in the infected or uninfected mosquito guts, can hybridize to the respective oligonucleotide probe in a quantitative manner. As shown in the bottom panel of Figure 6.3, probes in which Cy3-labelled or Cy5-labelled complementary RNA have hybridized, emit green (excitation/emission at 550/570 nm) or red (excitation/emission at 650/670 nm) fluourescence, respectively. If both Cy3 and Cy5 labelled complementary RNA oligonucleotides have hybridized to the probe, yellow fluourescence is emitted. Quantification of the emitted fluourescence intensity from each probe due to competitive hybridization of Cy3 and/or Cy5 labelled complementary RNA can be used to determine the respective hybridization ratio and thus the ratio of mRNA abundance of the respective transcript between infected and uninfected mosquito guts.

6.4 Principal component analysis of datasets corresponding to transcriptional responses following oral *S. marcescens* infection

To identify transcriptional responses following oral *S. marcescens* infection, six independent infection assays were carried out and total RNA from guts of *S. marcescens* infected mosquitoes as well as uninfected controls was used for reverse transcription, labelling and hybridization to custom Agilent 4x44k gene expression arrays. Infections were carried out in two batches (A-B), performed about one year apart from each other, with three performed infection assays in each batch. Therefore, considerable genetic variation was expected between mosquitoes used in each batch of infections as the employed *N'gousso* laboratory colony, maintained in a large population size, would be constantly evolving between the two batches of infections.

Furthermore, in some cases, total RNA from different pools of *S. marcescens* infected or uninfected mosquitoes, stemming from the same infection assay, were used for hybridization. This additional control allowed hybridization using genetically similar mosquitoes collected at the same time and treated in the same way, thus removing many variables that could influence the outcome of the microarray analysis. These hybridizations are closer to biological than technical replicates, as different mosquito pools were used for total RNA extraction, but allow the assessment of the effect of methodological variables or diverse mosquito genetic backgrounds on the observed analysis outcome.

Overall, eight hybridizations were carried out, four hybridizations in batch A, designated as *SmA1*, *SmA2*, *SmA3i* and *SmA3ii*, with *SmA3i* and *SmA3ii* stemming from the same independent infection and similarly, four hybridizations in batch B, designated as *SmB1*, *SmB2i*, *SmB2ii* and *SmB3*, with *SmB2i* and *SmB*ii stemming from the same independent infection. For all resulting datasets, probe fluourescence intensity was filtered to be >5, a threshold that was sufficient to remove background hybridization in *P. falciparum* probes also included in the array. Furthermore, to remove fluourescent dye intensity-dependent variation, Lowess normalization was applied to all datasets, removing such variation through a smoothing adjustment to the two-colour data (Berger et al., 2004).

To visualize the divergence of expression data in these eight datasets, a principal component analysis (PCA) was carried out (Abdi, 2010). PCA can capture expression data variation between datasets by transforming this variation in linearly uncorrelated variables called principal components. These principal components capture a pattern of similarity of the observed variation in decreasing order so that the first 2-3 principal components can visualize most of the divergence between the assayed datasets. The results of the PCA for the eight assayed datasets are shown in Figure 6.4. In the corresponding axes, the first three principal components are visualized, with the 1st principal component capturing 26.84% (X-axis), the 2nd principal component capturing 20.45% (Y-axis) and the 3rd principal component capturing 18.75% (Z-axis) of total variation.

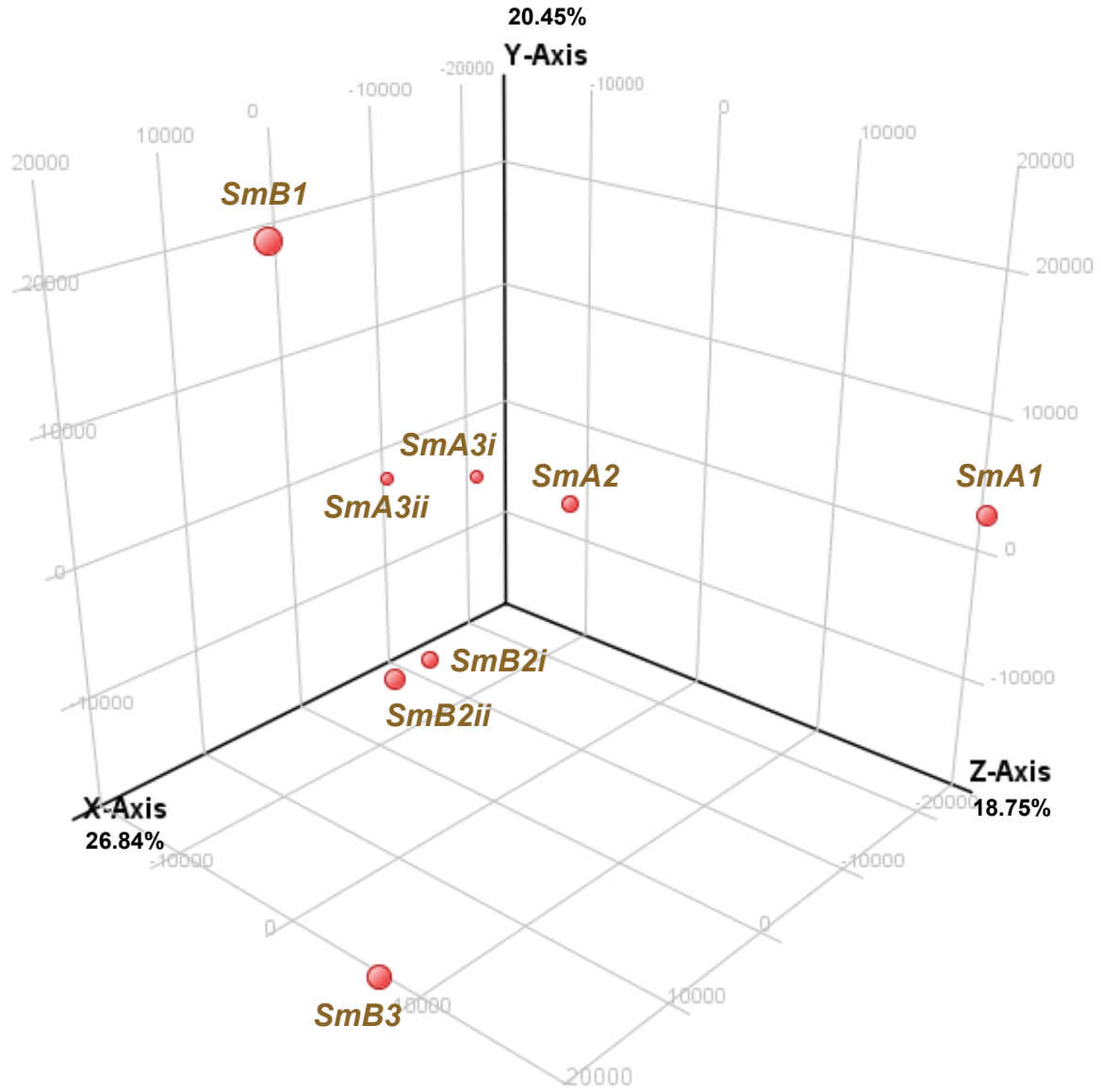


Figure 6.4: Principal component analysis of DNA microarray datasets corresponding to different hybridizations of complementary RNA stemming from *S. marcescens* infected and uninfected mosquitoes. Labelled complementary RNA, corresponding to the mRNA content in midguts of *S. marcescens* infected mosquitoes as well as uninfected controls, was used for hybridization on custom Agilent 4x44k DNA microarrays including oligonucleotide probes corresponding to all annotated transcripts of the *An. gambiae AgamP3.6* genome release. Following filtering with a probe fluourescence intensity cut-off of 5 and Lowess normalization, the log2-transformed fold-change transcriptional regulation between *S. marcescens* infected and uninfected mosquitoes was determined for each transcript, by averaging the respective fold-change regulation of the 2 oligonucleotide probes each transcript corresponded to. The acquired datasets were subsequently used in a principal component analysis so as to visualize data separation in the top three principal components, capturing most of the variation between datasets. Data separation in linear dimensions for the 1st, 2nd and 3rd principal component are plotted in the x-axis, y-axis and z-axis, respectively, with the % captured variation by each component indicated for each axis. Each data point corresponds to a dataset stemming from an individual hybridization, with the designated name for each dataset indicated for the data point it corresponds to.

Overall, considerable variation was observed between the assayed datasets for both batches. In both cases, though, in which hybridization was performed with total RNA stemming from the same infection, limited variation was observed, with *SmA3i* clustering with *SmA3ii* and *SmB2i* clustering with *SmB2ii*. Remarkably, though, *SmA2* clustered well with *SmA3i* and *SmA3ii*, with these three datasets forming a distinct cluster, especially with regard to the 1st principal component. In batch A, the remaining *SmA1* dataset was a distinct outlier, compared to all other datasets. In batch B, the *SmB2i* and *SmB2ii* datasets showed limited variation compared to the *SmA3i-SmA3ii-SmA2* cluster, with respect to the 1st and 3rd principal component, with most divergence attributed to the 2nd principal component. *SmB1* and *SmB3* were also distinct outliers, although *SmB3* variation compared to *SmB2i-SmB2ii* towards the 2nd principal component was limited.

Clustering of datasets stemming from the same infection assay suggests that the overall observed variation is most likely to be a result of the underlying genetic variation of mosquitoes used for each independent infection. The divergence of expression datasets between independent infections, with the exception of the *SmA3i-SmA3ii-SmA2* cluster, suggests that the underlying genetic variation might affect the observed transcriptional regulation, thus accounting for the observed divergence, while limited genetic variation between the *SmA3i-SmA3ii* and *SmB2i-SmB2ii* datasets most likely accounts for their respective observed clustering.

6.5 Clusters of differentially expressed transcripts following *S. marcescens* infection

As the averaging of the divergent replicate datasets corresponding to *S. marcescens* independent infections would be unlikely to result into any meaningful interpretation, clusters of co-regulated transcripts between independent infection datasets were identified using k-means clustering (Hartigan, 1979). The outlier datasets *SmA1* and *SmB1* were removed from further analysis while only one dataset from each of the pairs corresponding to the same independent infection (*SmA3i* and *SmB2i*) was further used, to avoid bias in favour of transcriptional programmes present in these independent infections. The k-means clustering algorithm implemented here partitioned the expression data for each respective dataset into 10 clusters, so that the log2-transformed differential expression value for each transcript was assigned to the cluster of the nearest mean. The results of k-means clustering for the *SmA2*, *SmA3i*, *SmB2i* and *SmB3* datasets can be seen in Figure 6.5. No single cluster showed clear co-regulation through all datasets but several clusters showed, partial or considerable, co-regulation between the *SmA2-SmA3i* or the *SmB2i-SmB3* datasets. In clusters 3 to 5, partial overlap was observed regarding co-regulation between

dataset pairs, while almost all transcripts in cluster 7 showed downregulation and in cluster 10 upregulation between the *SmA2-SmA3i* datasets. Finally, in cluster 8, upregulation was observed for most transcripts between the *SmB2i-SmB3* datasets.

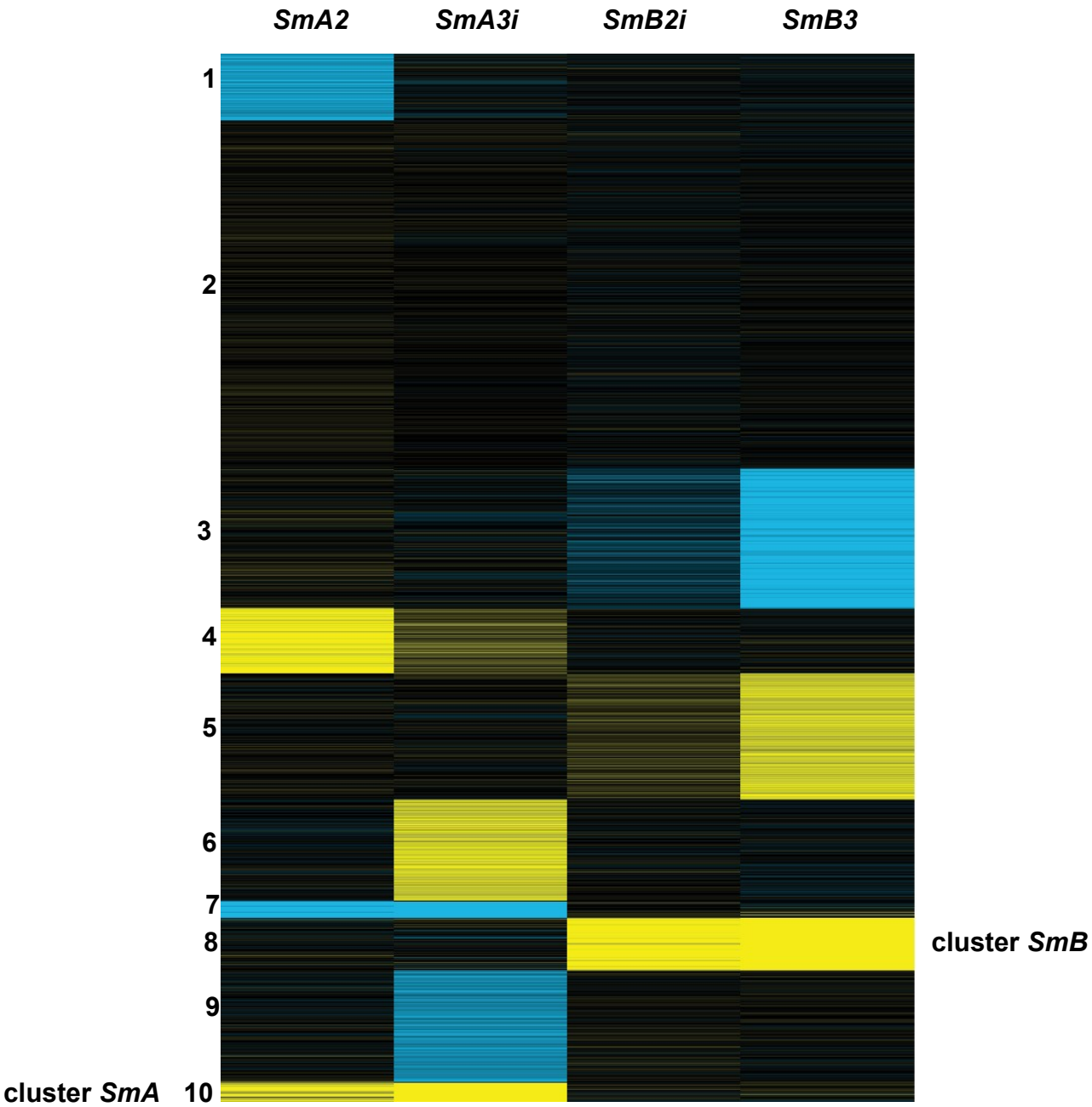


Figure 6.5: K-means clustering of DNA microarray datasets representing differential expression between *S. marcescens* infected mosquitoes and uninfected controls in 4 hybridizations stemming from independent infection assays. Lowess-normalized log2-transformed fold-change values, with a probe fluourescence intensity cut-off of 5, stemming from hybridizations of complementary RNA corresponding to the mRNA content of *S. marcescens* infected and uninfected mosquito midguts, indicated above each regulated transcript set, were used for k-means clustering, based on Euclidean distance of gene expression values. For each transcript, the fold-change value was assigned to 1 of 10 clusters based on proximity to the mean assigned to the cluster, following 10,000 iterations of the k-means algorithm. In every dataset, each transcript is represented by a single line, yellow for upregulation, blue for downregulation and black for no observed regulation. Clusters are numbered on the left side of the panel, with clusters *SmA* and *SmB* also indicated adjacent the respective clusters.

This level of co-regulation between independent infections that belong to the same batch was surprising given the distance between *SmB2i* and *SmB3* towards the 1st and 3rd principal component but also the proximity of *SmB2i* with *SmA2* and *SmA3i* towards the 1st principal component in the PCA conducted (Figure 6.4), suggesting that for *SmB3-SmB2i* and *SmA2-SmA3i*, the co-regulating transcripts lie within variation attributed to the 2nd principal component of the PCA. These results are consistent with the possibility of two distinct transcriptional programmes in independent infections between batches A-B which can, mostly, be ascribed to genetic differences between the two batches.

In cluster 7, showing consistent downregulation between the *SmA2-SmA3i* datasets, 197 transcripts were found. Genes of interest corresponding to the transcripts found in this cluster can be seen in Table 6.1. These included 7 genes with peptidase homologies, 5 MD2-lipid recognition receptors and 4 chitin-binding genes, with two of them being peritrophic matrix components, as determined by a previous proteomic analysis (Dinglasan et al., 2009). The only associated gene included was *IAP4*, a putative IMD/REL2 regulator (Gesellchen et al., 2005). Remarkably, though, despite the limited overlap, most gene families found in this cluster were also represented in the list of associated genes by different members. One exception was the downregulation of the serpins *SRPN6* and *SRPN10*, negative regulators of CLIP domain serine proteases, with *SRPN6* previously shown to be transcriptionally induced by *E. coli* challenge and mediate anti-*Plasmodium* defences (Abraham et al., 2005). Another remarkable property of genes included in this cluster was the representation of genes related to mosquito behaviour. These included the genes encoding the gustatory receptors Gr13, Gr25 and Gr59, 3 genes encoding haemolymph juvenile hormone binding domains as well as 3 odorant binding protein encoding genes.

Cluster 10, designated as cluster *SmA*, included 297 transcripts that mostly showed upregulation in the *SmA2-SmA3i* datasets (Figure 6.5). Genes of interest found in cluster *SmA* can be seen in Table 6.2. These included 25 genes with peptidase domains and 9 CLIP domain serine proteases, suggesting that the response to *S. marcescens* infection employs a proteolytic component related either to signalling cascades, as in the *Drosophila* Toll or insect melanization pathways, or a direct antibacterial proteolytic role for the encoded proteases. Of the identified CLIP domain serine proteases, *CLIPB14* has been previously shown to be upregulated following a bacterial challenge and be involved in anti-*Plasmodium* responses (Christophides et al., 2002; Volz et al., 2005). Furthermore, *CLIPA2* and *CLIPA8* have been previously implicated in the *An. gambiae* melanization cascade (Schnitger et al., 2007; Volz et al., 2006). Two C-type lectins as well as two serpins were also found in this upregulated cluster and may also be involved in functions similar to the ones hypothesized for aforementioned CLIP domain serine proteases.

A surprising finding was the presence of 7 cytochrome P450 genes in cluster *SmA*. P450 cytochromes are known to be involved in metabolic resistance to insecticides (Bariami et al., 2012; Riveron et al., 2013; Stevenson et al., 2012), but no link has yet to be established with antibacterial responses. P450 cytochromes have been shown to be differentially expressed following *P. berghei* infection, with a suggested role in detoxification processes, triggered by the induced oxidative stress following infection (Felix and Silveira, 2011). It is thus possible that *S. marcescens* infection also induces an oxidative stress response, with the P450 cytochromes involved in detoxification homeostatic processes. Importantly, the gene encoding the nitric oxide synthase NOS was also found in this cluster. *NOS* has been previously shown to be upregulated following a bacterial challenge (Gupta et al., 2009), and has been shown to be involved in anti-*Plasmodium* responses through an oxidative burst triggered by the generated nitric oxide (Kumar et al., 2010; Luckhart et al., 1998). Therefore, *S. marcescens* infection may also include the activation of NOS-mediated responses which are known to limit *Plasmodium* infection intensities but could also play a role in the outcome of *S. marcescens* infection.

Furthermore, cluster *SmA* included several genes known or suspected to be involved in haemolymph antibacterial or anti-*Plasmodium* responses. These included 4 genes encoding thioester-containing proteins: TEP1, TEP3, TEP4 and TEP12 and 6 LRIMs: LRIM1, APL1C, LRIM3, LRIM4, LRIM7 and LRIM10. The anti-*Plasmodium* effects of LRIM1, APL1C and TEP1 have been previously reported (Fraiture et al., 2009; Povelones et al., 2009) while TEP3 has also been implicated in anti-*Plasmodium* responses (Povelones et al., 2011). Therefore, *S. marcescens* infection also triggers a haemolymph immune response, indicated by the induced TEPs and LRIMs found in cluster *SmA*. As the design of the current analysis employed midgut tissue, differential expression of these genes can, most likely, be ascribed to haemocytes in connection to the midgut epithelium. Further analysis using fat body tissues could shed more light to the effect of *S. marcescens* oral infection in activation of systemic immune responses. Interestingly, the gene encoding the lipophorin Lp was also found in this cluster. Lp is a nutrient transporter known to influence TEP1-mediated complement responses (Rono et al., 2010).

Compared to genes associated with the outcome of *S. marcescens* infection, as in cluster 7, several associated gene families were represented in cluster *SmA* by different members. Genes belonging to families also detected in the SNP analysis include a gene with an ARID/BRIGHT DNA binding domain, several genes encoding CLIP-domain serine proteases, FREPs or glycoside hydrolases, the gustatory receptor Gr48, the GPCR GPROP1, 3 genes with juvenile hormone binding domains, 3 homeobox genes, LRIMs, the gene encoding the MD2-like receptor ML8, a gene including a ricin B lectin domain and *TOLL5B*,

whose paralogue was associated with the outcome of *S. marcescens* infection. *TOLL5B* polymorphisms have been associated with the outcome of *P. falciparum* infection and thus TOLL5B may be involved in immune signalling pathways related to anti-*Plasmodium* defences (Horton et al., 2010).

Interestingly, 4 genes found in cluster *SmA* were also associated with the outcome of *S. marcescens* infection: *PGRPLC*, AGAP008648, encoding a methyltransferase, AGAP0012395, encoding an oxide reductase and AGAP012252, encoding a protein kinase C. Except *PGRPLC*, the other genes have yet to be implicated in antibacterial responses and the combined identification of these genes both by the SNP and expression analyses suggests they might be involved in yet to be revealed processes that influence the outcome of *S. marcescens* infection.

As AGAP012252 is an orthologue of the *Drosophila PKC53E* (Chen et al., 2008), this *An. gambiae* protein kinase C might regulate NPF-mediated behavioural immune responses. The identification in cluster *SmA* of AGAP012252 but also of Gr48, OBP3 and genes with juvenile hormone binding domains suggests a behavioural immune component triggered following *S. marcescens* infection. Furthermore, the gene encoding takeout3 (TO3) was also found in this cluster, encoding a juvenile hormone binding protein. The *Drosophila* takeout homologue has been implicated in juvenile hormone signalling (Chamseddin et al., 2012) but also in regulation of feeding behaviour (Meunier et al., 2007) and, as the *Gr9* and *Gr10 Drosophila* orthologues (Bray and Amrein, 2003), in male courtship behaviour (Dauwalder et al., 2002). Therefore, TO3 could also be involved in similar circuits regulating behavioural immune responses following *S. marcescens* infection.

A TIR-domain containing gene was also found in this cluster, showing an orthologous relationship with *Drosophila Ect4*. Ect4 is an adaptor protein involved in a signalling pathway that antagonizes Imd activation upon oral bacterial infection in the *Drosophila* airway epithelium (Akhouayri et al., 2011). It is thus possible that its *Anopheles* counterpart could also be involved in similar regulatory functions following *S. marcescens* infection.

Finally, the genes encoding the C-type lysozymes LYSC1 and LYSC2 also showed upregulation and were found in the *SmA* cluster. Both *LYSC1* and *LYSC2* have been previously shown to be upregulated following a bacterial challenge (Li et al., 2005), while LYSC1 has been implicated in antibacterial and anti-*Plasmodium* effects (Kajla et al., 2010; Kajla et al., 2011; Lapcharoen et al., 2011; Li and Paskewitz, 2006).

The identification of at least 90 immune-related genes in this cluster of 297 upregulated genes suggests the induction of an immune transcriptional programme following oral *S. marcescens* infection in batch A

infections. Although the identification of these upregulated transcripts lacks statistical significance due to variation between the *SmA* datasets and the limitations of an array-based approach, these data provide a preliminary aspect of the complex expression of immune genes that shapes the outcome of *S. marcescens* infection. Based on these preliminary results, several layers of mosquito immune responses in response to oral *S. marcescens* infection can be identified. These include a proteolytic component, with copious induction of CLIP domain serine proteases or genes encoding peptidase domains, an oxidative stress response component that includes NOS but also P450 cytochrome detoxification enzymes, a haemolymph immune component comprising several LRIMs, TEPs or relevant regulators such as Lp and, importantly, a behavioural immune component comprising Gr48, OBP3, juvenile hormone binding proteins, including TO3, and a protein kinase C putatively involved in sensory perception.

Notably, several identified genes in cluster *SmA* are implicated in anti-*Plasmodium* defences, including *PGRPLC.* Furthermore, the putative regulators of the IMD/REL2 pathway *Ect4* and *IAP4*, the latter found in cluster 7 that comprises downregulated transcripts, were also found to be differentially expressed following *S. marcescens* infection. The IMD/REL2 pathway induction, in concert with anti-*Plasmodium* factors, induced independently or through the IMD/REL2 pathway, suggests that immune responses against *S. marcescens* can also be relevant for anti-*Plasmodium* defence.

Clustering of the datasets *SmB2i* and *SmB3* (Figure 6.5), stemming from the second batch of infections, was discrete from the *SmA2-SmA3i* clustering, suggesting the deployment of a different transcriptional programme for the second batch of infections. Cluster 8, designated as *SmB*, contained transcripts mostly showing upregulation in the *SmB2i-SmB3* datasets. Cluster *SmB* was considerably larger than *SmA*, containing 715 transcripts. Nevertheless, 125 immune-related genes were found in this cluster (Table 6.3), a number comparable to the 90 immune-related genes residing in the *SmA* cluster.

Interestingly, the *PGRPLC* splice variant AGAP005203-RC resided in cluster *SmB*, compared to AGAP005203-RA found in *SmA*. It is conceivable that IMD/REL2 regulation through different *PGRPLC* splice variants can, at least partly, account for the diverse differential expression between the batch A-B datasets. As in *SmA*, a large proteolytic component was found in *SmB*, with 35 genes encoding peptidase domains and three CLIP domain serine proteases, including *CLIPE6*, which was also associated with the outcome of *S. marcescens* infection. *CTL4* was also found in the *SmB* cluster, previously shown to exhibit an agonistic effect regarding the outcome of *P. berghei* infection (Osta et al., 2004a) but also an effect against Gram-negative bacteria (Schnitger et al., 2009). Although CTL4 has been shown to form a dimer with CTLMA2 (Schnitger et al., 2009), the genes encoding two other C-type lectins, CTLMA1 and CTLGA1

were also found upregulated in cluster *SmB*, raising the possibility of CTL4 engagement in multiple interactions with other C-type lectins.

The large haemolymph response component found in *SmA* was missing in cluster *SmB*, with only *APL1A* found in *SmB*. *APL1A* has been previously shown to be controlled by the IMD/REL2 pathway (Mitri et al., 2009). The induction of different LRIM family members between the batch A-B datasets could be attributed to differential regulation of the IMD/REL2 pathway between the respective datasets. The same possibility applies to the induction of oxidative stress responses. The genes encoding six P450 cytochromes and three haeme peroxidases, HPX5, HPX8 and HPX12 were found in cluster *SmB*, with *HPX8* previously shown to be upregulated following *Plasmodium* infection (Oliveira et al., 2012).

A large behavioural component was also present in cluster *SmB*, including the genes encoding five gustatory receptors, including Gr9, which was associated with the outcome of *S. marcescens* infection, but also the juvenile hormone binding protein takeout2 (TO2) and the olfactory receptors Or20, Or44 and Or70. Furthermore, the gene encoding the allatostatin receptor GPRALS1 was also found in cluster *SmB*. The *Drosophila* allatostatin A has been implicated in inhibition of feeding behaviour, in an antagonistic fashion compared to NPF (Hergarden et al., 2012). Furthermore, genes encoding putative neurotransmitter-triggered receptors found in cluster *SmB* included GPRNNA20, a paralogue of the serotonin receptor *GPR5HT7*, associated with the outcome of *S. marcescens* infection, GPRRK, a glycoprotein hormone receptor, also possibly involved in neuropeptide signalling, and GPRTYR, an octopamine/tyramine receptor. Octopamine has been indirectly implicated in aversion circuits in response to S*. marcescens* infection in *Manduca sexta*, the tobacco hornworm (Adamo, 2005).

The gene encoding the hypervariable pattern recognition receptor AgDscam, whose antibacterial and anti-*Plasmodium* activities have been previously reported (Dong et al., 2012; Dong et al., 2006b), was also found in cluster *SmB*. Furthermore, the gene encoding FN3D3, associated with the outcome of *S. marcescens* infection, along with three additional genes encoding FN3 domains were also found in cluster *SmB*. AgDscam has been recently implicated as an IMD/REL2 pathway triggered effector, with IMD/REL2-induced splicing factors leading to the expression of *AgDscam* splice variants that specifically target *P. berghei*, *P. falciparum* or commensal bacteria (Dong et al., 2012). This *AgDscam* induction is also consistent with the possibility that different IMD/REL2 regulation has led to *AgDscam* differential expression between batch A-B infections.

Overall, 16 genes associated with the outcome of *S. marcescens* infection were found in cluster *SmB*. Except the ones already mentioned, these included AGAP001111, encoding a glycoside hydrolase, along with six additional genes encoding glycoside hydrolases that were also found in *SmB*, AGAP001446, encoding a putative AMP-activated protein and AGAP001002, a *TOLL5B* paralogue, whose presence, along with the presence of *SPZ5*, encoding a spaetzle-like cytokine, in cluster *SmB*, raises the possibility of a Toll pathway involvement in responses to *S. marcescens* infection.

Finally, the gene encoding the antimicrobial peptide DEF1 was found in cluster *SmB*. The difference in antimicrobial peptide elicitation between batch A-B infections raises the possibility of diverse regulation of responses between the respective datasets.

Taken together, these data suggest that the response to *S. marcescens* infection, both in batch A and B infections, includes genes relevant to antibacterial immunity. The most likely explanation for the different expression profile between the respective batch A-B infections is that genetic variation, underlying the mosquito populations used in the respective infections, diverged between different infection batches and might have led to differential expression of diverse sets of immune genes between different infection batches. The presence in the *SmA-B* clusters of different *PGRPLC* splice variants raises the possibility of diverse IMD/REL2 pathway regulation between batch A-B infections. The presence of *APL1C* in *SmA* and *APL1A* in *SmB*, *TO3* in *SmA* and *TO2* in *SmB*, as well as *LYSC1* and *LYSC2* in *SmA* and DEF1 in *SmB*, further strengthen the possibility of the deployment of different transcription immune programmes between batch A-B infections.

The main differences between these responses, based on the genes present in the respective *SmA-B* clusters, is a much more limited number of haemolymph immunity genes combined with the presence of the genes encoding AgDscam and FN3D3 in *SmB*, but also a considerable behavioural component that included the differential expression of *Gr9* in batch B infections.

6.6 Core transcriptional responses following oral *S. marcescens* infection

To investigate whether there is a common immune response component between the diverse batch A and B patterns of transcriptional regulation following *S. marcescens* oral infection, transcriptional responses to oral *S. marcescens* infection were examined in three independent assays stemming from both batch A and batch B infections. Despite the observed variation, this approach incorporates immune

modules induced under diverse mosquito genetic backgrounds and is likely to include the most consistent and pronounced induced factors that shape the outcome of *S. marcescens* infection.

The microarray datasets *SmA2*, *SmA3i* and *SmB3*, stemming from the respective independent infections and capturing a considerable portion of the observed variation regarding differential expression following *S. marcescens* infection, were further used to identify core transcriptional responses to *S. marcescens* infection. The dataset *SmB2i*, showing a similar expression pattern as *SmB3* (Figure 6.5), was not further used so as to enhance statistical power by employing three instead of four datasets. As shown in Figure 6.6A and Table 6.4, a total of 99 transcripts showed more than 1.75-fold differential expression following oral *S. marcescens* infection, with 55 upregulated and 44 downregulated transcripts. The log2-transformed fold-change values from all probes in the three assayed datasets were used in a t-test against zero, where zero corresponds to no transcriptional regulation. Overall, 38 upregulated and 28 downregulated transcripts returned a significant p-value <0.05, indicating consistent differential expression between datasets that reached statistical significance (Figure 6.6A).

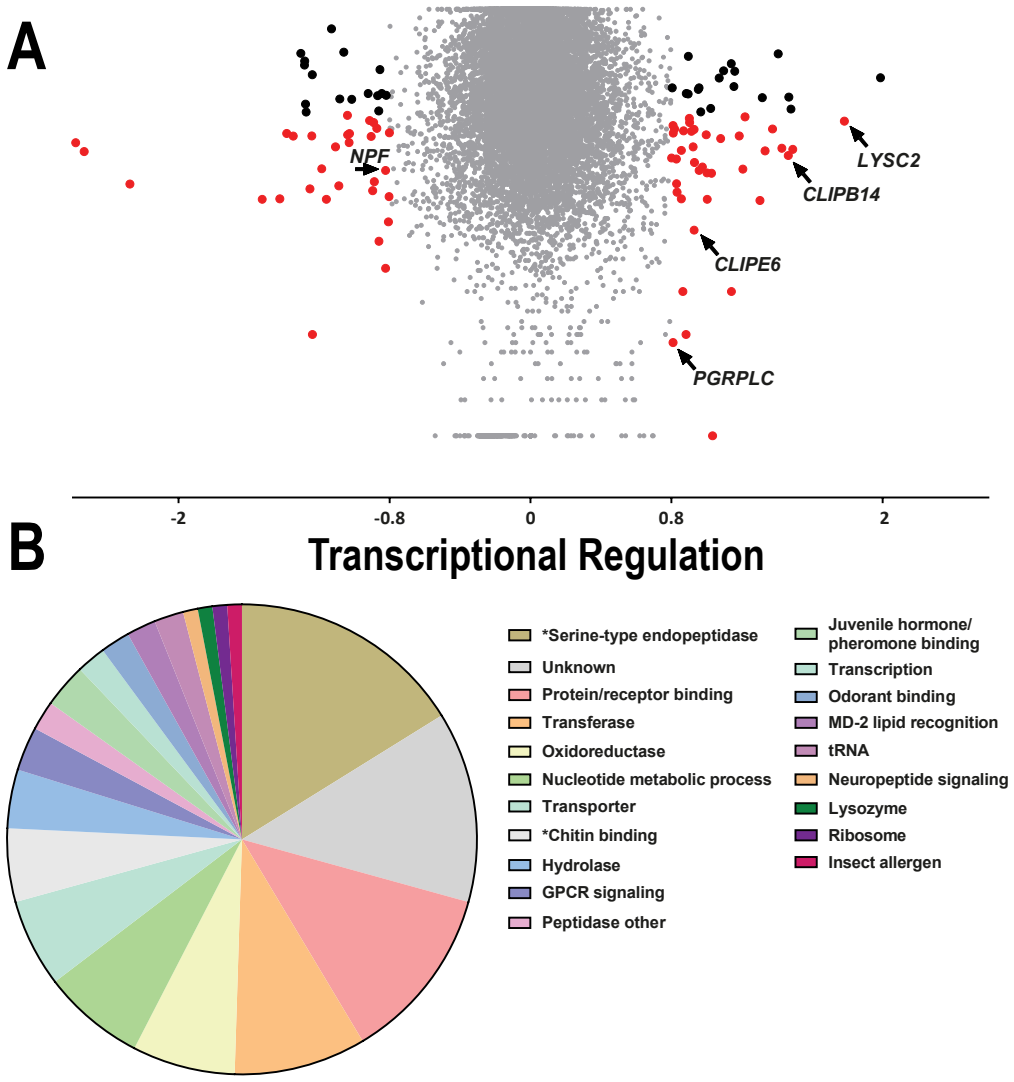


Figure 6.6: Core transcriptional regulation following oral *S. marcescens* infection. Core transcriptional responses to oral *S. marcescens* infection were identified by averaging the fold-change expression of each transcript as determined over three DNA microarray hybridizations stemming from independent infections. **A:** Manhattan plot of differential expression following oral *S. marcescens* infection. Each data point represents an *An. gambiae* transcript. The log2-transformed fold-change value for each transcript is plotted in the x-axis, as determined over three independent infections. The p-value in a t-test against zero, using the log2-transformed fold-change values for each probe corresponding to the specific transcript over the three independent infections used, is plotted in the y-axis. Transcripts regulated more than 1.75-fold (0.8-fold in log2-transformed values) are indicated by a red or black data point, depending on whether the computed p-value is <0.05 or >0.05, respectively. The remaining transcripts with <1.75-fold regulation are indicated by grey data points. Data points representing transcripts corresponding to PGRPLC, CLIPE6, CLIPB14, LYSC2 and NPF are indicated by arrows. **B:** Assigned functional classes for the 99 transcripts showing more than 1.75-fold transcriptional regulation following oral *S. marcescens* infection. The pie chart shows the relative representation of each functional class, based on the colour legend at the right side of the panel. Functional classes corresponding to GO terms significantly overrepresented in the set of more than 1.75-fold regulated transcripts are indicated by an asterisk.

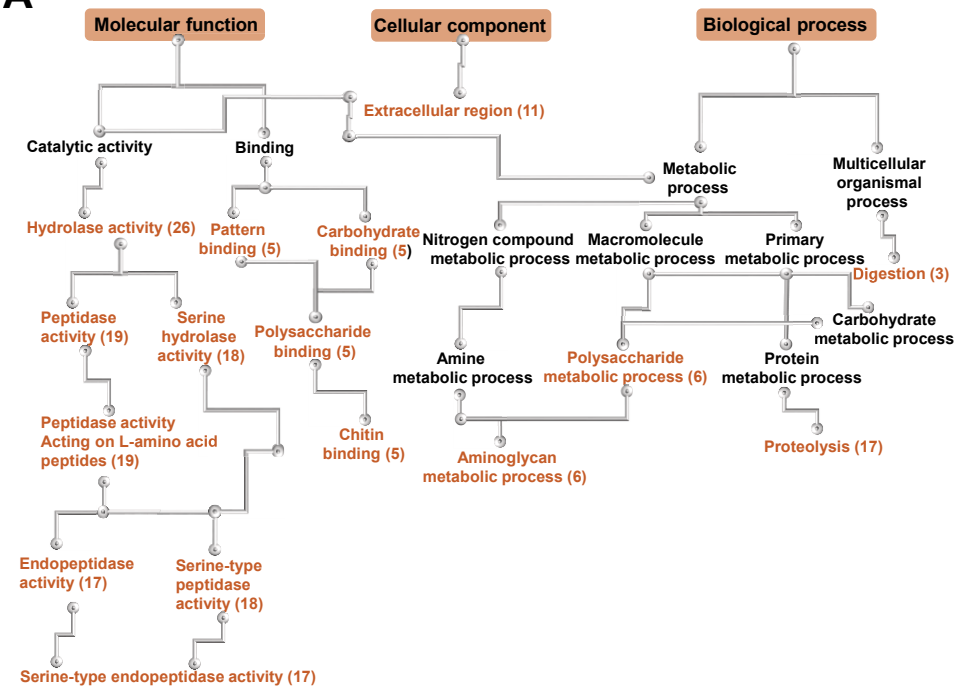
The 99 differentially expressed transcripts were assigned to their respective functional class, shown in Figure 6.6B. The most represented class comprised serine-type endopeptidases, with 16 members, including CLIPE6, also associated with the outcome of *S. marcescens* infection, CLIPB14, CLIPB17 and CLIPB20. The protein/receptor binding class comprised 12 members, including PGRPLC, LRIM1 but also four FREPs. The oxidoreductase class, comprising 7 members, included the P450 cytochromes CYP6N1 and CYP6P3, while the nucleotide metabolic process class, with six members, included five heat shock proteins, most likely involved in stress responses. The chitin-binding class comprised five members, including SCRASP1, known to be induced following bacterial infection (Christophides et al., 2002; Danielli et al., 2000), two downregulated peritrophic matrix components, as determined by a previous proteomic analysis (Dinglasan et al., 2009), and two remaining chitin-binding proteins, one upregulated and one downregulated. The hydrolase class comprised four members, including a glycoside hydrolase.

The gene encoding LYSC2 showed the most pronounced upregulation while two MD2-like lipid recognition receptors were downregulated. A core behavioural component was also present, including the downregulation of two transcripts of the gustatory receptor Gr13, but also transcripts corresponding to juvenile and pheromone binding genes, including the upregulation of *TO2* and the downregulation of *OBP54*, *OBP13*. Notably, the transcript corresponding to NPF, a neurotransmitter whose *Drosophila* orthologue has been implicated in modulation of various behavioural processes, including feeding behaviour (Hergarden et al., 2012), aversion (Wu et al., 2005b), aggression (Dierick and Greenspan, 2007) and regulation of reward systems (Shohat-Ophir et al., 2012), was downregulated following oral *S. marcescens* infection.

To identify gene classes significantly overrepresented in the set of 99 differentially expressed transcripts, a GO analysis was carried out. A hypergeometric test followed by Benjamini-Hochberg correction was used to determine enriched GO terms in the set of 97 genes, accounting for transcripts of splice variants corresponding to the same gene. The results of the GO analysis indicated 16 enriched GO terms (Figure 6.7A) that mainly corresponded to the serine-type endopeptidase activity GO term, with 17 members, and the chitin-binding GO term, with 5 members (Figure 6.7B). Genes corresponding to the serine-type endopeptidase activity GO term included the gene encoding the scavenger receptor SCRASP1, which also has a peptidase domain, the CLIP domain serine proteases *CLIPE6*, *CLIPB14*, *CLIPB17* and *CLIPB20*, as well as several other peptidases, including several putative trypsins. The proteolysis GO term mainly corresponded to the same genes as the serine-type endopeptidase GO term, with the addition of a serine carboxypeptidase and a microsomal dipeptidase. The aminoglycan metabolic process GO term also

corresponded to the same genes as the chitin-binding GO term, with the addition of *PGRPLC*. Finally, the significantly overrepresented digestion GO term corresponded to the genes encoding two trypsins and NPF while the extracellular region GO term included trypsins, chitin-binding genes, FREPs, *NPF* and a gene encoding the glutathione S transferase GSTD1.

A



B

Serine-type endopeptidase activity

Transcript ID	Name/Description	Fold change	Regulation
AGAP012037-RA	CLIPB20	1.75	up
AGAP005625-RB	SCRASP1	1.76	up
AGAP011785-RA	CLIPE6	1.90	up
AGAP010833-RA	CLIPB14	2.76	up
AGAP005669-RA	Serine collagenase	1.94	up
AGAP005689-RA	Serine collagenase	2.49	up
AGAP005690-RA	Serine collagenase	2.79	up
AGAP008290-RA	Trypsin-6	2.42	down
AGAP008291-RA	Trypsin-5	1.82	down
AGAP006707-RA	Serine-type endopeptidase	2.10	up
AGAP007142-RA	Serine-type endopeptidase	1.77	up
AGAP001198-RA	Trypsin putative	2.12	down
AGAP013221-RA	Trypsin putative	2.59	up
AGAP004316-RA	Serine-type endopeptidase Disulphide knot CLIP	1.85	up
AGAP012966-RA	Serine-type endopeptidase	1.95	up
AGAP001648-RA	CLIPB17	1.86	up
AGAP011325-RA	Coagulation factor	2.20	up

Proteolysis

Transcript ID	Name/Description	Fold change	Regulation
AGAP011442-RA	Serine carboxypeptidase	1.87	down
AGAP011654-RA	Microsomal dipeptidase	2.00	up
AGAP012037-RA	CLIPB20	1.75	up
AGAP005625-RB	SCRASP1	1.76	up
AGAP011785-RA	CLIPE6	1.90	up
AGAP010833-RA	CLIPB14	2.76	up
AGAP005669-RA	Serine collagenase	1.94	up
AGAP005689-RA	Serine collagenase	2.49	up
AGAP005690-RA	Serine collagenase	2.79	up
AGAP008290-RA	Trypsin-6	2.4	down
AGAP008291-RA	Trypsin-5	1.82	down
AGAP006707-RA	Serine-type endopeptidase	2.10	up
AGAP007142-RA	Serine-type endopeptidase	1.77	up
AGAP001198-RA	Trypsin putative	2.12	down
AGAP013221-RA	Trypsin putative	2.59	up
AGAP001648-RA	CLIPB17	1.86	up
AGAP011325-RA	Coagulation factor	2.20	up

Chitin-binding

Transcript ID	Name/Description	Fold change	Regulation
AGAP005625-RB	SCRASP1	1.76	up
AGAP006433-RA	AgAper34	1.83	down
AGAP006434-RA	AgAper57	2.36	down
AGAP006991-RA	Chitin binding	2.76	up
AGAP001203-RA	Chitin binding	2.05	down

Aminoglycan metabolic process

Transcript ID	Name/Description	Fold change	Regulation
AGAP005625-RB	SCRASP1	1.76	up
AGAP006433-RA	AgAper34	1.83	down
AGAP006434-RA	AgAper57	2.36	down
AGAP006991-RA	Chitin binding	2.76	up
AGAP001203-RA	Chitin binding	2.05	down
AGAP005203-RA	PGRPLC	1.75	up

Digestion

Transcript ID	Name/Description	Fold change	Regulation
AGAP008290-RA	Trypsin-6	2.42	down
AGAP008291-RA	Trypsin-5	1.82	down
AGAP004642-RA	Neuropeptide F	1.77	down

Extracellular region

Transcript ID	Name/Description	Fold change	Regulation
AGAP005625-RB	SCRASP1	1.76	up
AGAP008290-RA	Trypsin-6	2.42	down
AGAP008291-RA	Trypsin-5	1.82	down
AGAP005848-RA	FREP44	2.33	up
AGAP006433-RA	AgAper34	1.83	down
AGAP006434-RA	AgAper57	2.36	down
AGAP006991-RA	Chitin binding	2.76	up
AGAP001203-RA	Chitin binding	2.05	down
AGAP004164-RD	Glutathione S-transferase delta class GSTD1	2.31	up
AGAP004642-RA	Neuropeptide F	1.77	down
AGAP011228-RA	FREP24	1.87	up

Figure 6.7: GO terms significantly overrepresented in the set of more than 1.75-fold regulated transcripts following oral *S. marcescens* infection. A hypergeometric test with Benjamini-Hochberg correction was used to compare the representation of genes corresponding to the same GO term in the set of 97 more than 1.75-fold regulated genes following oral *S. marcescens* infection to the respective representation in the *An. gambiae* genome, as annotated in the *Pfalcip_Agamb2009* microarray design. 16 GO terms corresponding to significantly overrepresented groups of genes, using a corrected p-value cut-off of 0.1, are shown in orange type in the GO directed acyclic graph, with GO terms in the same path that did not meet the p-value cut-off shown in black type. For significantly overrepresented GO terms, the number of corresponding genes in the set of more than 1.75-fold regulated genes is shown in parenthesis. A table with the regulated transcripts corresponding to GO terms at each final leaf node is shown, including, for each transcript, the Transcript ID, an assigned name or description based on *Interpro*-predicted domains or homologies with *Drosophila* counterparts and the observed transcriptional regulation following oral *S. marcescens* infection.

Therefore, the core response to oral *S. marcescens* infection includes members of most classes found in the *SmA* and *SmB* clusters showing consistent differential expression throughout the assayed independent infections. The proteolytic component is dominant while several differentially expressed transcripts suggest the presence of a behavioural immune component. The haemolymph immune response and the oxidation stress component are less represented, although *LRIM1* and two P450 cytochrome genes are present. Finally, the identification of several genes as differentially expressed following *S. marcescens* infection that have been previously reported to be transcriptionally regulated following a bacterial challenge, including *CLIPB14* (Christophides et al., 2002), *SCRASP1* (Christophides et al., 2002; Danielli et al., 2000), *LYSC2* (Li et al., 2005) and *LRIM1* (Meister et al., 2005), further suggests that the combination of datasets stemming from both infection batches reveals the core response to oral *S. marcescens* infection, including transcripts with the more consistent and pronounced differential expression.

6.7 Transcriptional regulation following oral infection with bacteria of the genus *Asaia*

In a similar manner to oral infections with *S. marcescens*, antibiotic treated mosquitoes were orally infected with alpha-proteobacteria of the genus *Asaia*. Bacteria-fed mosquitoes were selected 2 days post infection and, the following day, dissected guts from infected mosquitoes, along with guts from uninfected mosquitoes kept on antibiotic treatment, were used for total RNA extraction.

A total of four independent infections were carried out, in two batches (A-B). For each of the two infections in the first batch, total RNA was extracted from two different pools of infected mosquitoes and used for separate DNA microarray hybridizations, designated as *AsaA1i*, *AsaA1ii*, *AsaA2i* and *AsaA2ii*. The two independent infections in the second batch were performed using mosquitoes from the same *N'gousso* laboratory colony but months apart from the infections of the first batch. Therefore, a level of genetic divergence between the mosquito pools used in the two infection batches was expected. The DNA microarray hybridizations using total RNA from the two independent infections of the second batch were designated as *AsaB1* and *AsaB2*. Therefore, a total of six DNA microarray hybridizations were carried out, corresponding to four independent infections: *AsaA1i*/*AsaA1ii*, *AsaA2i*/*AsaA2ii*, *AsaB1* and *AsaB2*.

6.8 Principal component analysis of datasets corresponding to transcriptional responses following oral infection with bacteria of the genus *Asaia*

The generated datasets from the six performed DNA microarray competitive hybridizations using total RNA from *Asaia* infected and uninfected mosquito guts, were used to perform a PCA so as to compare expression data variation between these datasets. The PCA results can be seen in Figure 6.8. The first three principal components are plotted, with the 1st principal component, in the X-axis, capturing 37.24%, the 2nd principal component, in the Y-axis, capturing 23.78% and the 3rd principal component, in the Z-axis, capturing 19.67% of the total variation. Interestingly, all datasets from the first batch (*AsaA1i*, *AsaA1ii*, *AsaA2i* and *AsaA2ii*), as well as *AsaB1*, showed limited variation with regard to the 1st principal component, which captured the largest proportion of the dataset variation. *AsaB2* was a distinct outlier relative to all other datasets, especially with regard to the 1st and 2nd principal components. Furthermore, the datasets stemming from the same independent infection, *AsaA1i-AsaA1ii* and *AsaA2ii-AsaA2ii*, did not show the same level of clustering compared to datasets stemming from the same *S. marcescens* infection (Figure 6.4), although this may be due to the limited variation between datasets from different independent infections. *AsaA1i* and *AsaA1ii* clustered well relatively to the 1st and 2nd but showed some divergence towards the 3rd principal component. *AsaA2i* and *AsaA2ii* clustered well relatively to the 1st and 3rd but showed divergence towards the 2nd principal component. Overall, inter-dataset variation was much more limited compared to the *S. marcescens* DNA microarray datasets, with only *AsaB2* showing extensive divergence compared to the other datasets.

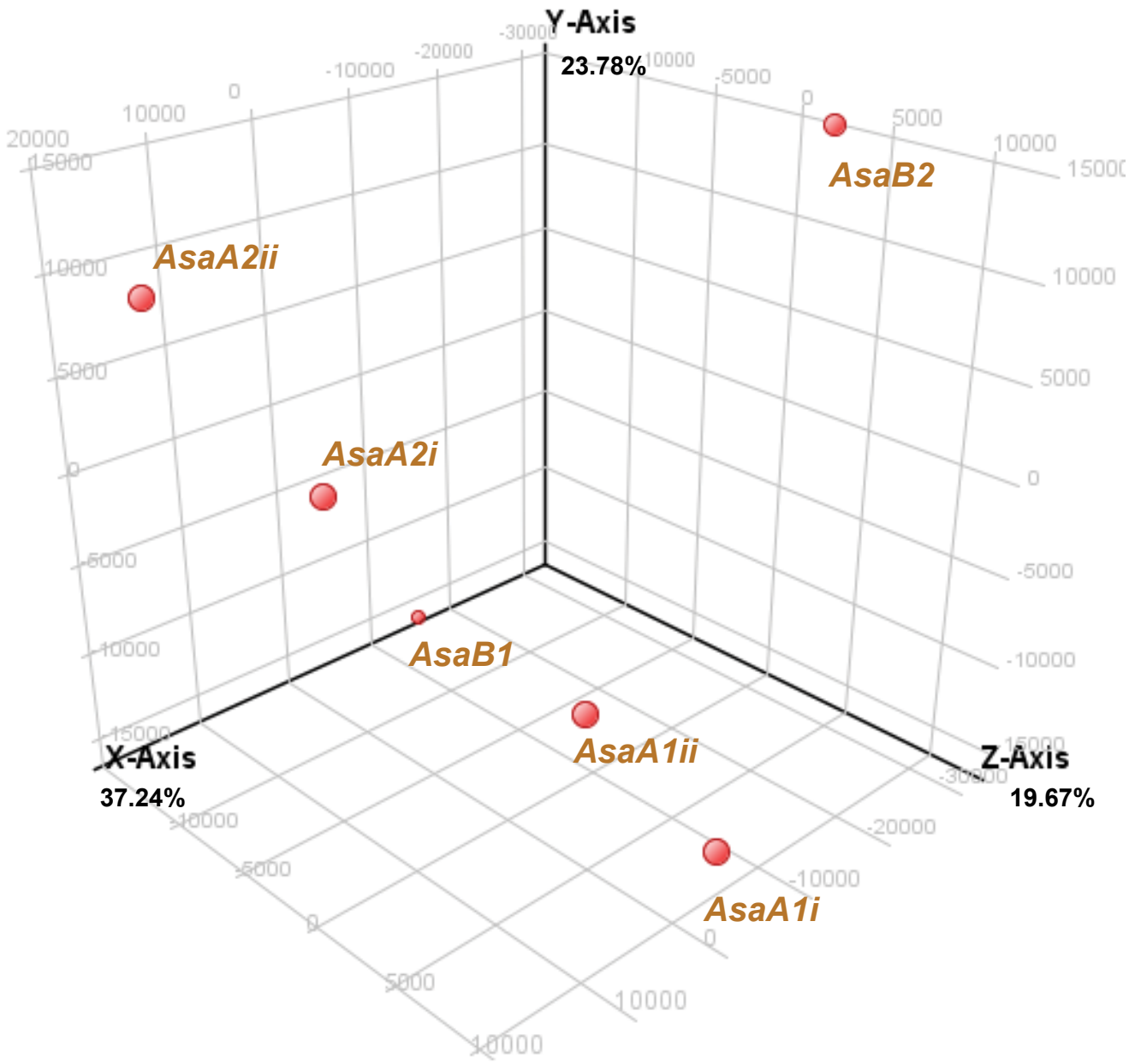


Figure 6.8: Principal component analysis of DNA microarray datasets corresponding to different hybridizations of complementary RNA stemming from *Asaia* infected and uninfected mosquitoes. A PCA, as described in Figure 6.4 for *S. marcescens* infections, was also conducted for DNA microarray datasets corresponding to hybridizations of complementary RNA stemming from guts of *Asaia* infected mosquitoes and uninfected controls. Data separation of linear dimensions for the 1st, 2nd and 3rd principal components is plotted in the x-axis, y-axis and z-axis, respectively. The proportions of captured variation by each principal component is indicated. The designated name of the DNA microarray dataset each data point corresponds to is also indicated.

6.9 Clusters of differentially expressed transcripts following *Asaia* infection

To identify clusters of co-regulated transcripts between *Asaia* DNA microarray datasets stemming from independent infections, a k-means clustering approach was used, partitioning the expression data of each dataset into 10 clusters based on proximity of the mean log2-transformed differential expression of each transcript with the mean of each cluster, following 10,000 algorithm iterations. For the dataset pairs stemming from the same infection, *AsaA1ii* and *AsaA2i* were further used based on their proximity in the PCA (Figure 6.8). The results of the k-means clustering between the *AsaA1ii*, *AsaA2i*, *AsaB1* and *AsaB2* datasets can be seen in Figure 6.9. Despite the proximity of *AsaA1ii*, *AsaA2i* and *AsaB1* in the PCA (Figure 6.8), k-means clustering revealed distinct clusters of up or down regulated transcripts between datasets stemming from the same batch of infections, indicating that, despite the more limited observed variation, distinct transcriptional programmes were most likely deployed in response to *Asaia* infection in the two infection batches, with the underlying mosquito genetic variation being the more conspicuous reason for this expression profile divergence.

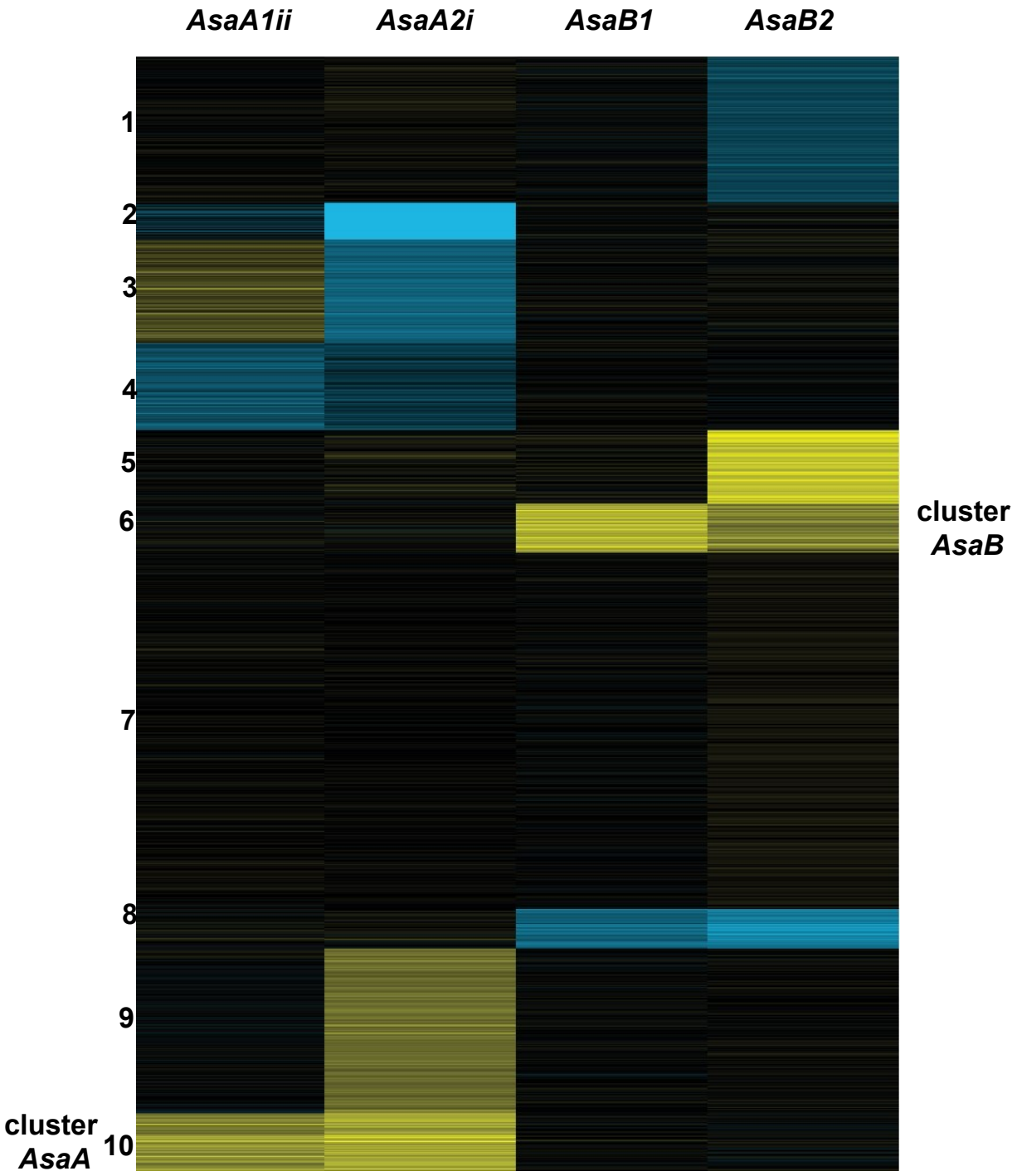


Figure 6.9: K-means clustering of DNA microarray datasets representing differential expression between *Asaia* infected mosquitoes and uninfected controls in hybridizations stemming from independent infection assays. DNA microarray datasets corresponding to individual hybridizations with complementary RNA stemming from guts of *Asaia* infected and uninfected mosquitoes were used for k-means clustering, as described in Figure 6.5 for the respective *S. marcescens* infections. The dataset used is indicated over each regulated transcript set. Clusters designated as *AsaA* and *AsaB* are indicated adjacent to the respective cluster.

Cluster 10, designated as cluster *AsaA*, contained 700 transcripts mostly showing upregulation in the *AsaA1ii-AsaA2i* datasets, stemming from the first infection batch. This cluster contained 121 immune-related transcripts, shown in Table 6.5. Based on the presence of upregulated transcripts in cluster *AsaA*, the response to *Asaia* infection in batch A infections comprised a large proteolytic component, with 16 genes encoding CLIP domain serine proteases and 22 genes encoding peptidase domains, including CLIPA2, CLIPA5 and CLIPA8, previously implicated in killing and melanization of *Plasmodium* parasites (Schnitger et al., 2007; Volz et al., 2006). At the same time, 8 genes encoding serpins were also found in cluster *AsaA*, suggesting the induction of tightly regulated proteolytic cascades.

Several genes residing in cluster *AsaA* also indicated the induction of a haemolymph immune component, as cluster *AsaA* contained the genes encoding the thioester-containing proteins TEP1, TEP3, TEP4, TEP12 and TEP14 as wells as 9 LRIMS, including the LRIM1- APL1C pair but also LRRD7 (LRIM17), previously shown to influence the outcome of *P. falciparum* infection (Dong et al., 2006a; Garver et al., 2012). The modulator of TEP1 responses, Lp (Rono et al., 2010), was also found in cluster *AsaA*, along with two genes encoding vitellinogen domains. Furthermore, the *AsaA* cluster also contained the genes encoding 8 FREPs but also ML1, also shown to influence the outcome of *P. falciparum* infection (Dong et al., 2006a).

Although the JAK/STAT pathway has been previously shown to be involved in *Drosophila* responses to Gram-negative bacteria, through regulation of stem cell proliferation and gut remodeling following oral bacterial infection (Buchon et al., 2009; Cronin et al., 2009; Jiang et al., 2009), no direct link of an *An. gambiae* JAK/STAT pathway involvement was provided by the SNP genotyping nor the expression analysis of *S. marcescens* infections. Interestingly, though, the gene encoding the *An. gambiae* orthologue of *Drosophila Socs44A*, a negative regulator of the JAK/STAT pathway (Rawlings et al., 2004), was found in cluster *AsaA*. The upregulation of the mosquito counterpart of *Socs44A* following oral *Asaia* infection suggests that, following oral infection with bacteria of the genus *Asaia*, the JAK/STAT pathway is actively repressed. Socs44A also functions as an activator of the EGFR pathway (Rawlings et al., 2004). A putative similar function of its mosquito counterpart, along with the association of *EGFR* with the outcome of *S. marcescens* infection, suggests an involvement of the EGFR but not the JAK/STAT pathway in mosquito responses to oral bacterial infection, both for *Asaia* and *S. marcescens*.

Furthermore, the gene encoding Cactus, a negative regulator of the Toll/REL1 pathway (Garver et al., 2012), was also found in cluster *AsaA*, along with the gene encoding TUBE, a putative Toll/REL1 pathway member, and TOLL5B, suggesting repression of the Toll/REL1 pathway following *Asaia* infection. An IMD/REL2 pathway member, IKK1/ird5 was also found in cluster *AsaA*. Its *Drosophila ird5* orthologue is

required for antibacterial responses through the Imd pathway (Lemaitre and Hoffmann, 2007; Lu et al., 2001). Therefore, the presence of IKK1/ird5 in cluster *AsaA* suggests that the IMD/REL2 pathway is likely to be active following *Asaia* infection.

Taken together, these data suggest that in batch A of *Asaia* infections, while the JAK/STAT and Toll/REL1 pathways are repressed, the EGFR and IMD/REL2 pathways are likely to be activated. It is also notable that, in contrast to *S. marcescens* infection, *Asaia* infection triggers regulators of all major pathways, indicating the repression of JAK/STAT and Toll/IMD and the activation of the EGFR and IMD/REL2 pathways. This level of regulation that most likely shapes the response to *Asaia* infection could be a result of a homeostatic balance achieved between mosquito immunity and a common, relatively innocuous or even beneficial member of the mosquito gut flora.

The induction of an oxidative burst following *Asaia* infection was suggested by the presence of 8 genes encoding P450 cytochromes, as well as the gene encoding DBLOX, a double oxidase, and the haeme peroxidase HPX7. The *AsaA* cluster also included the genes encoding the antimicrobial peptides LYSC1 and attacin C. AGAP011984, a gene with no predicted domains, and AGAP010012, encoding a leucine-rich repeat domain were the only genes associated with the outcome of *S. marcescens* infection found in cluster *AsaA*.

Furthermore, an induced behavioural component was suggested by the presence in cluster *AsaA* of the genes encoding TO3, a juvenile hormone binding protein, ITP, which includes a crust neurohormone domain and GPRGNR1, a putative gonadotrophin releasing hormone receptor. In contrast to the respective clusters identified following *S. marcescens* infection, no Grs were found in cluster *AsaA*.

Overall, the genes contained in cluster *AsaA* suggest that for the first batch of *Asaia* infections, the response to oral infection with bacteria of the genus *Asaia* includes considerable proteolytic and haemolymph immune components. The proteolytic component could possibly modulate the haemolymph responses, as the implication of SPCLIP1 in complement immunity suggests (Povelones et al., 2013). Furthermore, *Asaia* infections seem to elicit tight regulation towards the activation of the IMD/REL2 and EGFR and repression of the JAK/STAT and Toll/REL1 pathways. Members of oxidative stress or behavioural components are present but to a much lower scale compared to transcripts identified in the respective clusters following *S. marcescens* infection.

Clustering of the *AsaB1* and *AsaB2* datasets, stemming from independent infections of the second batch, formed two distinctive clusters (Figure 6.9). Cluster 6, designated as *AsaB*, contained mostly upregulated

transcripts, especially in the *AsaB1* dataset, and cluster 8 contained mostly downregulated transcripts. Cluster 8 contained a total of 453 transcripts, 59 of which possessed homologies suggesting involvement in immune responses, shown in Table 6.6. Notably, the gene encoding the catalytic peptidoglycan recognition protein PGRPLB was found in this cluster of downregulated transcripts. *Drosophila* PGRP-LB is a known modulator of the Imd pathway, which, following Imd-dependent activation, tapers the Imd immune response by degrading peptidoglycan fragments (Zaidman-Remy et al., 2006). Therefore, the downregulation of PGRPLB in the *AsaB1-AsaB2* infections is expected to enhance IMD/REL2-mediated responses against *Asaia*. At the same time, the gene encoding IAP2, another putative regulator of the IMD/REL2 pathway (Gesellchen et al., 2005), was also found in this cluster.

A total of 7 genes associated with the outcome of *S. marcescens* infection were found in this cluster of downregulated genes, including the gene encoding GPRNNB2 and AGAP005096, encoding a homeobox gene. One possibility that can be further investigated is whether this homeobox gene regulates IMD/REL2 responses along the lines of *Caudal*, another homeobox gene (Clayton et al., 2012).

Several genes related to mosquito behaviour were also found in this cluster, including the genes encoding the gustatory receptor Gr44, the allatostatin receptor GPRALS3, two nicotinic acetylholine receptors, a neurospecific receptor kinase and the olfactory receptors Or2, Or43 and Or54. As allatostatin signalling inhibits feeding behaviour in *Drosophila* (Hergarden et al., 2012), the downregulation of an allatostatin receptor would be expected to enhance feeding. The physiological relevance of the observed *GPRALS3* downregulation could thus be hypothesized to result in enhanced *Asaia* intake.

Interestingly, the genes encoding NOS but also the P450 cytochromes CYP9L2 and CYP9M1 and the haeme peroxidases HPX12 and HPX14 were also found in this cluster. As for GPRALS3, the downregulation of NOS would be expected to work in favour of tolerating the infection with *Asaia*, the physiological relevance of which remains to be determined. The same applies for the downregulation of the genes encoding several peptidases but also the lysozymes LYSC4 and LYSC5, found in this cluster.

Taken together, the genes found in this cluster of downregulated transcripts suggest that, for the infections corresponding to the *AsaB1-AsaB2* datasets, enhanced IMD/REL2 pathway activity but, at the same time, tapered oxidative burst and behavioural immune responses would be expected.

The cluster *AsaB* contained 647 transcripts, mostly showing upregulation in the *AsaB1-AsaB2* datasets. 98 of these transcripts possessed homologies suggesting participation in immune responses and are shown in Table 6.7. In contrast to datasets *AsaA1ii-AsaA2i*, where *Socs44A,* a negative regulator of the JAK/STAT

pathway was upregulated, the STAT2 transcription factor gene was found in cluster *AsaB*, suggesting activation of the JAK/STAT pathway (Gupta et al., 2009). Remarkably, while *Cactus* was found in the *AsaA* cluster, suggesting repression of the Toll/REL1 pathway, *REL1* was found in cluster *AsaB* along with three spaetzle-like cytokines, which could also be possibly involved in the same pathway, suggesting that *Asaia* infection in the *AsaB1-AsaB2* datasets triggered the Toll/REL1 pathway. The underlying mechanisms that regulate the Toll/REL1 pathway activation or repression following *Asaia* infection remain to be determined. Remarkably, though, *Asaia* infection, both in batches A-B, was shown to result in modulation of the JAK/STAT and Toll/REL1 immune pathways, a modulation that can either activate or repress the respective pathway.

As for the IMD/REL2 pathway, the gene encoding PGRPLC was found in cluster *AsaB*, through the transcript AGAP005203-RD. Interestingly, following *S. marcescens* infection, *PGRPLC* induction involved the AGAP005203-RA and AGAP005203-RC transcripts, found in different infection batches. It is possible that induction of alternative *PGRPLC* splice variants may be a key modulating event of the observed response, possibly also affecting the activation of other pathways.

Among 7 genes associated with the outcome of *S. marcescens* infection, also found in cluster *AsaB*, *PGRPLC* and *FN3D1* were found. Along *FN3D1*, 5 additional genes encoding FN3 domains were also found in cluster *AsaB*, including *sidestep* (AGAP001674), a paralogue of *FN3D3*. The *Drosophila sidestep* encodes an attractant involved, as Dscam, in motor axon guidance (Siebert et al., 2009; Sink et al., 2001). Interestingly, AGAP010184, another gene encoding a protein with an FN3 and immunoglobulin domain architecture found in the same cluster, is an orthologue of *Drosophila echinoid*, a negative regulator of the EGFR pathway (Bai et al., 2001).

Although NOS was found downregulated in cluster 8, a NOS paralogue, encoded by AGAP008257, was found upregulated in cluster *AsaB*, along with the genes encoding the haeme peroxidases HPX1, HPX3 and HPX5, and four P450 cytochromes, suggesting the induction of an oxidative burst in batch B infections.

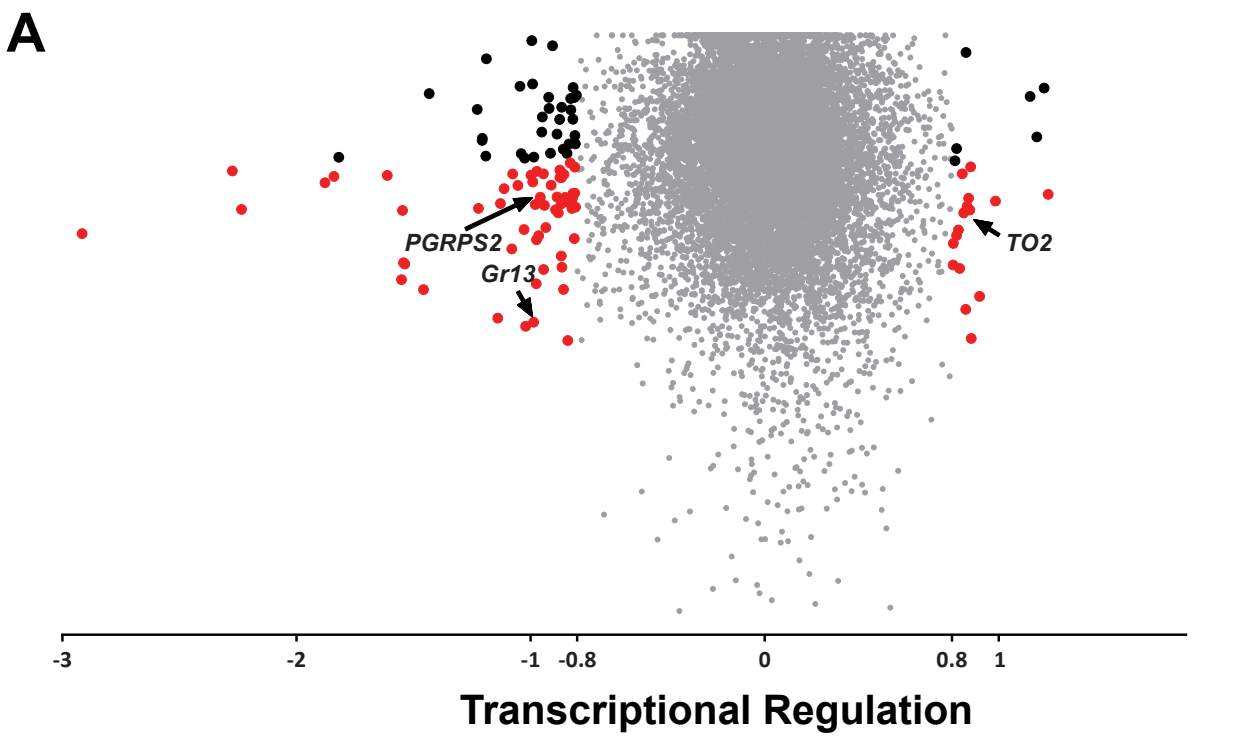
A behavioural component triggered by *Asaia* infection was also suggested by the presence in cluster *AsaB* of the genes encoding the gustatory receptors Gr16, Gr31, Gr37 and Gr56, the olfactory receptors Or20, Or45, Or47 and Or60, a juvenile hormone binding protein and the odorant binding proteins OBP12 and OBP51. The octopamine/tyramine receptor GPRTYR was also found in cluster *AsaB*. Finally, the presence of haemolymph immune components in cluster *AsaB* was limited, including only the gene encoding TEP8, four CLIP domain serine proteases and a C-type lectin.

Overall, the genes corresponding to transcripts residing in cluster *AsaB* suggest a complex layer of immune pathway regulation following *Asaia* infection, involving the IMD/REL2, Toll/REL1, EGFR and JAK/STAT pathways, which could be related to tolerance mechanisms developed during the *An. gambiae-Asaia* symbiosis. Furthermore, the presence of *FN3D1* and *sidestep* in cluster *AsaB* raise the possibility for involvement of these genes in regulatory mechanisms that shape the outcome of bacterial infections. Finally, the response to *Asaia* infection in the *AsaB1-AsaB2* datasets was shown to involve behavioural and oxidative components but a limited haemolymph immune component.

6.10 Core transcriptional responses following oral *Asaia* infection

To identify conserved transcriptional responses throughout the independent infections assayed, significantly regulated transcripts were identified using the *AsaA1i*, *AsaA2i* and *AsaB2* datasets, stemming from three independent infections and covering both infection batches. The *AsaB1* dataset, stemming from another independent infection, clustered well with *AsaB2* and was not further used so as to enhance statistical power by limiting the number of assayed datasets.

Overall, 111 transcripts showed more than 1.75-fold regulation, with 20 upregulated and 91 downregulated transcripts (Figure 6.10A and Table 6.8). The log2-transformed fold-change values from each probe across the assayed datasets for each transcript were used in a t-test against zero, where zero corresponds to no transcriptional regulation, yielding 73 transcripts in the set of more than 1.75-fold regulated transcripts with p-values <0.05, comprising 14 upregulated and 59 downregulated transcripts (Figure 6.10A). The preponderance of downregulated transcripts in the core transcriptional responses to *Asaia* infection was not observed in the respective core responses to *S. marcescens* infection and may indicate elicitation of tolerance mechanisms by the *Asaia* symbiont.


A
PGRPS2
Gr13
TO2
-3
-2
-1
-0.8
0
0.8
1
Transcriptional Regulation

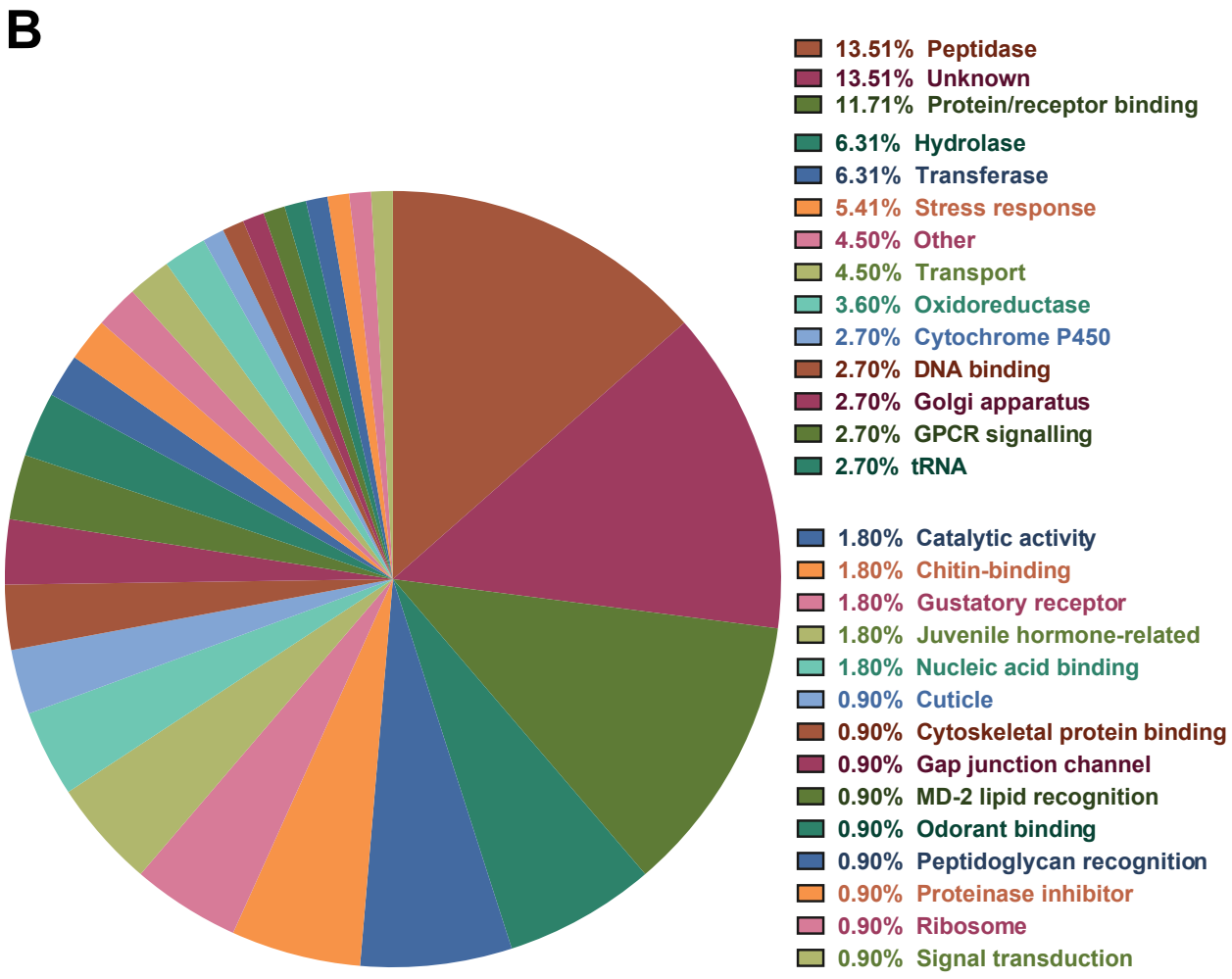

B
13.51% Peptidase
13.51% Unknown
11.71% Protein/receptor binding
6.31% Hydrolase
6.31% Transferase
5.41% Stress response
4.50% Other
4.50% Transport
3.60% Oxidoreductase
2.70% Cytochrome P450
2.70% DNA binding
2.70% Golgi apparatus
2.70% GPCR signalling
2.70% tRNA
1.80% Catalytic activity
1.80% Chitin-binding
1.80% Gustatory receptor
1.80% Juvenile hormone-related
1.80% Nucleic acid binding
0.90% Cuticle
0.90% Cytoskeletal protein binding
0.90% Gap junction channel
0.90% MD-2 lipid recognition
0.90% Odorant binding
0.90% Peptidoglycan recognition
0.90% Proteinase inhibitor
0.90% Ribosome
0.90% Signal transduction

Figure 6.10: Core transcriptional regulation following oral *Asaia* infection. Differential expression in *Asaia* infected mosquitoes compared to uninfected controls was determined over three independent infections. **A:** Manhattan plot of log2-transformed fold-change regulation determined for each transcript in relation to the p-value determined in a t-test against zero using the respective log2-transfomed transcriptional regulation values over three independent infections. Transcripts with more than 1.75-fold differential expression are indicated in red or black data points, depending on whether the computed p-value was <0.05 or >0.05, respectively. Data points corresponding to transcripts for PGRPS2, Gr13 and TO2 are indicated by arrows. **B:** Assigned functional classes for the 111 more than 1.75-fold regulated genes following *Asaia* infection. The % representation of each functional class is indicated and corresponds to the respective colour in the pie chart.

The genes corresponding to transcripts with more than 1.75-fold regulation were assigned to their respective functional class, based on predicted *Interpro* domains or homologies, mainly to *Drosophila* counterparts (Figure 6.10B). The most abundant functional class comprised 15 genes containing peptidase domains, corresponding to 13 downregulated and only 2 upregulated transcripts. No CLIP domain serine proteases were found in this group. The protein/receptor binding class included the genes encoding 4 FREPs and also IAP2, the latter being downregulated. The gene encoding the peptidoglycan recognition protein PGRPS2 was also downregulated by 1.94-fold. Its *Drosophila* orthologues, *PGRP-SC1* and *PGRP-SC2,* have been implicated in negative regulation of the Imd pathway (Bischoff et al., 2006). Therefore, the *PGRPS2* downregulation suggests an enhanced antibacterial response through activation of the IMD/REL2 pathway. The differential expression of immune pathway modulators following *Asaia* infection, a property shown in clusters of co-regulated transcripts between independent infections of different batches, is thus also evident in core responses to *Asaia* infection and may indicate a level of *Anopheles-Asaia* co-adaptation. Such tight regulation of immune pathway modulators was not seen in core responses to *S. marcescens* infection and may comprise a property that relies on characteristics found in *Asaia* but not *S. marcescens*.

The hydrolase functional class included two downregulated glycoside hydrolases while the stress response functional class included three downregulated heat shock protein genes. Downregulation of stress response components is indicative of a dampened stress response following *Asaia* infection that may lead to enhanced *Asaia* survival and reduced mosquito fitness cost, again suggesting a level of co-adaptation between *Asaia* and the mosquito host.

Interestingly, two different transcripts of the gustatory receptor Gr13 were downregulated following *Asaia* infection. *Gr13* was also downregulated as part of the core responses to *S. marcescens* infection. Gr13 has no known *Drosophila* homologues nor any *Anopheles* paralogues, so little can be surmised regarding its putative function. It is possible that its downregulation might be part of a behavioural immune response following *Asaia* or *S. marcescens* infection. Furthermore, the gene encoding the odorant binding protein OBP13 was downregulated while the gene encoding the juvenile hormone binding protein TO2 was upregulated, as shown also for *S. marcescens* infection. Therefore, *Asaia* or *S. marcescens* infection may induce a common response aimed to modulate feeding behaviour.

The transcript AGAP008703-RA, downregulated following *Asaia* infection, has been re-annotated since the DNA microarray platform design implemented here, in the current *AgamP3.7* (May 2013) *An. gambiae* gene annotation to the gene encoding the neuropeptide receptor GPRNPR2, associated with the outcome

of *S. marcescens* infection. Pending confirmation of the *GPRNPR2* downregulation following *Asaia* infection, this neuropeptide receptor may also participate in behavioural immune responses following *Asaia* or *S. marcescens* infection.

To identify significantly overrepresented groups of genes in the set of 111 more than 1.75-fold regulated transcripts following *Asaia* infection, a GO analysis was performed. Only two GO terms were found to be significantly enriched in this set of transcriptionally regulated genes, carbohydrate binding and digestion (Figure 6.11A). The carbohydrate binding GO term corresponded to 8 genes, including 2 chitin-binding genes, 3 peptidases and a gene encoding a ricin B lectin domain, all showing downregulation (Figure 6.11B). The digestion GO term corresponded to 4 downregulated trypsins.

A

B

Transcript ID	Name/Description	Functional class	Fold change	Regulation
AGAP011984-RA	Ricin_B_lectin	Transferase	1.78	down
AGAP009830-RA	Chitin-bd_dom	Chitin-binding	1.91	down
AGAP007165-RA	Late trypsin	Peptidase	2.31	down
AGAP007684-RA	Peptidase_C1A_C Wnt signalling	Peptidase	1.83	down
AGAP007684-RB	Peptidase_C1A_C Wnt signalling	Peptidase	1.88	down
AGAP001203-RA	Chitin-bd_dom	Chitin-binding	2.19	down
AGAP007611-RA	EGF-like_CS	Protein/receptor binding	1.80	down
AGAP004376-RA	aldose 1-epimerase	Carbohydrate binding	1.81	down

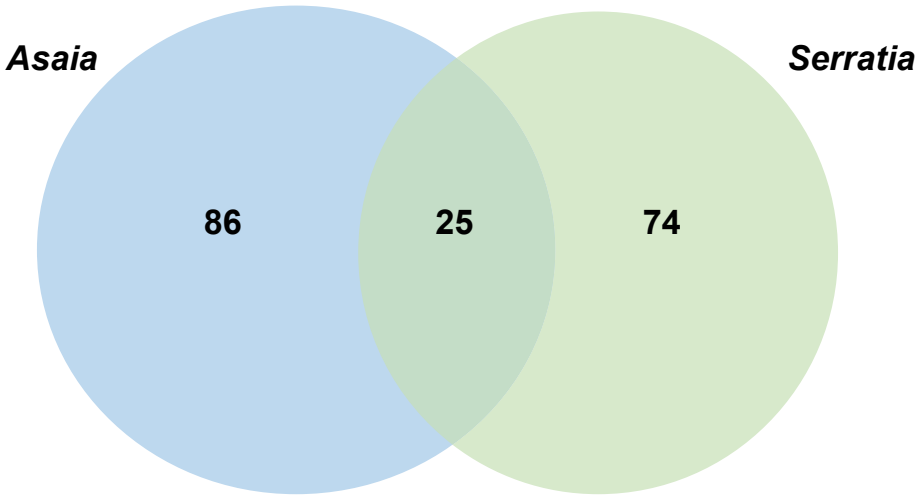
Transcript ID	Name/Description	Functional class	Fold change	Regulation
AGAP008290-RA	Trypsin-6	Peptidase	3.06	down
AGAP008291-RA	Trypsin-5	Peptidase	2.91	down
AGAP008292-RA	Trypsin-4	Peptidase	1.93	down
AGAP008293-RA	Trypsin-7	Peptidase	1.80	down

Figure 6.11: GO terms significantly overrepresented in the set of more than 1.75-fold regulated transcripts following oral *Asaia* infection. A: Two GO terms corresponded to significantly overrepresented groups of genes using a corrected p-value cut-off of 0.1, shown in the GO directed acyclic graph, along with GO terms in the same path that did not meet the p-value cut-off. For significantly overrepresented GO terms, the number of corresponding genes in the set of more than 1.75-fold regulated genes is shown in parenthesis. **B:** A table with the regulated transcripts corresponding to GO terms at each final leaf node of the acyclic graph is shown, including, for each transcript, the Transcript ID, an assigned name or description based on *Interpro*-predicted domains or homologies with *Drosophila* counterparts and the observed transcriptional regulation following *Asaia* infection.

6.11 Transcriptional regulation spanning both oral *Asaia* and *S. marcescens* infection

The comparison of the 99 transcripts showing more than 1.75-fold transcriptional regulation following *S. marcescens* infection (Table 6.4) and the 111 transcripts showing more than 1.75-fold regulation following *Asaia* infection (Table 6.8) identified 25 transcripts found in both sets (Figure 6.12A). This considerable overlap indicates that relevant genes were identified by the adopted DNA microarray approach to identify differential expression following oral *Asaia* or *S. marcescens* infection. Remarkably, all 25 transcripts showed consistent up or down regulation both after *Asaia* and *S. marcescens* infection (Figure 6.12B). *Gr13* and *OBP13* were downregulated while *TO2* was upregulated following both *Asaia* and *S. marcescens* infection, indicative of a conserved antibacterial behavioural immune response. Furthermore, 4 genes encoding heat shock proteins and 4 peptidase-domain containing genes were also downregulated following both *Asaia* and *S. marcescens* infection. This downregulation may indicate a homeostatic mechanism to limit gut damage due to the induced antibacterial response. The genes encoding 3 FREPs were also regulated following both *Asaia* and *S. marcescens* infection, suggesting a possible involvement in pattern recognition of a broad range of Gram-negative bacteria. Finally, the gene AGAP001203, encoding a chitin-binding gene, was also downregulated following *Asaia* and *S. marcescens* infection, also indicative of an involvement in antibacterial responses.

A



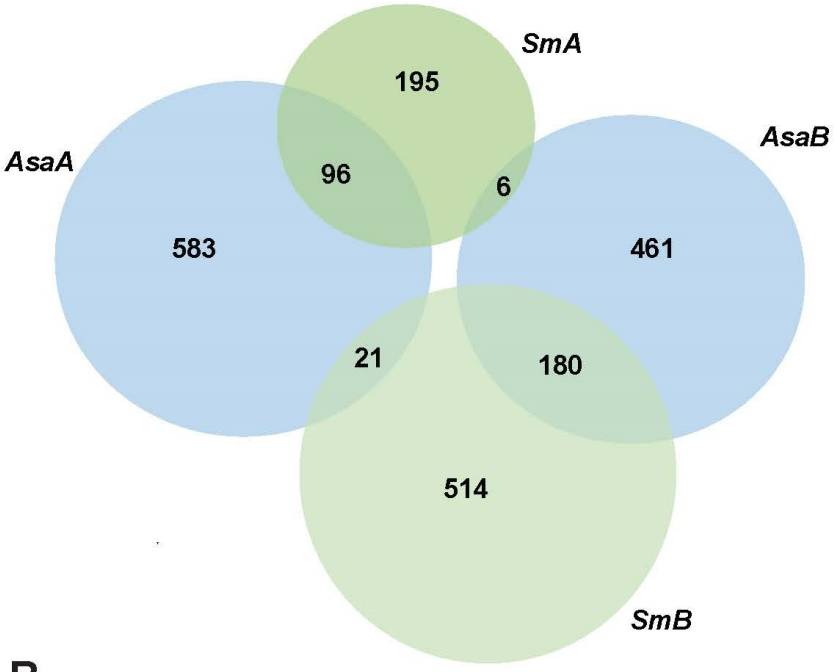
B

Transcript ID	Name/Description	Functional class	*Asaia* Fold change	*Asaia* Regulation	*Serratia* Fold change	*Serratia* Regulation
AGAP000570-RA	Unknown	Unknown	2.03	down	1.82	down
AGAP001198-RA	Trypsin putative	Serine-type endopeptidase	1.80	down	2.12	down
AGAP001203-RA	Chitin binding	Chitin binding	2.19	down	2.05	down
AGAP002635-RA	Gr13	GPCR signaling	2.08	down	2.68	down
AGAP002635-RB	Gr13	GPCR signaling	1.98	down	2.13	down
AGAP002905-RA	OBP13	Odorant binding	1.82	down	6.00	down
AGAP003209-RA	Sterol desaturase	Oxidoreductase	1.86	down	2.43	down
AGAP004581-RA	Heat shock protein 70	Nucleotide metabolic process	3.53	down	2.47	down
AGAP004582-RA	Heat shock protein 70	Nucleotide metabolic process	3.67	down	2.44	down
AGAP004583-RA	Heat shock protein 70	Nucleotide metabolic process	3.58	down	2.43	down
AGAP006689-RA	Protein binding BTB/POZ-like	Protein/receptor binding	1.82	down	1.83	down
AGAP006788-RA	Acyl CoA acyltransferase	Transferase	1.99	down	2.61	down
AGAP007267-RA	tRNA-Gly	tRNA	2.11	down	1.86	down
AGAP008290-RA	Trypsin-6	Serine-type endopeptidase	3.06	down	2.42	down
AGAP008291-RA	Trypsin-5	Serine-type endopeptidase	2.91	down	1.82	down
AGAP009146-RA	Unknown	Unknown	7.54	down	2.88	down
AGAP010760-RA	FREP69	Protein/receptor binding	1.79	down	2.04	down
AGAP010762-RA	FREP4	Protein/receptor binding	1.83	down	1.86	down
AGAP011228-RA	FREP24	Protein/receptor binding	2.32	up	1.87	up
AGAP011442-RA	Serine carboxypeptidase	Peptidase other	2.16	down	1.87	down
AGAP012552-RA	Unknown	Unknown	2.28	down	1.90	down
AGAP012703-RA	takeout2	Juvenile hormone/pheromone binding	1.80	up	1.78	up
AGAP012794-RA	tRNA-Arg	tRNA	1.92	down	2.37	down
AGAP012891-RA	Heat shock protein 70	Nucleotide metabolic process	2.92	down	2.02	down
AGAP013493-RA	Unknown	Unknown	2.93	down	2.28	down

Figure 6.12: Overlap between core transcriptional responses following oral *Asaia* and *S. marcescens* infection. A: Venn diagram of more than 1.75-fold regulated transcripts following *Asaia,* in a blue circle to the left, or *S. marcescens,* in a green circle to the right, infection. The number of transcripts included in each delineated part of the diagram is indicated. **B:** List of 25 transcripts found to be more than 1.75-fold regulated following both *Asaia* and *S. marcescens* infection. Each column shows the corresponding Transcript ID, the assigned name or a description of the transcript homologies, based on *Interpro*-predicted domains, assigned GO terms or *Drosophila* counterparts, the assigned functional class and the fold-change transcriptional regulation following *Asaia* and *S. marcescens* infection.

As shown both for *Asaia* (Figure 6.9) and *S. marcescens* (Figure 6.5) infections, datasets stemming from different batches of infections, most likely due to the underlying mosquito genetic variation, resulted in discrete transcription profiles. In the case of *S. marcescens* infection, two clusters of upregulated genes, *SmA* and *SmB*, were identified in datasets from different infection batches, both pointing to relevant immune genes and suggesting deployment of diverse transcriptional immune programmes. In the case of *Asaia* infections, the clusters of upregulated genes *AsaA* and *AsaB* were identified in a similar manner. To identify common components of transcriptionally regulated genes both after *Asaia* and *S. marcescens* infection, the genes corresponding to transcripts found in the *AsaA*, *AsaB*, *SmA* and *SmB* clusters were compared (Figure 6.13A). The *AsaA*-*SmA* clusters shared 96 genes found in both clusters while the *AsaB*-*SmB* clusters shared 180 genes. Overlap between *AsaA*-*SmB* and *AsaB*-*SmA* was much more limited, with 21 and 6 shared genes, respectively. These data suggest that the deployed transcriptional programmes following *Asaia* or *S. marcescens* infection in batches A-B share a common component which is most likely to be influenced by the underlying genetic variation in the respective batch of mosquitoes used.

A



B

AsaA-SmA

Gene ID	Name/Description
AGAP001826	Vitellinogen
AGAP002625	*CTL9*
AGAP003689	CLIP domain serine protease
AGAP004164	*GSTD1*
AGAP004262	*TO3*
AGAP005848	*FREP44*
AGAP006348	*LRIM1*
AGAP006910	*SRPN3*
AGAP007033	*APL1C*
AGAP007039	*LRIM4*
AGAP007347	*LYSC1*
AGAP007457	*LRIM7*
AGAP008210	*CYP6N1*
AGAP008213	*CYP6M3*
AGAP008382	nuclear receptor subfamily
AGAP008654	*TEP12*
AGAP009213	*SRPN16*
AGAP010669	*TOLL5B*
AGAP010731	*CLIPA8*
AGAP010812	*TEP4*
AGAP010815	*TEP1*
AGAP010816	*TEP3*
AGAP011228	*FREP24*
AGAP011790	*CLIPA2*
AGAP011791	*CLIPA1*

AsaB-SmB

Gene ID	Name/Description
AGAP000051	*HPX5*
AGAP000143	Glycoside hydrolase
AGAP001117	*Gr37*
AGAP001633	FN3
AGAP002519	*GPRTYR*
AGAP002813	*CLIPD6*
AGAP003254	*Gr16*
AGAP003871	Homeodomain
AGAP004847	*SCRB7*
AGAP005381	Glycoside hydrolase
AGAP006143	*Gr56*
AGAP007177	*SPZ5*
AGAP008118	Protein kinase
AGAP009288	*DER1*
AGAP009413	*Or20*
AGAP009594	Leucine-rich repeat
AGAP009621	Glycoside hydrolase
AGAP011225	*FREP28*
AGAP011294	*DEF1*
AGAP012422	Kinase-like domain
AGAP012591	*CLIPA3*

AsaA-SmB

Gene ID	Name/Description
AGAP001076	*CYP4G16*
AGAP003790	*ANXB9*
AGAP005281	Homeodomain
AGAP008369	Vitellinogen
AGAP010078	Ricin B lectin

AsaB-SmA

Gene ID	Name/Description
AGAP005203	*PGRPLC*
AGAP008782	Ricin B lectin

Figure 6.13: Overlap between clusters of co-regulated transcripts identified following *Asaia* or *S. marcescens* infection. A: Venn diagram of transcripts included in clusters *AsaA* and *AsaB*, identified in DNA microarray datasets stemming from oral *Asaia* infections, shown in blue circles, as well as clusters *SmA* and *SmB*, identified in DNA microarray datasets stemming from oral *S. marcescens* infections, shown in green circles. The number of transcripts included in each delineated part of the diagram is indicated. **B:** Genes of interest found to overlap between the clusters *AsaA-SmA*, *AsaB-SmB*, *AsaB-SmB* and *AsaB-SmA*. Each column shows the respective Gene ID and the assigned name or a description of the gene based on *Interpro*-predicted domains, assigned GO terms or *Drosophila* counterparts.

Relevant immune genes found to be shared between the *AsaA-SmA*, *AsaB-SmB*, *AsaA-SmB* and *AsaB-SmA* clusters are shown in Figure 6.13B. The common component of the *AsaA-SmA* clusters included genes involved in haemolymph responses, encoding the complement factors TEP1, LRIM1 and APL1C but also LRIM4, LRIM7, TEP3, TEP4 and TEP12. The genes encoding the CLIP domain serine proteases CLIPA2 and CLIPA8 were also shared between the *AsaA-SmA* clusters, along with CLIPA1, CTL9 and SRPN3. Therefore, the transcriptional programmes corresponding to the *AsaA-SmA* clusters included a considerable systemic immune component. Other genes shared between the *AsaA-SmA* clusters included the gene encoding the putative modulator of feeding behaviour TO3, suggesting induction of responses in these two clusters related to modulation of feeding behaviour.

The shared component between the *AsaB-SmB* clusters also included a considerable behavioural component, with genes encoding the gustatory receptors Gr16, Gr37 and Gr56, the olfactory receptor Or20 and the octopamine/tyramine receptor GPRTYR. Furthermore, 3 genes encoding glycoside hydrolases were also found in the *AsaB-SmB* shared component. Finally, the genes encoding the haeme peroxidase HPX5 and the spaetzle-like cytokine SPZ5 were also shared between the *AsaB-SmB* clusters. Interestingly, AGAP001633, a gene encoding both immunoglobulin and FN3 domains, was also shared between the *AsaB-SmB* clusters, whose function and relevance to FN3D1-3 remains to be determined. The overlap between *AsaA-SmB* and *AsaB-SmA* was limited but in both cases included a different gene encoding a ricin B lectin domain. *PGRPLC* was also found in both the *AsaB* and *SmA* clusters but, as mentioned previously, with different induced splice variants.

6.12 Microarray data validation using qRT-PCR

To examine the validity of the observed transcriptional regulation using a DNA microarray approach, microarray-based differential expression data were confirmed by employing a qRT-PCR approach. Total RNA pools from *S. marcescens* infected mosquito guts as well as from the respective uninfected controls, used for microarray hybridization in 5 of the datasets analyzed, *SmA2*, *SmA3i*, *SmB2i*, *SmB2ii* and *SmB3*, were used as template for cDNA synthesis. Transcript abundance, compared to the *AgS7* endogenous control was subsequently assayed using qRT-PCR. The datasets assayed included the 3 datasets used to identify core transcriptional responses following *S. marcescens* infection, *SmA2*, *SmA3i* and *SmB3*.

Differential expression between S*. marcescens* infected and uninfected pools was assayed by qRT-PCR for 5 transcripts: AGAP007343-RA, corresponding to LYSC2, AGAP006991-RA, corresponding to a chitin-binding gene, AGAP001648-RA, corresponding to CLIPB17, AGAP010833-RA, corresponding to CLIPB14,

and AGAP005901-RA, corresponding to Ect4. All assayed transcripts, except for the transcript corresponding to Ect4, were found to be more than 1.75-fold regulated in core *S. marcescens* transcriptional responses, with the transcript corresponding to Ect4 found in cluster *SmA* but not in the core response. The transcripts corresponding to LYSC2 and CLIPB14 yielded a significant (<0.05) p-value in a t-test against zero, suggestive of pronounced and consistent transcriptional regulation between the assayed microarray datasets. Furthermore, all transcripts, except AGAP006991-RA, were found in the *SmA* cluster of co-regulated transcripts between *SmA2* and *SmA3i*. This dataset and transcript selection, inevitable due to the resource and time consuming nature of qRT-PCR, infers that the validation process examined diverse aspects of the analyzed expression data.

For each dataset, abundance for the 5 assayed transcripts was determined relative to the *AgS7* endogenous control for infected and uninfected cDNA pools and was used to determine the fold-change differential expression of each transcript in *S. marcescens* infected mosquitoes. The obtained log2-transformed fold-change values are shown in Table 6.9 and Figure 6.14A, along with the respective log2-transformed fold-change values obtained by DNA microarrays. Overall, 25 pairs of microarray and qRT-PCR derived log2-tranformed fold-change values were obtained, corresponding to the transcriptional regulation of the assayed transcripts in each of the assayed pairs of total RNA samples corresponding to *S. marcescens* infected or uninfected mosquito guts.

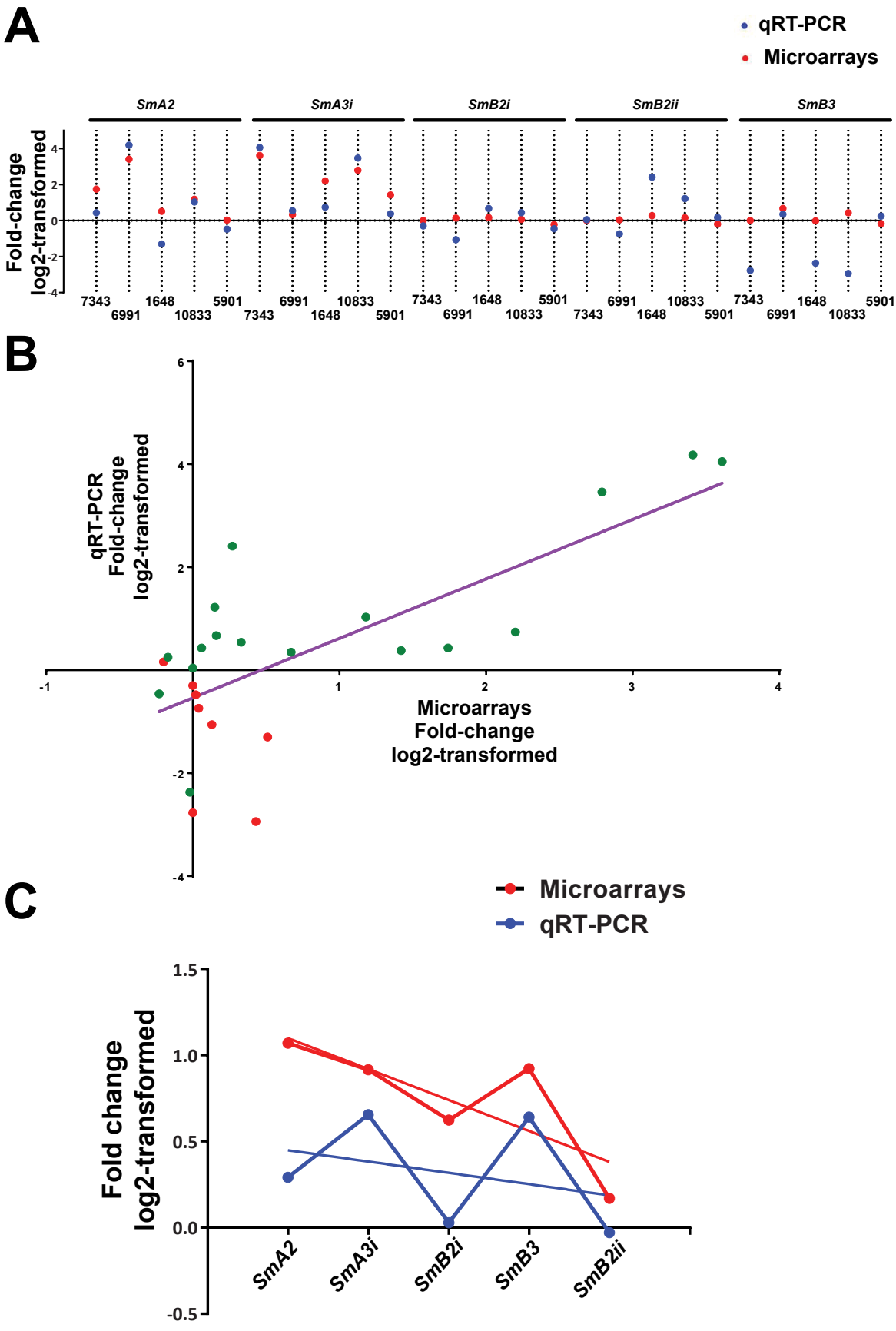

A
qRT-PCR
Microarrays
SmA2
SmA3i
SmB2i
SmB2ii
SmB3
Fold-change
log2-transformed
7343
6991
1648
10833
5901
B
qRT-PCR
Fold-change
log2-transformed
Microarrays
Fold-change
log2-transformed
C
Microarrays
qRT-PCR
Fold change
log2-transformed
SmA2
SmA3i
SmB2i
SmB3
SmB2ii

Figure 6.14: DNA microarray validation using a qRT-PCR-based differential expression approach. A: Differential expression was determined for the transcripts corresponding to LYSC2 (7343), a chitin-binding protein encoded by AGAP006991-RA (6991), CLIPB17 (1648), CLIPB14 (10833) and Ect4 (5901) by qRT-PCR. Total RNA stemming from midguts of *S. marcescens* infected mosquitoes or uninfected controls used in hybridizations corresponding to the datasets *SmA2*, *SmA3i*, *SmB2i*, *SmB2ii* and *SmB3*, was used for synthesis of cDNA that was further used as template. The resulting log2-transformed fold-change regulation between *S. marcescens* infected mosquitoes and uninfected controls is plotted for each of the 25 transcript-dataset combinations, as determined by qRT-PCR, with a blue data point. The respective log2-transformed fold-change regulation as determined by DNA microarrays is shown as a red data point. **B:** Each data point represents a transcript-dataset combination, with log2-transformed fold-change values derived from DNA microarrays and qRT-PCR plotted in the x-axis and the y-axis, respectively. Data points corresponding to transcript-dataset combinations with concordant up or down regulation as determined by DNA microarrays and qRT-PCR are in green, with the remaining non-concordant combinations in red. Based on the 25 plotted data points, the best-fit regression line has been plotted. **C:** Log2-transformed fold-change regulation for the examined transcripts, derived by DNA microarrays or qRT-PCR was used for each of the 5 examined datasets in a linear regression analysis. Average log2-transformed fold-change values are plotted for each examined dataset, with the derived regression line in red for microarray-obtained values and in blue for qRT-PCR –obtained values.

To examine the concordance of the pairs of fold-change values derived by microarray or qRT-PCR based methods, the respective log2-transformed fold-change values were plotted in Figure 6.14B. For each data point representing a fold-change value pair, the x-axis value corresponds to the microarray-derived fold-change value while the y-axis value corresponds to the qRT-PCR-derived fold-change value. 16 pairs of fold-change values showed concordance in the up or down regulation, derived by both the microarray and qRT-PCR analysis, while in the 9 pairs in which opposite regulation was identified by the microarray or qRT-PCR approaches, only 2 outliers were identified with >1.75-fold up or down regulation in the microarray or qRT-PCR analysis, with the remaining pairs showing marginal differential expression, which could have accounted for the observed divergence between the microarray and qRT-PCR analysis.

The 25 data points corresponding to each pair of microarray or qRT-PCR derived fold-change values were used in a linear regression analysis (Figure 6.14B). These 25 data points were used to fit a regression line that best predicts the qRT-PCR derived value, plotted on the Y axis, for any microarray derived value, plotted on the X axis. If such best-fit regression line is linear, perfect concordance is assumed between qRT-PCR and microarray fold-change values. Goodness of fit was quantified by the R square (R^2) value that was computed to be 0.5250. To assess deviation from linearity, a runs test was performed that yielded a non-significant p-value of 0.6675, indicating that the best-fit regression line is not significantly different from a best-fit linear line. These data suggest a high level of concordance between the qRT-PCR and microarray derived pairs of fold-change values.

As both qRT-PCR and microarray fold-change values are experimentally measured, it is considered appropriate to assess concordance between the pairs of qRT-PCR-microarray derived fold-change values using both linear regression and correlation approaches. Therefore, to further quantify the correlation between the microarray and qRT-PCR derived data, the Spearman's rank correlation co-efficient between the microarray and qRT-PCR derived values was measured. The Spearman's rank correlation co-efficient is a non-parametric assessment of the direction and magnitude of correlation between the assayed datasets, based on the 25 pairs of identified values. When all 25 pairs of qRT-PCR-microarray fold-change values were included in the analysis, the Spearman's rank correlation co-efficient between the microarray and qRT-PCR derived values was measured at 0.6105, in which 1 would designate a perfect correlation, 0 no correlation and -1 negative correlation, and a derived p-value of 0.0012, suggesting a statistically significant positive correlation between the microarray and qRT-PCR derived fold-change values. When the 9 pairs of values that showed opposite regulation between microarrays and qRT-PCR were excluded, the Spearman's rank correlation co-efficient for the remaining 16 pairs of values was measured at 0.6853,

with a derived p-value of 0.0044, suggesting an even stronger correlation that remained statistically significant despite the lower number of assayed data points.

To further investigate whether the correlation between the microarray and qRT-PCR data is affected by the magnitude of the fold-change regulation, the Spearman's rank correlation co-efficient was measured for pairs of microarray-qRT-PCR derived fold-change values corresponding to fold-change differences from 0 to 2-fold, in absolute terms, and, similarly, from 0 to 3-fold and over 3-fold change values.

For 10 pairs of fold-change values in which both values corresponded to fold-change differences up to 2-fold (0 to 1 in log2-transformed values), the Spearman's rank correlation co-efficient was computed at 0.4255. When 6 pairs with fold-change values up to 3-fold were included (up to 1.58 in log2-transformed values), the Spearman's rank correlation co-efficient was computed at 0.3929, indicating a slightly lower level of correlation compared to the inclusion of only up to 2-fold values. Finally, when values exceeding 3-fold were included, the Spearman's rank correlation co-efficient reached 0.6105, as described above. Indeed, the 9 pairs of microarray-qRT-PCR fold-change values that included values exceeding 3-fold (1.58 in log2-transformed values) reached a Spearman's rank correlation co-efficient of 0.7833, indicating a high level of correlation.

These data suggest that even low fold-change differences in pairs of qRT-PCR-microarray derived fold-change values show positive correlation, reaching a Spearman's rank correlation co-efficient of 0.3929 and 0.4255 for up to 3 and 2-fold change values, respectively. At the same time, pairs of fold-change values that include values exceeding 3-fold changes, show a tighter level of positive correlation, reaching a Spearman's rank correlation co-efficient of 0.7833 and thus increasing the overall correlation of the set of all pairs of fold-change values. This is consistent with the enhanced level of sensitivity that is expected in the detection of higher fold-change differences both in the microarray and qRT-PCR based approaches.

Further investigation of the concordance between microarray and qRT-PCR derived fold-change values at different levels of fold-change differences would require a higher number of pairs of fold-change values corresponding to low, mid or high fold-change differences to achieve statistical significance in demonstrating positive correlation between the respective pairs of fold-change values. Such investigation would exceed the scope of the presented analysis aiming at validation of the employed DNA microarray approach. Further sequencing-based analysis can shed more light in the identification of low-fold change differences compared to the array-based method used here.

The concordance of microarray and qRT-PCR derived values for each assayed dataset was also determined by a linear regression approach. The log2-transformed fold-change values of differential expression for the assayed transcripts were used in each of the 5 datasets to determine the best-fit slope and intercept of microarray and qRT-PCR derived data (Figure 6.14C). The resulting regression lines of microarray and qRT-PCR derived data were further compared to determine whether they are significantly different. An analysis of covariance (ANCOVA) method was employed to assess significant differences between the slope and intercept of the two regression lines. Slope comparison resulted in a p-value of 0.713 whereas intercept comparison resulted in a p-value of 0.3324, both suggesting that the differences of the slope and intercept between the two regression lines are not significant.

Taken together, these data suggest that the microarray data correlate well with the respective qRT-PCR data, thus confirming that the microarray approach provided useful information regarding differential transcript expression in the assayed datasets. A more comprehensive RNA sequencing-based analysis can further validate the microarray data revealed by the assayed datasets but also further refine aspects of the analysis that exceed the sensitivity of the presented array-based approach.

The approach adopted here for the validation of microarray data is consistent with previously reported array-based expression analyses in mosquitoes, with the positive correlation identified between microarray and qRT-PCR derived data being comparable with previously reported levels of correlation when similar platforms and biological samples were used (Dong et al., 2009; Kazura et al., 2011). Furthermore, the assessment of statistical significance in the observed level of correlation lends further credence to the validation of the assayed microarray datasets. The costly and time-consuming nature of such approach limits the number of assayed fold-change values, with some previously reported microarray-based analyses in mosquitoes limiting such validation to a small number of assayed transcripts, with no further statistical analysis (Magnusson et al., 2011; Rund et al., 2011).

6.13 Comparison of transcriptional programmes following oral bacterial infection between *Drosophila* and *An. gambiae*

This is the first reported expression analysis of mosquito responses to oral bacterial infections. This analysis provides several candidate genes that are possibly involved in *An. gambiae* antibacterial responses and can be followed up in future functional characterization efforts. It also provides several preliminary insights regarding the complexity of mosquito responses following oral bacterial infection.

The design of this microarray analysis mirrors similar *Drosophila* projects (Buchon et al., 2009; Chakrabarti et al., 2012), although several key differences include the use of different bacterial strains, which, as has been suggested in *Drosophila* (Chakrabarti et al., 2012), may elicit discrete responses. Methodological differences further include the initial mosquito antibiotic treatment and subsequent oral infection with bacterial solutions of lower concentration, and, most importantly, the use, instead of a genetically homogeneous fly line, of a recently established M form *An. gambiae* colony, thought to retain a considerable portion of genetic variation found in field mosquitoes.

The analysis presented here identified a core response of a limited number of regulated transcripts, both for *S. marcescens* and *Asaia* infections, compared to similar *Drosophila* microarray studies (Buchon et al., 2009; Chakrabarti et al., 2012). At the same time, clusters of co-regulated transcripts between microarray datasets included a much higher number of transcripts, commensurate to the number of regulated transcripts in *Drosophila*. Although in both cases the *Drosophila* Imd and the *An. gambiae* IMD/REL2 pathways play key roles in antibacterial responses, the involvement of the JAK/STAT pathway was shown only following *Asaia* infection, activated in batch B infections through the upregulation of *STAT2,* or actively repressed in batch A infections due to the upregulation of *Socs44A*, a factor whose *Drosophila* orthologue inhibits the JAK/STAT pathway but also activates the EGFR pathway, also involved in *Drosophila* epithelial renewal (Buchon et al., 2010; Rawlings et al., 2004). Although no role for the Toll pathway was identified in *Drosophila* (Buchon et al., 2009), a possible Toll/REL1 involvement was suggested following *Asaia* infection, again either through pathway repression or activation.

Both the *An. gambiae* and *Drosophila* microarray studies identified the involvement, both in *Drosophila* (Buchon et al., 2009) and *An. gambiae,* of chitin-binding genes following oral bacterial infection. A *Drosophila* chitin-binding gene has been further implicated in forming a barrier that plays a protective role following oral *S. marcescens* infection (Kuraishi et al., 2011). Chitin-binding genes which have been

identified as peritrophic matrix components in *An. gambiae (*Dinglasan et al., 2009*)*, a barrier formed after blood feeding, were downregulated following *S. marcescens* infection, while other chitin-binding genes, which most likely do not participate in peritrophic matrix formation, were upregulated. It is thus possible that a similar to the *Drosophila* barrier is also formed in mosquitoes in response to bacterial infections prior to blood feeding. A peroxidase/dual oxidase system has been implicated in *An. gambiae* in forming a dityrosine network barrier after blood feeding, discrete from the peritrophic matrix, which is important in limiting epithelial responses due to bacterial contact with triggers of epithelial immunity (Kumar et al., 2010). It thus also possible that these chitin-binding genes induced following *S. marcescens* infection, along with several induced peroxidases identified in this study, participate in formation of a similar barrier. Another suggested function for chitin-binding genes in mosquitoes, especially for scavenger receptors such as SCRASP1 that combine chitin-binding and peptidase domains, is the detection of danger signals through binding chitin released due to tissue damage (Danielli et al., 2000). This possibility can also be further investigated.

Both the *Drosophila* and *An. gambiae* transcription programmes following oral bacterial infection included the induction of several behavioural components. *Ecc15* oral infection in *Drosophila* induced several transcripts involved in neuronal transmission, including the upregulated gustatory receptor gene *Gr94a*, but also an allatostatin receptor and the odorant binding protein gene *Obp56d* (Buchon et al., 2009). It is also striking that infection with *P. entomophila*, a more pathogenic strain, elicited a behavioural immune response in cessation of feeding that had been ascribed to stress responses (Chakrabarti et al., 2012; Vodovar et al., 2005). Therefore, the identified behavioural component in response to oral *Asaia* or *S. marcescens* infection most likely parallels similar behavioural responses in *Drosophila*, although the latter passed relatively unremarkable in the respective reports (Buchon et al., 2009; Chakrabarti et al., 2012; Vodovar et al., 2005).

6.14 Influence of the underlying mosquito genetic variation on the observed expression profile following oral bacterial infection

The number of transcripts comprising the core responses to *Asaia* or *S. marcescens* infection is consistent with a previous microarray analysis in *An. gambiae* between antibiotic treated mosquitoes and mosquitoes retaining their natural gut microbiota (Dong et al., 2009). The limited number of 185 significantly regulated transcripts was ascribed to mosquito adaptation to commensal bacteria inhabiting the mosquito gut due to symbiotic interactions (Dong et al., 2009). At the same time, though, the

difference in the number of differentially expressed transcripts between the *An. gambiae* and *Drosophila* studies of differential expression following oral bacterial infection can be explained by the different levels of the underlying genetic variation between the assayed mosquito colonies and the genetically homogeneous fly lines.

The genetic basis of the outcome of *S. marcescens* infection suggests that the utilization of independent infections in different mosquito colonies or even in the same colony, especially in different time frames, could result in diverse transcriptional responses which are averaged out between independent infections. Therefore, due to reduced statistical power, a limited number of transcripts would be identified as differentially expressed while the fold-change regulation would also be reduced. Even within an independent infection, though, in a recently established laboratory colony retaining most of the genetic variation of field mosquito colonies, as the *N'gousso* colony used here, the high level of observed variation regarding the outcome of bacterial infection, entails that highly, lowly or non-infected mosquitoes are likely to exhibit diverse expression profiles that shape the observed outcome of infection, which are averaged out in the differential expression profile identified through the experimental design implemented here.

Therefore, the underlying genetic variation within each independent infection but also between infections most likely explains the diverse expression profiles, which, when averaged out, result in a limited number of transcriptionally regulated genes. In contrast, *Drosophila* expression analyses employ fly lines with limited genetic heterogeneity, therefore avoiding such expression profile variation and thus identifying a high number of regulated transcripts following a bacterial challenge.

It is very likely that core responses to *Asaia* or *S. marcescens* infection identified here include the most pronounced and consistently regulated transcripts that comprise the core response to oral bacterial infection. Indeed, several transcriptionally regulated transcripts identified here have been also identified by previous microarray analyses as differentially expressed following a bacterial challenge, including *CLIPB14* (Christophides et al., 2002; Volz et al., 2005) , *SCRASP1* (Christophides et al., 2002; Danielli et al., 2000), *LRIM1* (Dimopoulos et al., 2002; Meister et al., 2005) and *LYSC2* (Li et al., 2005).

While the averaging of three respective microarray datasets for *Asaia* or *S. marcescens* infections provides a number of core transcripts involved in responses to oral bacterial infection, each assayed dataset by itself provides a more detailed picture of differential expression following oral bacterial infection based on a limited genetic pool of mosquitoes, which, as they have been collected from the mosquito colony at

the same time, are expected to retain a large degree of genetic homogeneity. Therefore, the datasets stemming from independent infections identify differential expression following oral bacterial infection with much more limited genetic variation, which could influence the outcome of infection and thus the observed transcription profile.

Furthermore, comparison between datasets stemming from independent infections can provide insights into how a diverse genetic background may influence the outcome of infection by modulation of transcriptional regulation. For example, genetic variation in one of the putative transcription factors associated with the outcome of *S. marcescens* infection could possibly shape the outcome of infection by modulating the expression of effector genes, thus leading to transcription profile differences between datasets stemming from independent infections in which the genes encoding these transcription factors genetically diverge.

The main limitation of the adopted experimental design lies on the nature of microarray-based assays, in which statistically significant differential expression is typically expected following averaging of three or more microarray datasets stemming from biological replicate assays (Cui and Churchill, 2003; Yang et al., 2002). While this approach ensures reproducibility of array-based differential expression, in cases where some degree of variation is expected between biological replicates, this approach reduces statistical power. This limitation is to some degree circumvented here by clustering co-regulated transcripts between two or more microarray datasets. Furthermore, statistical confidence is also achieved by identifying significantly overrepresented groups of genes in the set of differentially expressed genes that corresponded to more than 1.75-fold regulated transcripts with significant or non-significant p-values in a t-test against zero assessing statistically significant differential expression.

The current microarray study on differential expression following oral *Asaia* or *S. marcescens* infection was essentially replicated in two discrete time periods, several months apart, yielding microarray datasets respectively assigned to batch A and B. Although the main reason for replicating the initial batch A infections was to enhance statistical power, a remarkable finding was that, both for *Asaia* and *S. marcescens* infections, discrete transcriptional profiles were identified in clusters of co-regulated transcripts, designated as *SmA-SmB* for the *S. marcescens* infections, and *AsaA-AsaB* for the *Asaia* infections. The main reason for this difference is most likely the diverse genetic background of mosquitoes used in the two infection batches, which, although they stem from the same laboratory colony, due to the constantly evolving nature of this colony, as it is kept in high numbers, with occasional population bottlenecks, they are expected to be genetically diverse.

Although the same procedures were used in both cases, as sample preparation, microarray labelling, hybridization and scanning were performed concomitantly for batch A infections and, at a later time point, for batch B infections, it is conceivable that a difference in the replication procedures may, at least partly, account for the observed diverse expression profiles. Two lines of evidence suggest that this is not the case. First, for both batch A and batch B infections, an internal control was implemented, which comprised the hybridization of samples stemming from the same independent infection, including *SmA3i-SmA3ii* and *SmB2i-SmB2ii* for *S. marcescens* infections, along with *AsaA1i-AsaA1ii* and *AsaA2i-AsaA2ii* for *Asaia* infections. In all of these cases, as shown in the PCA (Figure 6.4 for *S. marcescens* and Figure 6.8 for *Asaia*) and by clustering the respective datasets (data not shown), expression profile variation between these datasets was much more limited than variation between datasets from independent infections. Therefore, expression profile differences are most likely a result of the underlying genetic variation rather than a methodological issue related to sample preparation or the microarray analysis. Second, the identified co-regulated clusters between batch A and batch B infections point to relevant immune genes. It is remarkable that clusters of co-regulated genes from different infection batches point to functional differences in pathway regulation, which are unlikely to be a result of methodological errors. These include the JAK/STAT components Socs44A and STAT2, the Toll/REL1 components Cactus and REL1, the APL1 paralogues APL1A and APL1C or the takeout family members TO2 and TO3, found in different clusters. Furthermore, three different *PGRPLC* splice variants were found in three different clusters, suggesting that PGRPLC function may be a major determinant of the observed expression profile differences.

Another possibility is that epigenetic differences related to mosquito maintenance or composition of the gut flora could have influenced the outcome of the analysis in the different infection batches. Although any residual gut bacteria evading the antibiotic treatment would be expected to be found both in the infected and uninfected pools, with any effect being averaged out, it is possible that elicitation of an immune response by these bacteria or priming exerted by gut bacteria prior to antibiotic treatment might have influenced the observed expression profile following *Asaia* or *S. marcescens* infection. As similar studies in *Drosophila* did not apply any antibiotic treatment to clear the fly's gut flora (Buchon et al., 2009; Chakrabarti et al., 2012), any effect the natural gut microbiota might exert on the observed transcriptional regulation remains to be further investigated both in *Drosophila* or *Anopheles*.

6.15 Comparison of SNP genotyping and DNA microarray datasets following *S. marcescens* infection

Comparison of transcriptionally regulated genes following *S. marcescens* infection with genes associated with the outcome of infection showed limited overlap, with only *PGRPLC* and *CLIPE6* found in both sets. At the same time, though, a considerable number of gene families was represented in both sets by different members, including acyl-transferase, glycoside hydrolase, kinase, GPCR, LRIM, homeobox, zinc-finger, PGRP, peptidase, FREP, MD2-like and chitin-binding genes. An intriguing possibility is that the different gene family members found either differentially expressed or associated with the infection outcome might have functionally diverged. Indeed the role of different gene family members, especially for families such as FREPs that have shown considerable expansion in *An. gambiae* (Dong and Dimopoulos, 2009; Zdobnov et al., 2002), remains poorly understood. One possible explanation is that genes associated with the outcome of infection are constitutively expressed and play key regulatory roles while differentially expressed members of the same family are either induced later in the infection or show redundant function due to their concomitant induction by transcriptional programmes, therefore showing no association with the infection outcome. Interestingly, a previous transcriptomic analysis of immunoglobulin genes involved in antibacterial responses failed to identify *FN3D2* or *FN3D3* as differentially expressed following a bacterial challenge (Garver et al., 2008), further strengthening the complementarity of the SNP genotyping and expression analysis approaches.

6.16 Diverse transcriptional responses elicited by different bacteria

Comparison of core transcriptional responses to *Asaia* or *S. marcescens* infection with transcriptionally regulated genes between antibiotic treated and untreated mosquitoes (Dong et al., 2009), also shows limited overlap but also the induction of different members of common gene families. Such divergence would be expected given the diverse transcriptional programmes induced following infection of bacterial strains with differential pathogenicity in *Drosophila* (Chakrabarti et al., 2012). One attractive possibility is that diverse bacterial strains elicit the induction of different members of the same gene family, which entails that these differentially expressed gene family members show discrete functional characteristics in targeting diverse bacterial strains. Indeed, a considerable fraction of transcripts found in co-regulated clusters following *Asaia* or *S. marcescens* infection was restricted to oral infection by the one bacterial strain, suggesting bacterial strain-specific induction that remains to be further investigated.

6.17 Conserved transcriptional responses between *Asaia* and *S. marcescens* infections

Comparison of core transcriptional responses to *Asaia* and *S. marcescens* identified 25 conserved transcripts. Detection of differential expression for these transcripts by both assays suggests conserved roles in responses to oral infection by Gram-negative bacteria. Notably, this conserved component included transcripts related to mosquito behaviour, such as *Gr13* and *TO2*. The functional characterization of these genes will shed more light in their mode of action, but their differential expression following both *Asaia* and *S. marcescens* infection indicates the existence of complex behavioural networks in response to oral bacterial infection.

Overall, differentially expressed transcripts following *Asaia* and *S. marcescens* infection indicated the existence of several layers of immune responses to oral infection. A considerable proteolytic component was induced following oral bacterial infection. Especially for *S. marcescens* infection, this component included several CLIP domain serine proteases, likely to be involved in signalling cascades related to immune response activation. The identification of several differentially expressed serpins also points to this direction. Furthermore, oral bacterial infection induced systemic haemolymph immune responses, comprising several TEPs and LRIMs and including the well-studied TEP1, LRIM1 and APL1C (Fraiture et al., 2009; Povelones et al., 2009). The induction of a haemolymph immune component indicates that, at least, some bacteria cross the midgut epithelium, a phenomenon that has been reported for *S. marcescens* in *Drosophila* (Nehme et al., 2007*)*. The importance of the induction of these anti-*Plasmodium* immune factors prior to an infectious blood meal in their killing efficacy of malaria parasites remains to be further investigated. Nevertheless, enhanced *Plasmodium* killing upon reinfection through bacteria-induced immune priming relies on higher expression of such complement factors (Rodrigues et al., 2010). The induction of complement factors following oral *Asaia* or *S. marcescens* infection also further strengthens the possibility that these systemic immune responses may primarily target bacteria rather than malaria parasites.

An epithelial immune response based on an oxidative burst was also induced following oral bacterial infection. This response included NOS, differentially expressed following *Asaia* infection, but also several haeme peroxidases and P450 cytochromes, the latter most likely involved in detoxification processes. Such responses can also affect the outcome of *Plasmodium* infections (Gupta et al., 2009) and suggest that since their induction by gut bacteria may also influence the outcome of *Plasmodium* infection, gut bacteria may indirectly shape malaria transmission dynamics.

6.18 Intricate immune pathway regulation following *Asaia* infection

Strikingly, *Asaia* infection triggered the differential expression of several regulators of immune signalling pathways, including the IMD/REL2 regulators PGRPS2 and PGRPLB, the JAK/STAT components Socs44A and STAT2 as well as the Toll/REL1 Cactus and REL1 factors. One attractive explanation is that the *Anopheles-Asaia* symbiosis has led to immune response modulations that allow the mosquito gut colonization by bacteria of the genus *Asaia*, while at the same time targeting other strains that could possibly compete with *Asaia*.

One such example of indirect bacterial interactions is the modulation of DUOX responses in *Drosophila* by bacterial uracil. Uracil-producing pathogenic bacteria trigger DUOX responses that shape the fly's gut microbiota, while commensal bacteria that do not produce uracil inhabit the fly's gut as the DUOX pathway remains inactive (Lee et al., 2013a). It is possible that similar circuits might exist in the *Anopheles-Asaia* symbiosis through modulation of immune signalling pathway activation.

6.19 Induced behavioural responses following *Asaia* and *S. marcescens* infection

Interestingly, both oral *Asaia* and *S. marcescens* infection triggered a considerable behavioural component, including several differentially expressed gustatory and olfactory receptors, pheromone and juvenile hormone binding proteins, including TO2 and TO3, but also, for *S. marcescens* infection, the downregulation of *NPF*. NPF is involved is various behavioural processes, including alcohol sensitivity (Wen et al., 2005), modulation of reward systems (Shohat-Ophir et al., 2012) but also behavioural immune responses (Kacsoh et al., 2013). Strikingly, NPF seems to play, along with another neuropeptide, allatostatin A, a pivotal role in regulation of feeding behaviour in *Drosophila* (Hergarden et al., 2012). While allatostatin A neurons suppress feeding, NPF neurons suppress the allatostatin A inhibitory influence and thus promote feeding (Hergarden et al., 2012). Therefore, downregulation of *NPF* could conceivably modulate feeding behaviour into restricting the sugar meal intake and could thus constitute an aversive behavioural immune response. As NPF limits the intake of a sugar meal by suppressing feeding, it also limits the abundance of ingested *S. marcescens* included in the sugar meal. Further functional characterization of the NPF activity following *S. marcescens* infection, shown in chapter 8 in this study, sheds more light into the NPF mode of action.

Chapter 7

Three FN3 domain encoding genes shape the *Anopheles gambiae* gut microbiota composition by limiting the presence of *Enterobacteriaceae*

Introduction

7.1 Mosquitoes and gut *Enterobacteriaceae* communities

In mosquitoes, any insights regarding the mode of action of members of the mosquito gut microbial community that have been so far identified, involved bacteria belonging to the *Enterobacteriaceae* family. High-throughput sequencing of mosquito gut microbiota differing on the outcome of *P. falciparum* infection indicated that the presence of *Enterobacteriaceae* exhibited a significant correlation with mosquitoes susceptible to *P. falciparum* (Boissière et al., 2012). Furthermore, although the influence of specific members of the mosquito gut microbiota on the outcome of *Plasmodium* infections has been determined, the ecological context in which this influence is exerted remains poorly understood. A member of the *Enterobacteriaceae* family, *Esp_Z*, has been shown to generate reactive oxygen species that directly limit malaria parasites (Cirimotich et al., 2011b). The production of reactive oxygen species with anti-*Plasmodium* activity seems to be a quantitative trait among different bacterial strains, with several members of the mosquito gut microbial community exhibiting the ability to produce such anti-*Plasmodium* factors, at least to some degree (Cirimotich et al., 2011b). Another intriguing finding is the anti-*Plasmodium* activity of *S. marcescens* (Bando et al., 2013). Unlike *Esp_Z*, *S. marcescens* is a member of the *Enterobacteriaceae* family commonly found in field-collected or lab-reared mosquitoes (Boissière et al., 2012; Gonzalez-Ceron et al., 2003), therefore any effect of *S. marcescens* on the outcome of *Plasmodium* infections is directly relevant to malaria transmission dynamics in field mosquito populations.

To further investigate the effect of the three *FN3Ds* associated with the outcome of *S. marcescens* infection in mosquito antibacterial responses that limit *S. marcescens* and how these FN3Ds may influence the population structure of the mosquito gut microbial communities, a reverse genetics approach was implemented in which RNAi-mediated silencing of *FN3D1-3* was carried out in *An. gambiae* mosquitoes. Results indicate that silencing of any of *FN3D1-3* profoundly influences the outcome of *S. marcescens* infection. Furthermore, FN3D1-3 emerge as major modulators of the mosquito gut bacterial population structure as they limit the representation of *Enterobacteriaceae*, mainly *Serratia* or strains with similarity to *Serratia* reference sequences, but also, for FN3D3, bacteria of the genus *Burkholderia*. These data not only confirm the validity of the SNP genotyping approach used to identify the association of *FN3D1-3* with the outcome of S*. marcescens* infection but also reveal that FN3Ds participate in mosquito antibacterial responses that show much higher specificity than previously thought.

7.2 A reverse genetics approach in identification of *An. gambiae* immune factors involved in antibacterial responses

The association of *An. gambiae* genes with the outcome of *S. marcescens* infection or genes with identified transcriptional regulation following *Asaia* or *S. marcescens* infection, raised the possibility that these genes encode immune factors involved in responses to bacterial infection. For immune factors that play key roles in such responses, silencing the respective gene is expected to influence the outcome of infection, thus directly demonstrating the gene contribution in response to infection.

7.3 Prioritization of *An. gambiae* candidate genes putatively involved in antibacterial responses for reverse genetics follow-up

Given the limited time and resources available, candidate gene prioritization for reverse genetics follow-up was inevitable. The time-consuming nature of thoracic microinjections of every single assayed mosquito and further tracking of the outcome of infection in the injected mosquito for several days, severely limited the size of the screening for potential antibacterial factors to the most salient candidate genes. Such gene prioritization is not uncommon in genome-wide studies even in the genetically tractable *Drosophila* model system. In a somewhat similar study investigating *Drosophila* responses to *S. marcescens* infection using an RNAi knock-out approach, out of 885 implicated genes, only a small fraction of them was phenotypically characterized (Cronin et al., 2009).

Several criteria were used to identify candidate genes stemming from both the SNP genotyping and expression analysis for reverse genetics follow-up. The most important assessed feature of candidate genes was their homologies with known immune factors, strongly suggesting an involvement in similar responses, especially for homologies with factors known to modulate the outcome of bacterial infections in a species-specific manner or in other systems. In that respect, the homology of *FN3D2* with *Dscam*, known to be involved in antibacterial responses in *Anopheles* and *Drosophila* (Dong et al., 2012; Dong et al., 2006b; Watson et al., 2005), raised an intriguing speculation about the role of FN3D2 in antibacterial responses. Another interesting homology was that of the gustatory receptor Gr9 with the mammalian chemoattractant receptor GPR43, known to modulate responses to gut microbiota (Maslowski et al., 2009). Further work on Gr9 will be presented in chapter 8.

Another feature taken into account was the identification of several genes with similar homologies and domain architecture, especially in the SNP genotyping analysis. The identification of genes associated with the outcome of *S. marcescens* infection with similar architecture but in different peaks strengthens the possibility that this identification relates to a functional significance, suggesting a causal relationship of the underlying genetic variation with the infection outcome. Furthermore, such identification suggests a critical role for this family of genes in influencing the infection outcome, with genetic variation in individual peaks contributing to genetic association. Based on these criteria, the most prominent candidate genes emerging from the SNP genotyping analysis were the previously uncharacterized *FN3D1-3*. Finally, the novelty of a particular gene family with regard to involvement in antibacterial immune responses was also taken into account, given that the antibacterial activity for members of gene families such as CLIP domain serine proteases, FREPs, PGRPs, LRIMs or Toll-like receptors is highly suspected. Therefore, the phenotypic analysis carried out aimed principally for the identification of novel genes or gene families in responses that not only limit bacteria following oral infection but also shape the population structure of the mosquito gut bacterial community.

7.4 RNAi-mediated silencing of candidate genes differentially expressed following *S. marcescens* infection indicates no contribution to the infection outcome

To functionally characterize the involvement of *An. gambiae* genes associated with the outcome of *S. marcescens* infection or transcriptionally regulated following *S. marcescens* infection, in responses that shape the infection outcome, a reverse genetics approach was followed. Mosquitoes were treated at the day of emergence with dsRNA targeting the transcript corresponding to the gene of interest, leading to its silencing, as previously established (Blandin et al., 2002). At the same time, a control pool of mosquitoes was treated with dsRNA targeting the bacterial *LacZ* gene (*dsLacZ*), functioning as a control group with no silencing of the transcript of interest but also accounting for any effects caused by the thoracic microinjection used to introduce the dsRNA or by triggering the mosquito RNAi machinery. Following a 5-day antibiotic treatment with a gentamicin-penicillin-streptomycin cocktail, mosquitoes were starved overnight by replacing the antibiotics cocktail with dH_2O and were subsequently orally infected with *S. marcescens*. Mosquitoes that had ingested bacteria-containing sugar solution were selected 2 days post infection and, at day 5 post infection, bacterial load was determined in the mosquito gut in mosquitoes treated with the respective dsRNA or the *dsLacZ* control.

In each case, relative bacterial abundance was determined in dsRNA treated mosquitoes following *S. marcescens* infection, using qRT-PCR with derived cDNA used as template and primers targeting the 16S bacterial rRNA transcript, normalized to the *AgS7* endogenous control. This relative bacterial abundance is directly comparable between dsRNA treatments and can be used to assess changes in bacterial load between dsRNA treatments targeting a specific transcript and *dsLacZ* controls. Therefore, bacterial abundance for a dsRNA treatment targeting a specific transcript was normalized to the *dsLacZ* bacterial abundance. As a result, the normalized bacterial abundance for *dsLacZ* controls always equaled 1, while the normalized bacterial abundance in dsRNA treated pools directly expressed the fold-change difference in bacterial load compared to the *dsLacZ* treated control. Significant differences of log2-transformed normalized fold-change values were assessed in a t-test against zero, where zero corresponds to no difference from the *dsLacZ* control.

The influence of three genes upregulated following *S. marcescens* infection in shaping the outcome of infection was examined using RNAi-mediated silencing. Silencing of a chitin-binding domain encoding gene, AGAP006991 (Figure 7.1A), or Ect4, encoding a TIR domain containing putative IMD/REL2 regulator (Figure 7.1B), did not result in significant differences in bacterial load compared to *dsLacZ* treated controls in orally infected mosquitoes, in one assayed independent infection in each case. Silencing of *CLIPE6*, which was upregulated following *S. marcescens* infection but was also associated with the outcome of infection, also resulted in no significant change in bacterial load compared to the *dsLacZ*-treated control, in one independent infection (Figure 7.1C).

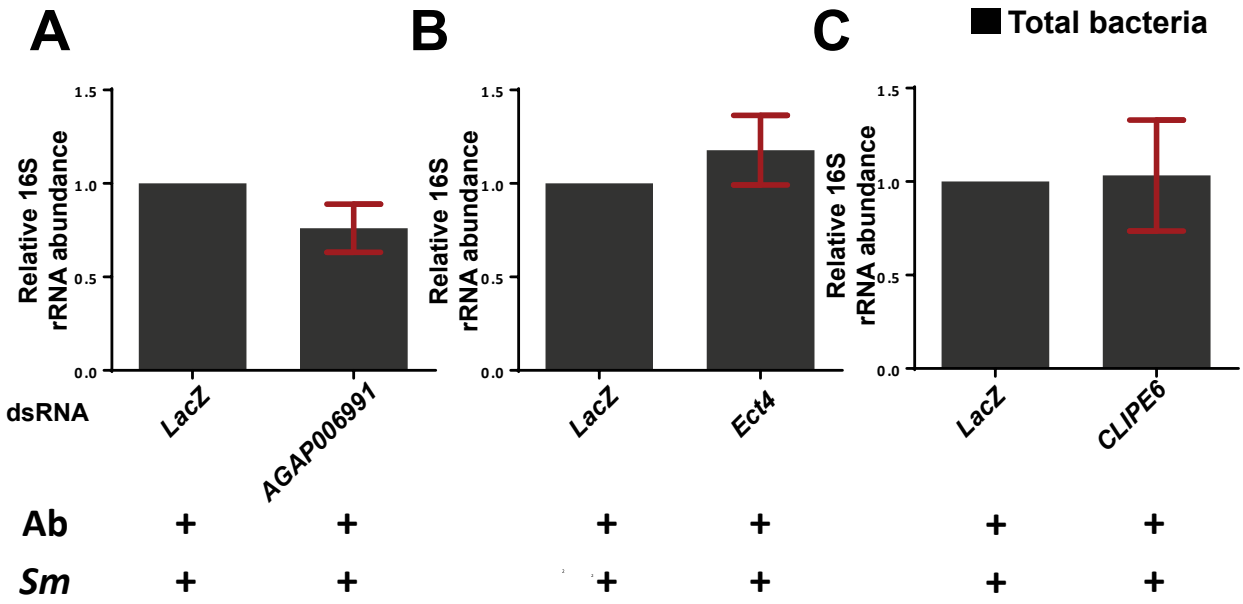


Figure 7.1: Modulation of *S. marcescens* levels following RNAi-mediated silencing of candidate genes. Antibiotic treated mosquitoes were treated with dsRNA targeting the transcript of a gene of interest or the *dsLacZ* control, as indicated below each bar, at the day of emergence. Mosquitoes were orally infected 5 days later with *S. marcescens* (indicated by *Ab+Sm+*), bacteria-fed mosquitoes were selected at day 2 post infection and, at day 5 post infection, bacterial load in guts of mosquitoes corresponding to each dsRNA treatment was determined. Total RNA was extracted from the guts of surface sterilized mosquitoes and used for cDNA synthesis. The respective cDNA was used as template in a qRT-PCR with broad-range bacterial 16S primers, while *AgS7* primers were used as an endogenous control. For each independent infection, qRT-PCR was performed at least in duplicate. **A-C:** The effect of RNAi-mediated silencing of AGAP006991 (panel A), *Ect4* (panel B) and *CLIPE6* (panel C) on modulating *S. marcescens* levels in orally infected mosquitoes was assayed. The relative bacterial abundance was determined in each case over a single independent infection. Bars indicate the average ±SEM of the determined relative bacterial abundance for each technical replicate, normalized to the *dsLacZ* treated control. Significant differences between the relative bacterial abundance corresponding to each dsRNA treatment and the respective *dsLacZ* control were assessed in a non-parametric Mann-Whitney test and were designated as non-significant, with a p-value of 0.3143 for AGAP006991 knockdown, 0.8857 for *Ect4* knockdown and 0.745 for *CLIPE6* knockdown.

The lack of significant changes in bacterial load following silencing of AGAP006991, *Ect4* and *CLIPE6* in *S. marcescens* infected mosquitoes is not conclusive, as only one independent infection assay was carried out in each case and the silencing efficiency of the respective dsRNA treatment was not determined. For *CLIPE6*, no effect of its silencing was observed in *S. marcescens* infected mosquitoes in a separate project (Chen, Y and Christophides, GK, Unpublished). Nevertheless, these data indicate that the contribution of transcriptionally regulated genes to the *S. marcescens* infection outcome might be limited, possibly due to transcriptional programmes inducing expression of factors with redundant function.

7.5 RNAi-mediated silencing of candidate genes associated with the *S. marcescens* infection outcome increases bacterial load in orally infected mosquitoes

The effect of silencing candidate genes associated with the outcome of *S. marcescens* infection was examined in shaping the *S. marcescens* infection outcome, over 2 independent infections in each case. Silencing of AGAP004375, encoding a ricin B lectin domain (Figure 7.2A), AGAP010012, encoding a leucine-rich repeat domain (Figure 7.2B), *LRIM15* (Figure 7.2C) or AGAP002492, encoding a DNA-binding domain (Figure 7.2D), resulted in moderate increases in bacterial load, normalized to the respective *dsLacZ* treated controls, which did not reach statistical significance in a t-test against zero. The observed fold-change increases were 2.1-fold for AGAP004375 silencing, 9.4-fold for AGAP010012 silencing, 4.6-fold for *LRIM15* silencing and 2.4-fold for AGAP002492 silencing. In all cases, the observed effect was consistent over 2 independent infections and might have reached statistical significance given a higher number of replicate infections, as would be required by the robustness of the adopted statistical approach. Interestingly, the effect of AGAP002492 silencing, showed a marginally non-significant 0.0736 p-value. This consistent effect indicates the involvement of this DNA-binding gene in responses that shape the *S. marcescens* infection outcome.

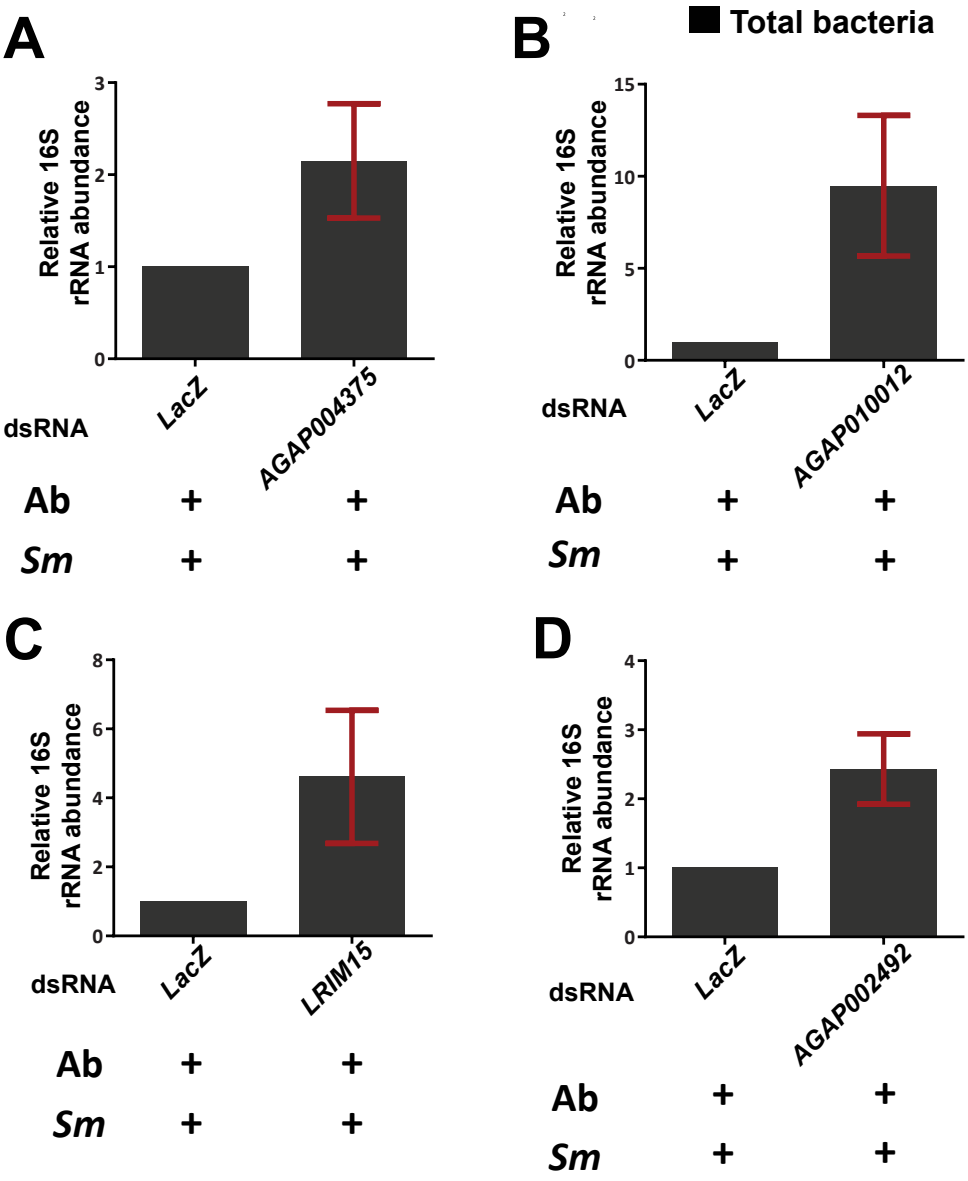


Figure 7.2: Modulation of *S. marcescens* levels following RNAi-mediated silencing of AGAP004375, AGAP010012, *LRIM15* and AGAP002492 in orally infected mosquitoes. Antibiotic treated mosquitoes were treated with dsRNA targeting the AGAP004375 (panel A), AGAP010012 (panel B), *LRIM15* (panel C) or AGAP002492 (panel D) transcript or with the *dsLacZ* control and, 5 days later, were orally infected with *S. marcescens*. The relative bacterial abundance in the gut of mosquitoes treated with the respective dsRNA was determined 5 days post infection, over 2 independent infections in each case, and normalized to the respective *dsLacZ* treated control. Bars indicate fold-change differences ±SEM for each dsRNA treatment. Significant differences were assessed using the log2-transformed fold-change values for each dsRNA treatment in a t-test against zero, where zero corresponds to no difference from *dsLacZ* treatment, deemed as non-significant in all cases, with a p-value of 0.2093 for AGAP004375 knockdown, 0.4249 for AGAP010012 knockdown, 0.5198 for *LRIM15* knockdown and 0.0736 for AGAP002492 knockdown.

7.6 *FN3D1-3* silencing precipitously increases bacterial load in the gut of *S. marcescens* infected mosquitoes

To assess the influence of FN3D1-3 in influencing the outcome of *S. marcescens* infection, modulation of bacterial load was determined following silencing of the respective *FN3D*, as well as *PGRPLC*, compared to *dsLacZ* treatment, in orally infected mosquitoes. Bacterial load was quantified using qRT-PCR with broad-range bacterial 16S primers but also *Serratia*-specific 16S primers, targeting 16S rRNA transcript sequences aligning with *Serratia* reference sequences. The normalized fold-change increases in bacterial load compared to the *dsLacZ* treated controls were determined over 11 independent infections for *PGRPLC* silencing (Figure 7.3A), 7 independent infections for *FN3D1* silencing (Figure 7.3B), 7 independent infections for *FN3D2* silencing (Figure 7.3C) and 8 independent infections for *FN3D3* silencing (Figure 7.3D). The respective log2-transformed fold-change values, for broad-range or *Serratia*-specific 16S primers, were used in a t-test against zero, where zero corresponds to no difference from *dsLacZ* treatment. In all cases, the fold-change increases were highly significant, suggesting that silencing of any of *FN3D1-3* or *PGRPLC* significantly increases bacterial load in the gut of *S. marcescens* infected mosquitoes compared to *dsLacZ* treated controls.

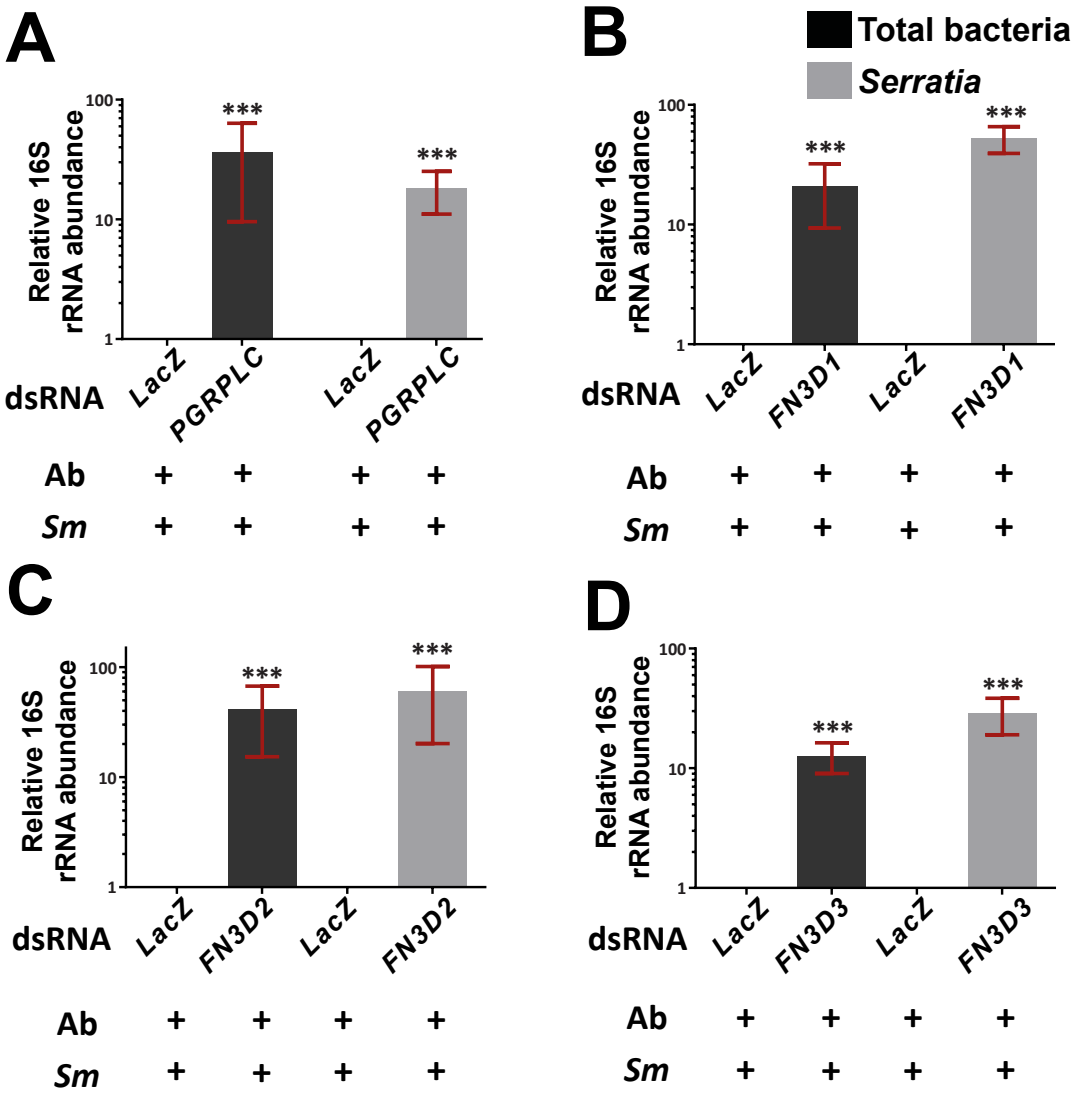


Figure 7.3: Silencing of *FN3D1-3*, associated with the outcome of *S. marcescens* infection, increases *S. marcescens* levels in the gut of orally infected mosquitoes. Antibiotic treated mosquitoes were treated with *PGRPLC* or *FN3D1-3* dsRNA or the *dsLacZ* control, as indicated below each bar, at the day of emergence. Mosquitoes were orally infected 5 days later with *S. marcescens* (indicated by *Ab+Sm+*), bacteria-fed mosquitoes were selected at day 2 post infection and, at day 5 post infection, bacterial load in the mosquito gut for each dsRNA treatment was determined. Total RNA was extracted from the guts of surface sterilized mosquitoes and used for cDNA synthesis. The respective cDNA was used as template in a qRT-PCR with broad-range or *Serratia*-specific bacterial 16S primers, while *AgS7* primers were used as an endogenous control. For each independent infection, qRT-PCR was performed at least in duplicate. The effect of *PGRPLC* (panel A), *FN3D1* (panel B), *FN3D2* (panel C) and *FN3D3* (panel D) knockdown was assessed. Relative bacterial abundance was determined in each case over 11, 7, 7 or 8 independent infections for *PGRPLC*, *FN3D1, FN3D2* or *FN3D3* knockdown, respectively, and normalized in each case to the corresponding *dsLacZ* treated control. Bars indicate fold-change differences ±SEM for each dsRNA treatment. Significant differences were assessed using the log2-transformed fold-change values for each dsRNA treatment in a t-test against zero, where zero corresponds to no difference from dsRNA treatment. Asterisks indicate significance, with a p-value of <0.0005.

Silencing of *PGRPLC* resulted in an average 36.4-fold increase of bacterial load using broad-range 16S primers and an average 18.1-fold increase in bacterial load using *Serratia*-specific primers (Figure 7.3A). The respective fold-change increases using broad-range or *Serratia*-specific 16S primers were 20.7-fold and 52.4-fold for *FN3D1* silencing (Figure 7.3B), 41.2 and 60.7-fold for *FN3D2* silencing (Figure 7.3C) and 12.7 and 28.7-fold for *FN3D3* silencing (Figure 7.3D). Interestingly, in *PGRPLC* knockdown mosquitoes, the average fold-change increase was higher using broad-range 16S primers while, in *FN3D1-3* knockdown mosquitoes, the average fold-change increase was consistently higher using *Serratia*-specific primers. Furthermore, for both broad-range and *Serratia*-specific 16S primers, the average fold-change increase following *FN3D1-3* silencing was the lowest in *FN3D3* knockdown and highest in *FN3D2* knockdown mosquitoes, with *FN3D1* knockdown mosquitoes found in the middle range for both sets of primers. Although no conclusive inferences can be made based on these observations, there is an indication that differences between the 16S and *Serratia*-specific primers might be a result of residual bacteria resistant to antibiotic treatment. *PGRPLC* silencing seems to result in increases of these residual bacteria, at least to a higher degree than *FN3D1-3* silencing, indicating a functional difference between PGRPLC and FN3D1-3. Furthermore, FN3D2 seems to have the most potent and FN3D3 the least potent antibacterial activity among FN3Ds. It has to be noted though, that the observed fold-change differences between *FN3D1-3* and *PGRPLC* silencing show sizable margins of error and are thus statistically non-significant. Furthermore, broad-range and *Serratia*-specific 16S primers might have differential binding efficacies which may depend on the used primer concentrations and template abundance. Finally, differences in the *FN3D1-3* or *PGRPLC* silencing efficiency might account for the observed fold-change differences between dsRNA treatments using broad-range or *Serratia*-specific 16S primers.

Taken together, these data provide compelling evidence that FN3D1-3 exert an antibacterial effect that shapes the outcome of *S. marcescens* infection. Silencing of *FN3D1-2* over 7 independent infections and *FN3D3* over 8 independent infections resulted in highly significant increases in bacterial load in the gut of *S. marcescens* infected mosquitoes compared to *dsLacZ* treated controls, which were comparable with the exerted effect of *PGRPLC* silencing.

7.7 Efficiency of *FN3D1-3* and *PGRPLC* RNAi-mediated silencing

To assess the efficiency of dsRNA treatment in silencing the respective *FN3D1-3* transcripts in the mosquito gut, the expression of the respective transcript was determined at the day of infection, 5 days post dsRNA treatment, in mosquitoes treated with the respective dsRNA, compared to *dsLacZ* treated

controls. Expression was determined using qRT-PCR, with primers targeting a part of the respective *FN3D1-3* transcript that is not targeted by the dsRNA, to avoid detection of residual dsRNA. The efficiency of *PGRPLC* silencing was also determined as an additional control. In each case, total RNA was extracted from a pool of guts from 8-10 dsRNA treated mosquitoes and used to synthesize cDNA representing transcripts present in the respective guts. Abundance of the respective transcript was normalized to the *AgS7* endogenous control. The relative abundance of the respective transcript normalized to the respective *dsLacZ* treated control was determined for *PGRPLC*, *FN3D1*, *FN3D2* and *FN3D3* dsRNA treated mosquitoes in 7, 12, 12 and 13 mosquito gut pools, respectively, over 3 independent assays (Figure 7.4A).

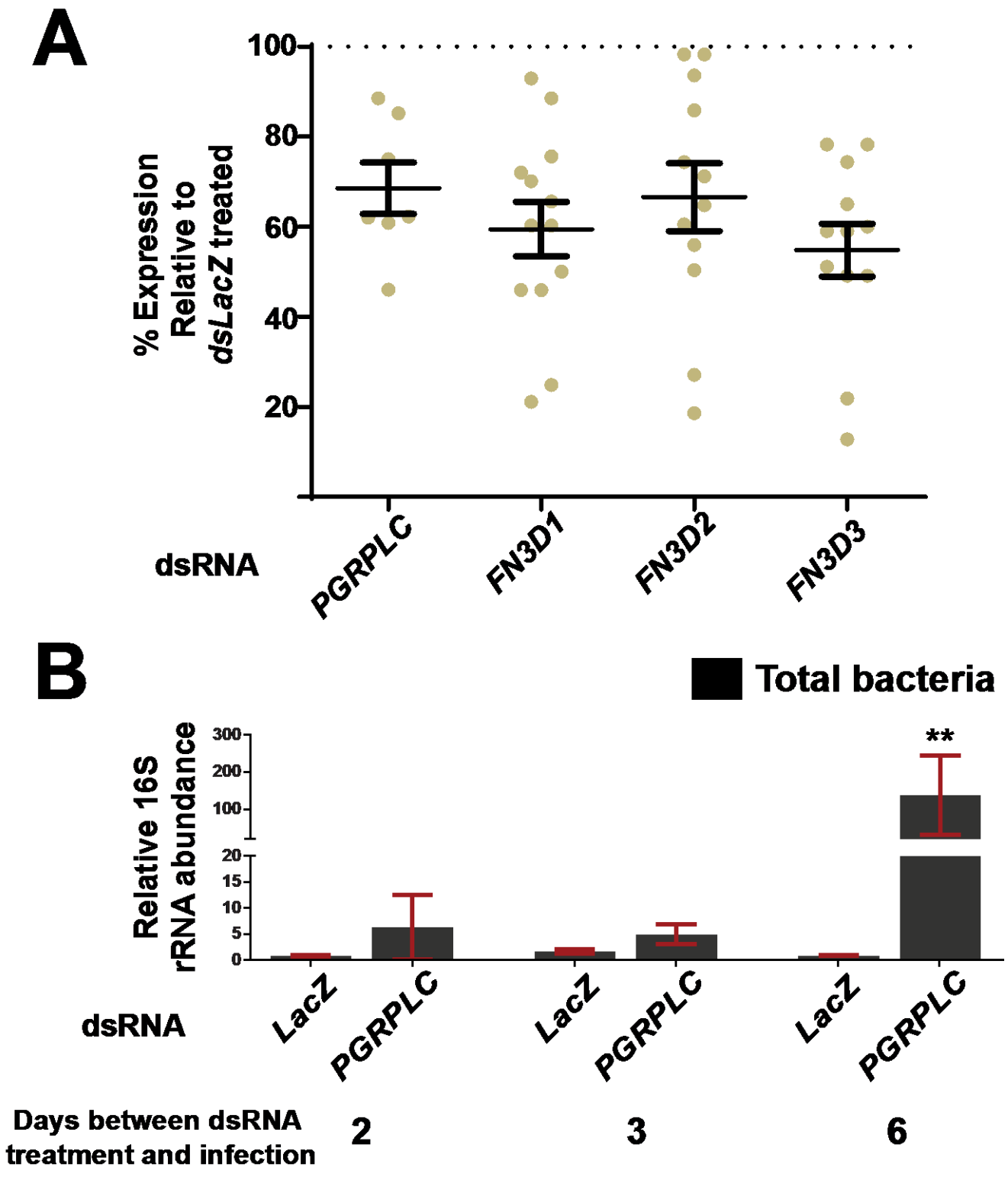

A
100
80
60
40
20
% Expression Relative to dsLacZ treated
dsRNA
PGRPLC
FN3D1
FN3D2
FN3D3
B
Total bacteria
300
200
100
20
15
10
5
0
Relative 16S rRNA abundance
**
dsRNA
LacZ
PGRPLC
LacZ
PGRPLC
LacZ
PGRPLC
Days between dsRNA treatment and infection
2
3
6

Figure 7.4: Efficiency of *PGRPLC* and *FN3D1-3* RNAi-mediated silencing. A: Antibiotic treated mosquitoes were microinjected with *PGRPLC* or *FN3D1-3* dsRNA or the *dsLacZ* control at the day of emergence. The guts of 8-10 mosquitoes were dissected 5 days later and used for total RNA extraction and cDNA synthesis. The expression levels of *PGRPLC*, *FN3D1*, *FN3D2* or *FN3D3* in mosquitoes treated with the respective dsRNA or the *dsLacZ* control were determined using primers targeting the respective transcript in a region not targeted by the dsRNA. Expression of the housekeeping transcript *AgS7* was used as an endogenous control and, in each case, qRT-PCR was performed at least in duplicate. Relative abundance for each transcript was determined for 7 pools of *PGRPLC*, 13 pools of *FN3D1*, 12 pools of *FN3D2* and 12 pools of *FN3D3* knockdown mosquitoes, over 3 independent assays, further normalized to the relative abundance in the respective *dsLacZ* treated control. In each case, the % expression level of the respective transcript for each pool is plotted, along with the determined average ±SEM using all pools. The average expression levels, compared to the *dsLacZ* treated control, were determined at 68.57%, 59.49%, 66.59% and 54.85% following *PGRPLC*, *FN3D1*, *FN3D2* and *FN3D3* silencing, respectively. **B:** Antibiotic treated mosquitoes were microinjected with *PGRPLC* dsRNA or the *dsLacZ* control and were subsequently orally infected with *S. marcescens* 2, 3 or 6 days post dsRNA treatment. In all cases, antibiotic treatment lasted for at least 5 days. Bacterial load was determined in the guts of mosquitoes corresponding to each dsRNA treatment 5 days post infection using broad-range 16S bacterial primers. The average ±SEM of the relative bacterial abundance for the pooled technical replicates can be seen for *PGRPLC* or *LacZ* dsRNA treatment in each case. Normalized to the *dsLacZ* relative bacterial abundance, *PGRPLC* silencing showed a 7.96, 2.9 and 160.3-fold increase in bacterial load when dsRNA treatment was performed 2, 3 or 6 days before infection, respectively. Significant differences were assessed in a non-parametric Mann-Whitney test between the *PGRPLC* knockdown relative bacterial abundance values of the performed technical replicates and the respective relative bacterial abundance of the *dsLacZ* treated control. The calculated p-values were 0.6, 0.21 and 0.0022 for dsRNA treatment 2, 3 or 6 days before infection, respectively. Statistical significance for the dsRNA treatment at 6 days before infection is indicated by 2 asterisks.

In all cases, the expression level of the respective transcript was reduced in the guts of dsRNA treated mosquitoes, compared to the *dsLacZ* treated control, showing an average expression of 68.6%, 60.3%, 66.6% and 54.8% in *PGRPLC*, *FN3D1*, *FN3D2* and *FN3D3* dsRNA treated mosquitoes, respectively (Figure 7.4A). Silencing efficiency in individual dsRNA treated pools was relatively variable, with some outliers showing either a >80% expression level despite dsRNA treatment or highly efficient silencing, with <20% expression levels. Such variable levels of variation might be due to stochastic differences between individual assays or the nature of the implemented methodology of manually microinjecting individual mosquitoes.

The efficiency of *PGPRLC* silencing shown here is consistent with previous reports showing relative *PGRPLC* expression levels from 40% to 80% following *PGRPLC* knockdown (Dong et al., 2011; Meister et al., 2009). The relatively moderate silencing efficiency for *PGRPLC* but also *FN3D1-3* indicates the difficulty in achieving highly efficacious silencing of genes expressed in the midgut, compared to fat body expressed genes, where dsRNA is initially microinjected. Previously reported silencing efficiency for midgut expressed genes indicated highly efficient silencing, with <20% relative expression, for *IMPer* and *NOS* at 24-36 hours post dsRNA treatment and for *Duox* at 3-4 days post dsRNA treatment (Kumar et al., 2010). Interestingly, silencing efficiency was less efficacious in antibiotic treated mosquitoes (Kumar et al., 2010). Similarly, efficient silencing of *dsMC1* midgut expression was reported 2 days post dsRNA treatment (Goncalves et al., 2012), and at 36 hours post dsRNA treatment for *AgESP* midgut expression (Rodrigues et al., 2012). Therefore, it is possible that, as the *FN3D1-3* silencing efficiency was examined 5 days post dsRNA treatment, more efficacious silencing might have been achieved at 1-3 days post dsRNA treatment, a possibility that remains to be confirmed.

Nevertheless, when *S. marcescens* infection was conducted 2-3 days post dsRNA treatment following *PGRPLC* silencing, a trend for lower fold-change bacterial load increases was observed, compared to experiments in which *PGRPLC* silencing was carried out 6 days prior to infection (Figure 7.4B). It remains to be determined whether the relative PGRPLC or FN3D1-3 protein expression, which is the main phenotypic driver, follows the dynamics of transcript expression or whether a lag time of 1-3 days is required for clearance of the respective protein product for a potent observed phenotype.

7.8 Modulation of gene expression following *PGRPLC* or *FN3D1-3* silencing in *S. marcescens* infected mosquitoes

PGRPLC activation has been previously shown to modulate the expression of antimicrobial peptides via the IMD/REL2 pathway, including the cecropin *CEC1* and the gambicin *GAM1* in response to *E. coli* bacterial challenge in mosquito cell lines (Lin et al., 2007). Furthermore, the expression of *CEC1* but also *CEC3*, *GAM1*, *CLIPB14* and *LRIM1* in mosquito cell lines has also been shown to be mediated by the IMD/REL2 pathway (Meister et al., 2005), while *CEC1* and *DEF1* expression has been shown to be controlled by REL2 in *An. gambiae* adults (Garver et al., 2009). At the same time, though, induction of *CEC1* and *DEF1* in response to *E. coli* bacterial injection in the mosquito haemolymph is not controlled by PGRPLC (Meister et al., 2009). Midgut expression of antimicrobial peptides has been previously shown to be triggered by the Imd pathway following oral bacterial infection in *Drosophila* (Vodovar et al., 2005), with the expression of several antimicrobial peptides, including diptericin and cecropin, controlled by the Imd pathway and modulated in the midgut by the homeobox transcription factor Caudal (Ryu et al., 2008).

Upregulation of antimicrobial peptides was detected following oral *S. marcescens* infection, including the upregulation of *DEF1* but also *LYSC2*, the latter showing the most pronounced differential expression in core transcriptional responses to *S. marcescens* infection. Thus, one examined possibility was that the expression of these antimicrobial peptides is mediated by FN3D1-3 or PGRPLC, through the IMD/REL2 pathway.

LYSC2 expression was tracked in *PGRPLC* knockdown mosquitoes orally infected with *S. marcescens*, compared to *dsLacZ* treated controls. *PGRPLC* silencing resulted in a significant 1.75-fold downregulation of *LYSC2* expression in orally infected mosquitoes compared to *dsLacZ* treated controls, as determined over 5 independent infections (Figure 7.5A). These data suggest that the upregulation of *LYSC2* following oral *S. marcescens* infection is at least partly controlled by *PGRPLC*.

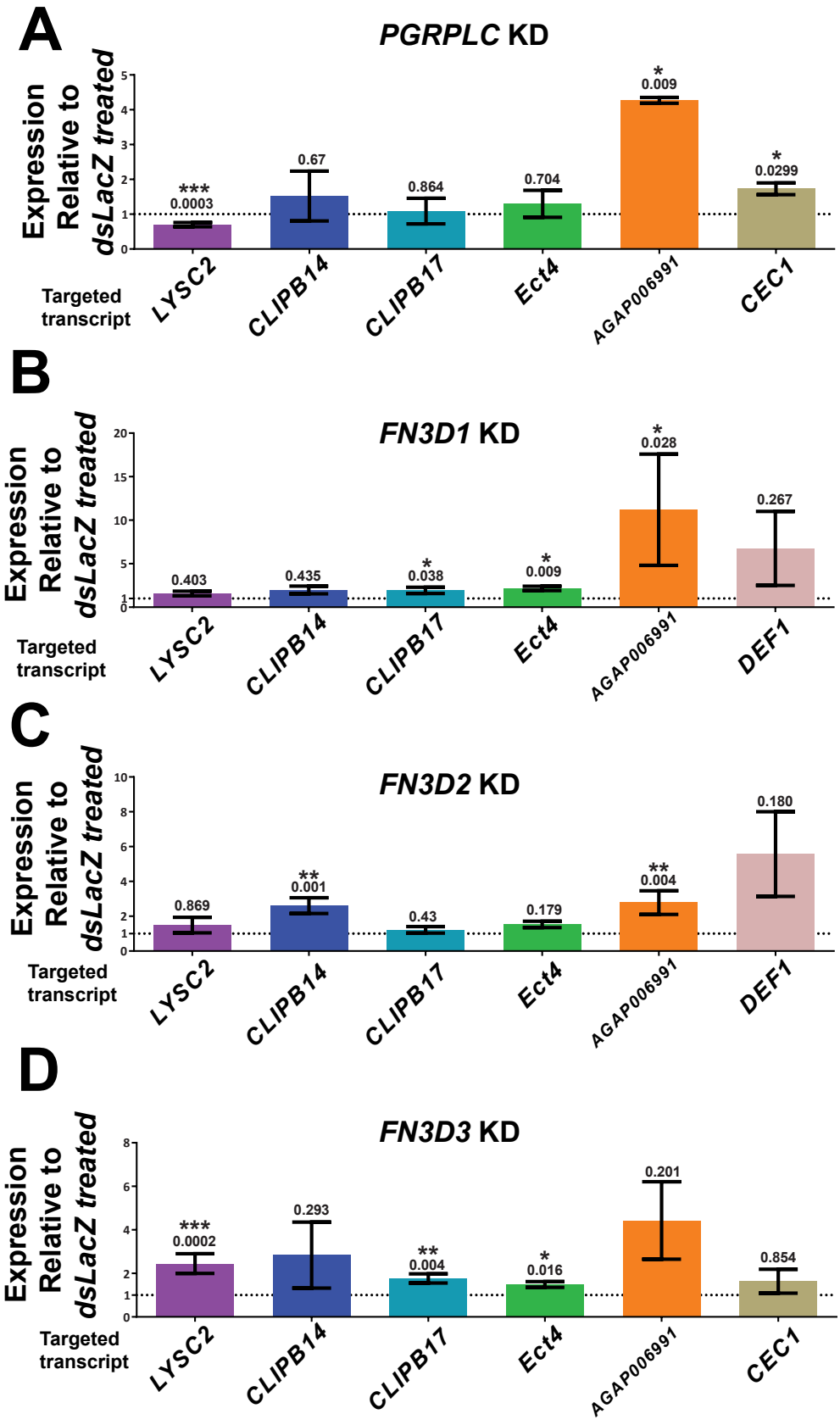

A
PGRPLC KD
Expression Relative to dsLacZ treated
***
0.0003
0.67
0.864
0.704
*
0.009
*
0.0299
Targeted transcript
LYSC2
CLIPB14
CLIPB17
Ect4
AGAP006991
CEC1
B
FN3D1 KD
Expression Relative to dsLacZ treated
0.403
0.435
*
0.038
*
0.009
*
0.028
0.267
Targeted transcript
LYSC2
CLIPB14
CLIPB17
Ect4
AGAP006991
DEF1
C
FN3D2 KD
Expression Relative to dsLacZ treated
0.869
**
0.001
0.43
0.179
**
0.004
0.180
Targeted transcript
LYSC2
CLIPB14
CLIPB17
Ect4
AGAP006991
DEF1
D
FN3D3 KD
Expression Relative to dsLacZ treated
***
0.0002
0.293
**
0.004
*
0.016
0.201
0.854
Targeted transcript
LYSC2
CLIPB14
CLIPB17
Ect4
AGAP006991
CEC1

Figure 7.5: Modulation of gene expression following *PGRPLC* or *FN3D1-3* silencing in *S. marcescens* infected mosquitoes. Antibiotic treated mosquitoes were treated with *PGRPLC*, *FN3D1-3* or *LacZ* dsRNA at the day of emergence and, 5 days later, were orally infected with *S. marcescens*. Bacteria-fed mosquitoes were selected at day 2 post infection and, at day 5 post infection, total RNA was extracted from dissected guts for each dsRNA treated pool and used for cDNA synthesis. The expression of *LYSC2*, *CLIPB14*, *CLIPB17*, *Ect4*, AGAP006991, *CEC1* and *DEF1* was determined by qRT-PCR using primers targeting the respective transcript while *AgS7* primers were used as an endogenous control. Relative abundance for each transcript in *PGRPLC* or *FN3D1-3* dsRNA treated pools was normalized to the respective relative abundance in the respective *dsLacZ* treated control. Bars indicate the average ±SEM fold-change difference for each transcript, as indicated below each bar, in *PGRPLC* (panel A), *FN3D1* (panel B), *FN3D2* (panel C) or *FN3D3* (panel D) knockdown pools, relative to the respective *dsLacZ* treated pool, as determined over 4-5 independent infections in each case. Significant differences were assessed by a t-test against zero, where zero corresponds to no difference from the *dsLacZ* treated pool, using the derived log2-transformed fold-change values. The obtained p-values are indicated over each bar. Asterisks indicate significance, with a p-value <0.05, <0.005 or <0.0005 corresponding to 1-3 asterisks, respectively.

To identify additional genes whose differential expression following oral *S. marcescens* infection may be controlled by PGRPLC, the midgut expression in orally infected mosquitoes was determined for *CLIPB14*, *CLIPB17*, *Ect4*, AGAP006991 and *CEC1*, both in *PGRPLC* and *LacZ* dsRNA treated mosquitoes. The assayed transcripts, except *CEC1*, have been shown to be upregulated following oral *S. marcescens* infection, while *CEC1* expression has been previously reported to be controlled by the IMD/REL2 pathway (Meister et al., 2005). For each transcript, relative expression in *PGRPLC* knockdown mosquitoes was normalized to the respective expression in the *dsLacZ* treated controls, over 4 independent infections, and the log2-transformed fold-change values were used to assess statistical significance in a t-test against zero, where zero corresponds to no difference from the *dsLacZ* treated control.

As can be seen in Figure 7.5A, *CLIPB14*, *CLIB17* and *Ect4* expression showed a modest 1.52, 1.09 and 1.29-fold upregulation, respectively, in *PGRPLC* dsRNA treated mosquitoes compared to the *dsLacZ* treated controls, which did not reach statistical significance. It is possible that PGRPLC may partly control the expression of these genes in orally infected mosquitoes although no firm conclusions can be reached regarding the PGRPLC involvement based on these data. Interestingly, the chitin-binding gene AGAP006991 showed a significant 4.27-fold upregulation in *PGRPLC* dsRNA treated mosquitoes, compared to the respective *dsLacZ* treated control. This pronounced upregulation suggests that the expression of this chitin-binding gene following *S. marcescens* infection is not controlled by PGRPLC but also that the increased bacterial load due to *PGRPLC* silencing might further trigger AGAP006991 expression, most likely in an IMD/REL2-independent manner. Finally, *CEC1* expression showed a significant 1.73-fold upregulation following *PGRPLC* silencing compared to the *dsLacZ* treated controls, suggesting that *CEC1* expression following *S. marcescens* infection is not dependent on PGRPLC. These data are consistent with a previous report showing that *CEC1* expression following injection with the Gram-negative bacterium *E. coli* is not mediated by PGRPLC (Meister et al., 2009).

The differential expression of *LYSC2*, *CLIPB14*, *CLIPB17*, *Ect4*, AGAP006991, but also of *DEF1*, whose expression has been reported to be controlled by the IMD/REL2 pathway (Garver et al., 2009), was tracked in *FN3D1* knockdown mosquitoes, compared to *dsLacZ* treated controls, in *S. marcescens* infected mosquitoes (Figure 7.5B). *LYSC2* and *CLIPB14* expression showed a non-significant 1.57 and 1.98-fold upregulation, respectively, while the levels of *CLIPB17*, *Ect4* and AGAP006991 showed a 1.92, 2.16 and 11.19-fold upregulation, respectively, that reached statistical significance. These data indicate differences between the effect of *PGRPLC* and *FN3D1* silencing in modulation of gene expression following oral *S. marcescens* infection. In contrast to the *LYSC2* downregulation following *PGRPLC* silencing, a non-

significant *LYSC2* upregulation suggests that PGRPLC-mediated responses are still active following *FN3D1* silencing. The same indication can be derived by the significant upregulation of *CLIPB17* and *Ect4* following *FN3D1* silencing, compared to marginal, non-significant upregulation following *PGRPLC* silencing. Furthermore, the AGAP006991 upregulation following *FN3D1* silencing suggests that its induction is not controlled by FN3D1. The same conclusion can be made for the non-significant 6.75-fold upregulation of *DEF1* following *FN3D1* silencing.

FN3D2 silencing in *S. marcescens* infected mosquitoes led to a non-significant 1.49-fold upregulation of *LYSC2*, compared to *dsLacZ* treated controls (Figure 7.5C), suggesting that PGRPLC-mediated responses are also active in *FN3D2* knockdown mosquitoes. Interestingly, differences were also observed compared to modulation of gene expression following *FN3D1* silencing in orally infected mosquitoes. *FN3D2* silencing led to a significant 2.6-fold increase in *CLIPB14* expression, while *CLIPB17* and *Ect4* expression showed non-significant 1.21 and 1.52-fold increases, respectively (Figure 7.5C), while in *FN3D1* knockdown mosquitoes the inverse pattern was observed, with a moderate non-significant increase in *CLIPB14* expression and significant upregulation of *CLIPB17* and *Ect4* (Figure 7.5B). *FN3D2* knockdown also led to a significant 11.19-fold AGAP006991 and a non-significant 6.75-fold *DEF1* upregulation (Figure 7.5C), suggesting that their respective differential expression is not controlled by FN3D2.

Finally, *FN3D3* silencing in *S. marcescens* infected mosquitoes triggered a significant 2.45-fold upregulation in *LYSC2* expression compared to *dsLacZ* treated controls (Figure 7.5D), indicating that IMD/REL2-mediated responses triggered by PGRPLC remain active in *FN3D3* knockdown mosquitoes, while they are likely further triggered due to increased bacterial load following *FN3D3* silencing. Differential expression of *CLIPB14*, *CLIPB17* and *Ect4* in *FN3D3* knockdown mosquitoes mirrored the respective differential expression following *FN3D1* silencing, with a non-significant 2.84-fold *CLIPB14* and a significant 1.76-fold *CLIPB17* and 1.49-fold *Ect4* upregulation, compared to *dsLacZ* treated controls (Figure 7.5D). AGAP006991 and *CEC1* also showed non-significant 4.43 and 1.64-fold upregulation, respectively, in *FN3D3* knockdown mosquitoes (Figure 7.5D), suggesting that their expression is not mediated by FN3D3.

These data indicate the existence of differences in the expression profile of orally infected mosquitoes following *PGRPLC* or *FN3D1-3* silencing, suggesting that the FN3D antibacterial effect is discrete from PGRPLC-mediated responses. These differences reached statistical significance in the modulation of *LYSC2* expression following *PGRPLC* and *FN3D3* silencing, with *LYSC2* expression showing significant downregulation and upregulation, respectively, compared to *dsLacZ* treated controls.

The low fold-change and in some cases non-significant differential expression shown following *PGRPLC* or *FN3D1-3* knockdown can be ascribed to variation in differential expression, also observed by DNA microarrays, most likely related to genetic variation between independent infections but also between highly, lowly or non-infected mosquitoes. Another complicating factor is the partially efficient silencing of *PGRPLC* or *FN3D1-3*. The robustness of the current analysis could have been enhanced by tracking gene expression in mosquitoes stemming from the same independent infection, thus limiting genetic variation, but also by discriminating between highly, lowly or non-infected mosquitoes. Another possibility is the use of mosquito cell lines for more robust tracking of IMD/REL2-mediated responses, as has been previously shown (Dong et al., 2012; Lin et al., 2007; Meister et al., 2005).

Nevertheless, the differential expression shown here in orally infected mosquitoes following *PGRPLC* or *FN3D1-3* silencing, especially for AGAP006991 and *DEF1* expression, which showed robust and consistent upregulation, further strengthens the validity of the DNA microarray approach in detecting transcriptional responses following oral *S. marcescens* infection.

7.9 FN3D1-3 modulate the presence of commensal *Serratia* populations in mosquitoes retaining their natural gut microbiota

As *FN3D1-3* silencing precipitously increased the abundance of *S. marcescens* following oral infection, the possibility that FN3D1-3 may also modulate the abundance of commensal *Serratia* naturally found in the mosquito gut emerged. To further investigate this possibility, mosquitoes from the laboratory *N'gousso* colony used in this study were treated with the respective *FN3D1-3* dsRNA, *PGRPLC* dsRNA or the *dsLacZ* control. These mosquitoes were neither treated with antibiotics nor were orally infected with *S. marcescens*, thus they retained their natural gut microbial communities.

The abundance of *Serratia* communities in the gut of *FN3D1-3*, *PGRPLC* or *LacZ* dsRNA treated mosquitoes was determined 5 days post dsRNA treatment in a qRT-PCR with cDNA representing the host-bacterial transcripts in the mosquito gut used as template and *Serratia*-specific primers targeting the bacterial 16S rRNA transcript. *PGRPLC* silencing led to a significant 5.3-fold increase in *Serratia* abundance, as determined over 2 independent infections (Figure 7.6A). These data are consistent with a previously reported increase of bacterial load in the gut of mosquitoes retaining their natural gut microbiota following *PGRPLC* silencing (Meister et al., 2009). Similarly, silencing of *FN3D1* (Figure 7.6B), *FN3D2* (Figure 7.6C) or *FN3D3* (Figure 7.6D) led to significant 3.8, 7.9 and 4.6-fold increases in *Serratia* load, respectively,

compared to *dsLacZ* treated mosquitoes also retaining their natural gut microbiota, over 4 independent infections in each case.

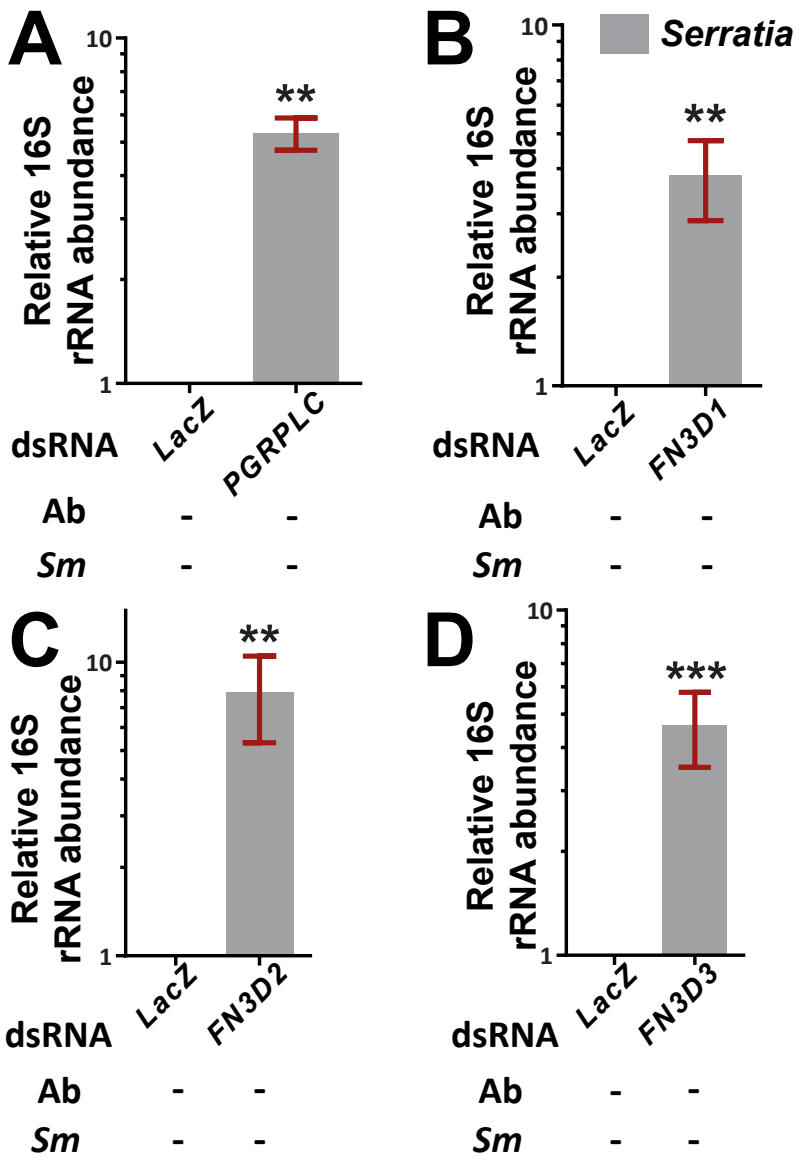


Figure 7.6: *PGRPLC* or *FN3D1-3* silencing modulates the abundance of commensal *Serratia* in the gut of mosquitoes retaining their natural gut microbiota. Mosquitoes retaining their natural gut microbiota, without antibiotic treatment or infection with *S. marcescens*, indicated as *Ab-Sm-*, were treated with *PGRPLC*, *FN3D1-3* or *LacZ* dsRNA and, 5 days post dsRNA treatment, total RNA was extracted from mosquito guts corresponding to each dsRNA treatment and was used for cDNA synthesis. *Serratia* abundance was determined by qRT-PCR with *Serratia*-specific 16S primers, while *AgS7* primers were used as an endogenous control. Relative *Serratia* abundance for each dsRNA treatment was normalized to the respective *dsLacZ* treated control. Bars indicate the average ±SEM fold-change of *Serratia* increase in *PGRPLC* (panel A), *FN3D1* (panel B), *FN3D2* (panel C) or *FN3D3* (panel D) knockdown mosquitoes, relative to the respective *dsLacZ* treated control, as determined over 2 independent assays for *PGRPLC* or over 4 independent assays for *FN3D1-3* knockdown mosquitoes. Significant differences were assessed by a t-test against zero, where zero corresponds to no difference from *dsLacZ* treatment, using the log2-transformed fold-change values for each respective dsRNA treatment. Asterisks indicate significance, with a p-value <0.005 or <0.0005 indicated by 2 or 3 asterisks, respectively.

These data confirm the FN3D1-3 antibacterial effect and suggest that FN3D1-3 are involved in constitutive antibacterial responses that shape the mosquito gut microbial community.

Interestingly, when the effect of *FN3D1-3* silencing was examined on the total bacterial load in the gut of mosquitoes retaining their natural gut microbiota, a non-uninform effect was observed both between independent assays but also compared to *Serratia* abundance. Total bacterial load was determined using broad-range bacterial 16S primers in each of the 4 independent assays previously used to determine *Serratia* abundance following *FN3D1-3* silencing (Figure 7.6). In all 4 replicates for *FN3D1* (Figure 7.7A), *FN3D2* (Figure 7.7B) or *FN3D3* (Figure 7.7C) knockdown, *Serratia* abundance showed a consistent increase from about 2-fold to about 8-fold, compared to *dsLacZ* treated mosquitoes. Pooling of these replicates resulted in statistically significant increases of *Serratia* abundance following *FN3D1* (Figure 7.6B), *FN3D2* (Figure 7.6C) or *FN3D3* (Figure 7.6D) silencing.

In the same assays, modulation of total bacterial abundance, as determined by broad-range 16S primers, following *FN3D1* (Figure 7.7A), *FN3D2* (Figure 7.7B) or *FN3D3* (Figure 7.7C) silencing, was not consistent between the 4 replicates for each dsRNA treatment nor was tracking the respective modulation of *Serratia* abundance in the respective assay. In *FN3D1* knockdown mosquitoes (Figure 7.7A), in replicates 1 and 2, total bacteria showed a marginal 1.04 and 1.02-fold decrease, respectively, compared to the *dsLacZ* treated controls, while *Serratia* increased by 8 and 2-fold, respectively. In replicates 3 and 4, though, a different trend was observed, with both total bacteria and *Serratia* showing modest increases, 2.71 and 2.84-fold, respectively, in replicate 3 and 2.5 and 2.29-fold, respectively, in replicate 4. Similarly, in the *FN3D2* knockdown assays (Figure 7.7B), in replicates 1 and 4, total bacteria showed a marginal 1.29 and 1.11-fold increase, respectively, while *Serratia* increased by 8.84 and 1.65-fold, respectively. In replicates 2 and 3, total bacteria showed a modest 2 and 1.57-fold increase, respectively, while *Serratia* increased in the respective assays by 3.42 and 4.87-fold. In the *FN3D3* knockdown assays (Figure 7.7C), in replicate 1, total bacteria decreased by 1.66-fold, in replicate 2 modestly increased by 1.26-fold while in replicate 3 total bacteria showed a marginal 1.02-fold decrease. In the same assays *Serratia* increased by 4.47, 7.59 and 3.40-fold, respectively. In replicate 4, both total bacteria and *Serratia* showed a moderate 1.80 and 2.13-fold increase, respectively. Therefore, in some cases both total bacteria and *Serratia* increased moderately, by 1.5 to 2.5-fold, while in other cases total bacteria showed a marginal increase or decrease while *Serratia* increased by up to 8-fold.

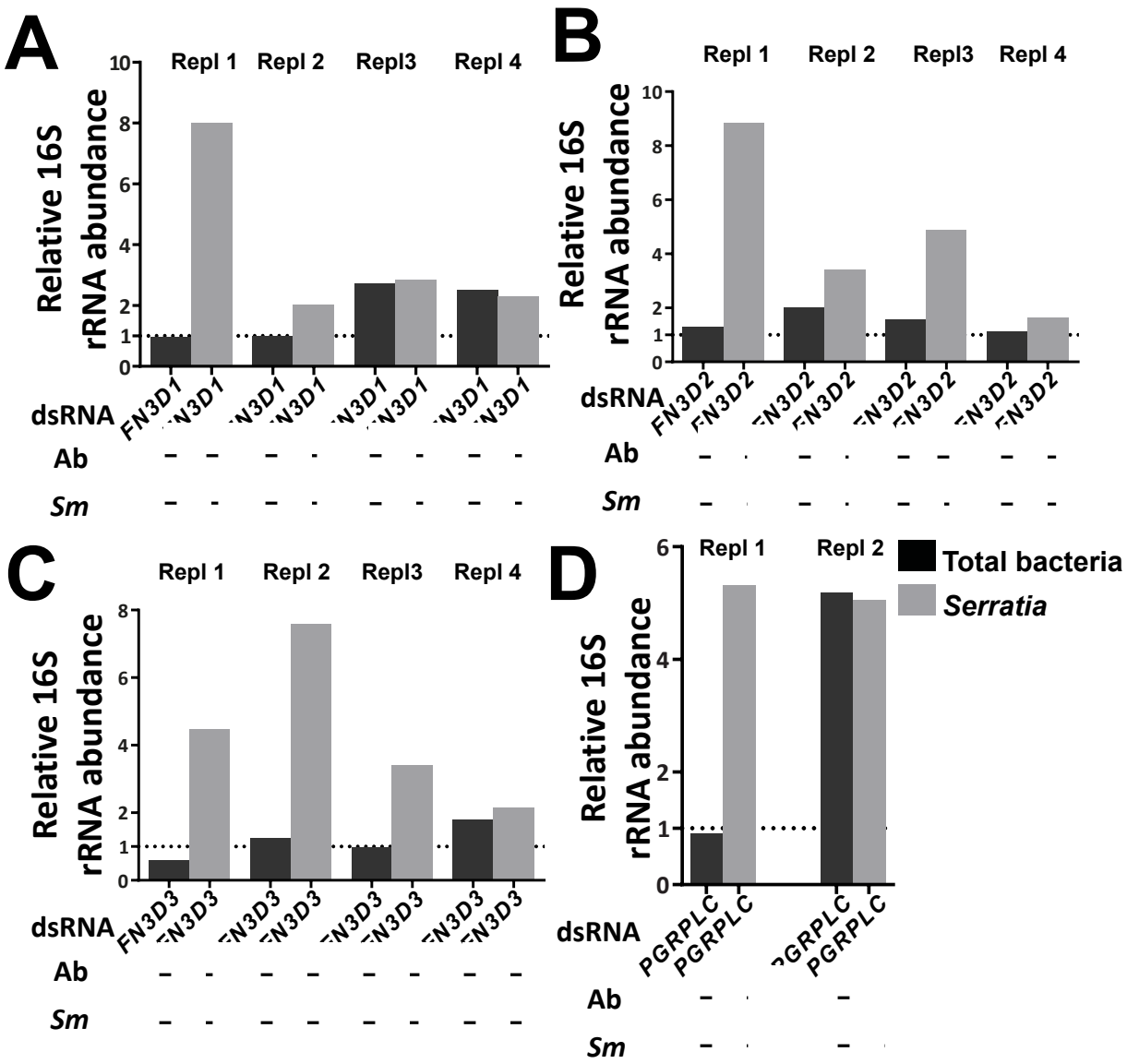


Figure 7.7: *PGRPLC* or *FN3D1-3* silencing non-uniformly influences total bacterial abundance, between independent replicates or compared to the respective *Serratia* abundance, in mosquitoes retaining their natural gut microbiota. Mosquitoes retaining their natural gut microbiota, without antibiotic treatment or infection with *S. marcescens*, indicated as *Ab-Sm-*, were treated with *PGRPLC*, *FN3D1-3* or *LacZ* dsRNA and, 5 days post dsRNA treatment, total RNA was extracted from mosquito guts corresponding to each dsRNA treatment and was used for cDNA synthesis. Abundance of total bacteria or *Serratia* was determined by qRT-PCR and broad-range or *Serratia*-specific 16S rRNA bacterial primers, respectively. The *AgS7* housekeeping gene primers were used as an endogenous control. Relative total bacterial or *Serratia* abundance was normalized for each dsRNA treatment to the respective *dsLacZ* treated control. Bars indicate the fold-change differences of total bacteria and *Serratia*, compared to *dsLacZ* treated mosquitoes, shown for *FN3D1* (panel A), *FN3D2* (panel B), *FN3D3* (panel C) or *PGRPLC* (panel D) knockdown mosquitoes over 4 independent assays for *FN3D1-3* or 2 independent assays for *PGRPLC* knockdown mosquitoes. Total bacteria or *Serratia* fold-change differences are shown individually for each independent infection, designated as replicate (Repl) 1-4 for *FN3D1-3* knockdown or replicate 1-2 for *PGRPLC* knockdown assays, as indicated above the respective bars.

The observed differences between broad-range and *Serratia*-specific 16S bacterial primers are unlikely to be related to a markedly different template-binding efficiency between the primer sets, as both primer sets were previously shown to detect increases in *S. marcescens* levels following oral infection (Figure 7.3). This non-uniform effect of *FN3D1-3* silencing between *Serratia* and total bacteria could be explained by an FN3D antibacterial effect exerted on a subset of the mosquito gut bacterial population. In that case, the initial *Serratia* representation in the mosquito gut could influence the observed FN3D effect on total bacteria. Therefore, if the initial *Serratia* representation was low in the mosquito gut, a specific FN3D antibacterial effect directed against *Serratia* would have little influence on the load of the total bacterial population. At the same time, it would be much easier for *FN3D* silencing to trigger higher fold-change *Serratia* increases, starting from an initial limited *Serratia* population. On the other hand, if *Serratia* dominated the initial mosquito gut microbial population, *FN3D* silencing would more severely influence the total bacterial population. At the same time, though, it would be more difficult for *FN3D* silencing to exert multiple fold-change increases on an initially extensive *Serratia* population, with the initial *Serratia* predominance entailing that total bacterial load should more closely track the *Serratia* population. Indeed, replicates in which *FN3D1-3* knockdown results in marginal modulation of total bacteria but a pronounced increase of *Serratia* abundance or replicates in which both total bacteria and *Serratia* show modest increases in abundance (Figure 7.7) are consistent with the hypothesis of an FN3D1-3 effect on a subset of the mosquito gut bacterial community, which includes *Serratia*. Based on these assumptions, these data raise the possibility that the FN3D antibacterial effect shows specificity against *Serratia* or a subset of, the predominantly Gram-negative, mosquito gut bacterial population.

The effect of *PGRPLC* knockdown on total bacteria was also tracked in the 2 assays used to determine an increased *Serratia* abundance following *PGRPLC* silencing in mosquitoes retaining their natural gut microbiota. As shown in Figure 7.7D, in replicate 1, total bacteria showed a marginal 1.09-fold decrease while *Serratia* increased by 5.31-fold following *PGRPLC* silencing, while in replicate 2 both total bacteria and *Serratia* increased by 5.18 and 5.05-fold, respectively.

The *PGRPLC* knockdown effect differs from the observed *FN3D1-3* effect in which *Serratia* increases were generally more pronounced when total bacteria showed marginal changes, compared to replicates where both *Serratia* and total bacteria increased. Further replicates are required to draw any firm conclusions, although these data suggest that PGRPLC also exerts an antibacterial effect on a subset of the mosquito gut microbial population. Although an increase of total bacteria following *PGRPLC* silencing has been previously reported, functional differences were also observed between different *PGRPLC* isoforms

(Meister et al., 2009). The significance of *PGRPLC* alternative splicing remains poorly understood and could be related to regulation of IMD/REL2 responses that shape the gut microbial communities by targeting or tolerating a subset of mosquito gut bacteria. In *Drosophila*, regulation of the Imd pathway by Caudal has been shown to exert antibacterial effects on a subset of gut bacteria, thus shaping the population structure of the fly's gut microbiota (Ryu et al., 2008). Therefore, it remains to be determined whether such level of IMD/REL2 regulation can be also found at the PGRPLC level.

7.10 FN3D1-3 modulate the mosquito gut microbial population structure by limiting *Enterobacteriaceae*

The observed non-uniform effect of FN3D1-3 between total bacteria and *Serratia* populations in the mosquito gut suggested that the FN3D1-3 antibacterial effect may be exerted on a subset of the mosquito gut bacterial population. To further investigate the FN3D1-3 mode of action in shaping the mosquito gut population structure, a metagenomic analysis was carried out in which the bacterial populations inhabiting the mosquito gut in *FN3D1-3* dsRNA treated mosquitoes and the respective *dsLacZ* treated controls were identified by sequencing their 16S V4-V6 hypervariable regions using 454 pyrosequencing.

As a template, cDNA pools were used corresponding to the host-bacterial transcripts in the mosquito gut of *FN3D1-3* or *LacZ* dsRNA treated mosquitoes retaining their natural gut microbiota, 5 days post dsRNA treatment. These cDNA pools corresponded to 4 independent assays, in which pools of mosquitoes retaining their natural gut microbiota were treated either with *LacZ* dsRNA or dsRNA targeting the *FN3D* transcript. These independent assays were carried out over a period of about 6 months, so a degree of genetic variation underlying the respective mosquito population was expected, along with stochastic environmental variables related to maintenance of the laboratory *N'gousso* mosquito colony that could conceivably also influence the mosquito gut microbiota composition.

In these independent assays, the total bacterial and *Serratia* abundance was established using broad-range or *Serratia*-specific 16S primers, as shown previously (Figure 7.7A-C). The respective cDNA pools were subsequently used as template for PCR amplification of the bacterial 16S V4-V6 variable region with suitable primers. The flanking sequences of this region show an elevated degree of conservation among bacterial taxa and its 560 bp length is compatible with the sequencing capabilities of the *Roche GS FLX* technology platform used for pyrosequencing, rendering amplification of this region suitable for such metagenomic analysis (Baker et al., 2003; Ge et al., 2013; Hamady and Knight, 2009).

The pyrosequencing technology implemented allows high-throughput next-generation sequencing of the amplified copies corresponding to the V4-V6 region of the 16S rRNA bacterial transcripts present in the mosquito gut (Novais and Thorstenson, 2011). The derived amplicons are rendered single-stranded and are further hybridized by primers recognizing sequences incorporated at the two ends of the amplicons during the PCR amplification step. Subsequently, each dNTP is added to the reaction, one at the time. If the added dNTP is complementary to the template, an added DNA polymerase catalyzes its incorporation and releases a pyrophosphate, which is converted to ATP and eventually leads to light emission through a luciferase-catalyzed reaction. Unincorporated dNTPs are removed and the next dNTP is added to the reaction sequentially. This process is massively parallel, with up to 300,000 amplicons sequenced at once. Furthermore, amplicons corresponding to different cDNA pools added concomitantly with different barcode sequences are recognized by different sequencing primers. Typically an up to 500 bp region is sequenced in this way from both ends of the amplicon, subsequently combined into a consensus sequence (Ahmadian et al., 2006; Gharizadeh et al., 2006).

The resulting sequence reads corresponding to each cDNA pool were filtered to a minimum length of 250 bp and were subsequently blasted against the NCBI *16SMicrobial* database (Madden, 2002). Each sequence read was then assigned to a bacterial family, based on the generated sequence alignment using the *blastn* nucleotide blast algorithm. Furthermore, for sequence reads assigned to a bacterial family, alignment of each sequence read with reference sequences of taxa that belong to the bacterial family were determined. As sequence reads typically aligned with multiple reference sequences, the top hit was determined for each taxon, i.e. the reference sequence with most aligned sequence reads. In this way, sequence reads assigned to the *Enterobacteriaceae* bacterial family were aligned to the respective reference sequences of *Enterobacteriaceae* genera and reads aligning to *Serratia* reference sequences were separately categorized while the remaining reads were designated as other *Enterobacteriaceae*. These sequence reads most likely represent bacteria belonging to the genus *Serratia* but also bacterial strains that show similarity to *Serratia* reference sequences, as *Enterobacteriaceae* sequence reads aligned to multiple reference sequences, belonging to different bacterial taxa of the *Enterobacteriaceae* family. Further analysis using a higher read depth, i.e. more total sequence reads for each pool, but also a higher 16S amplicon read length, possibly including additional 16S variable regions, can be used to more accurately determine the taxonomic status of *Enterobacteriaceae* present in the mosquito gut and thus precisely quantify bacteria of the genus *Serratia*. Given the limitations of this analysis, the alignment of sequence reads assigned to the *Enterobacteriaceae* family with *Serratia* reference sequences can provide

a measure of quantification of bacteria of the genus *Serratia* along with bacteria with similarity to *Serratia* reference sequences, which is directly comparable between the different sequenced pools.

The 454 pyrosequencing results can be found in Figure 7.8 and Table 7.1. In panel 7.8A, the sequence reads from 3 cDNA pools corresponding to *LacZ*, *FN3D1* and *FN3D2* dsRNA treated mosquitoes from the same independent assay can be seen categorized per assigned bacterial family. These pools are designated as group A. Similarly, in panel 7.8B, bacterial family categorization based on sequencing corresponding to *LacZ, FN3D2* and *FN3D3* dsRNA treated mosquitoes of a second independent assay can be seen, designated as group B. In panel 7.8C, the sequencing results using cDNA from *dsLacZ* and *FN3D1* dsRNA treated mosquitoes in another independent assay are shown, designated as group C, and, finally, in panel 7.8D, the sequencing results from a fourth independent assay with cDNA stemming from *dsLacZ* or *FN3D3* dsRNA treated mosquitoes can be seen, designated as group D. As each designated group comprised *dsLacZ* or *FN3D* dsRNA treated pools stemming from the same independent assay, the sequencing results within each group are thus directly comparable to each other.

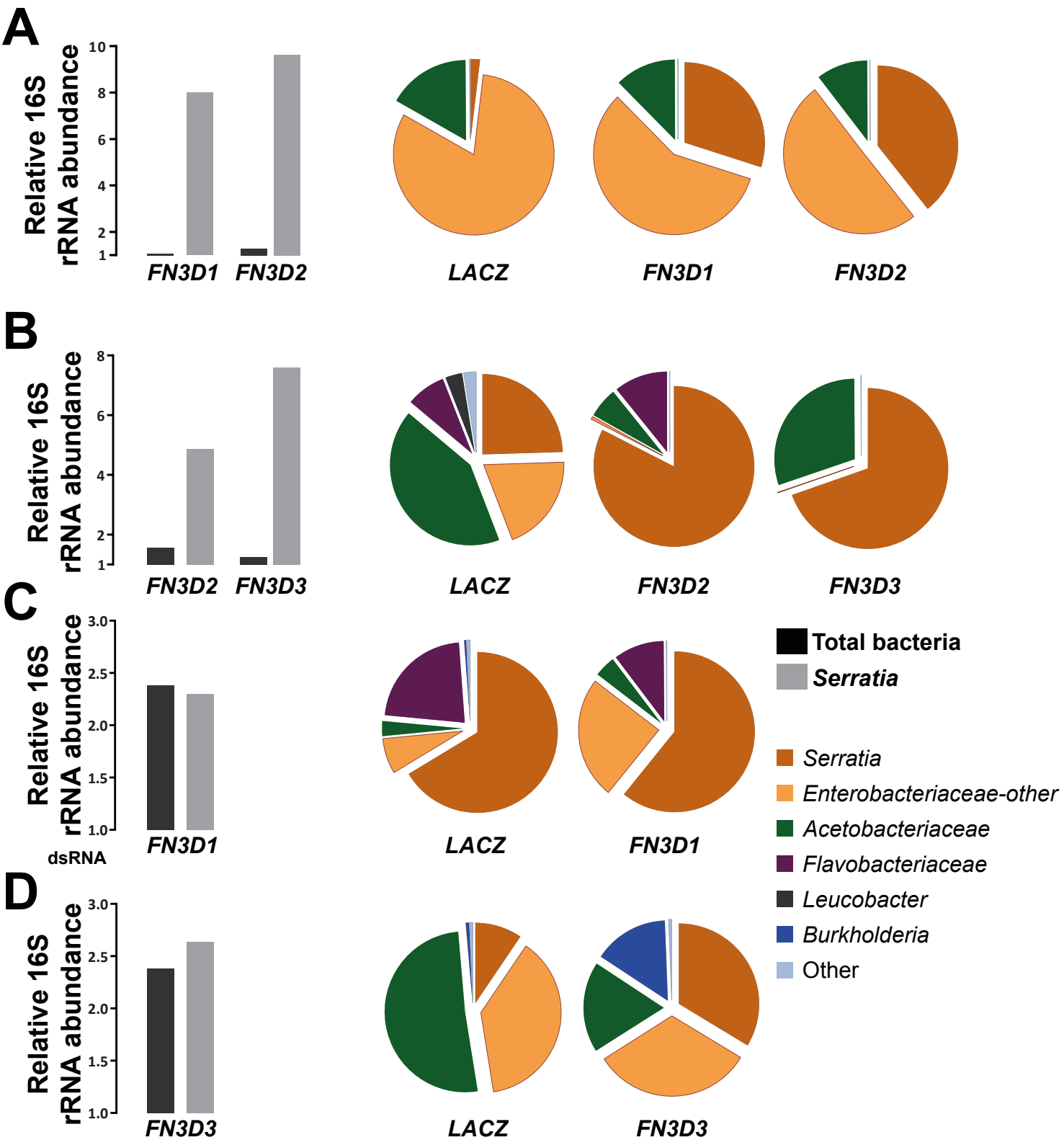

A
Relative 16S rRNA abundance
10
8
6
4
2
1
FN3D1
FN3D2
LACZ
FN3D1
FN3D2
B
Relative 16S rRNA abundance
8
6
4
2
1
FN3D2
FN3D3
LACZ
FN3D2
FN3D3
C
Relative 16S rRNA abundance
3.0
2.5
2.0
1.5
1.0
dsRNA
FN3D1
LACZ
FN3D1
Total bacteria
Serratia
Serratia
Enterobacteriaceae-other
Acetobacteriaceae
Flavobacteriaceae
Leucobacter
Burkholderia
Other
D
Relative 16S rRNA abundance
3.0
2.5
2.0
1.5
1.0
FN3D3
LACZ
FN3D3

Figure 7.8: *FN3D1-3* silencing modulates the mosquito gut bacterial population structure in favour of *Enterobacteriaceae*, mainly *Serratia* or strains that show similarity to *Serratia* reference sequences. Mosquitoes retaining their natural gut microbiota, without antibiotic treatment or infection with *S. marcescens*, were treated with *FN3D1-3* or *LacZ* dsRNA and, 5 days post dsRNA treatment, total RNA was extracted from mosquito guts corresponding to each dsRNA treatment and was used for cDNA synthesis. Total bacteria and *Serratia* abundance was determined by qRT-PCR, as shown in the bars to the left of each panel, using broad-range or *Serratia*-specific 16S bacterial primers. The cDNA pools from 4 independent assays, shown in panels A-D, were used for PCR amplification of the bacterial 16S V4-V6 region and the resulting amplicons were sequenced using 454 pyrosequencing. Sequence reads corresponding to *FN3D1-3* or *LacZ* dsRNA treated pools were blasted against the NCBI *16SMicrobial* database and assigned to their respective bacterial families. For *Enterobacteriaceae*, sequence reads that aligned to *Serratia* reference sequences were separately categorized from the remaining *Enterobacteriaceae*. For each dsRNA treatment, a pie chart represents the relative proportions of sequence reads assigned to each bacterial family, based on the colour legend included. Below each pie chart, the corresponding dsRNA treatment is indicated.

The 4 *dsLacZ* treated control pools in groups A-D (Figure 7.8A-D), stemming from 4 independent assays, showed considerable variation regarding their gut microbiota composition. This level of variation is consistent with previously reported metagenomic analyses both in lab-reared and field-collected mosquitoes showing extensive diversity of the mosquito gut bacterial communities both at the individual and population levels (Boissière et al., 2012; Osei-Poku et al., 2012; Wang et al., 2011b). *Enterobacteriaceae,* including both *Serratia* and other *Enterobacteriaceae*, were prevalent in all *dsLacZ* treated control pools, comprising 83.2% of total sequence reads in group A, 44.1% in group B, 73.5% in group C and 47.5% in group D (Table 7.1).

Interestingly, *Serratia*, i.e. *Enterobacteriaceae* sequence reads aligning to *Serratia* reference sequences, showed considerable variation regarding their representation in the 4 *dsLacZ* treated control pools corresponding to different independent assays. *Serratia* representation was determined at 1.88% in the *dsLacZ* treated pool of group A (Figure 7.8A), 9.46% in group D (Figure 7.8D), 24.52% in group B (Figure 7.8B) and 66.35% in group C (Figure 7.8C). This level of variation parallels the observed variation in the levels of *S. marcescens* infection achieved following oral infection, with highly, lowly or non-infected mosquitoes (Figure 4.4), suggesting that the observed level of *Serratia* variation between the *dsLacZ* treated pools may also reflect the underlying mosquito genetic variation.

Other bacterial families also traced in the *dsLacZ* treated pools included *Acetobacteriaceae*, representing 16.66% of total sequence reads of the *dsLacZ* treated pool in group A (Figure 7.8A), 41.95% in group B (Figure 7.8B), 3.09% in group C (Figure 7.8C) and 51.16% in group D (Figure 7.8D). Alignment of *Acetobacteriaceae* sequence reads with the respective reference sequences indicated that the *Acetobacteriaceae* sequence reads most likely correspond to bacteria of the genus *Swaminathania, Asaia*, *Gluconobacter* or *Gluconacetobacter* (Table 7.1). Bacteria of the genus *Asaia* are known to efficiently colonize *An. gambiae* (Damiani et al., 2010), while *Acetobacteriaceae* were also found to be prevalent in the *An. gambiae* gut microbial communities by previous metagenomic analyses (Boissière et al., 2012; Osei-Poku et al., 2012).

The *Flavobacteriaceae* bacterial family was also detected in *dsLacZ* treated control pools of group B (Figure 7.8B) and group C (Figure 7.8C), at 7.65% and 22.3% of total sequence reads, respectively. Sequence reads assigned to *Flavobacteriaceae* aligned with *Chryseobacterium* and *Elizabethkingia* reference sequences. Bacteria of the genus *Elizabethkingia* are common colonizers, especially of lab-reared mosquito guts, and display multiple antibiotic resistance (Kampfer et al., 2010). Bacteria of the

genus *Elizabethkingia* can be vertically transmitted in *An. gambiae* and affect mosquito longevity most likely through involvement in melanization processes (Akhouayri et al., 2013).

Other detected bacterial taxa in the *dsLacZ* treated control pools included bacteria of the genus *Leucobacter* but also, in the *dsLacZ* treated pools of group C (Figure 7.8C) and D (Figure 7.8D), bacteria of the genus *Burkholderia*, assigned mainly under the *Burkholderia cepacia* complex, at 0.43% and 0.71% of total sequence reads, respectively. Bacteria of the genus *Burkholderia* were also detected in the gut of field-collected *An. gambiae* in a previous metagenomic analysis (Boissière et al., 2012). *Burkholderia* are classified as beta-proteobacteria (Lessie et al., 1996; Santini et al., 2013) and they are common symbionts of insects (Kikuchi et al., 2011b; Kikuchi et al., 2005) and plants (Caballero-Mellado et al., 2004), while in humans they have been detected as opportunistic pathogens especially in cystic fibrosis patients (Coutinho et al., 2011). Interestingly, *Burkholderia* colonization is prevalent in the stinkbug *Riptorus clavatus* and has been associated with a fitness benefit (Kikuchi et al., 2007, 2011a). *Burkholderia* have also been shown to produce anti-fungal agents that can inhibit the germination of entomopathogenic fungi including *Beauveria bassiana* and *Metarhizium anisopliae* (Santos et al., 2004), known to colonize mosquitoes, induce immune responses and reduce their lifespan (Fang et al., 2011; Luz et al., 2011; Yassine et al., 2012).

Silencing of *FN3D1* or *FN3D2* in group A assay (Figure 7.8A) resulted in an increased representation of *Enterobacteriaceae* from 83.2% in the *dsLacZ* treated pool to 87.6% and 89.5% in the *FN3D1* and *FN3D2* knockdown pools, respectively. Remarkably, in both cases, the representation of *Serratia* showed a dramatic increase from 1.88% in the *dsLacZ* treated pool to 29.9% and 39.3% in the *FN3D1* and *FN3D2* knockdown pools, respectively. In both cases, qRT-PCR data indicated an about 8-fold increase of *Serratia* load with a concomitant marginal increase of total bacteria (Figure 7.8A, bars to the left of the panel). The representation of *Acetobacteriaceae* was decreased following both *FN3D1* and *FN3D2* silencing from 16.6% to 12.3% and 10.3%, respectively.

In the group B assay, *FN3D2* or *FN3D3* knockdown resulted in a much more pronounced increase in *Serratia* abundance, compared to total bacteria (Figure 7.8B, bars to the left of the panel). Silencing of *FN3D2* or *FN3D3* led to a precipitous increase of *Enterobacteriaceae* from an intermediate 44.1% representation in the *dsLacZ* treated pool to 83.1%, following *FN3D2*, and 69.8%, following *FN3D3* silencing (Figure 7.8B). At the same time, *Acetobacteriaceae* showed a dramatic decrease in representation following *FN3D2* silencing, from 41.95% in the *dsLacZ* treated pool to 6.1%. *FN3D3* knockdown, though, resulted in a modest *Acetobacteriaceae* representation decrease to 30.1%.

Furthermore, *FN3D2* knockdown resulted in a marginal *Flavobacteriaceae* representation increase from 7.95% to 9.8% while *FN3D3* knockdown completely eliminated the *Flavobacteriaceae* presence. In both cases, *FN3D2* or *FN3D3* silencing resulted in a precipitous increase of *Serratia* representation, from 24.5% of total sequence reads in the *dsLacZ* treated pool to almost all *Enterobacteriaceae* sequence reads aligning with *Serratia* reference sequences following *FN3D2* or *FN3D3* silencing (Figure 7.8B).

Taken together, these data suggest that FN3Ds play a major role in shaping the population structure of the mosquito gut microbial communities by limiting the representation of *Enterobacteriaceae*, mainly *Serratia* or strains with similarity to *Serratia* reference sequences. *FN3D1-3* silencing increased the abundance of *Serratia* but also shifted the mosquito gut bacterial population structure in favour of *Enterobacteriaceae*. This shift in the population structure of the mosquito gut microbial communities may be a result of an FN3D specificity in targeting *Enterobacteriaceae*, mainly *Serratia* or strains with similarity to *Serratia*.

To shed more light into the FN3D mode of action, shifts in bacterial population structure were determined in cases where *FN3D* silencing resulted in moderate increases of both *Serratia* and total bacteria (Figure 7.8C-D, bars to the left of the respective panel). In group C, *FN3D1* silencing resulted in an increase of *Enterobacteriaceae* representation from 73.5% in the *dsLacZ* treated pool to 85.4% in the *FN3D1* dsRNA treated pool (Figure 7.8C). At the same time, *FN3D1* silencing reduced *Flavobacteriaceae* representation from 22.3% in the *dsLacZ* treated pool to 10.2% while marginally increasing *Acetobacteriaceae* representation from 3.02% to 4.27%. Although bacteria of the genus *Burkholderia* were traced in the control *dsLacZ* treated pool at 0.43% of total sequence reads, they were completely eliminated in the *FN3D1* knockdown pool. *Enterobacteriaceae* sequence reads that aligned with *Serratia* reference sequences were prevalent in the *dsLacZ* treated control pool, at 66.3% of total sequence reads, by dominating *Enterobacteriaceae*, which represented 73.5% of total sequence reads. Therefore, the modest increase in abundance of both total bacteria and *Serratia* following *FN3D1* silencing is most likely to be a result of the initial *Serratia* dominance in the *dsLacZ* treated pool.

Although *FN3D1* silencing increased the abundance of *Serratia* in the group C assay, as determined by qRT-PCR (Figure 7.8C, bars to the left of the panel), *Serratia* representation was reduced to 60.8% of total sequence reads. At the same time, sequence reads aligning to reference sequences of other *Enterobacteriaceae* strains were increased compared to the *dsLacZ* treated controls and accounted for the overall increased *Enterobacteriaceae* representation in the *FN3D1* knockdown pool (Table 7.1).

Among them, alignment of sequence reads with *Citrobacter*, *Salmonella* and *Raoultella* reference sequences was increased by 6 to 9 percentage points in the *FN3D1* knockdown pool (Table 7.1).

Comparison of the shifts in the mosquito gut bacterial population structure due to *FN3D1* silencing in group A and C assays confirm that FN3D1 limits the representation of *Enterobacteriaceae*, as in both cases *FN3D1* knockdown resulted in increased *Enterobacteriaceae* representation in the mosquito gut microbiota. In both cases, *FN3D1* silencing increased the abundance of *Serratia*, although the increase in the group C assay was more modest, most likely due to the increased *Serratia* presence in the control *dsLacZ* treated pool. The decreased *Serratia* representation following *FN3D1* silencing in the group C assay suggests that FN3D1 additionally targets *Enterobacteriaceae* other than *Serratia*.

The difference of the *FN3D1* knockdown effect in modulating *Serratia* representation between group A and C assays also suggests that FN3D1 might target a subset of *Enterobacteriaceae* with different efficiency. Such differential FN3D1 efficiency might also underlie the differences of the bacterial population structure of the *dsLacZ* control pools between group A and C assays. Furthermore, the underlying mosquito genetic variation in loci other than *FN3D1* may also influence the FN3D1 efficiency in targeting *Serratia* or a subset of *Enterobacteriaceae* between group A and C assays. In that case, the prevalence of *Serratia* in the *dsLacZ* treated control pools would be influenced by genetic variation in the *FN3D1* locus or in other loci that also determine the mosquito gut microbial population structure. Therefore, a prevalent *Serratia* representation in the control pool of the group C assay would indicate a decreased efficiency of FN3D1 or other immune factors in limiting *Serratia*. Thus, *FN3D1* silencing would result in a less pronounced *Serratia* increase, with a concomitant decrease in *Serratia* representation due to increased abundance of other *Enterobacteriaceae* strains more efficiently targeted by FN3D1.

Furthermore, the differences in *Serratia* representation following *FN3D1* knockdown between group A and C assays discount the possibility that the observed shifts in gut bacterial population structure following *FN3D1-3* silencing are a result of an enhanced growth potential of *Serratia* or other *Enterobacteriaceae,* indicative of an FN3D uniform effect.

FN3D3 silencing in the group D assay increased *Enterobacteriaceae* representation from 47.5% in the *dsLacZ* treated control pool to 66%, with a concomitant increase of strains aligning to *Serratia* reference sequences from 9.4% to 33.7% (Figure 7.8D). At the same time, *Acetobacteriaceae* representation decreased from 51.1% to 18.2% of total sequence reads. These data confirm similar *FN3D3* knockdown effects in the group B assay (Figure 7.8B). Remarkably, though, bacteria of the genus *Burkholderia* traced

in the *dsLacZ* treated control pool at 0.71% of total sequence reads, showed a dramatic increase in their representation following *FN3D3* knockdown to 15.15% of total sequence reads (Figure 7.8D). Bacteria of the genus *Burkholderia* were not traced in the *dsLacZ* treated control pool of the group B assay (Figure 7.8B), in which the effect of *FN3D3* silencing was also assayed but were found at 0.43% of total sequence reads in the *dsLacZ* treated control pool in the group C assay, in which *FN3D1* silencing failed to result in a similar *Burkholderia* representation increase (Figure 7.8C). These data suggest that FN3D3 limits the representation of a subset of the mosquito gut bacterial population, including *Enterobacteriaceae* but also bacteria of the genus *Burkholderia*.

Interestingly, the different effect of *FN3D1* and *FN3D3* knockdown on bacteria of the genus *Burkholderia* found in the microbial communities of the *dsLacZ* treated control pools in group C and D assays, respectively, not only indicate a functional difference between FN3D1 and FN3D3 but also further discount the possibility that the observed effect is due to an enhanced *Burkholderia* growth potential or interactions between strains of the mosquito gut microbial community.

Taken together, these data suggest that FN3D1-3 are major modulators of the mosquito gut microbial population structure by limiting a subset of the bacterial population including *Enterobacteriaceae*, mainly *Serratia* or strains with similarity to *Serratia* reference sequences but also, for FN3D3, bacteria of the genus *Burkholderia*.

7.11 Shifts in the *Enterobacteriaceae* population structure following *FN3D1-3* silencing

To shed more light into the FN3D1-3 specificity in limiting *Enterobacteriaceae*, the shifts in the percentage of sequence reads aligning to reference sequences of bacterial taxa belonging to *Enterobacteriaceae* were determined in the sequenced *FN3D1-3* knockdown pools, relative to the respective *dsLacZ* treated pools. For each taxon, the corresponding shifts were determined in both independent assays in which the effect of *FN3D1-3* knockdown was determined. Shifts were expressed as percent changes in the number of sequence reads aligning with the respective reference sequence corresponding to each taxon, relative to the total number of sequence reads. As previously, alignment of the reference sequence with the highest number of aligned sequence reads, i.e. the top hit, was taken into account for each taxon. Furthermore, only taxa with aligned sequence reads amounting to at least 5% relative to the total number of sequence reads in the *dsLacZ* or *FN3D* knockdown pool were taken into account.

In the group A assay, *FN3D1* silencing strikingly reduced the representation of *Enterobacteriaceae* strains aligning to *Erwinia* reference sequences, from 79.65% in the *dsLacZ* treated pool to 24.25% in the *FN3D1* knockdown pool. The same applied to strains aligning to *Kluyvera* reference sequences, which were reduced, from 43.3%, to 16.1% in the *FN3D1* knockdown pool. Alignment with both of these strains was marginally traced in the *dsLacZ* treated control pool of the group C assay at about 1% of total sequence reads. *FN3D1* silencing in the group A assay also resulted in increased representation of *Serratia*, from 1.88% to 29.95%, but also increased the representation of strains aligning to *Enterobacter* and *Klebsiella* reference sequences, from 14.8% to 35.7% and from 22.3% to 43.4%, respectively (Table 7.1). In the group C assay, *FN3D1* silencing resulted in decreased *Serratia* representation but also led to a marginally increased *Enterobacter* representation, from 62% to 65.96%, and modestly decreased *Klebsiella* representation, from 54% to 48.3%. The *Enterobacteriaceae* strains that showed the more pronounced increase in representation following *FN3D1* silencing in the group C assay included *Citrobacter*, increased from 57.2% to 65.25%, *Raoultella*, increased from 23.7% to 33.6% and *Salmonella*, increased from 0.5% to 6.9%. In the group A assay, *FN3D1* silencing modestly reduced *Citrobacter* representation from 79.4% to 64.6%, no change was observed regarding the *Raoultella* representation, which stayed at about 41.5%, while *Salmonella* was not traced in the control pool and represented 2.24% of total sequence reads in the *FN3D1* knockdown pool (Table 7.1). The respective shifts in representation of each *Enterobacteriaceae* genus following *FN3D1* silencing can be found in Figure 7.9A.

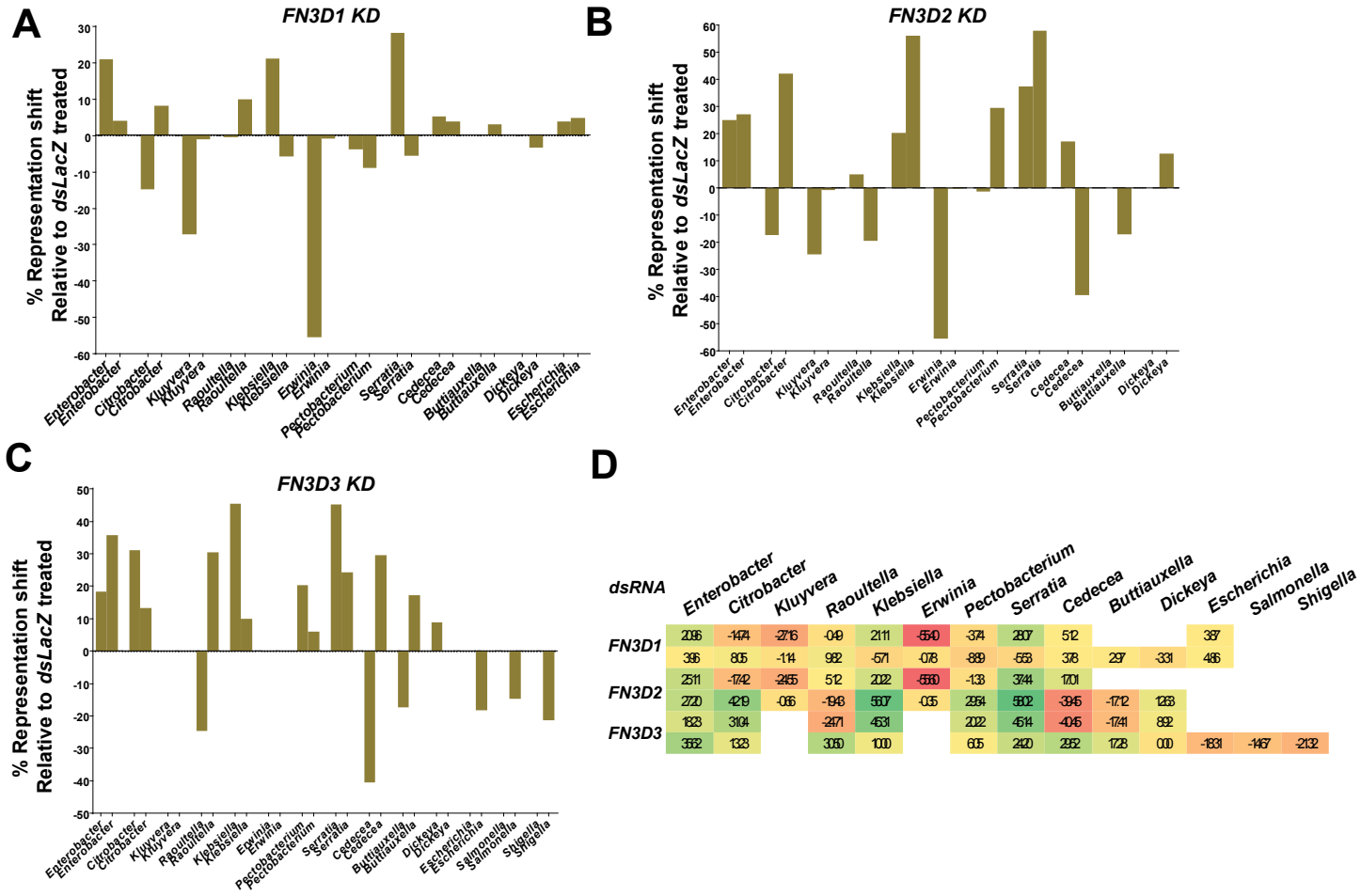


Figure 7.9: Shifts of *Enterobacteriaceae* genera representation following *FN3D1-3* silencing. The sequence reads assigned to *Enterobacteriaceae*, for each of the 2 sequenced independent assays following *FN3D1-3* silencing, were blasted against the NCBI *16SMicrobial* database, along with the sequence reads of the respective *dsLacZ* treated controls. The number of sequence reads aligning to reference sequences for each *Enterobacteriaceae* genus was determined in each case, with the reference sequence of each genus showing the highest number of aligned sequence reads taken into account, along with their relative proportions compared to total sequence reads, as shown in Table 7.1. The shift in representation of each *Enterobacteriaceae* genus in *FN3D1* (panel A), *FN3D2* (panel B) or *FN3D3* (panel C) knockdown pools, compared to the respective *dsLacZ* treated pool was determined. Bars indicate the % shift in representation of each genus, with positive values indicating an increase in *FN3D1-3* knockdown pools and negative values indicating a respective decrease. For each *Enterobacteriaceae* genus, the corresponding shifts are shown for both independent assays in which silencing of the respective *FN3D* was performed, indicated by an individual bar. Below each bar, the corresponding genus is indicated. *Enterobacteriaceae* genera with aligned sequence reads <5% of total reads in both the *FN3D* knockdown and *dsLacZ* treated pools are omitted. In panel D, the corresponding shifts for each *Enterobacteriaceae* genus are shown in a heat map. Each column corresponds to a different genus, as indicated above each column. Each row corresponds to the observed shifts in an independent assay, following *FN3D1-3* silencing, as indicated on the left side of each row.

These data indicate that FN3D1 targets a subset of *Enterobacteriaceae*, including strains aligning to *Serratia*, *Enterobacter*, *Citrobacter*, *Raoultella, Salmonella* and *Klebsiella* reference sequences, as *FN3D1* silencing led to increased representation of these strains and consequently of *Enterobacteriaceae*. The concomitant decreased representation of other *Enterobacteriaceae* strains, such as *Erwinia* or *Kluyvera,* most likely reflects increased growth of FN3D1-targeted *Enterobacteriaceae* or, possibly, of bacterial interactions or competition that limit the abundance of these strains. Another possibility is that differential FN3D1 efficiency in limiting these strains might have led to their decreased representation following *FN3D1* knockdown.

The respective profile of *Enterobacteriaceae* alignment shifts following *FN3D2* knockdown largely followed the *FN3D1* knockdown profile (Figure 7.9B). *Enterobacter*, *Klebsiella* and *Serratia* showed marked increases relative to *dsLacZ* treated pools, consistent in both assayed pools. *Citrobacter*, *Raoultella* and *Cedecea* showed both increases and decreases in the 2 respective assayed pools. As discussed previously for a similar trend following *FN3D1* silencing for *Serratia*, these differences between the assayed pools could be related to the underlying genetic variation that determines prevalence in the *dsLacZ* treated pool and further discount the possibility that the observed shifts are shaped by each taxon's growth potential or competitive capability. Finally, *Kluyvera*, *Erwinia* but also *Buttiauxella* showed consistent and pronounced alignment decreases in *FN3D2* knockdown pools.

In *FN3D3* knockdown mosquitoes, bacteria of the genus *Enterobacter*, *Citrobacter*, *Klebsiella*, *Pectobacterium* and *Serratia* showed consistent increases in alignment compared to *dsLacZ* treated controls, suggesting that FN3D3 limits the strains aligning to the respective reference sequences (Figure 7.9C). *Kluyvera* and *Erwinia* were traced at about 1% in one of the *dsLacZ* treated pools and no shifts led to a more than 5% representation in the *FN3D3* knockdown pools. Decreases in alignment, though, were detected for bacteria of the genus *Escherichia*, *Salmonella* and *Shigella*, found in the *dsLacZ* treated pool of group D for which only the *FN3D3* knockdown effect was assayed. Finally, *Raoultella*, *Cedecea* and *Buttiauxella* alignments showed both increases and decreases in the 2 assayed *FN3D3* knockdown pools.

These data indicate a relative consistency between *FN3D1*, *FN3D2* and *FN3D3* silencing in exerting shifts in alignment of *Enterobacteriaceae* sequence reads, compared to *dsLacZ* treated mosquitoes, shown for each detected *Enterobacteriaceae* genus in Figure 7.9D. These shifts are consistent with the possibility that FN3D1-3 target a subset of *Enterobacteriaceae* that most likely includes strains aligning with *Serratia* reference sequences. In many cases, for all 3 FN3Ds, both increases and decreases were observed in the 2 assayed pools. As shown for *Serratia*, a decrease in representation may be observed despite increased

abundance for the respective strain, suggesting that these strains may still be targeted by FN3Ds, perhaps with variable efficiency.

7.12 Phylogenetic distances of *Enterobacteriaceae* genera with differential representation shifts following *FN3D1-3* silencing

To identify a possible link between *Enterobacteriaceae* genera exhibiting relatively consistent representation shifts following *FN3D1-3* knockdown (Figure 7.9D), their phylogenetic relationships were investigated based on previously reported *Enterobacteriaceae* phylogenetic trees (Figure 7.10) (Paradis et al., 2005). As shown in Figure 7.9D, bacteria of the genus *Enterobacter*, *Klebsiella* and *Serratia* showed consistent and pronounced increases of aligned *Enterobacteriaceae* sequence reads over the 6 *FN3D1-3* knockdown assays, as *Enterobacter* representation increased in all 6 assays while *Klebsiella* and *Serratia* representation increased in 5 out of 6 assays. At the same time, bacteria of the genus *Kluyvera* and *Erwinia* showed a consistent decrease in aligned sequence reads in all assays they were present in the control pool at a more than 5% representation.

As can be seen in Figure 7.10, *Enterobacter* and *Klebsiella* form a cluster which is distinct but relatively close to *Kluyvera*. *Raoultella*, found between *Enterobacter* and *Klebsiella*, showed both increases and decreases in aligned sequence reads following *FN3D1-3* silencing. Between *Kluyvera* and the *Enterobacter*/*Klebsiella* cluster, *Salmonella*, *Escherichia* and *Shigella* are found, all showing decreases in aligned sequence reads following *FN3D3* silencing in one control pool they were present (Figure 7.9D). *Citrobacter*, found between *Kluyvera* and *Serratia*, showed both increases and decreases in aligned sequence reads following *FN3D1-3* knockdown. The same applies for *Cedecea*, found between *Erwinia* and the *Enterobacter*/*Klebsiella* cluster.

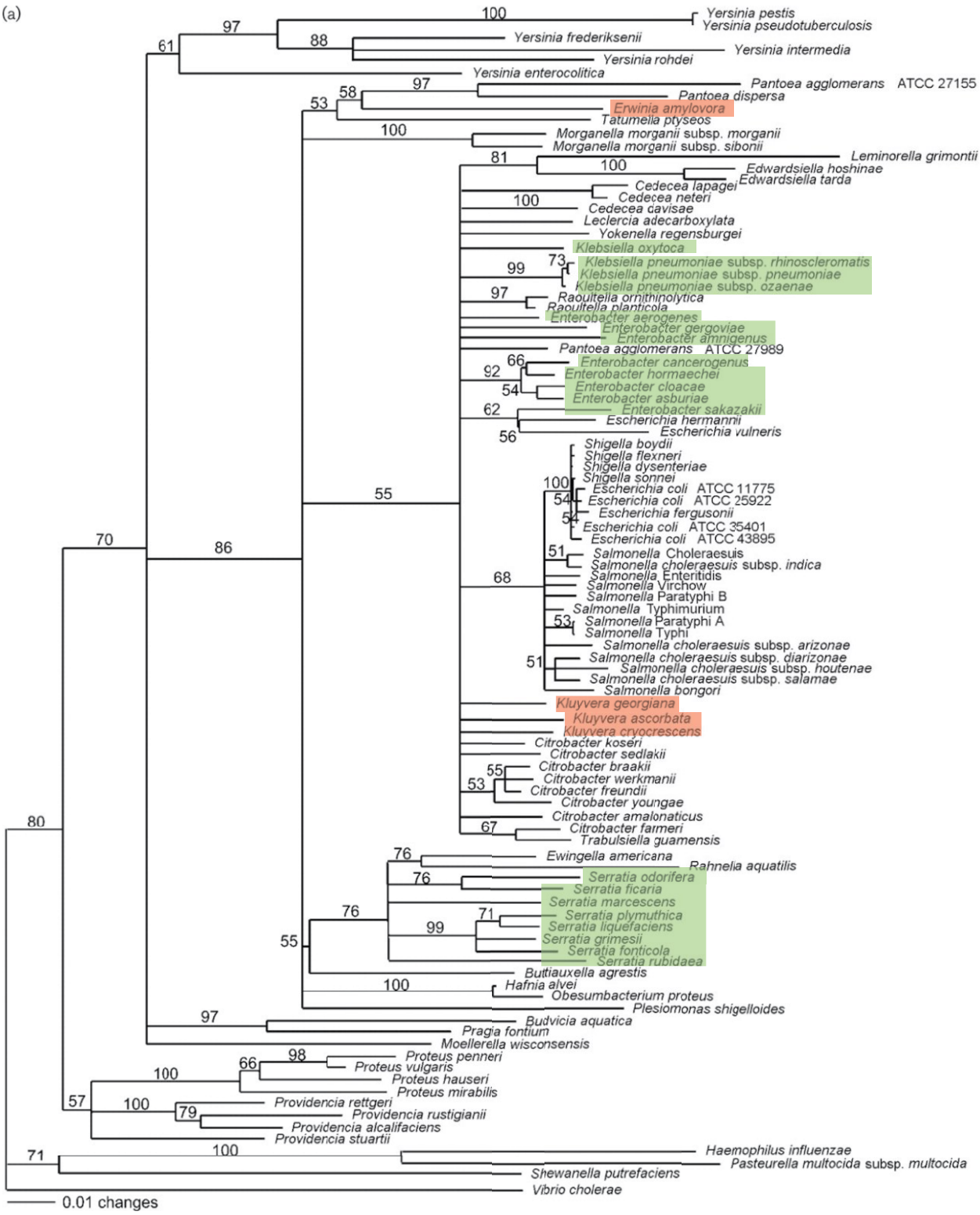

(a)
Yersinia pestis
Yersinia pseudotuberculosis
Yersinia frederiksenii
Yersinia intermedia
Yersinia rohdei
Yersinia enterocolitica
Pantoea agglomerans ATCC 27155
Pantoea dispersa
Erwinia amylovora
Tatumella ptyseos
Morganella morganii subsp. morganii
Morganella morganii subsp. sibonii
Leminorella grimontii
Edwardsiella hoshinae
Edwardsiella tarda
Cedecea lapagei
Cedecea neteri
Cedecea davisae
Leclercia adecarboxylata
Yokenella regensburgei
Klebsiella oxytoca
Klebsiella pneumoniae subsp. rhinoscleromatis
Klebsiella pneumoniae subsp. pneumoniae
Klebsiella pneumoniae subsp. ozaenae
Raoultella ornithinolytica
Raoultella planticola
Enterobacter aerogenes
Enterobacter gergoviae
Enterobacter amnigenus
Pantoea agglomerans ATCC 27989
Enterobacter cancerogenus
Enterobacter hormaechei
Enterobacter cloacae
Enterobacter asburiae
Enterobacter sakazakii
Escherichia hermannii
Escherichia vulneris
Shigella boydii
Shigella flexneri
Shigella dysenteriae
Shigella sonnei
Escherichia coli ATCC 11775
Escherichia coli ATCC 25922
Escherichia fergusonii
Escherichia coli ATCC 35401
Escherichia coli ATCC 43895
Salmonella Choleraesuis
Salmonella choleraesuis subsp. indica
Salmonella Enteritidis
Salmonella Virchow
Salmonella Paratyphi B
Salmonella Typhimurium
Salmonella Paratyphi A
Salmonella Typhi
Salmonella choleraesuis subsp. arizonae
Salmonella choleraesuis subsp. diarizonae
Salmonella choleraesuis subsp. houtenae
Salmonella choleraesuis subsp. salamae
Salmonella bongori
Kluyvera georgiana
Kluyvera ascorbata
Kluyvera cryocrescens
Citrobacter koseri
Citrobacter sedlakii
Citrobacter braakii
Citrobacter werkmanii
Citrobacter freundii
Citrobacter youngae
Citrobacter amalonaticus
Citrobacter farmeri
Trabulsiella guamensis
Ewingella americana
Rahnella aquatilis
Serratia odorifera
Serratia ficaria
Serratia marcescens
Serratia plymuthica
Serratia liquefaciens
Serratia grimesii
Serratia fonticola
Serratia rubidaea
Buttiauxella agrestis
Hafnia alvei
Obesumbacterium proteus
Plesiomonas shigelloides
Budvicia aquatica
Pragia fontium
Moellerella wisconsensis
Proteus penneri
Proteus vulgaris
Proteus hauseri
Proteus mirabilis
Providencia rettgeri
Providencia rustigianii
Providencia alcalifaciens
Providencia stuartii
Haemophilus influenzae
Pasteurella multocida subsp. multocida
Shewanella putrefaciens
Vibrio cholerae
0.01 changes

Figure 7.10: *Enterobacteriaceae* phylogenetic tree indicating phylogenetic distances between genera. A phylogenetic analysis of *Enterobacteriaceae* genera based on the neighbour-joining method, calculated using the Kimura two-parameter method, indicated phylogenetic distances based on sequencing of the *tuf* gene encoding the Tu elongation factor (Paradis et al., 2005). The *Enterobacter*, *Klebsiella* and *Serratia* genera, showing a consistent and pronounced increase in aligned sequence reads following *FN3D1-3* knockdown, compared to the respective *dsLacZ* treated controls, are highlighted in green. The *Erwinia* and *Kluyvera* genera, showing consistent and pronounced decreases in the corresponding aligned sequence reads of *FN3D1-3* knockdown pools, compared to the respective *dsLacZ* treated control, are highlighted in red. Adapted from: (Paradis et al., 2005).

Taken together, the phylogenetic distances of *Enterobacteriaceae* genera with differential representation shifts following *FN3D1-3* silencing show a trend for clustering of genera showing similar shifts but also for mixed behaviour regarding such shifts for genera with intermediate phylogenetic distances. One possibility that cannot be overruled is that a putative trait that determines the behaviour of *Enterobacteriaceae* genera following *FN3D1-3* silencing may have evolved independently in distant genera. Furthermore, bacterial interactions may also influence the observed shifts and further complicate the current analysis. As shown previously for *S. marcescens* (Bando et al., 2013) and *Pseudomonas aeruginosa* (Limmer et al., 2011), variation in bacterial loci can play key roles during pathogenesis. Therefore, bacterial genetic variation could also influence the observed shifts and sequencing-based strategies could shed more light in genetic differences between *Enterobacteriaceae* genera associated with representation increases or decreases following *FN3D1-3* silencing.

7.13 *FN3D1-3* silencing modestly modulates *Asaia* abundance following oral infection

To further demonstrate the specificity of the FN3D1-3 antibacterial activity in targeting a subset of the mosquito gut bacterial population, antibiotic treated mosquitoes were treated with *FN3D1-3* dsRNA or the *dsLacZ* control and were subsequently orally infected with bacteria of the genus *Asaia*. *Asaia* are alpha-proteobacteria belonging to the *Acetobacteriaceae* family (Yamada et al., 2000). *Acetobacteriaceae* were present in all *dsLacZ* treated control pools and *FN3D1-3* silencing resulted in decreased *Acetobacteriaceae* representation, with the exception of *FN3D1* knockdown in the group C assay, which resulted in a marginal increase of an initially low 3.02% *Acetobacteriaceae* representation to 4.26%.

The *Asaia* infection assay mirrored the corresponding *S. marcescens* infection assays. Bacterial load was determined in the guts of *FN3D1-3* or *LacZ* dsRNA treated mosquitoes using qRT-PCR with broad-range bacterial 16S primers. The relative bacterial abundance in *FN3D1-3* dsRNA treated mosquitoes was normalized to the corresponding relative bacterial abundance in *dsLacZ* treated mosquitoes, also orally infected with *Asaia*. The fold-change increases in bacterial load in *FN3D1-3* dsRNA treated mosquitoes, as determined over 3 independent infections in each case, were averaged and significant differences were assessed in a t-test against zero using the respective log2-transformed fold-change values, where zero corresponds to no difference from *dsLacZ* treatment. As can be seen in Figure 7.11, *FN3D1*, *FN3D2* or *FN3D3* silencing showed modest and non-significant bacterial load increases of 1.36, 1.16 and 1.64-fold, respectively, compared to *dsLacZ* treated mosquitoes.

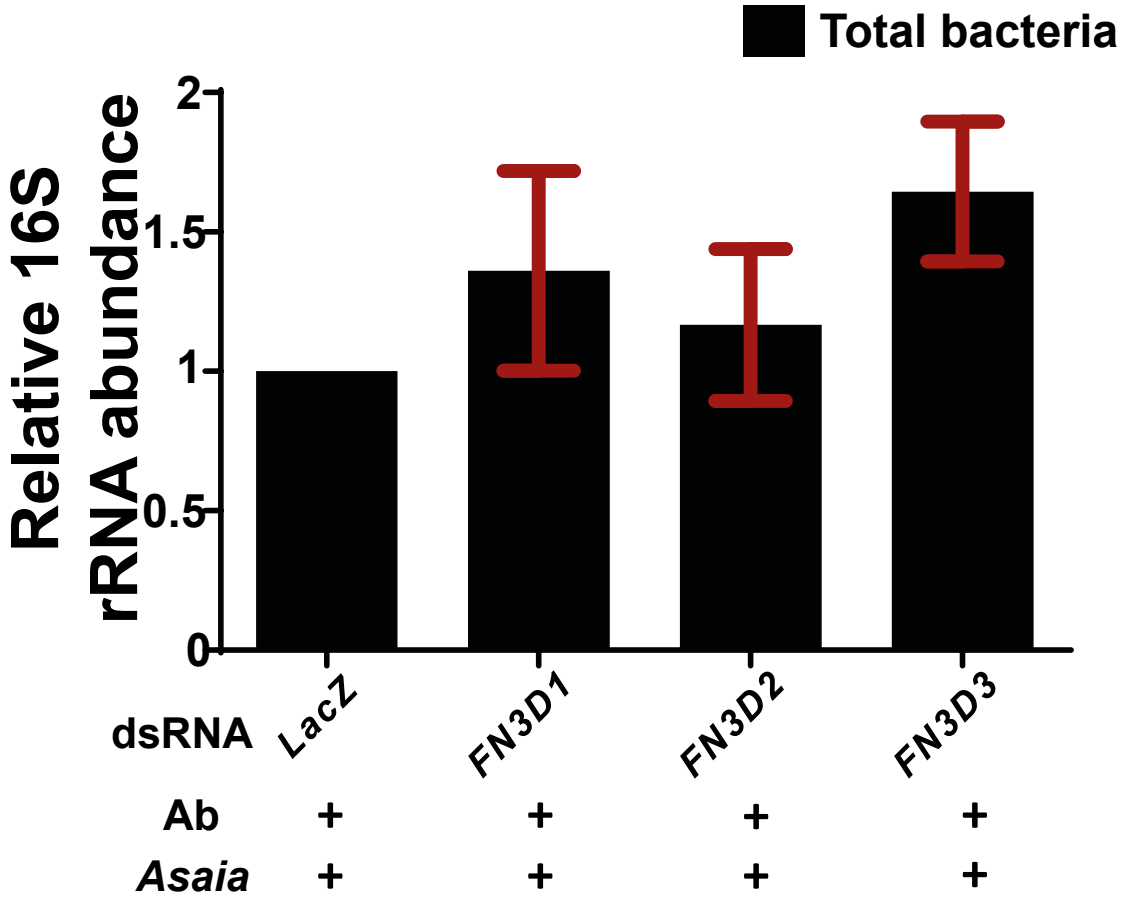


Figure 7.11: *FN3D1-3* silencing marginally modulates *Asaia* abundance in orally infected mosquitoes. Antibiotic treated mosquitoes were treated with *FN3D1-3* dsRNA or the *dsLacZ* control, as indicated below each bar, at the day of emergence. Mosquitoes were orally infected 5 days later with bacteria of the genus *Asaia* (indicated by *Ab+Asaia+*), bacteria-fed mosquitoes were selected at day 2 post infection and, at day 5 post infection, bacterial load in mosquito guts corresponding to each dsRNA treatment was determined. Total RNA was extracted from the guts of surface sterilized mosquitoes and used for cDNA synthesis. The respective cDNA was used as template in a qRT-PCR with broad-range bacterial 16S primers, while *AgS7* primers were used as an endogenous control. In each independent infection, qRT-PCR was performed at least in duplicate. The relative bacterial abundance was determined in each case over 3 independent infections and normalized to the corresponding *dsLacZ* treated control. Bars indicate fold-change differences ±SEM for each dsRNA treatment, compared to the respective *dsLacZ* control. Significant differences were assessed using the log2-transformed fold-change values for each *FN3D1-3* dsRNA treatment in a t-test against zero, where zero corresponds to no difference from *dsLacZ* treatment. The obtained p-values, 0.1664, 0.2416 and 0.2774 for *FN3D1*, *FN3D2* and *FN3D3* knockdown pools, respectively, indicate non-significant differences of fold-change increases in bacterial load compared to the *dsLacZ* control.

Interestingly, *FN3D3* knockdown not only showed the highest fold-change increase compared to the other *FN3D*s, but also the margins of error of the *FN3D3* knockdown fold-change increase, as determined over 3 independent infections, were consistently over 1, suggesting the presence of a subtle effect of *FN3D3* silencing on *Asaia* abundance that did not reach statistical significance. Further replicates are required to confirm this possibility. This observation is consistent, though, with a modest decrease of *Acetobacteriaceae* representation following *FN3D3* silencing from 41.9% to 30.1% of total sequence reads, compared to the corresponding decrease following *FN3D2* knockdown to 9.85% in the group B assay of sequenced *FN3D2-3* knockdown gut microbiota pools.

Overall, though, the effect of *FN3D1-3* silencing on *Asaia* abundance following oral infection was distinguishably lower than the previously observed effect of *FN3D1-3* silencing on *S. marcescens* levels following oral infection. Taken together, these data suggest that the FN3D1-3 antibacterial effect is not uniform upon all Gram-negative bacteria.

7.14 FN3D3 displays a discrete mode of action in modulation of the mosquito gut bacterial population structure

Another intriguing observation is the different effects exerted on non-*Enterobacteriaceae* strains by *FN3D1-3* knockdown in mosquitoes retaining their natural gut microbiota. In the group B assay, *FN3D2* knockdown resulted in a precipitous decrease of *Acetobacteriaceae* while *FN3D3* knockdown more modestly decreased *Acetobacteriaceae* but completely eliminated *Flavobacteriaceae* (Figure 7.8B). These data indicate functional differences between FN3D2 and FN3D3 that shape the mosquito gut bacterial population structure. Furthermore, these differences in the FN3D2-3 effect on non-*Enterobacteriaceae* strains further discount the possibility that the observed shifts are the result of differential growth potential or bacterial interactions that follow repression of a uniform FN3D antibacterial effect due to silencing of the respective *FN3D* transcript. Furthermore, a modest increase in *Asaia* abundance following *FN3D3* silencing, although it did not reach statistical significance compared to the *FN3D1-2* knockdown effect on *Asaia* abundance, further reinforces the suggestion of an FN3D3 discrete mode of action compared to the other FN3Ds.

These observed functional differences between FN3D2 and FN3D3, combined with the different effect of *FN3D1* and *FN3D3* silencing on *Burkholderia* populations (Figure 7.8C-D) and the similar *FN3D1* and *FN3D2* silencing effects in the group A assay (Figure 7.8A) are consistent with the possibility that FN3D1 and FN3D2 exhibit a similar mode of action in comparison to FN3D3, whose mode of action seems to be

discrete from FN3D1-2. This possibility is further strengthened by the identified gene expression modulation following *S. marcescens* infection, in which *LYSC2* upregulation in *FN3D3* knockdown mosquitoes was more pronounced than in *FN3D1-2* knockdown mosquitoes, compared to *dsLacZ* treated controls (Figure 7.5B-D).

7.15 Conclusions and future perspectives

The association with the outcome of *S. marcescens* infection of three previously uncharacterized genes encoding FN3 domains, in different peaks, suggested a functional significance for these genes in shaping the outcome of *S. marcescens* infection. Furthermore, the homology of FN3D2 with the hypervariable pattern recognition receptor Dscam, known to bind bacteria and participate in antibacterial responses (Dong et al., 2012; Dong et al., 2006b; Watson et al., 2005), strengthened the possibility that these FN3Ds may be also involved in antibacterial responses that limit *S. marcescens.* Furthermore, the FN3D2-Dscam homology opened the possibility that the specific recognition capacity shown for AgDscam through alternative splicing (Dong et al., 2012) may involve a broader family of immune factors. Therefore, these *FN3Ds* emerged as prime candidates from the SNP genotyping analysis as a group of novel immune factors, putatively participating in responses that limit *S. marcescens*. The RNAi-mediated silencing data presented in this chapter provide compelling evidence that *FN3D1-3* profusely influence the outcome of *S. marcescens* infection, thus confirming that the association of these *FN3D* genes with the outcome of *S. marcescens* infection pointed to genetic variation causally linked to the *S. marcescens* infection outcome through modulation of the FN3D1-3 antibacterial effect.

Furthermore, based on the presented metagenomic analysis of the *An. gambiae* gut bacterial community, FN3D1-3 emerge as major modulators of the mosquito gut bacterial population structure. FN3D1-3 were shown to limit the representation of *Enterobacteriaceae*, mainly *Serratia* or strains with similarity to *Serratia* reference sequences, but also, for FN3D3, bacteria of the genus *Burkholderia*. Silencing of any of *FN3D1-3* resulted in a dramatic shift of the gut bacterial community structure in favour of a subset of *Enterobacteriaceae*, but also, for *FN3D3* knockdown, bacteria of the genus *Burkholderia*. Shifts in gut bacterial population structure in various organisms, including humans, described under the term dysbiosis, can elicit gut pathology (Lee et al., 2013a; Ryu et al., 2008). The FN3D1-3 effects on mosquito fitness were not investigated here, though, and remain to be determined. Nevertheless, the pathogenicity of *Serratia*, both in mosquitoes (Seitz et al., 1987) and *Drosophila* (Flyg et al., 1980), along with the ability

of *Serratia* to cross the *Drosophila* midgut epithelium and evade systemic immunity (Nehme et al., 2007), suggest that an increased *Serratia* representation in the mosquito gut could influence mosquito fitness.

At the same time, the known anti-*Plasmodium* effect exerted by *Serratia* (Bando et al., 2013; Gonzalez-Ceron et al., 2003), suggests that an increased *Serratia* representation in the mosquito gut could influence the outcome of *Plasmodium* infection. Therefore, immune factors such as FN3D1-3, through shaping *Serratia* representation in the mosquito gut, could influence malaria transmission dynamics. The role of *Serratia* inhabiting the mosquito gut after blood feeding and the contribution of FN3Ds in influencing the outcome of *Plasmodium* infection, in a direct manner or indirectly through modulation of *Serratia* levels, should be further investigated along these lines.

The FN3D1-3 mode of action remains to be determined. The homology of FN3D2 with Dscam and the similar domain architecture of FN3D3 with these immune factors suggest a similar to Dscam role for FN3D2-3 in antibacterial, and, possibly, anti-*Plasmodium* responses. Further functional characterization of FN3D2-3 can shed more light not only in their mode of action but also on a possible interplay between them or with known immune pathways such as the IMD/REL2 pathway.

As mentioned previously, several lines of evidence indicate that FN3D1 and FN3D2 are functionally similar while FN3D3 seems to exhibit a discrete, compared to FN3D1-2, functionality, especially in limiting bacteria of the genus *Burkholderia*. Such relationships between FN3Ds remain to be confirmed, especially with regard to possible physical interactions. The possibility of FN3D1-3 binding to bacteria and whether this binding relies on recognition of molecular patterns found on a subset of the bacterial population or on secreted bacterial-derived metabolites can also be further investigated.

As the mosquito gut microbial community includes, in addition to commensal bacteria, fungi and viruses, the effect of FN3D1-3 on the composition of the mosquito gut bacterial population structure may also influence population dynamics of resident fungal or viral communities. Mosquito defences against entomopathogenic fungi include the immune factors TEP1 and CLIPA8 (Yassine et al., 2012), with TEP1 also involved in antibacterial responses (Levashina et al., 2001) and CLIPA8 known to participate in bacterial melanization (Schnitger et al., 2007). Therefore, shifts of the gut bacterial population structure or immune responses induced by bacteria may influence responses that limit such fungal entomopathogens. Such effects could profoundly influence mosquito physiology and longevity and also modulate the *Plasmodium* infection outcome, given that fungal pathogens can both reduce longevity and influence malaria transmission dynamics (Blanford et al., 2005). The possibility of direct interactions

between bacteria and fungi can be inferred by similar interactions in other organisms, as mouse intestinal microbiota have been shown to influence viral replication (Kuss et al., 2011) while bacteria of the genus *Serratia* inhabiting the *Ae. aegypti* gut have been shown to enhance susceptibility to Dengue-2 virus (Apte-Deshpande et al., 2012). Furthermore, interactions between mosquito gut bacteria, dengue viruses and innate immune responses, not only influence susceptibility to dengue virus but also influence gut microbial load due to the viral presence (Ramirez et al., 2012). Therefore, modulations of fungal or viral populations in FN3D1-3 mediated responses, due to shifts in the mosquito gut bacterial population or induced immune responses, should be taken into account while studying the complex mosquito gut microbial ecosystem.

Another possibility that could be further investigated is whether FN3D1-3 regulate antibacterial responses that indirectly shape the gut bacterial community through differential expression of specialized effectors or through responses that could be evaded by a subset of the bacterial population. An alternative possibility is that FN3D1-3 regulate bacterial interactions, e.g. by sequestering bacterial metabolites which lead to shifts in the bacterial population structure by modulating interactions between bacterial strains with differential capacity in producing such metabolites.

At the same time, it is possible that IMD/REL2 responses may also shape the mosquito gut bacterial population structure by targeting a subset of gut bacteria, possibly in conjunction with various levels of regulation or the existence of bacterial strains able to evade the relevant IMD/REL2-mediated responses. A metagenomic analysis of *PGRPLC* knockdown effects on the mosquito gut microbial population may reveal that *PGRPLC* or certain *PGRPLC* isoforms or alleles, could also shape the composition of the mosquito gut microbial community, as indicated here.

The conclusion to be drawn is that this initial characterization of FN3D1-3 as major modulators of the mosquito gut bacterial community structure raises many intriguing questions regarding the complexity and specificity of the mosquito antibacterial immune responses. Future research efforts could better characterize the FN3D1-3 mode of action and the ecological relevance of *FN3D1-3* genetic variation in mosquito gut homeostatic interactions and malaria transmission dynamics.

Chapter 8

The gustatory receptor Gr9 modulates *Serratia marcescens* levels following oral infection in an NPF-mediated behavioural response

Introduction

8.1 A behavioural component in mosquito responses to *S. marcescens*

The identification of *An. gambiae* genes implicated in responses to oral *S. marcescens* infection, both through genetic association with the infection outcome and transcriptional regulation in response to infection, indicated that oral *S. marcescens* infection triggers two discrete modes of immunity, an epithelial and a behavioural immune component, as several genes, either associated with the outcome of *S. marcescens* infection or differentially expressed in response to infection, suggested the existence of a behavioural immune component. Notably, genes encoding the putative neurotransmitter-triggered GPCRs, GPR5HT7, GPRGBB1 and GPRNPR2, but also the gustatory receptors Gr9 and Gr10, were associated with the outcome of *S. marcescens* infection. Furthermore, genes associated with the *S. marcescens* infection outcome also exhibiting homologies indicating involvement in behavioural modulation included AGAP006405, the orthologue of *Drosophila derailed2*, involved in establishment of olfactory circuits (Sakurai et al., 2009), AGAP012252, the orthologue of *Drosophila PKC53E*, implicated in NPF-mediated alcohol sensitivity (Chen et al., 2008, 2010) and AGAP010503, the orthologue of the *Drosophila SK* channel, implicated in behavioural courtship memory (Abou Tayoun et al., 2012).

Transcriptional responses to oral *S. marcescens* infection also suggested the existence of a behavioural immune component triggered following infection, as several transcripts related to mosquito behaviour showed differential expression following oral *S. marcescens* infection. These included two downregulated transcripts of the gustatory receptor Gr13, three upregulated juvenile-hormone inducible kinases, several up or down regulated juvenile hormone or pheromone binding protein transcripts, including the upregulation of *TO2*, whose *Drosophila* homologue, *takeout*, is known to regulate feeding behaviour (Meunier et al., 2007).

8.2 NPF, a modulator of feeding behaviour

Notably, the known modulator of *Drosophila* feeding behaviour, *NPF*, was downregulated following oral *S. marcescens* infection. Activation of NPF neurons in adult flies has been shown to increase feeding, in an interplay with allatostatin A neurons, whose activation inhibits feeding (Hergarden et al., 2012). Importantly, this neural circuit was shown to modulate *Drosophila* feeding behaviour without affecting gut motility, food excretion or inducing metabolic changes that could result in a state of satiety, suggesting that both NPF and allatostatin A are part of a mechanism that translates various cues, related to food

quality, satiety or hunger, in a direct alteration of feeding behaviour (Hergarden et al., 2012). Interestingly, it was also suggested that such alterations in feeding behaviour might be related to allatostatin A activation in neuroendocrine cells (Hergarden et al., 2012), where NPF is also expressed both in *Drosophila* (Brown et al., 1999) and *Ae. aegypti* (Stanek et al., 2002).

Similarly, in *Drosophila* larvae, NPF expression is related to larval attraction to food whereas *NPF* downregulation results in food aversion (Wu et al., 2003). Furthermore, NPF, along with the NPF receptor NPFR1, have been shown to regulate responses to noxious food in *Drosophila* larvae, with NPFR1 overexpression increasing and NPFR1 knock-out reducing intake of noxious food, in the latter case in hungry fly larvae that would be expected to more readily consume tainted food (Wu et al., 2005b). Therefore, based on the NPF functionality both in *Drosophila* larvae and adults, mosquito *NPF* downregulation might also result in an aversive behavioural response, which would disrupt food intake. Such NPF functionality is consistent with the possibility that ingestion of *S. marcescens* may trigger a behavioural immune response that results in *NPF* downregulation, which, by suppressing mosquito feeding, is expected to disrupt the intake of *S. marcescens* included in the sugar meal, thus limiting infection.

NPF activation affects several additional behavioural processes, including alcohol sensitivity (Wen et al., 2005) and aggression, the latter through an interplay with another neurotransmitter, serotonin (Dierick and Greenspan, 2007). The *Drosophila* NPF-NPFR system suppresses aversive responses triggered by various stressful stimuli, thus promoting a stress-resistant behaviour (Xu et al., 2010). NPF also mediates an anticipatory behavioural immune response in *Drosophila* against endoparasitoid wasps that depends on visual wasp identification, in which recognition of innate search images triggers a modulation of oviposition behaviour through NPF, which serves to protect the fly's offspring from the endoparasitoid predator (Kacsoh et al., 2013).

A fascinating question that remains poorly resolved is the identification of cues that are transduced through the NPF-NPFR neural circuit and exert a modulation in feeding behaviour. Food deprivation seems to be one such cue, and NPF expression in the brain is considered to represent the food-deprived state, regulating behavioural expression of food-associated memory, which is promoted in hungry and constrained in satiated flies (Krashes et al., 2009). Neuropeptide Y, the human homologue of NPF, has been shown to modulate feeding behaviour through circuits regulated by leptin and insulin (Clark et al., 1984; Figlewicz and Benoit, 2009; Saper et al., 2002; Schwartz et al., 1992). Hunger-driven behaviour also

seems to involve insulin-like peptides in *Drosophila*, acting on NPFR1-expressing neurons (Wu et al., 2005a).

Gustatory stimulation by sugar has been shown to activate NPF, suggesting that the *Drosophila* NPF-NPFR circuit may be part of a sensory system that mediates food signalling (Shen and Cai, 2001). NPF expression is also regulated by sexual experience, with the *Drosophila* NPF-NPFR circuit being part of a reward signalling pathway that regulates reward seeking activities such as alcohol consumption (Shohat-Ophir et al., 2012).

8.3 Gustatory receptors and feeding behaviour

The *Drosophila* gustatory receptor Gr43a has been recently shown to be involved in modulation of feeding behaviour in relation to nutrient availability, essentially functioning as a nutrient sensor (Miyamoto et al., 2012; Miyamoto et al., 2013). Gr43a recognizes haemolymph fructose in the brain and either promotes feeding in hungry flies or suppresses feeding in satiated flies (Miyamoto et al., 2012). By recognizing a relatively scarce haemolymph carbohydrate, Gr43a can robustly sense changes in the fly's nutritional status through steep fructose changes, thus acting in a similar fashion to the human insulin and glycagon signalling pathways (Miyamoto et al., 2012). Remarkably, Gr43a activation, reflecting the fly's haemolymph sugar levels, can exert changes in feeding behaviour depending on the fly's satiation status through mechanisms that remain to be determined. One possibility that has been proposed (Miyamoto et al., 2012) is that the NPF-NPFR circuit may modulate the Gr43a brain circuitry. Therefore, differential activation of NPF in hungry or satiated flies may affect the Gr43a-mediated change in feeding behaviour.

8.4 Mosquito host-seeking behaviour

The underlying mechanisms of mosquito behaviour have been studied mainly with regard to host-seeking behaviour, which can directly influence the mosquito vectorial capacity and may also be exploited in intervention efforts aiming to disrupt malaria transmission, as many aspects of mosquito behaviour not only influence vectorial capacity but may also interfere with interventions such as indoor insecticide spraying or release of genetically modified mosquitoes (Pates and Curtis, 2005). Such aspects of mosquito behaviour include the diverse zoophilic or anthropophilic proclivity of different *Anopheles* species (Habtewold et al., 2008; Kamali et al., 2012), i.e. their preference to blood feed from humans or other animals, but also their endophilic or exophilic proclivity (Reddy et al., 2011; Riehle et al., 2011), i.e. their

preference for indoor or outdoor resting, and feeding behaviour such as the indoor or outdoor blood feeding or their biting rate (Smits et al., 1996).

Host-seeking behaviour relies on external stimuli mosquitoes employ to track down their human hosts, including CO_2 exhaled by human hosts, warmth, related to the human host's normal temperature, or water vapour (Klun et al., 2013). Host odour plays a major role in mediating the mosquito's host–seeking behaviour (Takken and Knols, 1999). Various components of human odour, including ammonia, lactic acid and carboxylic acid, are thought to be recognized by mosquitoes and initiate a host-seeking behavioural response (Mukabana et al., 2012; Smallegange et al., 2009; Smallegange et al., 2005). Interestingly, attraction to human odour seems to be enhanced in malaria-infected mosquitoes, in what appears to be a parasite adaptation into completion of their life cycle (Smallegange et al., 2013).

As attraction of mosquitoes to their human hosts varies at the individual level (Mukabana et al., 2002), several human host-derived volatile chemicals have been shown to interfere with mosquito attraction (Logan et al., 2008). What is more, the composition of the human skin microbiota also influences the mosquito host-seeking behaviour, with bacterial-derived molecules thought to influence the mosquito behavioural response (Verhulst et al., 2010; Verhulst et al., 2011).

Despite the importance of the mosquito host-seeking behaviour in influencing malaria transmission dynamics, the molecular mechanisms underlying this behavioural response remain poorly understood. Increased levels of the neurotransmitter dopamine in the brain of *Ae. albopictus* have been shown to reduce host-seeking activity (Fukumitsu et al., 2012). Furthermore, a pheromonotropic neuropeptide, Head Peptide-I, has been shown to inhibit host-seeking behaviour (Matsumoto et al., 1992; Naccarati et al., 2012). This host-derived peptide is transferred to female mosquitoes during copulation by male mosquitoes (Naccarati et al., 2012). Although Head Peptide-I shows similarity to short neuropeptide F and also signals through neuropeptide receptors, the mechanism through which this inhibitory effect is exerted most likely comprises a more complex circuit that remains to be determined (Liesch et al., 2013).

As mosquito olfaction plays a pivotal role in directing the mosquito's host-seeking behaviour, olfactory receptors have been a subject of intense research interest (Hill et al., 2002; Sato et al., 2008; Spehr and Munger, 2009). A *Drosophila* odorant receptor complex has been shown to be a target, at least of one aspect, of the repellant activity of the potent insecticide DEET, with natural polymorphisms in the gene of the odorant receptor OR59B modulating the sensitivity of this complex to odour ligands or DEET (Pellegrino et al., 2011). Furthermore, disrupting a co-receptor of the same complex in *Ae. aegypti*, Orco,

influences the mosquito's host-seeking behaviour and further disrupts the olfactory-related functionality of DEET (Degennaro et al., 2013).

Gustatory receptors are also likely to modulate host-seeking behaviour. The *Drosophila* Gr21a and Gr63a have been shown to mediate CO_2 detection (Jones et al., 2007), while their *Ae. aegypti* counterparts, Gr1 and Gr3, also modulate mosquito sensitivity to CO_2 (Erdelyan et al., 2011).

8.5 Behavioural immune responses

No link has yet to be established between mosquito behaviour and innate immunity. Behavioural immune responses relate to pathogen-induced behavioural modifications aiming to limit or disrupt a pathogenic infection. Such response might be related to prophylactic or therapeutic medication such as oviposition on toxic plants or alcohol-containing food that reduces parasite growth (Kacsoh et al., 2013; Lefevre et al., 2010; Milan et al., 2012; Singer et al., 2009). Another mode of behavioural responses includes aversion mechanisms that limit or disrupt pathogen intake and thus render the infection much more controllable by the immune system. Pathogenic bacteria such as *P. aeruginosa* or *S. marcescens* have been shown to alter odour preference in *C. elegans* in an aversive olfactory learning response aiming to avoid odours related to pathogenic bacteria, while, at the same time, exhibiting attraction to odours from non-pathogenic bacteria, through changes in serotonin expression in chemosensory neurons (Zhang et al., 2005).

A behavioural immune response of *C. elegans* also includes the recognition and avoidance of *S. marcescens* through sensory mechanisms, most likely involving unidentified GPCRs, that recognize a *Serratia* natural product, cyclic lipodepsipentapeptide serrawettin W2 (Pradel et al., 2007). This response is mediated by the nematode Toll-like receptor tol-1 (Pujol et al., 2001).

Feeding suppression following pathogen ingestion has been previously reported in *Manduca sexta*, the tobacco hormworm, following ingestion of *Cotesia congreagata*, a parasitic wasp, or heat-killed *S. marcescens*, in a behavioural response which most likely relies on host-derived octopamine (Adamo, 2005). Interestingly, cessation of feeding has been reported following oral infection with *P. entomophila* in *Drosophila* (Chakrabarti et al., 2012; Liehl et al., 2006; Vodovar et al., 2005), although little is known about the molecular mechanisms underlying this response.

Here, the gustatory receptor Gr9, encoded by a gene associated with the outcome of *S. marcescens* infection, was shown to exert an antibacterial effect that influenced the outcome of *S. marcescens*

infection. Intriguingly, Gr9 was also shown to modulate the mosquito's feeding behaviour and influence the expression levels of NPF, suggesting a possible link between *S. marcescens* levels and mosquito behaviour. Indeed, the Gr9 antibacterial effect was shown to rely on changes in *NPF* expression, pointing to a behavioural immune response that limits *S. marcescens* oral infection.

8.6 Gr9 modulates *S. marcescens* levels following oral infection

The genomic loci encoding the gustatory receptors Gr9 and Gr10 were associated with the outcome of *S. marcescens* infection, found in peak 3R-11 within a 5 kb radius of a SNP with MAF difference >0.5 between the highly and non-infected *S. marcescens* infection phenotypic pools (Figure 5.3). The *Gr9* locus encodes 13 putative splice variants, based on the *AgamP3.7* annotation of the *An. gambiae* genome, compared to just one for *Gr10*. Furthermore, the alternative splicing of *Gr9* has been previously proposed (Hill et al., 2002). Only 5 out of 76 *An. gambiae* gustatory receptors are putatively alternatively spliced, with *Gr9* singled out for its intricate structure, with 14 NH_2-terminal exons, each hypothesized to be alternatively spliced to 2 COOH-terminal exons encoding 7-transmembrane regions (Figure 5.6) (Hill et al., 2002). These intriguing *Gr9* properties led to the hypothesis that causal polymorphisms driving SNP divergence in this genomic area associated with the outcome of *S. marcescens* infection are most likely related to the *Gr9* rather than the *Gr10* locus.

To further investigate any Gr9 effects on the outcome of *S. marcescens* infection, antibiotic treated mosquitoes were microinjected at the day of emergence with dsRNA targeting the 5' exon 1 of the *Gr9* transcript, mainly limiting the AGAP009805-RA and AGAP009805-RG splice variants, or with the *dsLacZ* control. Mosquitoes were orally infected with *S. marcescens*, bacteria-fed mosquitoes were selected at day 2 post infection and, at day 5 post infection, the bacterial load was determined in the guts of *Gr9* or *LacZ* dsRNA treated mosquitoes, using broad-range or *Serratia*-specific bacterial 16S rRNA primers. As can be seen in Figure 8.1, *Gr9* silencing resulted in a dramatic increase of *S. marcescens* abundance, compared to the *dsLacZ* treated control, showing an average 47 and 35.7-fold increase in bacterial load using 16S broad-range or *Serratia*-specific primers, respectively, as determined over 5 independent infections.

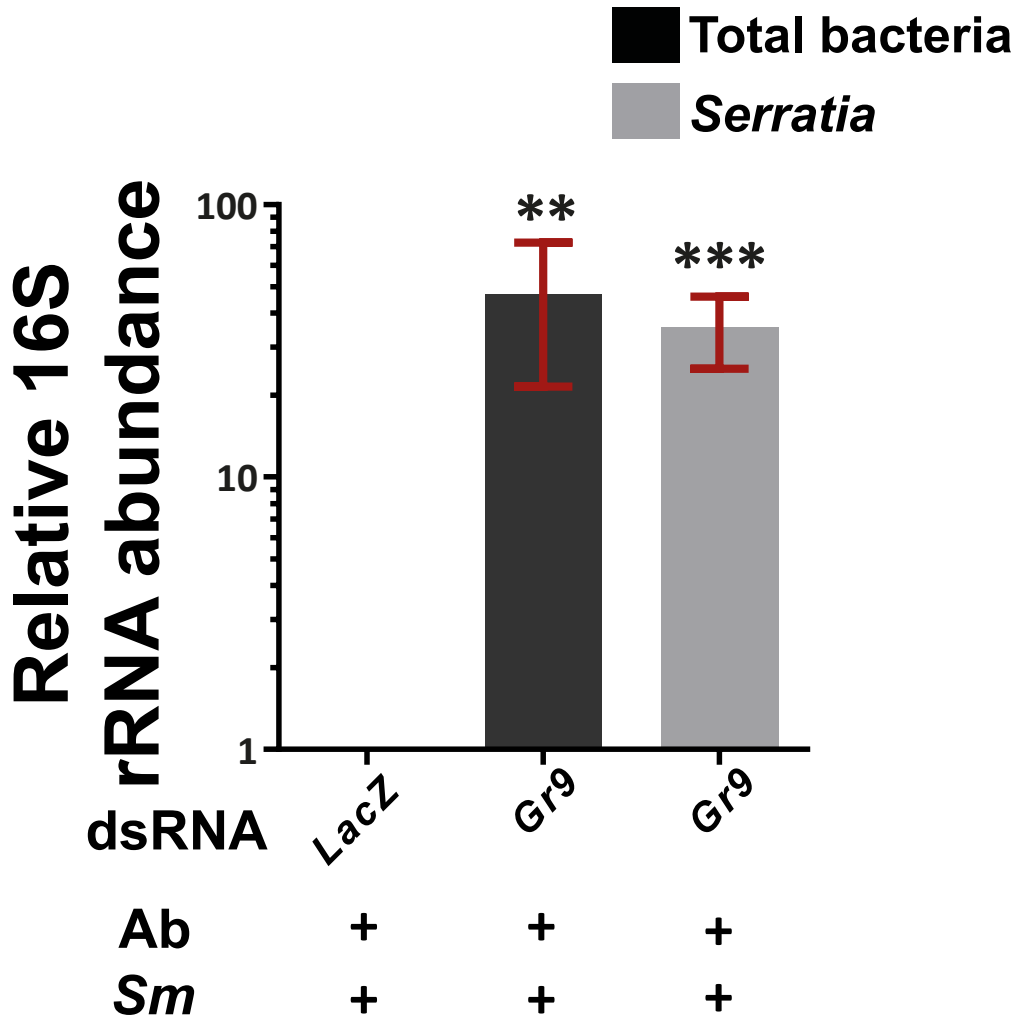


Figure 8.1: *Gr9* silencing increases *S. marcescens* abundance in orally infected mosquitoes. Antibiotic treated mosquitoes, microinjected with *Gr9* dsRNA or the *dsLacZ* control, were orally infected with *S. marcescens* (designated as *Ab+Sm+*). Bacteria-fed mosquitoes were selected 2 days post infection and bacterial load in the guts of infected mosquitoes was determined for each dsRNA treatment 5 days post infection. Extracted total RNA was used for cDNA synthesis, serving as template in a qRT-PCR using broad-range or *Serratia*-specific bacterial 16S rRNA primers as well as primers for the housekeeping *AgS7* transcript, used as an endogenous control. Relative bacterial abundance, as determined with both broad-range and *Serratia*-specific primers, was normalized to the respective *dsLacZ* control. The average ±SEM fold-change increase in bacterial load can be seen, as determined over 5 independent infections, with the qRT-PCR performed at least in duplicate in each case. Significant differences were assessed by a t-test against zero using the log2-transformed fold-change values for the *Gr9* knockdown treatments, determined either by broad-range or *Serratia*-specific primers, where zero corresponds to no difference from the *dsLacZ* treatment. Asterisks indicate significance with a p-value <0.005 or <0.0005, indicated by 2 or 3 asterisks, respectively.

The fold-change increase in bacterial abundance following *Gr9* knockdown in *S. marcescens* infected mosquitoes is comparable to the respective fold-change increases observed following *PGRPLC* or *FN3D1-3* knockdown following oral *S. marcescens* infection. Although differences in fold-change increases were not statistically significant and could be related to the RNAi-mediated silencing efficiency, it is notable that the achieved fold-change increases following *Gr9* knockdown were higher than any of the respective increases in *PGRPLC* or *FN3D1-3* knockdown mosquitoes, regarding the broad-range 16S primers. Using *Serratia*-specific primers, the *Gr9* knockdown fold-change increase was higher than *PGRPLC* or *FN3D3* knockdown but lower than the achieved fold-change increases in *FN3D1* or *FN3D2* knockdown mosquitoes.

As hypothesized previously for the *PGRPLC* and *FN3D1-3* knockdown assays, differences in fold-change increases between the broad-range and *Serratia*-specific bacterial 16S rRNA primers might be due to the existence of residual bacteria evading the antibiotic treatment, which might be increased along with *S. marcescens* following silencing of the respective transcript. Although binding efficiency differences might also account for any diverse observed effects between broad-range and *Serratia*-specific primers, it has been notable that *PGRPLC* knockdown resulted in higher fold-change increases using broad-range rather than *Serratia*-specific primers, while *FN3D1-3* knockdown resulted in higher fold-change increases using *Serratia*-specific rather than broad-range primers (Figure 7.3). Taking into account the FN3D1-3 effects on shaping the gut microbiota population structure by limiting mainly *Enterobacteriaceae* (Figure 7.8), this observation is consistent with the possibility that residual antibiotic-resistant bacteria might be targeted by PGRPLC but not FN3D1-3.

Interestingly, *Gr9* knockdown showed higher fold-change increases with broad-range rather than *Serratia*-specific primers in orally infected mosquitoes, as observed for *PGRPLC* knockdown. Therefore, the Gr9 antibacterial effect, in this regard, resembles the PGRPLC effect by being putatively exerted on a larger subset of gut bacteria than FN3D1-3. Although the effect of *Gr9* knockdown on mosquitoes retaining their natural gut microbiota was not examined, future assays, perhaps along with a metagenomic analysis of bacterial population shifts following *Gr9* silencing, can be used to confirm this possibility.

8.7 Gr9 is a modulator of mosquito feeding behaviour

The robust Gr9 effect in influencing the outcome of *S. marcescens* infection raised the possibility that this Gr9 antibacterial activity might be related to changes in feeding behaviour. Enhanced or suppressed feeding can decisively influence the *S. marcescens* abundance that the mosquito takes in. The abundance

of orally ingested *S. marcescens* may in turn influence the ability of the epithelial immune response to handle the infection. Therefore, an increased or reduced number of ingested bacteria, in conjunction with the efficiency of the epithelial immune component to tackle the infection, can influence the outcome of *S. marcescens* infection.

The *Gr9 Drosophila* orthologues, *Gr32a*, *Gr39a* and *Gr68a*, have been characterized as pheromone receptors (Bray and Amrein, 2003; Miyamoto and Amrein, 2008), involved in regulation of mating behaviour (Watanabe et al., 2011), with *Gr32a* also involved in modulation of aggressive behaviour following pheromone recognition (Wang et al., 2011a) but also feeding suppression related to recognition of the potent insecticide DEET or other antifeedants (Lee et al., 2010). These *Gr9* homologies suggest a Gr9 role in recognition of chemosensory cues that might lead to modulation of feeding behaviour, as is the case with Gr32a (Lee et al., 2010). Importantly, though, the function of the *Gr9 Drosophila* orthologues was studied in relation to their expression in external sensory organs located in the fly's forelegs (Bray and Amrein, 2003) or mouthpart (Lee et al., 2010). *Gr9*, on the other hand, shows significant overexpression in the midgut (Baker et al., 2011; Marinotti et al., 2006) and only moderate expression in external sensory organs (Pitts et al., 2011), suggesting a Gr9 functional divergence compared to its *Drosophila* orthologues.

Based on these *Gr9* homologies, two hypotheses were formed regarding a putative Gr9 role in modulation of feeding behaviour. As *Gr9* showed no homology with *Drosophila* gustatory receptors that recognize sugars (Jiao et al., 2007), it was considered more likely that Gr9 is triggered by chemosensory cues, being either bacterial-derived metabolites or mosquito-derived infection-induced molecules. The subsequent Gr9 activation might lead to an aversive behavioural response aimed to limit or disrupt pathogen intake. Alternatively, Gr9 might modulate feeding behaviour through recognition of nutrients or mosquito-derived molecules and only indirectly influence the outcome of *S. marcescens* infection, with *Gr9* variants affecting its physiological role in modulation of mosquito feeding behaviour.

To further investigate a possible Gr9 role in modulation of feeding behaviour, the effect of *Gr9* silencing on mosquito feeding behaviour was investigated through a modified capillary feeder (CAFE) assay. The CAFE assay has been established in *Drosophila*, aiming to study feeding behaviour through measuring ingestion by individual or groups of flies, on a scale from minutes to days (Ja et al., 2007). Modified CAFE assays have also been used to study feeding behaviour in mosquitoes (Liesch et al., 2013). The concept behind the CAFE assay can be seen in a schematic diagram in Figure 8.2A. Flies, or mosquitoes, are placed inside a vial or cup and offered to feed through a glass capillary. Water is placed in an outside chamber to

humidify the feeder while the outside opening of the capillary is sealed through mineral oil or parafilm. The solution inside the capillary contains a dye so that consumption can be assessed through the reduced abundance of the coloured solution, measured by the extent of the empty portion of the capillary. Although evaporation is thought to be limited due to the blockage of the outside opening of the capillary, any such evaporation is averaged out between treatments or, if one treatment is assayed, by a control vial with no fly or mosquito inside. Although feeding behaviour could conceivably be affected by restricted food accessibility through the capillary, no inhibition of feeding has been shown due to the conditions of the CAFE assay in *Drosophila* (Ja et al., 2007).

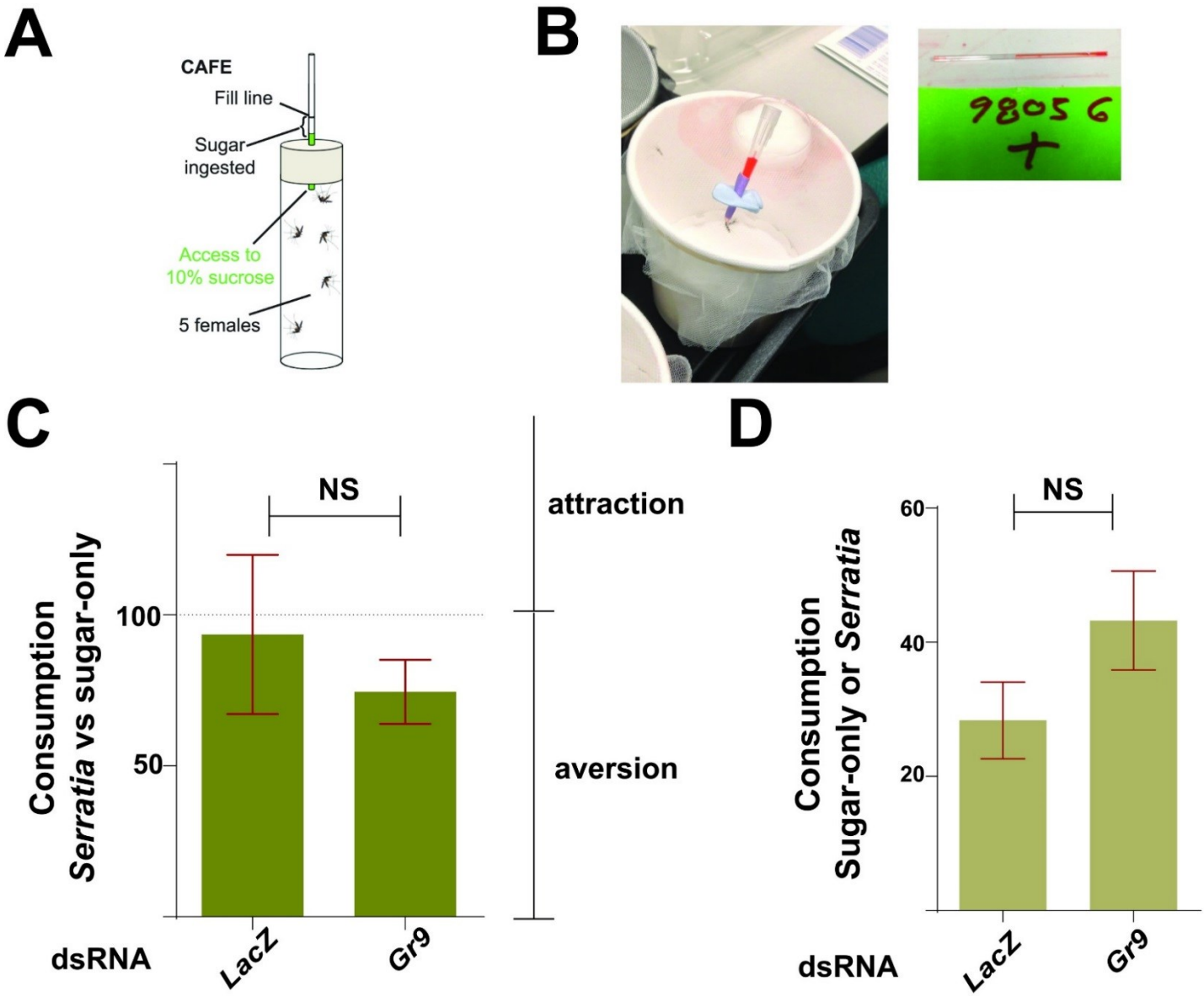

A
CAFE
Fill line
Sugar ingested
Access to 10% sucrose
5 females
B
9805 G
+
C
NS
attraction
aversion
Consumption Serratia vs sugar-only
100
50
dsRNA
LacZ
Gr9
D
NS
Consumption Sugar-only or Serratia
60
40
20
dsRNA
LacZ
Gr9

Figure 8.2: A two-choice preference assay to identify behavioural modulations in *Gr9* knockdown mosquitoes. A: Schematic diagram of the CAFE assay. Mosquitoes are placed inside a vial and are offered access to a 5 µl capillary containing a sugar solution. A dye also contained in the solution is used to determine consumption. The outside opening of the capillary is sealed with an oil layer on top of the sugar solution or with parafilm to minimize evaporation. Adapted from: (Liesch et al., 2013). **B:** The modified CAFE assay as employed here. A group of 8-11 mosquitoes were placed inside a pot sealed on top with a net to prevent mosquito flight. A smaller pot was used when placing a single mosquito inside instead of a group of mosquitoes. Mosquitoes were offered to feed from one or two capillaries containing a sugar solution with a dye, after being starved overnight. In the two-choice preference assay, one of the capillaries also contained *S. marcescens* while all capillaries also contained tetracycline and carbenicillin. A piece of cotton dampened with water was also placed on top of the pot to humidify the pot enclosure. Mosquitoes were let to feed for 16 hours. After that period, the presence of the dye-containing solution was evident in the gut of at least some mosquitoes, indicating that mosquitoes were able to ingest the solution through the capillary. Consumption was determined by measuring the proportion of the capillary rendered empty, as shown in the photo to the right of the panel. Possible reduction of the solution contained in the capillary due to evaporation was averaged out by normalization to the suitable control, either a capillary not containing *S. marcescens* or by comparison to the *dsLacZ* treatment. **C:** Antibiotic treated mosquitoes, treated either with *Gr9* or *LacZ* dsRNA, were starved and, 5 days post dsRNA treatment, were placed in groups of 8-11 mosquitoes and were offered to feed from 2 capillaries, one containing *S. marcescens* and one only sugar. The % consumption in the respective capillary was determined 16 hours later and the ratio of the % consumption in the *S. marcescens* containing capillary to the % consumption in the sugar-only capillary was determined for each *dsLacZ* or *Gr9* dsRNA treated pool. The average ±SEM of the ratio for each dsRNA treatment can be seen as determined in 6 *dsLacZ* and 11 *Gr9* dsRNA treated pools. Significant differences were assessed using the non-parametric Mann-Whitney test, with a resulting p-value of 0.7395, indicated as non-significant (NS). A percentage ratio of <100% would indicate aversion to the *S. marcescens* containing solution while a percentage ratio of >100% would indicate attraction to the *S. marcescens* containing solution for mosquitoes treated with the respective dsRNA, as indicated. **D:** The % consumption in each *S. marcescens* containing or sugar-only capillary in the two-choice preference assay of panel C was used to determine consumption in *dsLacZ* or *Gr9* dsRNA treated mosquitoes, irrespective of the presence of *S. marcescens*. The average ±SEM consumption for 12 and 22 *LacZ* or *Gr9* dsRNA treated mosquitoes, respectively, can be seen for each dsRNA treatment. Significant differences were assessed by a non-parametric Mann-Whitney test, resulting in a p-value of 0.288, indicated as non-significant (NS).

The modified CAFE assay used can be seen in Figure 8.2B. A group of 8 to 11 mosquitoes was placed inside a pot covered by a net, as used in all other performed assays in this study, and offered to feed from a glass capillary, sealed on the outside by parafilm. Distilled water (dH_2O) was provided through a wet piece of cotton. Mosquitoes were let to feed for 16 hours and, subsequently, the capillaries were removed and the empty portion of the capillary, corresponding to the consumed solution, was measured.

The CAFE assay was utilized to investigate whether *Gr9* silencing might modulate aversion to the *S. marcescens* containing sugar solution. This possibility was examined through a two-choice preference approach. Mosquitoes were treated at the day of emergence with *Gr9* dsRNA or the *dsLacZ* control and were antibiotic treated for 5 days. Subsequently, they were offered to feed from 2 capillaries, one containing a sugar solution with a dye and the relevant antibiotics and one also containing *S. marcescens*, at the same concentration as the solution used for oral infections described in chapter 4. The difference between consumption in the *S. marcescens* containing and the sugar-only capillary was determined for each *Gr9* or *LacZ* dsRNA treated group. Overall, the consumption difference was determined in 6 *dsLacZ* and 11 *Gr9* dsRNA treated pools, in 3 independent assays (Figure 8.2C), with no statistically significant differences identified between *Gr9* and *LacZ* dsRNA treated mosquitoes.

These data suggest that *Gr9* silencing does not affect food consumption due to the presence of *S. marcescens* in the sugar meal. Notably, though, both the *Gr9* and *LacZ* dsRNA treated pools showed an average consumption difference between *S. marcescens* containing and sugar-only capillaries which indicated reduced consumption for the *S. marcescens* containing capillary (Figure 8.2C), thus indicating a Gr9-independent aversion mechanism. Although the margins of error were considerable for the *dsLacZ* treatment due to an outlier showing increased consumption for the *S. marcescens*-containing capillary, any differences to the *Gr9* dsRNA treatment were non-significant and might also be related to the lower number of assayed pools for the *dsLacZ* treatment.

Interestingly, observation of *Gr9* dsRNA treated mosquitoes which had ingested the dye-containing sugar solution, irrespective of the *S. marcescens* presence, indicated that some mosquitoes seemed to have ingested an inordinate amount of sugar solution, evident by the intensity of the dye in their guts. This observation prompted an alternative analysis of the same two-choice preference assay, shown in Figure 8.2D. The sugar solution consumption from both the sugar-only and the *S. marcescens* containing capillaries was determined for *Gr9* and *LacZ* dsRNA treated mosquitoes. While *dsLacZ* treated mosquitoes showed an average 28.4% consumption of sugar solution from the capillaries, irrespective of the presence of *S. marcescens*, *Gr9* dsRNA treated mosquitoes consumed an average of 43.25% of sugar solution from

their capillaries. Assessed significance between the *Gr9* and *LacZ* dsRNA treated pools returned a non-significant p-value of 0.288, most likely due to the limited number of assessed data points. These data provide an initial indication that Gr9 might modulate sugar consumption irrespective of the presence of *S. marcescens*.

To further investigate a possible Gr9 effect on sugar consumption, mosquitoes were treated with *Gr9* or *LacZ* dsRNA at the day of emergence and were antibiotic treated for 5 days, as previously. Subsequently, they were placed in individual pots and were offered a sugar meal through a single capillary. The identification of sugar consumption in individual rather than groups of mosquitoes aimed to increase the statistical power of the assay. No water-containing piece of cotton was placed in these pots, so as to increase consumption through the capillary for individual mosquitoes but also due to steric constraints. Sugar consumption in the capillaries of *Gr9* or *LacZ* dsRNA treated mosquitoes was determined 16 hours later through the reduction of the dye-containing sugar solution in each respective capillary.

Overall, consumption was determined in 38 *LacZ* and 55 *Gr9* dsRNA treated mosquitoes (Figure 8.3). Remarkably, *Gr9* dsRNA treated mosquitoes showed a significantly higher sugar consumption, at an average of 86.5%, compared to *dsLacZ* treated mosquitoes, which showed sugar consumption at an average of 59.6%. The higher consumption proportions of sugar solution, compared to the previous two-choice preference assay, could be related to the use of water-containing pieces of cotton in the two-choice preference assay but also to differences in the experimental design, mostly related to the use of individual or groups of mosquitoes, although the influence of stochastic environmental variables or fitness differences between mosquito batches cannot be excluded.

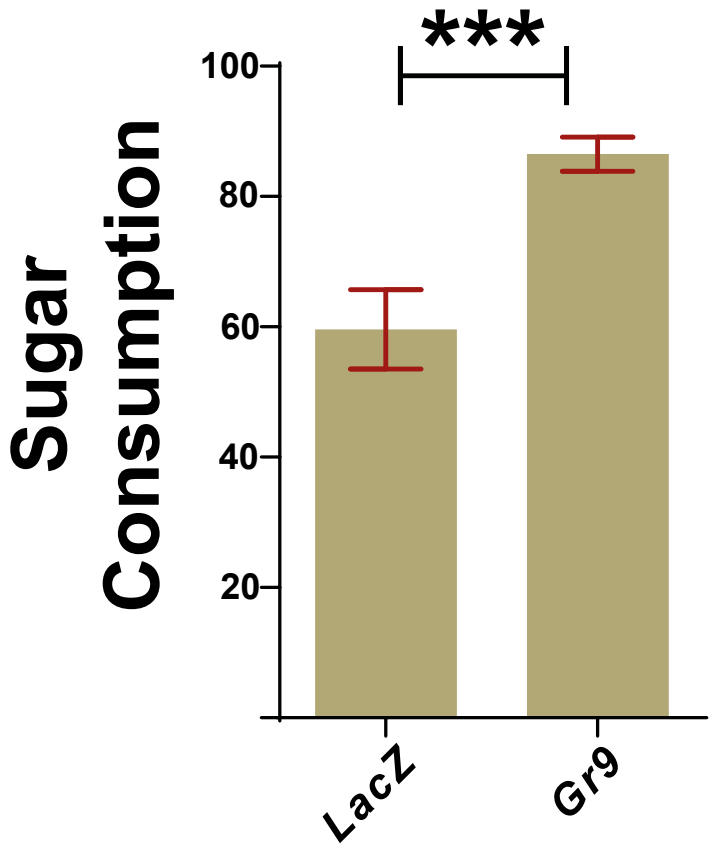


Figure 8.3: *Gr9* silencing increases mosquito sugar consumption. Antibiotic treated mosquitoes were treated with *Gr9* or *LacZ* dsRNA and, 5 days later, they were starved overnight and were subsequently placed in individual pots in which they were offered to feed from a capillary containing a sugar solution along with a dye. Consumption was determined 16 hours later through the reduction of the sugar solution in each capillary. Pots showing mosquito mortality were discarded. The average ±SEM of % consumption of the sugar solution, as determined in 38 *dsLacZ* and 55 *Gr9* dsRNA treated mosquitoes can be seen for each dsRNA treatment. Significant differences were assessed by the non-parametric Mann-Whitney test. Asterisks indicate a p-value <0.0005.

These data suggest that *Gr9* silencing increases sugar consumption, irrespective of the presence of bacteria. This Gr9 effect on feeding behaviour is consistent with the increase in *S. marcescens* levels following *Gr9* silencing. As *Gr9* knockdown mosquitoes ingest a higher abundance of sugar solution, they also ingest a higher abundance of *S. marcescens* contained in the sugar solution. It is likely that this increased *S. marcescens* intake can affect the efficiency of the epithelial immune component to handle the infection, and, possibly beyond a threshold, after which epithelial responses cannot efficiently reduce *S. marcescens* levels, this increased *S. marcescens* intake may influence the outcome of *S. marcescens* infection.

8.8 The Gr9 antibacterial effect mostly relies on changes in *NPF* expression

The downregulation of the known modulator of feeding behaviour, *NPF*, following oral *S. marcescens* infection, raised the possibility of a behavioural immune response, in which *S. marcescens* infection triggers the downregulation of *NPF*, resulting in feeding suppression, as would be expected based on the NPF functionality in *Drosophila* (Hergarden et al., 2012), thus limiting or disrupting the *S. marcescens* intake through the oral route.

As *Gr9* knockdown was shown to increase sugar consumption, an intriguing possibility of a functional interplay between Gr9 and NPF emerged. As Gr9 functionality is expected to inhibit feeding, such inhibition could be related to *NPF* downregulation, also expected to inhibit feeding (Hergarden et al., 2012). Both *Gr9* (Baker et al., 2011; Marinotti et al., 2005) and *NPF* (Brown et al., 1999; Stanek et al., 2002) are expected to be expressed in the mosquito midgut. What is more, the *Gr9 Drosophila* orthologue, *Gr39a*, has been shown to be expressed in enteroendocrine chemosensory cells in the fly's intestine, where Gr39a co-localizes with NPF (Park and Kwon, 2011), further strengthening the possibility of a functional interplay between Gr9 and NPF.

To further investigate a Gr9-NPF functional link, the possibility that *Gr9* silencing might modulate *NPF* expression in mosquitoes orally infected with *S. marcescens* was examined. Antibiotic treated mosquitoes were treated with *Gr9* dsRNA or the *dsLacZ* control and, 5 days later, they were orally infected with *S. marcescens*. Bacteria-fed mosquitoes were selected 2 days post infection and, at day 5 post infection, total RNA was extracted from mosquito guts and used for cDNA synthesis, representing the transcripts present in the gut of orally infected mosquitoes, treated either with *Gr9* or *LacZ* dsRNA, which were used to determine *NPF* levels by qRT-PCR. Indeed, *Gr9* knockdown mosquitoes orally infected with *S. marcescens* showed a significant 1.77-fold increase in *NPF* transcript levels, compared to *dsLacZ* treated

mosquitoes also infected with *S. marcescens*, as determined over 8 independent infections (Figure 8.4). These data suggest that the enhanced feeding behaviour due to *Gr9* knockdown could be related to increased *NPF* levels. Furthermore, as *NPF* was downregulated following *S. marcescens* infection, it is possible that Gr9 might mediate *NPF* downregulation in response to *S. marcescens* infection.

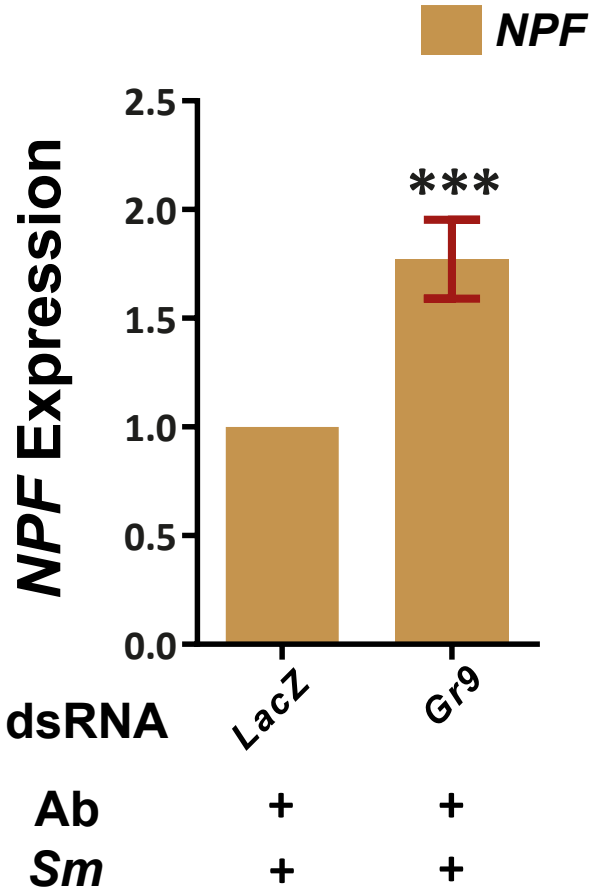


Figure 8.4: *NPF* levels are elevated in *Gr9* knockdown mosquitoes following oral *S. marcescens* infection. Mosquitoes were treated with *Gr9* or *LacZ* dsRNA at the day of emergence and were further antibiotic treated for 5 days. Subsequently, they were orally infected with *S. marcescens* and, 5 days post infection, *NPF* levels were determined in their guts, using qRT-PCR with primers targeting the *NPF* transcript and primers for the *AgS7* housekeeping transcript, used as an endogenous control. Relative *NPF* abundance in *Gr9* knockdown pools was normalized to the *NPF* abundance in the respective *dsLacZ* treated control. The average ±SEM of the normalized *NPF* fold-change values can be seen, as determined over 8 independent infections. Statistically significant differences were assessed in a t-test against zero, where zero corresponds to no change from *dsLacZ* treatment, using the log2-transformed fold-change values of *NPF* expression in *Gr9* knockdown mosquitoes. Asterisks indicate statistical significance with a p-value <0.0005.

The possibility that the Gr9 effect on feeding behaviour, and most likely also on the outcome of *S. marcescens* infection, is exerted through modulation of *NPF* expression, entails that the level of *NPF* expression also influences the outcome of *S. marcescens* infection. To further demonstrate that the Gr9 antibacterial effect relies on changes in NPF expression, antibiotic treated mosquitoes were treated either with a 50%-50% mix of dsRNA targeting the *Gr9* and *NPF* transcripts, or with dsRNA targeting the *Gr9* transcript and the *dsLacZ* control, or just with *dsLacZ*. Mosquitoes were orally infected with *S. marcescens*, bacteria-fed mosquitoes were selected 2 days post infection and, at day 5 post infection, *S. marcescens* levels were determined in the mosquito gut. Remarkably, concomitant silencing of *Gr9* and *NPF* resulted in a statistically significant, about 10-fold, decrease of bacterial load compared to *Gr9* silencing alone, as determined by *Serratia*-specific 16S bacterial rRNA primers, over 3 independent infections (Figure 8.5A). While *Gr9* silencing increased *S. marcescens* levels by an average of 104.6-fold, as determined over 3 independent infections, concomitant *Gr9* and *NPF* silencing increased *S. marcescens* abundance by only an average of 9.1-fold.

Bacterial load determined by broad-range 16S bacterial rRNA primers exhibited the same trend but did not result in statistically significant differences between dsRNA treatments due to considerable margins of error in the *Gr9-LacZ* dsRNA treatment (Figure 8.5B). It is possible that this difference between broad-range and *Serratia*-specific primers might be related to primer efficiency at high levels of bacterial 16S rRNA template abundance. The observed trend, though, suggests that additional replicates would most likely lead to statistically significant differences between the *Gr9-LacZ* and *Gr9-NPF* dsRNA treatments, also by using broad-range bacterial 16S rRNA primers.

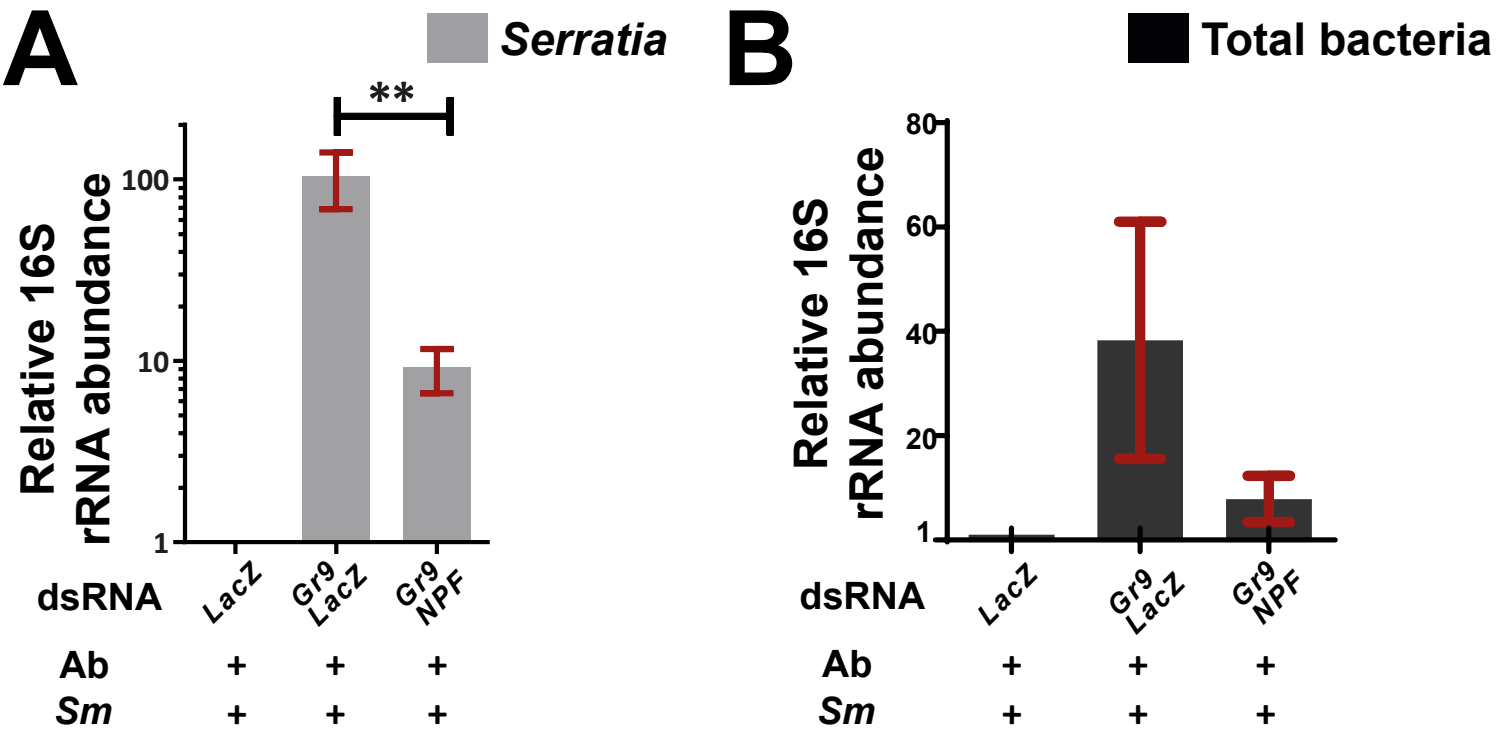


Figure 8.5: The Gr9 antibacterial effect mostly relies on changes in *NPF* expression. Mosquitoes were treated at the day of emergence with a 50%-50% mix of *Gr9* and *NPF* dsRNA or a 50%-50% mix of *Gr9* and *LacZ* dsRNA or just with the *dsLacZ* control. After 5 days of antibiotic treatment, mosquitoes were orally infected with *S. marcescens* (designated as *Ab+Sm+*), bacteria-fed mosquitoes were selected 2 days post infection and, at day 5 post infection, bacterial load was determined in the guts of surface sterilized mosquitoes using qRT-PCR with *Serratia*-specific (panel A) or broad-range (panel B) bacterial 16S rRNA primers. Relative bacterial abundance, for *Serratia*-specific or broad-range bacterial 16S rRNA primers, was normalized to the corresponding bacterial abundance in *dsLacZ* treated mosquitoes. The average ±SEM of fold-change increases can be seen for each dsRNA treatment, as determined over 3 independent infections, with the qRT-PCR performed at least in duplicate in each case. Statistically significant differences between the *Gr9-LacZ* and *Gr9-NPF* dsRNA treatment fold-change increases were assessed by the non-parametric Mann-Whitney test. Asterisks indicate statistical significance, with a p-value <0.005.

Taken together, these data point to a behavioural immune response involving Gr9 and exerted through modulation of NPF expression levels. As the Gr9 antibacterial effect relies on changes in NPF expression, the possibility that emerges is that the Gr9-dependent modulation of feeding behaviour is transduced through the NPF-NPFR neural circuit, thus explaining the NPF mediation of the Gr9 antibacterial effect. Although it was not demonstrated here, concomitant silencing of *Gr9* and *NPF* would be expected to reverse the enhanced sugar consumption shown following *Gr9* silencing alone (Figure 8.3), in line with the previously reported functionality of NPF in *Drosophila* (Hergarden et al., 2012). Silencing *NPF* alone, though, is unlikely to produce any changes in the outcome of *S. marcescens* infection, as *NPF* is downregulated in *S. marcescens* infected mosquitoes. *NPF* overexpression would be expected, though, to increase *S. marcescens* levels through enhanced feeding behaviour, although overexpression or tissue-specific modulation of gene expression cannot be routinely performed in *An. gambiae* so as to confirm this possibility

8.9 Efficiency of *Gr9-NPF* RNAi-mediated silencing

As the *Gr9-NPF* double-knockdown showed an average of 9.1-fold increase in bacterial load, using *Serratia*-specific 16S rRNA primers, compared to the *dsLacZ* treatment in *S. marcescens* infected mosquitoes, it is possible that Gr9 might also exert an antibacterial effect in an NPF-independent manner. Another possibility is that the increase in *S. marcescens* abundance when both *Gr9* and *NPF* are silenced is related to inefficient *NPF* silencing. To further investigate this possibility, the *Gr9* and *NPF* transcript levels were determined 5 days post dsRNA treatment in mosquitoes treated with *Gr9* or a 50%-50% mix of *Gr9* and *NPF* dsRNA, respectively (Figure 8.6). In both cases, treatment with *Gr9* or *Gr9-NPF* dsRNA moderately decreased the respective *Gr9* or *NPF* transcript levels in the mosquito gut to an average of 67.9% and 70.6%, respectively, compared to *dsLacZ* treated mosquitoes. The efficiency of RNAi-mediated silencing of *Gr9* and *NPF* is comparable to the efficiency of *PGRPLC* or *FN3D1-3* knockdown, also assessed in the midgut. Furthermore, these data confirm the expression of *Gr9* and *NPF* in the mosquito gut.

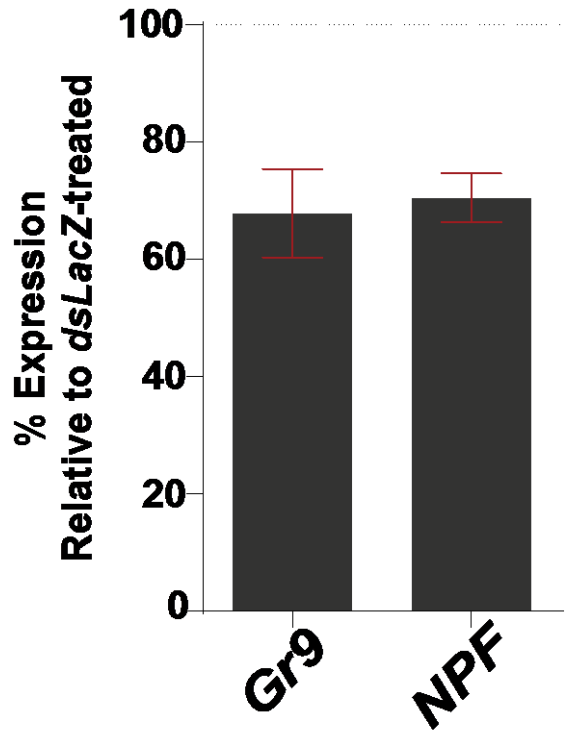


Figure 8.6: Efficiency of *Gr9* and *NPF* RNAi-mediated silencing. Mosquitoes were treated with *Gr9* or *LacZ* dsRNA or a 50%-50% mix of *Gr9* and *NPF* dsRNA. Following 5 days of antibiotic treatment, dissected guts from groups of 8-10 mosquitoes, corresponding to each dsRNA treatment, were used for total RNA extraction and, subsequently, for cDNA synthesis. The level of *Gr9* transcript expression was determined in *Gr9* and *LacZ* dsRNA mosquitoes using qRT-PCR with primers targeting the *Gr9* transcript, in a 5' exon 1 region outside the region targeted by the *Gr9* dsRNA. Primers targeting the *AgS7* housekeeping transcript were used as an endogenous control. Relative *Gr9* transcript abundance in *Gr9* knockdown mosquitoes was normalized to the corresponding *Gr9* expression in *dsLacZ* treated mosquitoes. The average ±SEM of the % *Gr9* expression in *Gr9* knockdown mosquitoes, relative to *dsLacZ* treated mosquitoes, can be seen, as determined over 3 independent assays. Similarly, the *NPF* relative abundance was determined, using primers targeting the *NPF* transcript, in *Gr9-NPF* knockdown mosquitoes and was normalized to the corresponding *NPF* expression in *dsLacZ* treated mosquitoes. The average ±SEM of the % *NPF* expression in *Gr9-NPF* knockdown mosquitoes, relative to the *dsLacZ* control, can be seen, as determined over 3 independent assays.

8.10 A constitutive or infection-induced Gr9-NPF interplay in modulation of mosquito feeding behaviour

These data are consistent with the possibility that Gr9 constitutively limits sugar consumption through modulation of mosquito feeding behaviour. *Gr9* variants with differential efficiency in limiting sugar consumption may affect the abundance of *S. marcescens* intake through the oral route, thus affecting the outcome of *S. marcescens* infection. This possibility could explain the *Gr9* association with the infection outcome but also the increased *S. marcescens* levels in *Gr9* knockdown mosquitoes. It is also possible that the Gr9-mediated modulation of feeding behaviour could rely on the mosquito satiation status or could be further triggered following *S. marcescens* infection. Based on the data presented here, the Gr9 trigger is likely to be a mosquito-derived molecule.

The *NPF* downregulation following oral *S. marcescens* infection suggests an NPF role in response to infection. The increased *NPF* levels in *Gr9* knockdown mosquitoes orally infected with *S. marcescens* suggest that Gr9 might signal the *NPF* downregulation following oral *S. marcescens* infection. Alternatively, a different cue might lead to *NPF* downregulation following oral infection, as NPF is known to integrate various signals that modulate diverse behavioural aspects (Shohat-Ophir et al., 2012). The possibility of an NPF role in integrating satiation and aversion signals in *An. gambiae* cannot be ruled out and might entail both an indirect NPF-mediated influence on the outcome of *S. marcescens* infection but also an infection-induced effect aimed to disrupt pathogen intake.

Discussion

8.11 Gustatory receptor-mediated behavioural modulations

The data presented here suggest that Gr9 exerts an antibacterial effect through modulation of mosquito feeding behaviour that relies on changes in *NPF* expression. Gustatory receptors, both in *Drosophila* and *Anopheles*, constitute a diverse family of receptors that recognize a variety of chemosensory cues, resulting in behavioural modulations (Montell, 2013). Many questions regarding the gustatory receptor mode of action remain unanswered, including not only the cues that selectively trigger some gustatory receptors but also the possible interplay between different gustatory receptors in the formation of multimeric complexes and the way gustatory receptor activation relays the signal that modifies the fly's or the mosquito's behaviour. The Gr43a satiation-dependent feeding enhancement or suppression suggests that complex interactions are at play that influence gustatory receptor functionality (Miyamoto et al., 2012).

8.12 Future perspectives on Gr9-mediated modulations of bacterial load and mosquito feeding behaviour

The association of *Gr9* with the outcome of *S. marcescens* infection and the potent Gr9 antibacterial effect shown through reverse genetics, led to the formation of two hypotheses regarding the Gr9 mode of action: a putative Gr9 involvement in epithelial immunity, along the lines of mammalian chemoattractant receptor involvement in immune responses, leading, in some cases, to gut microbiota modulation (Maslowski et al., 2009; Serhan et al., 2008), or a Gr9-mediated behavioural response related to modulation of feeding behaviour. Both the Gr9 effect on sugar consumption and the reliance on *NPF* expression levels for the Gr9 antibacterial effect, point to a Gr9 involvement in a behavioural response. Although the Gr9 modulation of feeding behaviour does not seem to rely on the presence of bacteria, it remains to be determined whether Gr9 is triggered following *S. marcescens* infection and whether Gr9 is involved in the mediation of the *NPF* downregulation. Alternatively, Gr9 might be constitutively triggered, most likely in an interplay with other modulators of feeding behaviour, with the *NPF* downregulation following *S. marcescens* infection relying on integration of other, Gr9-independent, aversion circuits.

Further research is required to identify whether the Gr9 behavioural effect is due to increased meal size or an increased number of meals, two modes of feeding behaviour which are controlled by different neural circuits (Al-Anzi et al., 2010; Ja et al., 2007). Further confirmation of the Gr9 effect can also be

provided by the quantification of food consumption using colourimetric methods (Hergarden et al., 2012) or by quantifying ingested fructoce, using thin layer chromatography (Miyamoto et al., 2012; Nunes et al., 2008) or mass spectrometry, in *Gr9* knockdown mosquitoes and suitable controls.

Since the two-choice preference assay indicated that the presence of *S. marcescens* does not significantly affect sugar consumption following *Gr9* silencing (Figure 8.2C), it is very likely that *Gr9* knockdown also increases intake of a *S. marcescens* containing sugar solution. Nevertheless, Gr9-independent aversion circuits could conceivably taper overall sugar consumption both in the *dsLacZ* and *Gr9* dsRNA treated pools, accounting for the non-significant increase in sugar consumption in *Gr9* knockdown mosquitoes, as determined in the two-choice preference assay based on consumption of both sugar-only and *S. marcescens* containing solutions (Figure 8.2D). It is also possible, though, that such an aversion circuit may not be relevant in the employed infection model, in which mosquitoes are offered with a bacteria-containing solution and have no choice between water-only, sugar-only or bacteria containing solutions. Indeed, the need to starve mosquitoes prior to infection, as mentioned previously, served not only to clear mosquitoes from the antibiotic treatment solution but also to increase the percentage of fed mosquitoes, indicative of an aversion circuit to consumption of a bacteria containing sugar solution (data not shown). Therefore, hunger seems to override this aversion circuit, at least in some cases.

The Gr9 behavioural modifications could also be further investigated regarding their influence in inducing anorexia, in addition to modulating the sugar meal size. The experimental design employed here, in which mosquitoes had been selected at day 2 post infection based on the presence of bacteria containing solution in their gut, precludes the use of mosquitoes averse to bacteria containing sugar but allows mosquitoes with differential levels of consumption of bacteria containing sugar. Whether or not Gr9 may also influence mosquito aversion to a bacteria containing sugar meal, leading to no feeding at all, requires a modified experimental design than the one employed here and thus remains to be determined.

Another aspect of the Gr9 effect on feeding behaviour that needs to be further investigated is whether Gr9 modulates consumption of sugar solutions or any solution. Consumption assays as the one described above, in which the *Gr9* knockdown effect on consumption of water or solutions that include different sugars from fructose that was used here, such as sucrose or trehalose, can shed more light on the Gr9 mode of action. Furthermore, any Gr9 effects on DEET-mediated aversion could also be further investigated in the same way, along the lines of the effect of its *Drosophila* orthologue Gr32a (Lee et al., 2010). Gr9 could also influence blood feeding behaviour, an aspect of mosquito behaviour that was not investigated here and could be further pursued in the future.

Further research should also aim at the identification of the trigger for Gr9, being either a nutrient, a mosquito-derived molecule or a bacterial-derived metabolite, although the latter seems unlikely based on the results of the two-choice preference assay showing no significant changes in feeding behaviour due to the presence of *S. marcescens*. It is possible that, as *Drosophila* Gr43a, Gr9 might act as a nutrient sensor and modulate feeding behaviour based on the mosquito satiation status. Gr43a suppresses feeding in satiated flies but promotes feeding in hungry flies (Miyamoto et al., 2012). As the *Gr9* knockdown effect was assayed here in mosquitoes that had been previously starved, a nutrient sensor analogous to *Drosophila* Gr43a would be expected to promote rather than restrict feeding, as shown here for Gr9. One possibility that could be further investigated, though, is whether the Gr9 antibacterial effect is influenced by the mosquito satiation status.

Although the function of the *Gr9 Drosophila* orthologues *Gr32a*, *Gr39a* and *Gr68a* in mating behaviour, acting as pheromone receptors (Bray and Amrein, 2003; Miyamoto and Amrein, 2008; Watanabe et al., 2011), appears to be incongruent with a Gr9 role in antibacterial immunity through modulation of feeding behaviour, Gr32a has also been shown to be involved in feeding suppression in response to DEET or other antifeedants (Lee et al., 2010), showing that gustatory receptors can have multiple functions, perhaps through recognition of various chemosensory cues or through their expression in different body tissues. For example, the *Drosophila* Gr43a function in nutrient sensing relies on its expression in the brain (Miyamoto et al., 2012) while other sugar receptors such as Gr5a rely on external sensory organ expression (Chyb et al., 2003). As tissue-specific *Gr9* knockdown or over-expression is not possible based on routinely used *An. gambiae* genetic tools, it remains to be determined whether the Gr9 effect on feeding behaviour relies on its midgut expression, as suggested by the *Gr9* midgut overexpression compared to other tissues (Baker et al., 2011; Marinotti et al., 2005). It is also possible that Gr9 mediates different functions by its expression in different tissues. Gustatory receptors in *Drosophila,* including the *Gr9* orthologue *Gr39a* (Park and Kwon, 2011), have been shown to be expressed in enteroendocrine chemosensory cells in the midgut, but their role due to their midgut expression remains to be determined. Therefore, multiple Gr9 functions, related to *Gr9* expression in different tissues, that influence mosquito feeding behaviour or antibacterial effects cannot be ruled out.

Further validation of the Gr9 influence on sugar meal size, irrespective of the presence of bacteria, could also involve the influence of *Gr9* knockdown on the outcome of oral infection with other bacterial strains, such as *Asaia*. Based on the data presented here, *Gr9* knockdown is expected to increase the abundance of any bacterial strain included in the mosquito sugar meal. At the same time, though, Gr9-independent

aversion circuits may obfuscate the identification of the Gr9 behavioural role in such cases, by tapering the ingestion of bacteria containing sugar both in *Gr9* knockdown and *dsLacZ* control mosquitoes.

The interplay between Gr9 and NPF could also be further investigated. The co-localization of Gr9 and NPF can be confirmed in mosquito enteroendocrine cells, along with the influence of the *Gr9-NPF* double-knockdown on mosquito feeding behaviour. Based on the data presented here, silencing of *NPF* is expected to reverse the increased sugar consumption shown for *Gr9* knockdown. Finally, the effect of the *Gr9-NPF* double-knockdown can be investigated in influencing the outcome of *P. falciparum* infections. Interestingly, silencing of the *An. gambiae* orthologue of allatostatin A, whose interplay with NPF shapes the *Drosophila* feeding behaviour (Hergarden et al., 2012), was shown to increase *P. berghei* and *P. falciparum* infection intensity (Vlachou, D; EMBO meeting "Molecular and Population Biology of Mosquitoes and Other Disease Vectors", July 2013). These data suggest that *Gr9* knockdown would be expected to also increase *Plasmodium* levels, a possibility that remains to be confirmed.

8.13 Multiple roles for *Gr9* splice variants?

The *Gr9* locus encodes 13 splice variants which share two common 3' exons, E035669 and E035668, 123 and 162 bp in length, respectively, joined by a larger 5' exon through alternative splicing, different for each splice variant (Figure 5.6 and VectorBase VB-2013-10 release). The knockdown approach employed here targeted the large 5' exon, which, as determined by blasting the respective dsRNA sequence, mainly limited the AGAP009805-RA and AGAP009805-RG splice variants, although the silencing efficiency of individual splice variants was not assayed and knockdown of additional *Gr9* variants cannot be ruled out. Remarkably, the *Drosophila Gr28b* locus encodes 5 discrete gustatory receptor transcripts, *Gr28b.a* to *Gr28b.e*, through 5 different transcriptional start sites that result in different 5' exons joined by alternative splicing to common 3' exons (Montell, 2013). *Gr28b* has been implicated in light avoidance through its expression in class IV dendritic arborization neurons in *Drosophila* larvae (Xiang et al., 2010). Although the Gr28b mode of action was not directly demonstrated, its similarity with LITE-1, a *C. elegans* light receptor (Liu et al., 2010a), indicates a similar function for Gr28b (Montell, 2013). The *Gr28b* splice variant specificity responsible for this light recognition effect was not identified (Xiang et al., 2010).

Recently, *Gr28b.d* has been implicated in thermosensation, suggestive of a putative function as a warmth sensor (Ni et al., 2013). The discrete functionality for fly *Gr28b* splice variants (Montell, 2013), in relation to the similar *Gr9* alternative splicing modality in *An. gambiae*, suggest that different *Gr9* variants may also exert different functionality. It remains to be determined which *Gr9* splice variant or variants are

responsible for modulation of the mosquito's feeding behaviour and whether the Gr9 antibacterial effect relies solely on this Gr9 functionality or whether different splice variants may function in different capacities, affecting the outcome of *S. marcescens* infection. A broader analysis of silencing individual *Gr9* splice variants can thus further explicate the functional significance of the *Gr9* alternative splicing in influencing the outcome of *S. marcescens* infection or modulating mosquito feeding behaviour.

Interestingly, a reverse genetics screen of *Drosophila* mutants with increased susceptibility to infection with the intracellular pathogen *Listeria monocytogenes* identified *Gr28b* as a gene influencing sensitivity to infection (Ayres et al., 2008). *Gr28b* mutants also showed increased sensitivity to injection with the Gram-positive bacterium *Staphylococcus aureus* but also increased resistance to the Gram-negative bacterium *Salmonella typhimurium* (Ayres et al., 2008). The Gr28b influence on *Drosophila* immune responses was shown to be related to modulation of feeding behaviour, with *Gr28b* mutants inducing anorexia, a response that is also infection-induced following a bacterial challenge (Ayres and Schneider, 2009). Indeed, anorexia was shown to be responsible for altered immunity to infection by *L. monocytogenes* or *S. typhimurium*, as diet restricted flies showed the same effect as *Gr28b* mutants, while this effect relied on modulation of *Drosophila* innate immune responses related to antimicrobial peptide production and melanization that differentially affected the outcome of infection with different bacteria (Ayres and Schneider, 2009).

Therefore, Gr28b modulated *Drosophila* feeding behaviour in a manner opposite to Gr9 in *Anopheles*, resulting in opposite infection outcomes following systemic infection with different bacteria. As restricted feeding in *Gr28b* mutant flies was shown to influence fly systemic immunity, the differential outcome of bacterial infection most likely relies on different systemic immune responses targeting different bacterial strains. Although the effect of Gr9 on mosquito feeding behaviour can explain the outcome of infection through increased bacterial abundance taken in by the mosquito, another possibility is that, in a similar manner with *Drosophila* Gr28b, Gr9-mediated modulation of feeding behaviour can alter mosquito innate immune responses that affect the outcome of infection, most likely resulting in different infection outcomes following infection with different bacterial strains.

8.14 NPF signal integration and mosquito behavioural responses

Regulation of feeding behaviour in *Drosophila* appears to be a complex process that integrates different signals. This can be attested by an interplay between allatostatin A and NPF in regulation of feeding behaviour (Hergarden et al., 2012) or by modulation of the Gr43a effect on feeding behaviour by the fly's

satiation status (Miyamoto et al., 2012). Such signal integration can lead to phenomenally paradoxical correlations. One such example is a result of NPF integration of natural and drug reward systems that leads to increased alcohol consumption due to sexual deprivation (Shohat-Ophir et al., 2012).

Another paradox, whose underlying mechanism remains a mystery, is the way mating preference in *Drosophila* relies on the composition of the fly's gut microbiota (Sharon et al., 2010). One explanation for this correlation of the gut microbiota with mating preference that has been posited, is that bacterial-derived volatile compounds may act as pheromones, recognized by the mating partner (Sharon et al., 2010). Assuming that pheromone receptors recognize bacterial-derived molecules that influence mating behaviour, it is conceivable that these pheromone receptors might have been also recruited, possibly through their expression in the gut, to recognize these bacterial-derived metabolites and trigger responses to infection that most likely relate to feeding suppression.

The *Gr9* homologies with *Drosophila* pheromone receptors leave open the possibility that Gr9 might have been recruited during the *An. gambiae* evolutionary process to participate in antibacterial responses in a similar manner. Although the Gr9 inhibitory effect on sugar consumption seems to rely on a mosquito-derived molecule, as it is exerted irrespective of the presence of bacteria, it is still possible that *Gr9* might also function, perhaps through a different splice variant, in an infection-induced way. At the same time, other gustatory receptors, such as *Gr13*, downregulated following *Asaia* or *S. marcescens* infection, might function in recognition of bacterial-derived metabolites. Recently, uracil produced by pathogenic bacteria was shown to trigger the DUOX pathway in *Drosophila*, through a yet to be identified GPCR (Lee et al., 2013a). It is conceivable that gustatory receptors might function in recognition of such metabolites, either to trigger a behavioural response or to mount epithelial responses.

8.15 Complex behavioural processes following mosquito oral bacterial infection

The identification of several *An. gambiae* genes related to mosquito behaviour, either associated with the outcome of *S. marcescens* infection or transcriptionally regulated following *Asaia* or *S. marcescens* infection, indicates the existence of complex behavioural processes in response to bacterial infection, which are yet to be revealed. These include the neurotransmitter-triggered GPCRs, whose encoding genes were associated with the outcome of *S. marcescens* infection, GPR5HT7, GPRGBB1 and GPRNPR2, but also the downregulation of *Gr13* and the upregulation of *takeout2* following *Asaia* or *S. marcescens* infection. The influence of these factors in the outcome of *S. marcescens* infection, along with possible influences on mosquito behaviour or an interplay with Gr9 or NPF, remain to be determined. Furthermore, the

indication of Gr9-independent aversion circuits, along with the possibility of Gr9-independent *NPF* downregulation following *S. marcescens* infection suggest the existence of such circuits that remain to be further elucidated.

8.16 Prophylactic or infection-induced mosquito behavioural components

The identification of the Gr9 and NPF involvement in influencing the outcome of *S. marcescens* infection suggests that the response to oral bacterial infection involves two discrete but inextricably linked modes of immunity, a behavioural and an epithelial response. The behavioural component that includes Gr9 may constitutively limit pathogen intake, in a mode of immune response conceptually similar to barrier responses that inhibit contact of pathogens with triggers of epithelial responses, such as the dityrosine network formed by a peroxidase/dual oxidase system following a blood meal (Kumar et al., 2010). Disruption of such a behavioural barrier can affect the outcome of infection. At the same time, genetic variation that might influence this behavioural component, e.g. *Gr9* alleles with differential efficiency in modulating feeding behaviour, can affect the infection outcome but also influence constitutive processes that determine the mosquito gut bacterial population structure. This possibility is consistent with the association of *Gr9* with the outcome of *S. marcescens* infection.

Another possibility is that an infection-induced behavioural component might limit or disrupt pathogen intake following oral infection. The infection-induced transcriptional regulation of *NPF*, *Gr13* and *takeout2* indicates the existence of such behavioural circuit, although the influence of such behavioural response to the outcome of infection remains to be determined.

8.17 An interplay between epithelial and behavioural modes of immunity

In both cases, though, of a constitutive or an infection-induced behavioural component, the outcome of infection relies on the efficiency of the epithelial component to handle the infection. This co-dependence render the behavioural and epithelial components inextricably linked. A surfeit of pathogen abundance due to an impaired behavioural component might render the epithelial component unable to handle the infection. On the other hand, a highly efficient epithelial response might offset increased pathogen abundance due to an impaired behavioural component. This possibility implies that a pathogen abundance threshold exists, after which epithelial immunity cannot efficiently handle the infection. This aspect of epithelial immunity remains poorly understood. Nevertheless, in mosquito infections with *P.*

falciparum through membrane feeding, using different parasite intensities, it has been shown that the intensity of infection correlates with the efficacy of different IMD/REL2 pathway induced components to limit the outcome of *P. falciparum* infection, suggesting that different elicitors might be deployed or might be more efficient in determination of the outcome of infection in low, mid or high infection intensities (Garver et al., 2012).

As the intensity of infection can be correlated to pathogen abundance, also relying on feeding behaviour in natural infections, this observation entails that an interplay between the epithelial and behavioural modes of immunity could shape both responses. As different epithelial effectors may be deployed under different infection intensities, the behavioural component can shape the epithelial response. In this way, the use of an oral infection model allows not only the study of the behavioural component but also of aspects of the epithelial component that rely on different infection intensities. On the other hand, it is possible that epithelial responses might also trigger a behavioural response, a possibility that can be further investigated in the future.

The conclusion to be drawn from the identified mosquito behavioural component influencing the outcome of *S. marcescens* infection is that the mosquito can deploy several layers of immunity to fight a bacterial infection. We have barely scratched the surface in uncovering the complexity of such responses, which most likely also shape the gut bacterial population structure, thus influencing malaria transmission dynamics.

Chapter 9

Concluding remarks and future perspectives

9.1 A brief recapitulation of the research field status and formation of the hypotheses tested in this study

The study of commensal bacteria inhabiting the mosquito gut has recently gained considerable attention, mainly due to the identification of tripartite interactions between the mosquito immune system, its gut microbiota and infections with *Plasmodium*, with such interactions thought to shape the outcome of *Plasmodium* infection and thus malaria transmission dynamics. Although the influence of mosquito gut bacteria on the outcome of *Plasmodium* infections has been known for decades (Gonzalez-Ceron et al., 2003; Pumpuni et al., 1993; Seitz et al., 1987), little attention had been paid by the research community to the underlying mechanisms of this effect, mainly as to whether there is a direct or indirect influence of mosquito gut bacteria on malaria parasites. Most attention by the research community on the mosquito gut microbiota was related to the possibility of paratransgenic approaches, i.e. the use of bacteria as vehicles for the delivery of anti-*Plasmodium* factors (Wang and Jacobs-Lorena, 2013).

Importantly, studies investigating the genetic basis of the outcome of *Plasmodium* infections disregarded the influence of mosquito gut bacteria in shaping the *Plasmodium* infection outcome or driving genetic variation which in turn determines susceptibility and refractoriness to malaria transmission (Blandin et al., 2009; Harris et al., 2010a; Niare et al., 2002; Riehle et al., 2006). Despite being previously proposed (Niare et al., 2002), the effect of *Plasmodium* in driving genetic variation in *An. gambiae* remains a matter of debate (Sangare et al., 2013). Although components of the complement response, such as *APL1* (Riehle et al., 2006) and *TEP1* (Blandin et al., 2009), have been genetically associated with the *Plasmodium* infection outcome, most studies on the mechanistic basis of this response have utilized the rodent malaria model *P. berghei*, the relevance of which on infections with the human malaria parasite *P. falciparum* has been questioned (Cohuet et al., 2006). Furthermore, recent studies have suggested that no specific recognition takes place in *An. gambiae* complement responses to *Plasmodium* and some *P. falciparum* strains largely evade this response (Molina-Cruz et al., 2013; Oliveira et al., 2012).

On the other hand, as bacteria can affect survival (Lemaitre et al., 1995; Schnitger et al., 2007), they are considered a major evolutionary force that can drive genetic variation in various organisms, especially with regard to immune genes (Grossman et al., 2013). Therefore, the hypothesis emerging is that genetic variation associated with the outcome of bacterial infections can directly or indirectly influence malaria transmission dynamics.

A more detailed study of the tripartite interactions between mosquito gut bacteria, immune responses they elicit and *Plasmodium* infections has identified direct and indirect interactions in this intricate

biological system. Antibiotic treatment of mosquitoes has been shown to increase the infection intensity of *P. berghei* or *P. falciparum* (Dong et al., 2009; Meister et al., 2009*)*. Furthermore, induction of immune-related genes by gut bacteria (Dong et al., 2009) and the influence of PGRPLC on the outcome of *P. falciparum* infection only in the presence of bacteria (Meister et al., 2009), suggest that the effect of gut bacteria on the *Plasmodium* infection outcome is indirect and relies on activation of mosquito immune responses that target malaria parasites. Indeed, the IMD/REL2 pathway, whose only identified elicitor is DAP-type peptidoglycan, common to all Gram-negative bacteria, recognized by PGRPLC, has been shown to target *P. falciparum* through the elicitation of immune factors that include LRIM1, APL1A and TEP1 (Frolet et al., 2006; Garver et al., 2012; Garver et al., 2009; Meister et al., 2005; Mitri et al., 2009). Therefore, the possibility that emerges is that differential activation or regulation of the IMD/REL2 pathway by gut bacteria, which show great variation at the individual and population level (Boissière et al., 2012; Osei-Poku et al., 2012; Wang et al., 2011b), can drive malaria transmission dynamics.

At the same time, mosquito gut bacteria were shown to participate in formation of immune memory that limits *Plasmodium* infection intensity upon reinfection (Rodrigues et al., 2010). Immune memory formation could, conceivably, also affect malaria transmission dynamics, although the contribution of discrete gut bacterial strains, especially in field mosquito populations, in influencing the extent of immune memory formation and thus affecting susceptibility and refractoriness to malaria transmission, remain to be further investigated.

Gut bacteria were also shown to directly limit *P. falciparum* populations through the production of reactive oxygen species (Cirimotich et al., 2011b). As different bacteria showed differential propensity in their anti-*Plasmodium* effect, with some members of the *Enterobacteriaceae* family, including *Esp_Z*, showing enhanced reactive oxygen species production that could lead to refractoriness to malaria transmission (Cirimotich et al., 2011b), the hypothesis that emerges is that the population structure of the mosquito gut microbiota can influence malaria transmission dynamics, through the differential representation of bacteria with differential ability to target *Plasmodium*. It is thus possible that certain mosquito gut microbiota composition “fingerprints”, or enterotypes, can be correlated with susceptibility and refractoriness to malaria transmission. This possibility is further strengthened by the correlation of the presence of *Enterobacteriaceae* with susceptibility to *P. falciparum* infection (Boissière et al., 2012). These data are consistent with the possibility that mosquito genetic variation can shape malaria transmission dynamics through shaping the mosquito gut bacteria population structure.

Indeed, the possibility that the composition of the mosquito gut microbiota influences the outcome of *Plasmodium* infection is further strengthened by the identification of intra-specific variation in *S. marcescens* that affects the ability of this common member of the *Enterobacteriaceae* family to influence the outcome of *Plasmodium* infection in an IMD/REL2-independent manner (Bando et al., 2013).

The much more detailed knowledge of the tripartite interactions between the mosquito gut microbiota, elicited immune responses and *Plasmodium* infections, gained in recent years, suggests that the observed variation in susceptibility and refractoriness to malaria transmission in field mosquito populations, correlated with the underlying mosquito genetic variation, can be driven by variation in the load and composition of the mosquito gut microbiota and their influence in shaping the underlying mosquito genetic variation through the *An. gambiae* evolutionary process.

9.2 Formation of the PhD project experimental design and specific aims addressed

A summary of the research questions aiming to be addressed in this PhD project, as formed in the initial phase of this PhD project, are shown below:

My project explores the tripartite interactions between the mosquito gut microbial communities, the PGRPLC-mediated antibacterial defence reactions and infections with *Plasmodium*, and how these might be exploited in future interventions aiming to control malaria transmission. Bacteria in the mosquito midgut have been previously shown to affect *Plasmodium* development. The dramatic bacterial proliferation after a blood meal coincides with the invasion of the mosquito midgut by *Plasmodium* ookinetes, with *An. gambiae* immune responses to limit bacterial proliferation affecting the infection intensity of human and rodent *Plasmodium* parasites. The aim of the project is to investigate the diversity and dynamics of the mosquito gut microbiota, elucidate constitutive immune responses triggered by commensal bacteria that may affect susceptibility and refractoriness against *Plasmodium* infections, characterize bacterial populations responsible for inducible immunity and explore the possibility of the use of genetically modified bacteria for blocking malaria transmission.

Except for the part about genetically modified mosquitoes that was never addressed, the specific aims of the PhD project involved the identification of mosquito antibacterial responses, elicited by diverse bacteria, how these affect the outcome of *Plasmodium* infections and, therefore, how a subset of the mosquito gut bacterial population can affect malaria transmission dynamics and can be thus exploited in paratransgenic efforts. Therefore, the initial aims of the project envisioned two distinct phases, a phase

which would mirror *Drosophila* immunity projects in identifying antibacterial immune responses following oral bacterial infection and a second phase in which identified immune factors would be correlated with the *Plasmodium* infection outcome but also with bacterial communities eliciting expression of these factors.

The first part of the project was designed to establish oral infections in *An. gambiae* with three distinct bacterial strains differing in their pathogenicity, *Asaia*, which was considered an innocuous commensal strain found in a wide variety of insects, *S. marcescens*, a more pathogenic strain which is also common in mosquito gut bacterial communities and *P. aeruginosa PA14,* a pathogenic strain which is scarcely found in mosquito populations (Gendrin and Christophides, 2013). Oral infections with these strains could have been subsequently used to identify transcriptional responses following oral infection and thus immune factors elicited by all or a subset of assayed bacteria. This experimental design closely resembles a recently published study in *Drosophila*, assaying transcriptional responses to oral infection with bacteria of different pathogenicity (Chakrabarti et al., 2012). The second part of the project would aim to assay the influence of the identified immune factors in influencing the outcome of *P. berghei* but also *P. falciparum* infections.

Implementation of these specific aims differed in many aspects of the project. The *P. aeruginosa PA14* strain was never used due to Home Office safety regulations. The part related to *P. falciparum* infections was also never implemented as during the whole duration of the PhD project a *P. falciparum* culture was not established in the group, a culture that requires extensive technical support to maintain. Furthermore, the part of the project related to *P. berghei* infections was also rife with technical problems related to *Plasmodium* infection intensities and mosquito mortality. These issues and the lack of technical support prevented any meaningful analysis. In any case, any results through *P. berghei* infections would have been necessary to be followed up in *P. falciparum* to derive any meaningful conclusions.

A part of the project that was not included in the initial plan of specific aims, was the identification of SNP divergence associated with the outcome of bacterial infections. Although this part of the project eventually formed the backbone of most aspects of the study presented here, the identification of genetic variation associated with the outcome of bacterial infections was not a premise previously investigated and efforts to map genetic divergence in *An. gambiae* were focused by our group and others in the outcome of *Plasmodium* infections.

The experimental design of the SNP genotyping assay first of all required a conceptual shift, related to the realization that the outcome of bacterial infections can be variable. Oral bacterial infections in *Drosophila* essentially inundated flies with bacteria, without antibiotic treatment or tracking of the infection outcome. In mosquitoes, an infection outcome that could have been tracked might have comprised the infection-related mortality. Although conceivable, such mortality could be related to many other aspects of mosquito physiology and could be influenced by fitness differences between mosquito batches or stochastic environmental variables between infections. Therefore, mosquito mortality as a proxy for the bacterial infection outcome would have been very difficult to implement.

The model of oral bacterial infections employed here, using antibiotic treatment and subsequent feeding with a relatively low concentration of bacteria, the separation of bacteria-fed mosquitoes to ensure feeding and the tracking of the level of infection through microscopic observation of fluourescent bacteria present in the mosquito gut innovated in several of its aspects and served as a useful proxy to establish the infection outcome.

As initial oral bacterial infections indicated that there is variation regarding the infection outcome, the possibility emerged that this variation could be quantified and the existence of an underlying genetic variation between phenotypic pools could be investigated. Parallel research efforts aimed at identifying *An. gambiae* SNP divergence associated with the outcome of *P. falciparum* infections using an array-based approach. In this way, the possibility of mapping genetic divergence associated with the outcome of bacterial infections was also pursued. The identification of associated genes relevant to antibacterial responses, including *PGRPLC* and *EGFR*, validated the experimental design approach. Phenotypic characterization of newly identified candidate genes also further validated the findings of this SNP genotyping array assay.

9.3 The ecological aspect of fly and mosquito immunity

Both the SNP genotyping assay, but also the FN3D1-3 phenotypic analysis regarding the modulation of the population structure of the mosquito gut bacterial communities, bring forth an ecological aspect of antibacterial immunity that was missing in the initial experimental design but proved central in interpreting the results of such an intricate biological system that encompassed a high degree of variation both in the mosquito genetic makeup and the composition of the mosquito gut microbiota.

Harnessing the natural genetic variation of the mosquito colony employed here, allowed the identification of aspects of epithelial immunity previously unknown both in the mosquito but also in the much more genetically tractable *Drosophila* system, including the phenotypic characterization of FN3D1-3 as major modulators of the *An. gambiae* gut bacterial population structure, but also the identification of a behavioural immune component, suggestive of an immune role for *Gr9* and other genes that shape mosquito behaviour.

Studies of epithelial immunity in *Drosophila* have long used genetically homogeneous fly lines to uncover finely tuned aspects of immune responses to bacteria and other pathogens. Although such a biological system can robustly and reproducibly uncover mechanistic insights of complex and highly regulated signalling pathways, the context in which these layers of regulation function in field populations remains poorly understood. Therefore, the study of immunity in biological systems encompassing a higher level of genetic variation, not only can be used for the identification of novel components involved in such responses through genetic associations, but also uncover the context in which such components operate. Recently, the study of many aspects of *Drosophila* physiology within the context of genetic variation has started to gain traction, most notably with the establishment of the *Drosophila* genetic reference panel (Mackay et al., 2012). Although such studies related to immunity are limited (Magwire et al., 2012; Weber et al., 2012), it is expected that antibacterial immune responses in *Drosophila* will also be studied with regard to a broader ecological context.

Especially for the study of behavioural responses, the use of homogeneous fly lines can severely limit the identification of critical elements that shape such responses. As has been suggested for the study of flight behaviour, to best leverage the most sophisticated *Drosophila* genetic approaches, fast, small in scale and easily replicated for parallel design measurements have been favoured, leaving us quite ignorant about many aspects of fly behaviour (Dickinson, 2013). The same can probably be asserted for the study of immune responses, as immunity also responds and adapts to outside environmental pressures. As the *Drosophila* neurobiologist Michael Dickinson recently said, the laboratory elicits only a limited range of behaviour from flies; “Most biologists study them in this incredibly benign environment. But studying something as marvelous as a fruit fly in the lab is like having a BMW and driving it around the block” (Gorman, 2013).

In mosquitoes, the study of ecological aspects of physiology and immunity, in particular, has been necessitated by the drive to correlate laboratory findings with malaria intervention efforts in field mosquito populations. Therefore, a considerable aspect of the mosquito research community is related

to the genetic structure of field populations (Coluzzi et al., 2002; della Torre et al., 2001; Neafsey et al., 2010) and how field mosquito genetic variation can be associated with the *Plasmodium* infection outcome (Niare et al., 2002; Riehle et al., 2006). The compromise between detailed mechanistic analysis of immunity and relevance to field mosquito populations has not been easy. This includes the use of *P. berghei* as a model system and findings that are not always validated in the field (Cohuet et al., 2006). Another issue is the use of highly inbred laboratory mosquito strains which differ substantially in their elicited effects to field mosquito populations (Aguilar et al., 2005). On the other hand, the use of recently established mosquito colonies, as the *N'gousso* colony employed here, combined with the mosquito rearing methodology entailing constant variation of the underlying genetic makeup, can reduce reproducibility and question the relevance of any findings. One example is the anti-*Plasmodium* effect of *APL1* paralogues, tested in different mosquito laboratory populations (Garver et al., 2012; Mitri et al., 2009) or the effect of *LRRD7* (*LRIM17*) (Dong et al., 2006a; Garver et al., 2012), another LRIM, on *P. berghei* infections in the *N'gousso* colony used here (Christophides, GK and Upton, LM, Unpublished).

9.4 The presented expression analysis and future research directions

The findings presented in this study can be largely categorized in three parts. The first part involved the establishment of a model of *An. gambiae* oral bacterial infections with *Asaia* and *S. marcescens*, the quantification of the achieved infection levels and a survival analysis of orally infected mosquitoes. The second part of the project involved genome-wide SNP genotyping and expression analyses, aiming to identify genes involved in antibacterial responses, while the third part of the project aimed at the phenotypic analysis of candidate genes emerging from the genome-wide analyses, based on a reverse genetics approach.

The first part of the project, although time consuming and largely descriptive, formed the basis of any further analysis. Therefore, the quantification of infection levels and the segregation of phenotypic pools led to the identification of the underlying genetic variation through a SNP genotyping approach, with the analysis of the infection outcome dynamics forming the basis of the hypothesis that the variable infection outcome could depend on the underlying mosquito genetic variation. As the design of the expression analysis, utilizing DNA microarrays, preceded this realization, the expression analysis design did not incorporate these insights, but only in the phase of result interpretation. In this respect, the expression analysis would have been much more effective if bacterial load had been determined in individual

mosquitoes and expression analysis had been carried out using pools of highly, lowly or non-infected mosquitoes.

The expression analysis was essentially carried out in two replicate batches, several months apart. Surprisingly, the two replicates, despite sharing the same experimental design, deviated in the identified sets of differentially expressed transcripts, with both replicate batches pointing to relevant genes. These differences are most likely due to the underlying genetic variation of assayed mosquitoes between individual infections and to the averaging out of highly, lowly and non-infected mosquitoes in each respective independent infection. Replication of the expression analysis indicated that the diverse expression profile between different batches of infections could be related to differential regulation of the IMD/REL2, JAK/STAT, EGFR and Toll/REL1 pathways, with positive or negative modulators of these pathways showing differential expression in different batches. Strikingly, differential expression of different *PGRPLC* isoforms in different independent infections points to a key regulatory role for *PGRPLC* alternative splicing that is yet to be revealed.

At the same time, core responses to *An. gambiae* infection with *Asaia* or *S. marcescens*, encompassing independent infections from both replicate batches, revealed differentially expressed transcripts with the more pronounced and consistent transcriptional regulation that can be further followed up regarding their involvement in antibacterial responses. These include the gustatory receptor Gr13 and the juvenile hormone binding protein takeout2, putatively involved in mosquito behavioural immune responses to infection with *Asaia* or *S. marcescens*.

This level of observed variation regarding differential expression between replicate mosquito batches, differing in their underlying genetic makeup, presents a key reproducibility issue that could not have been addressed in the current array-based approach. A question that may emerge in the design of future expression analyses would be whether to limit the assayed variation, whether it involved genetic variation of the mosquito colony or stochastic environmental variables, by conducting replicate hybridizations from the same independent mosquito infection, or examine several independent infections, preferably months apart, to capture a larger mosquito genetic component and better validate the results by independent replication of the whole process. The complicating factor in array-based approaches is that a number of, typically 3 to 4, independent replicate hybridizations is required to achieve statistical significance for the observed differential expression. Indeed, hybridizations stemming from the same independent infection largely clustered well together in the expression analysis shown in this study, thus confirming the reproducibility of the employed methodology.

At the same time, independent replication of these mosquito infections under different genetic backgrounds can further confirm the observed differential expression, or introduce additional aspects of differential expression not revealed by the initial analysis, but also identify core responses consistent under variable mosquito genetic backgrounds. Therefore, the current expression analysis could have benefited by replication of 3 to 4 hybridizations from the same infection, while a more robust analysis of mosquito differential gene expression under variable genetic backgrounds would require a sequencing-based approach, correlating the observed transcription profile with the underlying genetic makeup and thus making the whole analysis not only more reproducible but also uncovering polymorphisms leading to regulatory processes that shape the observed mosquito gene expression profile.

9.5 Future research directions based on the SNP genotyping assay presented here

The SNP genotyping approach employed here investigated genetic variation in ~400,000 previously identified SNPs in the *An. gambiae* genome (Lawniczak et al., 2010; Neafsey et al., 2010), identifying SNPs that diverged between phenotypic pools of highly or non-infected *S. marcescens* infected mosquitoes. Using a relatively conservative statistical analysis, compared e.g. to the use of a permutation analysis with 50-SNP windows (Neafsey et al., 2010), a large number of 138 genes were characterized as associated with the outcome of *S. marcescens* infection, as would be expected for a complex quantitative trait as the bacterial infection outcome. The association of genes with known involvement in antibacterial responses but also the phenotypic analysis of FN3D1-3 and Gr9 suggest that additional genes associated with the outcome of *S. marcescens* infection could also be involved in antibacterial responses through mechanisms that have yet to be characterized. Prominent candidates for further phenotypic analysis include three neurotransmitter-triggered GPCRs but also other genes with homologies that suggest involvement in behavioural immune responses such as AGAP012252, the *Drosophila PKC53E* orthologue, possibly involved in NPF-mediated alcohol sensitivity (Chen et al., 2010) and AGAP010503, the *Drosophila SK* channel orthologue, possibly involved in behavioural courtship memory (Abou Tayoun et al., 2012). The involvement of two glycoside hydrolase genes but also of several putative transcription factors could also be investigated regarding a possible involvement in processes that regulate the response to *S. marcescens* infection.

Avenues of further research based on the SNP genotyping assay results presented here would be the identification of causal polymorphisms that would define allele variants associated with the *S. marcescens* infection outcome. Whole-genome sequencing could serve not only to validate the array-based results

and refine the list of associated genes but most likely expand this list by assaying genetic divergence at variable positions other than the ~400,000 SNPs assayed here.

As has been shown in *Drosophila*, bacteria of different pathogenicity may elicit diverse transcriptional responses (Chakrabarti et al., 2012). Therefore, transcriptional analysis following oral infection of additional bacterial strains is not suitable to serve as confirmation of the differential expression following *Asaia* and *S. marcescens* infection presented here, but could serve to identify additional aspects of antibacterial responses as well as conserved responses to a variety of bacteria. Importantly, such diverse expression profiles most likely rely on the involvement of different regulatory factors following mosquito infection with diverse bacterial strains. Therefore, association studies using *Asaia* or other bacterial strains may uncover novel aspects of antibacterial immunity that shape the observed transcription profile following oral infection.

Furthermore, as the FN3D1-3 mode of action shows a level of specificity towards a subset of gut bacteria that includes *Serratia*, association with the infection outcome using different bacterial strains is expected to identify a different set of associated genes, possibly showing specificity against these bacterial strains. These could involve additional FN3D encoding genes, such as AGAP001674, the *Drosophila sidestep* orthologue that is also an *FN3D3* paralogue, but also AGAP010184, the *Drosophila echinoid* orthologue, a negative regulator of the EGFR pathway exhibiting immunoglobulin and FN3 domains (Bai et al., 2001), both found to be transcriptionally regulated following *Asaia* infection.

Another research question that could be addressed in the future is the allelic variant structure of associated genes in field mosquito populations and how these can be correlated with the gut microbial population structure, and, more importantly, with susceptibility and refractoriness to malaria transmission. Whole-genome sequencing approaches could also be used to identify associations with certain “fingerprints” of mosquito gut microbiota composition but also with the presence of certain bacterial strains.

Finally, research efforts related to mosquito genetic associations with the outcome of *P. falciparum* infections, in field or laboratory mosquito populations, using the same or similar SNP genotypic arrays or whole-genome sequencing approaches could elucidate the relation of genetic variation that drives the gut microbiota population structure and malaria transmission dynamics. Such efforts to identify a genetic basis on the outcome of *P. falciparum* infection could shed more light on the possibility that the underlying genetic variation that determines susceptibility and refractoriness to malaria transmission is driven by

mosquito responses that target gut bacteria. Therefore, it is possible that immune factors associated with the outcome of bacterial infections could also show association with the outcome of *P. falciparum* infections in future *An. gambiae* association studies.

9.6 Novel research avenues based on the FN3D1-3 phenotypic analysis

Phenotypic analysis of four genes that emerged from the SNP genotyping analysis as associated with the outcome of *S. marcescens* infection, *FN3D1-3* and *Gr9,* indicated a key role for these genes in influencing the *S. marcescens* infection outcome. The antibacterial effect of these novel immune factors suggests that mosquito epithelial responses to oral bacterial infections are more complex than previously thought.

Further insights into the mode of action of these immune factors were provided in the final part of this study. FN3D1-3 were shown to modulate the mosquito gut bacterial community structure by limiting a subset of gut bacteria that included *Enterobacteriaceae*, mainly *Serratia* or strains that show similarity to *Serratia* reference sequences. The combination of the FN3D1-3 specificity with the influence of the underlying mosquito genetic variation into their mode of action, inferred by their association with the outcome of *S. marcescens* infection, suggests that natural genetic variation can determine microbiome dynamics, which in turn may also influence malaria transmission dynamics. Further research of such intricate interactions in field mosquito populations can uncover ecological aspects underlying mosquito susceptibility and refractoriness to *P. falciparum* infection, essential for effective interventions aiming to disrupt or eliminate malaria transmission.

The FN3D1-3 specificity in limiting *Enterobacteriaceae*, combined with the FD3D2 homology with the hypervariable pattern recognition receptor Dscam, suggests that pathogen-specific recognition shown for AgDscam (Dong et al., 2012) may involve a broader family of immune factors. Indeed, a major paradigm shift is underway in insect immunity. Although it was previously thought that insect innate immune responses can only discriminate between Gram-negative and Gram-positive bacteria through the Imd (IMD/REL2) and Toll (Toll/REL1) pathways, respectively, with tight regulation of such responses responsible for the determination of gut bacterial population structure, several lines of evidence suggest otherwise, including the IMD/REL2-triggered pathogen-specific effector function of AgDscam (Dong et al., 2012) but, more importantly, of uracil recognition that leads to DUOX pathway activation (Lee et al., 2013a).

Further research on the FN3D1-3 specificity can shed more light on the mechanisms leading to the observed shift in mosquito gut bacterial populations, which could possibly involve specific recognition of bacterial patterns, elicitation of targeted effectors, regulation of generalist IMD/REL2 responses, even regulation of bacterial interactions. Furthermore, as *S. marcescens* (Bando et al., 2013) but also other *Enterobacteriaceae* strains as *Esp_Z* (Cirimotich et al., 2011b) are known to influence the *Plasmodium* infection outcome, the possibility that *FN3D1-3* alleles can be correlated with malaria transmission dynamics can be further investigated in field mosquito populations.

Although the conducted phenotypic analysis indicated considerable similarities between FN3D1-3, FN3D3 showed a discrete mode of action compared to FN3D1 and FN3D2, mainly in limiting bacteria of the genus *Burkholderia*. Future research efforts could aim to better discriminate between the modes of action of these three FN3Ds. Biochemical analysis could also better demonstrate the immune specificity of these FN3Ds, as to whether they bind specifically to *Serratia* or a subset of bacteria. A reverse genetics analysis at the bacterial level could also be used to identify bacterial genes that possibly make these strains visible to FN3Ds. Finally, an epistatic analysis can better characterize the relationship of FN3Ds with other immune pathways. The possibility for alternative splicing and the binding specificity of any splice variants can be also further addressed for *FN3D2* or *FN3D3*. Furthermore, any FN3D2 or FN3D3 effect towards *P. falciparum* can be a subject of further research. Interestingly, a recent report indicated the transcriptional regulation of *FN3D2* by *P. falciparum* irrespective of the presence of midgut bacteria (Blumberg et al., 2013), suggesting a possible involvement in *P. falciparum* defence.

As FN3D1 most likely functions in regulation of gene expression, one possibility is that FN3D1 might control the expression of specific effector genes. The repertoire of genes whose expression FN3D1 putatively controls and the interplay with the other two FN3Ds, mainly FN3D2 with which FN3D1 showed a highly similar mode of action, can be further addressed in future research efforts.

9.7 Future directions in the study of mosquito behavioural immunity

The association of *Gr9* with the outcome of *S. marcescens* infection and the employed phenotypic analysis through *Gr9* RNAi-mediated silencing revealed a surprising role in antibacterial immunity for Gr9 that could be conceivably extended to other gustatory receptors.

Phenotypic analysis of the Gr9 antibacterial effect indicated that Gr9 influences the *S. marcescens* infection outcome through modulation of feeding behaviour that relies on changes in *NPF* expression. The

role of behavioural responses in antibacterial immunity is poorly understood, although such responses might play key roles in influencing the infection outcome. Indeed, the hypothesis presented here posits that *Gr9* genetic variation, related to modulation of feeding behaviour, could profoundly shape the mosquito gut bacterial population structure and thus malaria transmission dynamics. Future research efforts using next-generation sequencing in field mosquito populations can further explicate such possibility.

Furthermore, the identification of a wide array of behaviour-related genes, associated with the *S. marcescens* infection outcome or transcriptionally regulated following *S. marcescens* or *Asaia* infection, point to complex behavioural processes whose elucidation can be of a profound biological interest. At the same time, modulation of mosquito behaviour can lead to novel interventions to eliminate malaria transmission and a deep understanding of such behavioural circuits is of vital importance towards such goal.

Another possibility emerging from the implication of gustatory receptors such as Gr9 and Gr13 in antibacterial immunity is that pathogen-specific immune responses might rely on recognition of bacterial-derived molecules by gustatory receptors. The GPCR hypothesized to be a uracil receptor that triggers the DUOX pathway in *Drosophila* (Lee et al., 2013a) might as well be a gustatory receptor. Gustatory receptors, resembling olfactory receptors, exhibit a reverse topology compared to canonical GPCRs and most likely function as ligand-gated cation channels, although they might still retain cyclic-nucleotide activities (Hill et al., 2002; Sato et al., 2011; Zhang et al., 2011). Furthermore, the possibility of the involvement of bacterial-derived pheromones in modulation of *Drosophila* mating preference (Sharon et al., 2010) entails that gustatory receptors acting as pheromone receptors, such as the *Gr9 Drosophila* orthologues *Gr32a*, *Gr39a* and *Gr68a* (Bray and Amrein, 2003; Miyamoto and Amrein, 2008), may also function in specific bacterial recognition. The role of gustatory receptors goes well beyond taste perception and studies on their involvement in various physiological processes has recently gained considerable traction (Miyamoto et al., 2013; Montell, 2013; Ni et al., 2013). Therefore, the involvement of mosquito gustatory receptors in antibacterial or anti-*Plasmodium* responses merits further research efforts.

9.8 *An. gambiae* as a model organism in the study of epithelial immunity

As the current study mainly focused on mosquito antibacterial responses, a field that has been extensively studied in *Drosophila* (Lemaitre and Hoffmann, 2007), a commonly used and genetically tractable insect

model system, the analysis presented here exemplifies the limitations but also the possibilities for the use of mosquitoes as a model system to study immune mechanisms. Mosquitoes, especially *An. gambiae*, lack the genetic tools for complete, tissue-specific gene silencing or overexpression, combined with the difficulty to maintain any transgenic lines. Despite some recent progress (Windbichler et al., 2011), especially in *Ae. aegypti* (Degennaro et al., 2013; Liesch et al., 2013*)*, the genetic tools available in *Drosophila* and decades of accumulated knowledge, render *Drosophila* a much superior model system, especially for the study of mechanistic aspects of immunity.

The main motivation in the study of *An. gambiae* immunity is the immense public health significance the study of the main malaria vector presents. As it has been made clear, the main incentive in the study of mosquito gut epithelial responses was the identification of tripartite interactions between gut bacteria, mosquito immunity and *Plasmodium* infections, likely to affect malaria transmission dynamics. Indeed, the main significance of the findings presented in this study is their contribution in understanding aspects of mosquito immunity that influence malaria transmission.

At the same time, as the research field expands into ecological aspects of immunity, the extensive study of the field mosquito population structure (Reidenbach et al., 2012) and the intense research efforts on population genetics (Mitri and Vernick, 2012; Niare et al., 2002) bring the mosquito into the vanguard of this aspect of research efforts compared to *Drosophila* or other model organisms. One such investigated aspect is the recent SNP genotyping efforts based on sequencing of the M/S forms of *An. gambiae* (Lawniczak et al., 2010; Neafsey et al., 2010), on which the association study presented here relied upon. The phenotypic analysis presented here, despite the limited genetic tools available, presents novel aspects of epithelial immunity irrespective of the insect organism used.

Another advantage of the use of the mosquito as a model organism is both the mosquito haematophagy lifestyle and the natural *Plasmodium* infections, which can reveal aspects of immunity that cannot be investigated in *Drosophila*. Although this aspect of immunity was scarcely addressed in the current study, future research efforts in an integrative model of infections with *Plasmodium* parasites and bacteria following blood feeding, in conjunction with available genetic association tools, in laboratory or field mosquito populations, can reveal previously unknown immune components. Hopefully, this aspect of research into mosquito immunity has only just begun.

Chapter 10

References

1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061-1073.
Aballay, A. (2009). Neural regulation of immunity: role of NPR-1 in pathogen avoidance and regulation of innate immunity. Cell Cycle *8*, 966-969.
Abdi, H.W., L.J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics *2*, 433-459.
Abou Tayoun, A.N., Pikielny, C., and Dolph, P.J. (2012). Roles of the Drosophila SK channel (dSK) in courtship memory. PLoS One *7*, e34665.
Abraham, E.G., Pinto, S.B., Ghosh, A., Vanlandingham, D.L., Budd, A., Higgs, S., Kafatos, F.C., Jacobs-Lorena, M., and Michel, K. (2005). An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. Proc Natl Acad Sci U S A *102*, 16327-16332.
Adamo, S.A. (2005). Parasitic suppression of feeding in the tobacco hornworm, Manduca sexta: parallels with feeding depression after an immune challenge. Arch Insect Biochem Physiol *60*, 185-197.
Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.*, et al.* (2000). The genome sequence of Drosophila melanogaster. Science *287*, 2185-2195.
Agnew, C., Borodina, E., Zaccai, N.R., Conners, R., Burton, N.M., Vicary, J.A., Cole, D.K., Antognozzi, M., Virji, M., and Brady, R.L. (2011). Correlation of in situ mechanosensitive responses of the Moraxella catarrhalis adhesin UspA1 with fibronectin and receptor CEACAM1 binding. Proc Natl Acad Sci U S A *108*, 15174-15178.
Aguilar, R., Dong, Y., Warr, E., and Dimopoulos, G. (2005). Anopheles infection responses; laboratory models versus field malaria transmission systems. Acta Trop *95*, 285-291.
Ahmadian, A., Ehn, M., and Hober, S. (2006). Pyrosequencing: history, biochemistry and future. Clinica chimica acta; international journal of clinical chemistry *363*, 83-94.
Akbari, O.S., Matzen, K.D., Marshall, J.M., Huang, H., Ward, C.M., and Hay, B.A. (2013). A synthetic gene drive system for local, reversible modification and suppression of insect populations. Curr Biol *23*, 671-677.
Akhouayri, I., Turc, C., Royet, J., and Charroux, B. (2011). Toll-8/tollo negatively regulates antimicrobial response in the Drosophila respiratory epithelium. PLoS Pathog *7*, e1002319.
Akhouayri, I.G., Habtewold, T., and Christophides, G.K. (2013). Melanotic Pathology and Vertical Transmission of the Gut Commensal in the Major Malaria Vector. PLoS One *8*, e77619.
Akman Gunduz, E., and Douglas, A.E. (2009). Symbiotic bacteria enable insect to use a nutritionally inadequate diet. Proc Biol Sci *276*, 987-991.
Al-Anzi, B., Armand, E., Nagamei, P., Olszewski, M., Sapin, V., Waters, C., Zinn, K., Wyman, R.J., and Benzer, S. (2010). The leucokinin pathway and its neurons regulate meal size in Drosophila. Curr Biol *20*, 969-978.
Alavi, Y., Arai, M., Mendoza, J., Tufet-Bayona, M., Sinha, R., Fowler, K., Billker, O., Franke-Fayard, B., Janse, C.J., Waters, A.*, et al.* (2003). The dynamics of interactions between Plasmodium and the mosquito: a study of the infectivity of Plasmodium berghei and Plasmodium gallinaceum, and their transmission by Anopheles stephensi, Anopheles gambiae and Aedes aegypti. Int J Parasitol *33*, 933-943.
Alonso, P.L., Ballou, R., Brown, G., Chitnis, C., Loucq, C., Moorthy, V., Saul, A., and Wirth, D. (2011). A research agenda for malaria eradication: vaccines. PLoS Med *8*, e1000398.
Alphey, L., Beard, C.B., Billingsley, P., Coetzee, M., Crisanti, A., Curtis, C., Eggleston, P., Godfray, C., Hemingway, J., Jacobs-Lorena, M.*, et al.* (2002). Malaria control with genetically manipulated insect vectors. Science *298*, 119-121.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.
Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.
Anders, K.L., and Hay, S.I. (2012). Lessons from malaria control to help meet the rising challenge of dengue. Lancet Infect Dis *12*, 977-984.
Anderson, A.A., Kendal, C.E., Garcia-Maya, M., Kenny, A.V., Morris-Triggs, S.A., Wu, T., Reynolds, R., Hohenester, E., and Saffell, J.L. (2005). A peptide from the first fibronectin domain of NCAM acts as an inverse agonist and stimulates FGF receptor activation, neurite outgrowth and survival. J Neurochem *95*, 570-583.
Andrews, G.L., Tanglao, S., Farmer, W.T., Morin, S., Brotman, S., Berberoglu, M.A., Price, H., Fernandez, G.C., Mastick, G.S., Charron, F.*, et al.* (2008). Dscam guides embryonic axons by Netrin-dependent and -independent functions. Development *135*, 3839-3848.
Apte-Deshpande, A., Paingankar, M., Gokhale, M.D., and Deobagkar, D.N. (2012). Serratia odorifera a midgut inhabitant of Aedes aegypti mosquito enhances its susceptibility to dengue-2 virus. PLoS One *7*, e40401.
Arensburger, P., Megy, K., Waterhouse, R.M., Abrudan, J., Amedeo, P., Antelo, B., Bartholomay, L., Bidwell, S., Caler, E., Camara, F.*, et al.* (2010). Sequencing of Culex quinquefasciatus establishes a platform for mosquito comparative genomics. Science *330*, 86-88.
Arita, M., Ohira, T., Sun, Y.P., Elangovan, S., Chiang, N., and Serhan, C.N. (2007). Resolvin E1 selectively interacts with leukotriene B4 receptor BLT1 and ChemR23 to regulate inflammation. J Immunol *178*, 3912-3917.
Artavanis-Tsakonas, K., Tongren, J.E., and Riley, E.M. (2003). The war between the malaria parasite and the immune system: immunity, immunoregulation and immunopathology. Clin Exp Immunol *133*, 145-152.
Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M.*, et al.* (2011). Enterotypes of the human gut microbiome. Nature *473*, 174-180.
Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.
Ayala, D., Fontaine, M.C., Cohuet, A., Fontenille, D., Vitalis, R., and Simard, F. (2010). Chromosomal inversions, natural selection and adaptation in the malaria vector Anopheles funestus. Mol Biol Evol.
Ayres, J.S., Freitag, N., and Schneider, D.S. (2008). Identification of Drosophila mutants altering defense of and endurance to Listeria monocytogenes infection. Genetics *178*, 1807-1815.
Ayres, J.S., and Schneider, D.S. (2009). The role of anorexia in resistance and tolerance to infections in Drosophila. PLoS Biol *7*, e1000150.
Azambuja, P., Feder, D., and Garcia, E.S. (2004). Isolation of Serratia marcescens in the midgut of Rhodnius prolixus: impact on the establishment of the parasite Trypanosoma cruzi in the vector. Exp Parasitol *107*, 89-96.
Azambuja, P., Garcia, E.S., and Ratcliffe, N.A. (2005). Gut microbiota and parasite transmission by insect vectors. Trends Parasitol *21*, 568-572.
Backhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005). Host-bacterial mutualism in the human intestine. Science *307*, 1915-1920.

Baganz, N.L., and Blakely, R.D. (2013). A dialogue between the immune system and brain, spoken in the language of serotonin. ACS chemical neuroscience *4*, 48-63.
Bahia, A.C., Kubota, M.S., Tempone, A.J., Araujo, H.R., Guedes, B.A., Orfano, A.S., Tadei, W.P., Rios-Velasquez, C.M., Han, Y.S., Secundino, N.F., *et al.* (2011). The JAK-STAT Pathway Controls Plasmodium vivax Load in Early Stages of Anopheles aquasalis Infection. PLoS Negl Trop Dis *5*, e1317.
Bai, J., Chiu, W., Wang, J., Tzeng, T., Perrimon, N., and Hsu, J. (2001). The cell adhesion molecule Echinoid defines a new pathway that antagonizes the Drosophila EGF receptor signaling pathway. Development *128*, 591-601.
Baker, D.A., Nolan, T., Fischer, B., Pinder, A., Crisanti, A., and Russell, S. (2011). A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, Anopheles gambiae. BMC Genomics *12*, 296.
Baker, G.C., Smith, J.J., and Cowan, D.A. (2003). Review and re-analysis of domain-specific 16S primers. J Microbiol Methods *55*, 541-555.
Baker, M. (2010a). Genomics: The search for association. Nature *467*, 1135-1138.
Baker, M. (2010b). Structural biology: The gatekeepers revealed. Nature *465*, 823-826.
Balter, M. (2012). Taking stock of the human microbiome and disease. Science *336*, 1246-1247.
Bando, H., Okado, K., Guelbeogo, W.M., Badolo, A., Aonuma, H., Nelson, B., Fukumoto, S., Xuan, X., Sagnon, N., and Kanuka, H. (2013). Intra-specific diversity of Serratia marcescens in Anopheles mosquito midgut defines Plasmodium transmission capacity. Scientific reports *3*, 1641.
Bariami, V., Jones, C.M., Poupardin, R., Vontas, J., and Ranson, H. (2012). Gene Amplification, ABC Transporters and Cytochrome P450s: Unraveling the Molecular Basis of Pyrethroid Resistance in the Dengue Vector, Aedes aegypti. PLoS Negl Trop Dis *6*, e1692.
Barillas-Mury, C. (2007). CLIP proteases and Plasmodium melanization in Anopheles gambiae. Trends Parasitol *23*, 297-299.
Baton, L.A., Robertson, A., Warr, E., Strand, M.R., and Dimopoulos, G. (2009). Genome-wide transcriptomic profiling of Anopheles gambiae hemocytes reveals pathogen-specific signatures upon bacterial challenge and Plasmodium berghei infection. BMC Genomics *10*, 257.
Beaty, B.J., Prager, D.J., James, A.A., Jacobs-Lorena, M., Miller, L.H., Law, J.H., Collins, F.H., and Kafatos, F.C. (2009). From Tucson to genomics and transgenics: the vector biology network and the emergence of modern vector biology. PLoS Negl Trop Dis *3*, e343.
Becnel, J., Johnson, O., Luo, J., Nassel, D.R., and Nichols, C.D. (2011). The serotonin 5-HT7Dro receptor is expressed in the brain of Drosophila, and is essential for normal courtship and mating. PLoS One *6*, e20800.
Beebe, K., Lee, W.C., and Micchelli, C.A. (2009). JAK/STAT signaling coordinates stem cell proliferation and multilineage differentiation in the Drosophila intestinal stem cell lineage. Dev Biol.
Beerntsen, B.T., James, A.A., and Christensen, B.M. (2000). Genetics of mosquito vector competence. Microbiol Mol Biol Rev *64*, 115-137.
Behrens, M., and Meyerhof, W. (2011). Gustatory and extragustatory functions of mammalian taste receptors. Physiology & behavior *105*, 4-13.
Beier, M.S., Pumpuni, C.B., Beier, J.C., and Davis, J.R. (1994). Effects of para-aminobenzoic acid, insulin, and gentamicin on Plasmodium falciparum development in anopheline mosquitoes (Diptera: Culicidae). J Med Entomol *31*, 561-565.
Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K., *et al.* (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc Natl Acad Sci U S A *107*, 18933-18938.
Benton, R., Sachse, S., Michnick, S.W., and Vosshall, L.B. (2006). Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo. PLoS Biol *4*, e20.
Berger, J., Suzuki, T., Senti, K.A., Stubbs, J., Schaffner, G., and Dickson, B.J. (2001). Genetic mapping with SNP markers in Drosophila. Nat Genet *29*, 475-481.
Berger, J.A., Hautaniemi, S., Jarvinen, A.K., Edgren, H., Mitra, S.K., and Astola, J. (2004). Optimized LOWESS normalization parameter selection for DNA microarray data. BMC Bioinformatics *5*, 194.
Besansky, N.J., Powell, J.R., Caccone, A., Hamm, D.M., Scott, J.A., and Collins, F.H. (1994). Molecular phylogeny of the Anopheles gambiae complex suggests genetic introgression between principal malaria vectors. Proc Natl Acad Sci U S A *91*, 6885-6888.
Bian, G., Joshi, D., Dong, Y., Lu, P., Zhou, G., Pan, X., Xu, Y., Dimopoulos, G., and Xi, Z. (2013). Wolbachia invades Anopheles stephensi populations and induces refractoriness to Plasmodium infection. Science *340*, 748-751.
Bischoff, V., Vignal, C., Duvic, B., Boneca, I.G., Hoffmann, J.A., and Royet, J. (2006). Downregulation of the Drosophila immune response by peptidoglycan-recognition proteins SC1 and SC2. PLoS Pathog *2*, e14.
Blagborough, A.M., Churcher, T.S., Upton, L.M., Ghani, A.C., Gething, P.W., and Sinden, R.E. (2013a). Transmission-blocking interventions eliminate malaria from laboratory populations. Nature communications *4*, 1812.
Blagborough, A.M., Delves, M.J., Ramakrishnan, C., Lal, K., Butcher, G., and Sinden, R.E. (2013b). Assessing transmission blockade in Plasmodium spp. Methods Mol Biol *923*, 577-600.
Blandin, S., Moita, L.F., Kocher, T., Wilm, M., Kafatos, F.C., and Levashina, E.A. (2002). Reverse genetics in the mosquito Anopheles gambiae: targeted disruption of the Defensin gene. EMBO Rep *3*, 852-856.
Blandin, S., Shiao, S.H., Moita, L.F., Janse, C.J., Waters, A.P., Kafatos, F.C., and Levashina, E.A. (2004). Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector Anopheles gambiae. Cell *116*, 661-670.
Blandin, S.A., Wang-Sattler, R., Lamacchia, M., Gagneur, J., Lycett, G., Ning, Y., Levashina, E.A., and Steinmetz, L.M. (2009). Dissecting the genetic basis of resistance to malaria parasites in Anopheles gambiae. Science *326*, 147-150.
Blanford, S., Chan, B.H., Jenkins, N., Sim, D., Turner, R.J., Read, A.F., and Thomas, M.B. (2005). Fungal pathogen reduces potential for malaria transmission. Science *308*, 1638-1641.
Blumberg, B.J., Trop, S., Das, S., and Dimopoulos, G. (2013). Bacteria- and IMD Pathway-Independent Immune Defenses against Plasmodium falciparum in Anopheles gambiae. PLoS One *8*, e72130.
Boissière, A., Tchioffo, M.T., Bachar, D., Abate, L., Marie, A., Nsango, S.E., Shahbazkia, H.R., Awono-Ambene, P.H., Levashina, E.A., Christen, R., *et al.* (2012). Midgut Microbiota of the Malaria Mosquito Vector Anopheles gambiae and Interactions with Plasmodium falciparum Infection. PLoS Pathogens *8*, e1002742.
Boseley, S. (2011). Malaria vaccine set to save millions of lives, but who will fund it? In The Guardian (London: Guardian News and Media Limited).
Bray, S., and Amrein, H. (2003). A putative Drosophila pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. Neuron *39*, 1019-1029.

Broderick, N.A., Raffa, K.F., Goodman, R.M., and Handelsman, J. (2004). Census of the bacterial community of the gypsy moth larval midgut by using culturing and culture-independent methods. Appl Environ Microbiol *70*, 293-300.
Brown, M.R., Crim, J.W., Arata, R.C., Cai, H.N., Chun, C., and Shen, P. (1999). Identification of a Drosophila brain-gut peptide related to the neuropeptide Y family. Peptides *20*, 1035-1042.
Buchon, N., Broderick, N.A., Kuraishi, T., and Lemaitre, B. (2010). Drosophila EGFR pathway coordinates stem cell proliferation and gut remodeling following infection. BMC Biol *8*, 152.
Buchon, N., Broderick, N.A., Poidevin, M., Pradervand, S., and Lemaitre, B. (2009). Drosophila intestinal response to bacterial infection: activation of host defense and stem cell proliferation. Cell Host Microbe *5*, 200-211.
Caballero-Mellado, J., Martinez-Aguilar, L., Paredes-Valdez, G., and Santos, P.E. (2004). Burkholderia unamae sp. nov., an N2-fixing rhizospheric and endophytic species. Int J Syst Evol Microbiol *54*, 1165-1172.
Capone, A., Ricci, I., Damiani, C., Mosca, M., Rossi, P., Scuppa, P., Crotti, E., Epis, S., Angeletti, M., Valzano, M*., et al.* (2013). Interactions between Asaia, Plasmodium and Anopheles: new insights into mosquito symbiosis and implications in Malaria Symbiotic Control. Parasit Vectors *6*, 182.
Caputo, B., Nwakanma, D., Jawara, M., Adiamoh, M., Dia, I., Konate, L., Petrarca, V., Conway, D.J., and della Torre, A. (2008). Anopheles gambiae complex along The Gambia river, with particular reference to the molecular forms of An. gambiae s.s. Malar J *7*, 182.
Caputo, B., Santolamazza, F., Vicente, J.L., Nwakanma, D.C., Jawara, M., Palsson, K., Jaenson, T., White, B.J., Mancini, E., Petrarca, V*., et al.* (2011). The "far-west" of Anopheles gambiae molecular forms. PLoS One *6*, e16415.
Carafoli, F., Saffell, J.L., and Hohenester, E. (2008). Structure of the tandem fibronectin type 3 domains of neural cell adhesion molecule. J Mol Biol *377*, 524-534.
Carey, A.F., Wang, G., Su, C.Y., Zwiebel, L.J., and Carlson, J.R. (2010). Odorant reception in the malaria mosquito Anopheles gambiae. Nature *464*, 66-71.
Carter, R., and Mendis, K.N. (2002). Evolutionary and historical aspects of the burden of malaria. Clin Microbiol Rev *15*, 564-594.
Cassone, B.J., Mouline, K., Hahn, M.W., White, B.J., Pombi, M., Simard, F., Costantini, C., and Besansky, N.J. (2008). Differential gene expression in incipient species of Anopheles gambiae. Mol Ecol *17*, 2491-2504.
Cator, L.J., George, J., Blanford, S., Murdock, C.C., Baker, T.C., Read, A.F., and Thomas, M.B. (2013). 'Manipulation' without the parasite: altered feeding behaviour of mosquitoes is not dependent on infection with malaria parasites. Proc Biol Sci *280*, 20130711.
Catteruccia, F., Crisanti, A., and Wimmer, E.A. (2009). Transgenic technologies to induce sterility. Malar J *8 Suppl 2*, S7.
Catteruccia, F., Godfray, H.C., and Crisanti, A. (2003). Impact of genetic manipulation on the fitness of Anopheles stephensi mosquitoes. Science *299*, 1225-1227.
Catteruccia, F., Nolan, T., Loukeris, T.G., Blass, C., Savakis, C., Kafatos, F.C., and Crisanti, A. (2000). Stable germline transformation of the malaria mosquito Anopheles stephensi. Nature *405*, 959-962.
Chakrabarti, S., Liehl, P., Buchon, N., and Lemaitre, B. (2012). Infection-induced host translational blockage inhibits immune responses and epithelial renewal in the Drosophila gut. Cell Host Microbe *12*, 60-70.
Chamseddin, K.H., Khan, S.Q., Nguyen, M.L., Antosh, M., Morris, S.N., Kolli, S., Neretti, N., Helfand, S.L., and Bauer, J.H. (2012). takeout-dependent longevity is associated with altered Juvenile Hormone signaling. Mechanisms of ageing and development *133*, 637-646.
Chanana, B., Steigemann, P., Jackle, H., and Vorbruggen, G. (2009). Reception of Slit requires only the chondroitin-sulphate-modified extracellular domain of Syndecan at the target cell surface. Proc Natl Acad Sci U S A *106*, 11984-11988.
Chandler, J.A., Morgan Lang, J., Bhatnagar, S., Eisen, J.A., and Kopp, A. (2011). Bacterial communities of diverse Drosophila species: ecological context of a host-microbe model system. PLoS Genet *7*, e1002272.
Chang, C.I., Pili-Floury, S., Herve, M., Parquet, C., Chelliah, Y., Lemaitre, B., Mengin-Lecreulx, D., and Deisenhofer, J. (2004). A Drosophila pattern recognition receptor contains a peptidoglycan docking groove and unusual L,D-carboxypeptidase activity. PLoS Biol *2*, E277.
Chasan, R., and Anderson, K.V. (1989). The role of easter, an apparent serine protease, in organizing the dorsal-ventral pattern of the Drosophila embryo. Cell *56*, 391-400.
Chen, B.E., Kondo, M., Garnier, A., Watson, F.L., Puettmann-Holgado, R., Lamar, D.R., and Schmucker, D. (2006). The molecular diversity of Dscam is functionally required for neuronal wiring specificity in Drosophila. Cell *125*, 607-620.
Chen, C.H., Huang, H., Ward, C.M., Su, J.T., Schaeffer, L.V., Guo, M., and Hay, B.A. (2007). A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila. Science *316*, 597-600.
Chen, J., Zhang, Y., and Shen, P. (2008). A protein kinase C activity localized to neuropeptide Y-like neurons mediates ethanol intoxication in Drosophila melanogaster. Neuroscience *156*, 42-47.
Chen, J., Zhang, Y., and Shen, P. (2010). Protein kinase C deficiency-induced alcohol insensitivity and underlying cellular targets in Drosophila. Neuroscience *166*, 34-39.
Choe, K.M., Lee, H., and Anderson, K.V. (2005). Drosophila peptidoglycan recognition protein LC (PGRP-LC) acts as a signal-transducing innate immune receptor. Proc Natl Acad Sci U S A *102*, 1122-1126.
Choe, K.M., Werner, T., Stoven, S., Hultmark, D., and Anderson, K.V. (2002). Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in Drosophila. Science *296*, 359-362.
Chouaia, B., Rossi, P., Epis, S., Mosca, M., Ricci, I., Damiani, C., Ulissi, U., Crotti, E., Daffonchio, D., Bandi, C*., et al.* (2012). Delayed larval development in Anopheles mosquitoes deprived of Asaia bacterial symbionts. BMC Microbiol *12 Suppl 1*, S2.
Chouaia, B., Rossi, P., Montagna, M., Ricci, I., Crotti, E., Damiani, C., Epis, S., Faye, I., Sagnon, N., Alma, A*., et al.* (2010). Typing of Asaia spp. bacterial symbionts in four mosquito species: molecular evidence for multiple infections. Appl Environ Microbiol.
Chowdhury, D.R., Angov, E., Kariuki, T., and Kumar, N. (2009). A potent malaria transmission blocking vaccine based on codon harmonized full length Pfs48/45 expressed in Escherichia coli. PLoS One *4*, e6352.
Christophides, G.K., Vlachou, D., and Kafatos, F.C. (2004). Comparative and functional genomics of the innate immune system in the malaria vector Anopheles gambiae. Immunol Rev *198*, 127-148.
Christophides, G.K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P.T., Collins, F.H., Danielli, A., Dimopoulos, G*., et al.* (2002). Immunity-related genes and gene families in Anopheles gambiae. Science *298*, 159-165.

Chung, H., Pamp, S.J., Hill, J.A., Surana, N.K., Edelman, S.M., Troy, E.B., Reading, N.C., Villablanca, E.J., Wang, S., Mora, J.R., *et al.* (2012). Gut immune maturation depends on colonization with a host-specific microbiota. Cell *149*, 1578-1593.
Chyb, S., Dahanukar, A., Wickens, A., and Carlson, J.R. (2003). Drosophila Gr5a encodes a taste receptor tuned to trehalose. Proc Natl Acad Sci U S A *100 Suppl 2*, 14526-14530.
Cirillo, C., Vanden Berghe, P., and Tack, J. (2011). Role of serotonin in gastrointestinal physiology and pathology. Minerva endocrinologica *36*, 311-324.
Cirimotich, C.M., Clayton, A.M., and Dimopoulos, G. (2011a). Low- and high-tech approaches to control Plasmodium parasite transmission by anopheles mosquitoes. Journal of tropical medicine *2011*, 891342.
Cirimotich, C.M., Dong, Y., Clayton, A.M., Sandiford, S.L., Souza-Neto, J.A., Mulenga, M., and Dimopoulos, G. (2011b). Natural microbe-mediated refractoriness to Plasmodium infection in Anopheles gambiae. Science *332*, 855-858.
Clark, J.T., Kalra, P.S., Crowley, W.R., and Kalra, S.P. (1984). Neuropeptide Y and human pancreatic polypeptide stimulate feeding behavior in rats. Endocrinology *115*, 427-429.
Clayton, A.M., Cirimotich, C.M., Dong, Y., and Dimopoulos, G. (2012). Caudal is a negative regulator of the Anopheles IMD Pathway that controls resistance to P. falciparum infection. Dev Comp Immunol.
Clemente, Jose C., Ursell, Luke K., Parfrey, Laura W., and Knight, R. (2012). The Impact of the Gut Microbiota on Human Health: An Integrative View. Cell *148*, 1258-1270.
Coates, C.J., Jasinskiene, N., Miyashiro, L., and James, A.A. (1998). Mariner transposition and transformation of the yellow fever mosquito, Aedes aegypti. Proc Natl Acad Sci U S A *95*, 3748-3751.
Coates, J.C., and de Bono, M. (2002). Antagonistic pathways in neurons exposed to body fluid regulate social feeding in Caenorhabditis elegans. Nature *419*, 925-929.
Coetzee, M., Craig, M., and le Sueur, D. (2000). Distribution of African malaria mosquitoes belonging to the Anopheles gambiae complex. Parasitol Today *16*, 74-77.
Coetzee M, H.R., Wilkerson R, Della Torre A, Coulibaly MB, Besansky NJ (2013). Anopheles coluzzii and Anopheles amharicus, new members of the Anopheles gambiae complex. Zootaxa *3619*, 246-274.
Cohen, J.M., Smith, D.L., Cotter, C., Ward, A., Yamey, G., Sabot, O.J., and Moonen, B. (2012). Malaria resurgence: a systematic review and assessment of its causes. Malar J *11*, 122.
Cohuet, A., Harris, C., Robert, V., and Fontenille, D. (2010). Evolutionary forces on Anopheles: what makes a malaria vector? Trends Parasitol *26*, 130-136.
Cohuet, A., Krishnakumar, S., Simard, F., Morlais, I., Koutsos, A., Fontenille, D., Mindrinos, M., and Kafatos, F.C. (2008). SNP discovery and molecular evolution in Anopheles gambiae, with special emphasis on innate immune system. BMC Genomics *9*, 227.
Cohuet, A., Osta, M.A., Morlais, I., Awono-Ambene, P.H., Michel, K., Simard, F., Christophides, G.K., Fontenille, D., and Kafatos, F.C. (2006). Anopheles and Plasmodium: from laboratory models to natural systems in the field. EMBO Rep *7*, 1285-1289.
Collins, F.H., Sakai, R.K., Vernick, K.D., Paskewitz, S., Seeley, D.C., Miller, L.H., Collins, W.E., Campbell, C.C., and Gwadz, R.W. (1986). Genetic selection of a Plasmodium-refractory strain of the malaria vector Anopheles gambiae. Science *234*, 607-610.
Coluzzi, M. (1982). Mechanisms of speciation (New York: Liss).
Coluzzi, M. (1992). Malaria vector analysis and control. Parasitol Today *8*, 113-118.
Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M.A., and Petrarca, V. (2002). A polytene chromosome analysis of the Anopheles gambiae species complex. Science *298*, 1415-1418.
Coluzzi, M., Sabatini, A., Petrarca, V., and Di Deco, M.A. (1979). Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex. Trans R Soc Trop Med Hyg *73*, 483-497.
Coluzzi, M.P., V. ; Di Deco, M.A. (1985). Chromosomal inversion intergradation and incipient speciation in Anopheles gambiae. Boll Zool *52*, 45-63.
Cook, G.C., and Webb, A.J. (2000). Perceptions of malaria transmission before Ross' discovery in 1897. Postgrad Med J *76*, 738-740.
Corby-Harris, V., Drexler, A., Watkins de Jong, L., Antonova, Y., Pakpour, N., Ziegler, R., Ramberg, F., Lewis, E.E., Brown, J.M., Luckhart, S., *et al.* (2010). Activation of Akt Signaling Reduces the Prevalence and Intensity of Malaria Parasite Infection and Lifespan in Anopheles stephensi Mosquitoes. PLoS Pathog *6*, e1001003.
Corby-Harris, V., Pontaroli, A.C., Shimkets, L.J., Bennetzen, J.L., Habel, K.E., and Promislow, D.E. (2007). Geographical distribution and diversity of bacteria associated with natural populations of Drosophila melanogaster. Appl Environ Microbiol *73*, 3470-3479.
Costantini, C., Ayala, D., Guelbeogo, W.M., Pombi, M., Some, C.Y., Bassole, I.H., Ose, K., Fotsing, J.M., Sagnon, N., Fontenille, D., *et al.* (2009). Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in Anopheles gambiae. BMC ecology *9*, 16.
Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. Science *326*, 1694-1697.
Costello, E.K., Stagaman, K., Dethlefsen, L., Bohannan, B.J., and Relman, D.A. (2012). The application of ecological theory toward an understanding of the human microbiome. Science *336*, 1255-1262.
Coutelis, J.B., and Ephrussi, A. (2007). Rab6 mediates membrane organization and determinant localization during Drosophila oogenesis. Development *134*, 1419-1430.
Coutinho, C.P., Dos Santos, S.C., Madeira, A., Mira, N.P., Moreira, A.S., and Sa-Correia, I. (2011). Long-term colonization of the cystic fibrosis lung by Burkholderia cepacia complex bacteria: epidemiology, clonal variation, and genome-wide expression alterations. Frontiers in cellular and infection microbiology *1*, 12.
Cowman, A.F., Galatis, D., and Thompson, J.K. (1994). Selection for mefloquine resistance in Plasmodium falciparum is linked to amplification of the pfmdr1 gene and cross-resistance to halofantrine and quinine. Proc Natl Acad Sci U S A *91*, 1143-1147.
Cox, C.R., and Gilmore, M.S. (2007). Native microbial colonization of Drosophila melanogaster and its use as a model of Enterococcus faecalis pathogenesis. Infect Immun *75*, 1565-1576.
Cox, F.E. (2010). History of the discovery of the malaria parasites and their vectors. Parasit Vectors *3*, 5.

Crampton, J.M., Beard, C.B., Louis, C., and World Health Organization. (1997). The molecular biology of insect disease vectors : a methods manual (London: Chapman & Hall).
Cronin, S.J., Nehme, N.T., Limmer, S., Liegeois, S., Pospisilik, J.A., Schramek, D., Leibbrandt, A., Simoes Rde, M., Gruber, S., Puc, U*., et al.* (2009). Genome-wide RNAi screen identifies genes involved in intestinal pathogenic bacterial infection. Science *325*, 340-343.
Crotti, E., Damiani, C., Pajoro, M., Gonella, E., Rizzi, A., Ricci, I., Negri, I., Scuppa, P., Rossi, P., Ballarini, P*., et al.* (2009). Asaia, a versatile acetic acid bacterial symbiont, capable of cross-colonizing insects of phylogenetically distant genera and orders. Environ Microbiol.
Crotti, E., Rizzi, A., Chouaia, B., Ricci, I., Favia, G., Alma, A., Sacchi, L., Bourtzis, K., Mandrioli, M., Cherif, A*., et al.* (2010). Acetic acid bacteria, new emerging symbionts of insects. Appl Environ Microbiol.
Cui, X., and Churchill, G.A. (2003). Statistical tests for differential expression in cDNA microarray experiments. Genome Biol *4*, 210.
Cunha, C.B., and Cunha, B.A. (2008). Brief history of the clinical diagnosis of malaria: from Hippocrates to Osler. J Vector Borne Dis *45*, 194-199.
Curtis, M.A., Zenobia, C., and Darveau, R.P. (2011). The relationship of the oral microbiotia to periodontal health and disease. Cell Host Microbe *10*, 302-306.
Dahanukar, A., Foster, K., van der Goes van Naters, W.M., and Carlson, J.R. (2001). A Gr receptor is required for response to the sugar trehalose in taste neurons of Drosophila. Nat Neurosci *4*, 1182-1186.
Dalma-Weiszhausz, D.D., Warrington, J., Tanimoto, E.Y., and Miyada, C.G. (2006). The affymetrix GeneChip platform: an overview. Methods Enzymol *410*, 3-28.
Damiani, C., Ricci, I., Crotti, E., Rossi, P., Rizzi, A., Scuppa, P., Capone, A., Ulissi, U., Epis, S., Genchi, M*., et al.* (2010). Mosquito-Bacteria Symbiosis: The Case of Anopheles gambiae and Asaia. Microb Ecol.
Damiani, C., Ricci, I., Crotti, E., Rossi, P., Rizzi, A., Scuppa, P., Esposito, F., Bandi, C., Daffonchio, D., and Favia, G. (2008). Paternal transmission of symbiotic bacteria in malaria vectors. Curr Biol *18*, R1087-1088.
Danielli, A., Loukeris, T.G., Lagueux, M., Muller, H.M., Richman, A., and Kafatos, F.C. (2000). A modular chitin-binding protease associated with hemocytes and hemolymph in the mosquito Anopheles gambiae. Proc Natl Acad Sci U S A *97*, 7136-7141.
Dauwalder, B., Tsujimoto, S., Moss, J., and Mattox, W. (2002). The Drosophila takeout gene is regulated by the somatic sex-determination pathway and affects male courtship behavior. Genes Dev *16*, 2879-2892.
De Graeve, F., Bahr, A., Chatton, B., and Kedinger, C. (2000). A murine ATFa-associated factor with transcriptional repressing activity. Oncogene *19*, 1807-1819.
De Gregorio, E., Spellman, P.T., Tzou, P., Rubin, G.M., and Lemaitre, B. (2002). The Toll and Imd pathways are the major regulators of the immune response in Drosophila. EMBO J *21*, 2568-2579.
Dearolf, C.R., Topol, J., and Parker, C.S. (1989). The caudal gene product is a direct activator of fushi tarazu transcription during Drosophila embryogenesis. Nature *341*, 340-343.
Degennaro, M., McBride, C.S., Seeholzer, L., Nakagawa, T., Dennis, E.J., Goldman, C., Jasinskiene, N., James, A.A., and Vosshall, L.B. (2013). orco mutant mosquitoes lose strong preference for humans and are not repelled by volatile DEET. Nature.
Dela Cruz, C.S., Liu, W., He, C.H., Jacoby, A., Gornitzky, A., Ma, B., Flavell, R., Lee, C.G., and Elias, J.A. (2012). Chitinase 3-like-1 Promotes Streptococcus pneumoniae Killing and Augments Host Tolerance to Lung Antibacterial Responses. Cell Host Microbe *12*, 34-46.
della Torre, A., Costantini, C., Besansky, N.J., Caccone, A., Petrarca, V., Powell, J.R., and Coluzzi, M. (2002). Speciation within Anopheles gambiae--the glass is half full. Science *298*, 115-117.
della Torre, A., Fanello, C., Akogbeto, M., Dossou-yovo, J., Favia, G., Petrarca, V., and Coluzzi, M. (2001). Molecular evidence of incipient speciation within Anopheles gambiae s.s. in West Africa. Insect Mol Biol *10*, 9-18.
Demaio, J., Pumpuni, C.B., Kent, M., and Beier, J.C. (1996). The midgut bacterial flora of wild Aedes triseriatus, Culex pipiens, and Psorophora columbiae mosquitoes. Am J Trop Med Hyg *54*, 219-223.
Deng, L., Luo, M., Velikovsky, A., and Mariuzza, R.A. (2013). Structural insights into the evolution of the adaptive immune system. Annual review of biophysics *42*, 191-215.
Dennes, A., Cromme, C., Suresh, K., Kumar, N.S., Eble, J.A., Hahnenkamp, A., and Pohlmann, R. (2005). The novel Drosophila lysosomal enzyme receptor protein mediates lysosomal sorting in mammalian cells and binds mammalian and Drosophila GGA adaptors. J Biol Chem *280*, 12849-12857.
Dethlefsen, L., McFall-Ngai, M., and Relman, D.A. (2007). An ecological and evolutionary perspective on human-microbe mutualism and disease. Nature *449*, 811-818.
Diabate, A., Baldet, T., Chandre, C., Dabire, K.R., Kengne, P., Guiguemde, T.R., Simard, F., Guillet, P., Hemingway, J., and Hougard, J.M. (2003). KDR mutation, a genetic marker to assess events of introgression between the molecular M and S forms of Anopheles gambiae (Diptera: Culicidae) in the tropical savannah area of West Africa. J Med Entomol *40*, 195-198.
Dickinson, M.H. (2013). Death Valley, Drosophila, and the Devonian Toolkit. Annu Rev Entomol.
Dickson, B.J., and Gilestro, G.F. (2006). Regulation of Commissural Axon Pathfinding by Slit and its Robo Receptors. Annual review of cell and developmental biology *22*, 651-675.
Dierick, H.A., and Greenspan, R.J. (2007). Serotonin and neuropeptide F have opposite modulatory effects on fly aggression. Nat Genet *39*, 678-682.
Dillon, R., and Charnley, K. (2002). Mutualism between the desert locust Schistocerca gregaria and its gut microbiota. Res Microbiol *153*, 503-509.
Dillon, R.J., and Dillon, V.M. (2004). The gut bacteria of insects: nonpathogenic interactions. Annu Rev Entomol *49*, 71-92.
Dimopoulos, G., Casavant, T.L., Chang, S., Scheetz, T., Roberts, C., Donohue, M., Schultz, J., Benes, V., Bork, P., Ansorge, W*., et al.* (2000). Anopheles gambiae pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines. Proc Natl Acad Sci U S A *97*, 6619-6624.
Dimopoulos, G., Christophides, G.K., Meister, S., Schultz, J., White, K.P., Barillas-Mury, C., and Kafatos, F.C. (2002). Genome expression analysis of Anopheles gambiae: responses to injury, bacterial challenge, and malaria infection. Proc Natl Acad Sci U S A *99*, 8814-8819.
Dinglasan, R.R., Devenport, M., Florens, L., Johnson, J.R., McHugh, C.A., Donnelly-Doman, M., Carucci, D.J., Yates, J.R., 3rd, and Jacobs-Lorena, M. (2009). The Anopheles gambiae adult midgut peritrophic matrix proteome. Insect Biochem Mol Biol *39*, 125-134.
Dinglasan, R.R., and Jacobs-Lorena, M. (2008). Flipping the paradigm on malaria transmission-blocking vaccines. Trends Parasitol *24*, 364-370.

Ditch, L.M., Shirangi, T., Pitman, J.L., Latham, K.L., Finley, K.D., Edeen, P.T., Taylor, B.J., and McKeown, M. (2005). Drosophila retained/dead ringer is necessary for neuronal pathfinding, female receptivity and repression of fruitless independent male courtship behaviors. Development *132*, 155-164.
Dondorp, A.M., Nosten, F., Yi, P., Das, D., Phyo, A.P., Tarning, J., Lwin, K.M., Ariey, F., Hanpithakpong, W., Lee, S.J., *et al.* (2009). Artemisinin resistance in Plasmodium falciparum malaria. N Engl J Med *361*, 455-467.
Dong, Y., Aguilar, R., Xi, Z., Warr, E., Mongin, E., and Dimopoulos, G. (2006a). Anopheles gambiae immune responses to human and rodent Plasmodium parasite species. PLoS Pathog *2*, e52.
Dong, Y., Cirimotich, C.M., Pike, A., Chandra, R., and Dimopoulos, G. (2012). Anopheles NF-kappaB-Regulated Splicing Factors Direct Pathogen-Specific Repertoires of the Hypervariable Pattern Recognition Receptor AgDscam. Cell Host Microbe *12*, 521-530.
Dong, Y., Das, S., Cirimotich, C., Souza-Neto, J.A., McLean, K.J., and Dimopoulos, G. (2011). Engineered anopheles immunity to Plasmodium infection. PLoS Pathog *7*, e1002458.
Dong, Y., and Dimopoulos, G. (2009). Anopheles fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. J Biol Chem *284*, 9835-9844.
Dong, Y., Manfredini, F., and Dimopoulos, G. (2009). Implication of the mosquito midgut microbiota in the defense against malaria parasites. PLoS Pathog *5*, e1000423.
Dong, Y., Taylor, H.E., and Dimopoulos, G. (2006b). AgDscam, a hypervariable immunoglobulin domain-containing receptor of the Anopheles gambiae innate immune system. PLoS Biol *4*, e229.
Donnelly, M.J., Corbel, V., Weetman, D., Wilding, C.S., Williamson, M.S., and Black, W.C.t. (2009). Does kdr genotype predict insecticide-resistance phenotype in mosquitoes? Trends Parasitol *25*, 213-219.
Douglas, A.E. (2011). Lessons from studying insect symbioses. Cell Host Microbe *10*, 359-367.
Drosophila 12 Genomes, C., Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., *et al.* (2007). Evolution of genes and genomes on the Drosophila phylogeny. Nature *450*, 203-218.
Dubrovsky, E.B., Dubrovskaya, V.A., Bernardo, T., Otte, V., DiFilippo, R., and Bryan, H. (2011). The Drosophila FTZ-F1 nuclear receptor mediates juvenile hormone activation of E75A gene expression through an intracellular pathway. J Biol Chem *286*, 33689-33700.
Duerschmied, D., Suidan, G.L., Demers, M., Herr, N., Carbo, C., Brill, A., Cifuni, S.M., Mauler, M., Cicko, S., Bader, M., *et al.* (2013). Platelet serotonin promotes the recruitment of neutrophils to sites of acute inflammation in mice. Blood *121*, 1008-1015.
Duffy, J.B. (2002). GAL4 system in Drosophila: a fly geneticist's Swiss army knife. Genesis *34*, 1-15.
Dzitoyeva, S., Dimitrijevic, N., and Manev, H. (2003). Gamma-aminobutyric acid B receptor 1 mediates behavior-impairing actions of alcohol in Drosophila: adult RNA interference and pharmacological evidence. Proc Natl Acad Sci U S A *100*, 5485-5490.
Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., and Relman, D.A. (2005). Diversity of the human intestinal microbial flora. Science *308*, 1635-1638.
Endo, Y., Matsushita, M., and Fujita, T. (2011). The role of ficolins in the lectin pathway of innate immunity. Int J Biochem Cell Biol *43*, 705-712.
Erdelyan, C.N., Mahood, T.H., Bader, T.S., and Whyard, S. (2011). Functional validation of the carbon dioxide receptor genes in Aedes aegypti mosquitoes using RNA interference. Insect Mol Biol.
Erickson, S.M., Xi, Z., Mayhew, G.F., Ramirez, J.L., Aliota, M.T., Christensen, B.M., and Dimopoulos, G. (2009). Mosquito infection responses to developing filarial worms. PLoS Negl Trop Dis *3*, e529.
Erkosar, B., Storelli, G., Defaye, A., and Leulier, F. (2013). Host-intestinal microbiota mutualism: "learning on the fly". Cell Host Microbe *13*, 8-14.
Escalante, A.A., and Ayala, F.J. (1994). Phylogeny of the malarial genus Plasmodium, derived from rRNA gene sequences. Proc Natl Acad Sci U S A *91*, 11373-11377.
Evans, D.L., Lynch, K.G., Benton, T., Dube, B., Gettes, D.R., Tustin, N.B., Lai, J.P., Metzger, D., and Douglas, S.D. (2008). Selective serotonin reuptake inhibitor and substance P antagonist enhancement of natural killer cell innate immunity in human immunodeficiency virus/acquired immunodeficiency syndrome. Biol Psychiatry *63*, 899-905.
Evans, T.A., and Bashaw, G.J. (2010). Axon guidance at the midline: of mice and flies. Current opinion in neurobiology *20*, 79-85.
Ezenwa, V.O., Gerardo, N.M., Inouye, D.W., Medina, M., and Xavier, J.B. (2012). Microbiology. Animal behavior and the microbiome. Science *338*, 198-199.
Faith, J.J., McNulty, N.P., Rey, F.E., and Gordon, J.I. (2011). Predicting a human gut microbiota's response to diet in gnotobiotic mice. Science *333*, 101-104.
Falk, P.G., Hooper, L.V., Midtvedt, T., and Gordon, J.I. (1998). Creating and maintaining the gastrointestinal ecosystem: what we know and need to know from gnotobiology. Microbiol Mol Biol Rev *62*, 1157-1170.
Fang, W., Vega-Rodriguez, J., Ghosh, A.K., Jacobs-Lorena, M., Kang, A., and St Leger, R.J. (2011). Development of transgenic fungi that kill human malaria parasites in mosquitoes. Science *331*, 1074-1077.
Favia, G., della Torre, A., Bagayoko, M., Lanfrancotti, A., Sagnon, N., Toure, Y.T., and Coluzzi, M. (1997). Molecular identification of sympatric chromosomal forms of Anopheles gambiae and further evidence of their reproductive isolation. Insect Mol Biol *6*, 377-383.
Favia, G., Dimopoulos, G., della Torre, A., Toure, Y.T., Coluzzi, M., and Louis, C. (1994). Polymorphisms detected by random PCR distinguish between different chromosomal forms of Anopheles gambiae. Proc Natl Acad Sci U S A *91*, 10315-10319.
Favia, G., Lanfrancotti, A., Spanos, L., Siden-Kiamos, I., and Louis, C. (2001). Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of Anopheles gambiae s.s. Insect Mol Biol *10*, 19-23.
Favia, G., Ricci, I., Damiani, C., Raddadi, N., Crotti, E., Marzorati, M., Rizzi, A., Urso, R., Brusetti, L., Borin, S., *et al.* (2007). Bacteria of the genus Asaia stably associate with Anopheles stephensi, an Asian malarial mosquito vector. Proc Natl Acad Sci U S A *104*, 9047-9051.
Favia, G., Ricci, I., Marzorati, M., Negri, I., Alma, A., Sacchi, L., Bandi, C., and Daffonchio, D. (2008). Bacteria of the genus Asaia: a potential paratransgenic weapon against malaria. Adv Exp Med Biol *627*, 49-59.
Felix, R.C., and Silveira, H. (2011). The Interplay between Tubulins and P450 Cytochromes during Plasmodium berghei Invasion of Anopheles gambiae Midgut. PLoS One *6*, e24181.
Ferreira, C., and Veldhoen, M. (2012). Host and microbes date exclusively. Cell *149*, 1428-1430.
Figlewicz, D.P., and Benoit, S.C. (2009). Insulin, leptin, and food reward: update 2008. American journal of physiology Regulatory, integrative and comparative physiology *296*, R9-R19.

Fleckenstein, E.D., Hans G. (1996). Elimination of Mycoplasma Contamination in Cell Cultures. Biochemica *1*, 48-51.
Flower, D.R. (1999). Modelling G-protein-coupled receptors for drug design. Biochim Biophys Acta *1422*, 207-234.
Flyg, C., Kenne, K., and Boman, H.G. (1980). Insect pathogenic properties of Serratia marcescens: phage-resistant mutants with a decreased resistance to Cecropia immunity and a decreased virulence to Drosophila. J Gen Microbiol *120*, 173-181.
Foley, D.H., Rueda, L.M., and Wilkerson, R.C. (2007). Insight into global mosquito biogeography from country species records. J Med Entomol *44*, 554-567.
Foussard, H., Ferrer, P., Valenti, P., Polesello, C., Carreno, S., and Payre, F. (2010). LRCH proteins: a novel family of cytoskeletal regulators. PLoS One *5*, e12257.
Fraiture, M., Baxter, R.H., Steinert, S., Chelliah, Y., Frolet, C., Quispe-Tintaya, W., Hoffmann, J.A., Blandin, S.A., and Levashina, E.A. (2009). Two mosquito LRR proteins function as complement control factors in the TEP1-mediated killing of Plasmodium. Cell Host Microbe *5*, 273-284.
Fraser, M.J. (2012). Insect transgenesis: current applications and future prospects. Annu Rev Entomol *57*, 267-289.
Fredriksson, R., Lagerstrom, M.C., Lundin, L.G., and Schioth, H.B. (2003). The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Molecular pharmacology *63*, 1256-1272.
Frolet, C., Thoma, M., Blandin, S., Hoffmann, J.A., and Levashina, E.A. (2006). Boosting NF-kappaB-dependent basal immunity of Anopheles gambiae aborts development of Plasmodium berghei. Immunity *25*, 677-685.
Fukumitsu, Y., Irie, K., Satho, T., Aonuma, H., Dieng, H., Ahmad, A.H., Nakashima, Y., Mishima, K., Kashige, N., and Miake, F. (2012). Elevation of dopamine level reduces host-seeking activity in the adult female mosquito Aedes albopictus. Parasit Vectors *5*, 92.
Funkhouser, J.D., and Aronson, N.N., Jr. (2007). Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. BMC Evol Biol *7*, 96.
Gaio Ade, O., Gusmao, D.S., Santos, A.V., Berbert-Molina, M.A., Pimenta, P.F., and Lemos, F.J. (2011). Contribution of midgut bacteria to blood digestion and egg production in aedes aegypti (diptera: culicidae) (L.). Parasit Vectors *4*, 105.
Ganal, S.C., Sanos, S.L., Kallfass, C., Oberle, K., Johner, C., Kirschning, C., Lienenklaus, S., Weiss, S., Staeheli, P., Aichele, P.*, et al.* (2012). Priming of natural killer cells by nonmucosal mononuclear phagocytes requires instructive signals from commensal microbiota. Immunity *37*, 171-186.
Garbe, D.S., and Bashaw, G.J. (2007). Independent Functions of Slit-Robo Repulsion and Netrin-Frazzled Attraction Regulate Axon Crossing at the Midline in Drosophila. Journal of Neuroscience *27*, 3584-3592.
Garver, L.S., Bahia, A.C., Das, S., Souza-Neto, J.A., Shiao, J., Dong, Y., and Dimopoulos, G. (2012). Anopheles imd pathway factors and effectors in infection intensity-dependent anti-Plasmodium action. PLoS Pathog *8*, e1002737.
Garver, L.S., Dong, Y., and Dimopoulos, G. (2009). Caspar controls resistance to Plasmodium falciparum in diverse anopheline species. PLoS Pathog *5*, e1000335.
Garver, L.S., Xi, Z., and Dimopoulos, G. (2008). Immunoglobulin superfamily members play an important role in the mosquito immune system. Dev Comp Immunol *32*, 519-531.
Ge, X., Rodriguez, R., Trinh, M., Gunsolley, J., and Xu, P. (2013). Oral microbiome of deep and shallow dental pockets in chronic periodontitis. PLoS One *8*, e65520.
Gendrin, M., and Christophides, G.K. (2013). The Anopheles Mosquito Microbiota and Their Impact on Pathogen Transmission.
Gendrin, M., Zaidman-Rémy, A., Broderick, N.A., Paredes, J., Poidevin, M., Roussel, A., and Lemaitre, B. (2013). Functional Analysis of PGRP-LA in Drosophila Immunity. PLoS ONE *8*, e69742.
Gentile, G., Della Torre, A., Maegga, B., Powell, J.R., and Caccone, A. (2002). Genetic differentiation in the African malaria vector, Anopheles gambiae s.s., and the problem of taxonomic status. Genetics *161*, 1561-1578.
Gentile, G., Slotman, M., Ketmaier, V., Powell, J.R., and Caccone, A. (2001). Attempts to molecularly distinguish cryptic taxa in Anopheles gambiae s.s. Insect Mol Biol *10*, 25-32.
Gesellchen, V., Kuttenkeuler, D., Steckel, M., Pelte, N., and Boutros, M. (2005). An RNA interference screen identifies Inhibitor of Apoptosis Protein 2 as a regulator of innate immune signalling in Drosophila. EMBO Rep *6*, 979-984.
Gharizadeh, B., Akhras, M., Nourizad, N., Ghaderi, M., Yasuda, K., Nyren, P., and Pourmand, N. (2006). Methodological improvements of pyrosequencing technology. Journal of biotechnology *124*, 504-511.
Ghosh, A., Edwards, M.J., and Jacobs-Lorena, M. (2000). The journey of the malaria parasite in the mosquito: hopes for the new century. Parasitol Today *16*, 196-201.
Gneme, A., Guelbeogo, W.M., Riehle, M.M., Sanou, A., Traore, A., Zongo, S., Eiglmeier, K., Kabre, G.B., Sagnon, N., and Vernick, K.D. (2013). Equivalent susceptibility of Anopheles gambiae M and S molecular forms and Anopheles arabiensis to Plasmodium falciparum infection in Burkina Faso. Malar J *12*, 204.
Goecks, J., Nekrutenko, A., Taylor, J., and Galaxy, T. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol *11*, R86.
Goncalves, R.L., Oliveira, J.H., Oliveira, G.A., Andersen, J.F., Oliveira, M.F., Oliveira, P.L., and Barillas-Mury, C. (2012). Mitochondrial reactive oxygen species modulate mosquito susceptibility to Plasmodium infection. PLoS One *7*, e41083.
Gonzalez-Ceron, L., Santillan, F., Rodriguez, M.H., Mendez, D., and Hernandez-Avila, J.E. (2003). Bacteria in midguts of field-collected Anopheles albimanus block Plasmodium vivax sporogonic development. J Med Entomol *40*, 371-374.
Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., Knight, R., and Gordon, J.I. (2009). Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe *6*, 279-289.
Gordon, J.I. (2012). Honor thy gut symbionts redux. Science *336*, 1251-1253.
Gorman, J. (2013). Focusing on Fruit Flies, Curiosity Takes Flight. In The New York Times (New York, NY, USA: The New York Times Company).
Gorman, M.J., Severson, D.W., Cornel, A.J., Collins, F.H., and Paskewitz, S.M. (1997). Mapping a quantitative trait locus involved in melanotic encapsulation of foreign bodies in the malaria vector, Anopheles gambiae. Genetics *146*, 965-971.
Gottar, M., Gobert, V., Michel, T., Belvin, M., Duyk, G., Hoffmann, J.A., Ferrandon, D., and Royet, J. (2002). The Drosophila immune response against Gram-negative bacteria is mediated by a peptidoglycan recognition protein. Nature *416*, 640-644.
Granier, S., Manglik, A., Kruse, A.C., Kobilka, T.S., Thian, F.S., Weis, W.I., and Kobilka, B.K. (2012). Structure of the delta-opioid receptor bound to naltrindole. Nature *485*, 400-404.

Graveley, B.R., Kaur, A., Gunning, D., Zipursky, S.L., Rowen, L., and Clemens, J.C. (2004). The organization and evolution of the dipteran and hymenopteran Down syndrome cell adhesion molecule (Dscam) genes. RNA *10*, 1499-1506.
Gray, E.M., Rocca, K.A., Costantini, C., and Besansky, N.J. (2009). Inversion 2La is associated with enhanced desiccation resistance in Anopheles gambiae. Malar J *8*, 215.
Gray, J.M., Karow, D.S., Lu, H., Chang, A.J., Chang, J.S., Ellis, R.E., Marletta, M.A., and Bargmann, C.I. (2004). Oxygen sensation and social feeding mediated by a C. elegans guanylate cyclase homologue. Nature *430*, 317-322.
Greenwood, B., Bhasin, A., and Targett, G. (2012). The Gates Malaria Partnership: a consortium approach to malaria research and capacity development. Trop Med Int Health *17*, 558-563.
Grimaldi, D.E., MS (2005). Evolution of the Insects (Cambridge: Cambridge University Press).
Grimont, P.A., and Grimont, F. (1978). The genus Serratia. Annual review of microbiology *32*, 221-248.
Grimont, P.A., Grimont, F., and De Rosnay, H.L. (1977). Taxonomy of the genus Serratia. J Gen Microbiol *98*, 39-66.
Grossman, G.L., Rafferty, C.S., Clayton, J.R., Stevens, T.K., Mukabayire, O., and Benedict, M.Q. (2001). Germline transformation of the malaria vector, Anopheles gambiae, with the piggyBac transposable element. Insect Mol Biol *10*, 597-604.
Grossman, S.R., Andersen, K.G., Shlyakhter, I., Tabrizi, S., Winnicki, S., Yen, A., Park, D.J., Griesemer, D., Karlsson, E.K., Wong, S.H*., et al.* (2013). Identifying recent adaptations in large-scale genomic data. Cell *152*, 703-713.
Grubbs, N., Leach, M., Su, X., Petrisko, T., Rosario, J.B., and Mahaffey, J.W. (2013). New Components of Drosophila Leg Development Identified through Genome Wide Association Studies. PLoS One *8*, e60261.
Gudmundsson, J., Sulem, P., Gudbjartsson, D.F., Masson, G., Agnarsson, B.A., Benediktsdottir, K.R., Sigurdsson, A., Magnusson, O.T., Gudjonsson, S.A., Magnusdottir, D.N*., et al.* (2012). A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. Nat Genet *44*, 1326-1329.
Guerra, C.A., Snow, R.W., and Hay, S.I. (2006). Mapping the global extent of malaria in 2005. Trends Parasitol *22*, 353-358.
Gupta, L., Molina-Cruz, A., Kumar, S., Rodrigues, J., Dixit, R., Zamora, R.E., and Barillas-Mury, C. (2009). The STAT pathway mediates late-phase immunity against Plasmodium in the mosquito Anopheles gambiae. Cell Host Microbe *5*, 498-507.
Ha, E.M., Oh, C.T., Ryu, J.H., Bae, Y.S., Kang, S.W., Jang, I.H., Brey, P.T., and Lee, W.J. (2005). An antioxidant system required for host protection against gut infection in Drosophila. Dev Cell *8*, 125-132.
Habtewold, T., Povelones, M., Blagborough, A.M., and Christophides, G.K. (2008). Transmission blocking immunity in the malaria non-vector mosquito Anopheles quadriannulatus species A. PLoS Pathog *4*, e1000070.
Haga, K., Kruse, A.C., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., Zhang, C., Weis, W.I., Okada, T., Kobilka, B.K., Haga, T*., et al.* (2012). Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. Nature *482*, 547-551.
Hallem, E.A., Dahanukar, A., and Carlson, J.R. (2006). Insect odor and taste receptors. Annu Rev Entomol *51*, 113-135.
Hamady, M., and Knight, R. (2009). Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. Genome Res *19*, 1141-1152.
Han, Y.S., Thompson, J., Kafatos, F.C., and Barillas-Mury, C. (2000). Molecular interactions between Anopheles stephensi midgut cells and Plasmodium berghei: the time bomb theory of ookinete invasion of mosquitoes. EMBO J *19*, 6030-6040.
Hand, T.W., Dos Santos, L.M., Bouladoux, N., Molloy, M.J., Pagan, A.J., Pepper, M., Maynard, C.L., Elson, C.O., 3rd, and Belkaid, Y. (2012). Acute gastrointestinal infection induces long-lived microbiota-specific T cell responses. Science *337*, 1553-1556.
Harari-Steinberg, O., Cantera, R., Denti, S., Bianchi, E., Oron, E., Segal, D., and Chamovitz, D.A. (2007). COP9 signalosome subunit 5 (CSN5/Jab1) regulates the development of the Drosophila immune system: effects on Cactus, Dorsal and hematopoiesis. Genes to cells : devoted to molecular & cellular mechanisms *12*, 183-195.
Harbach, R.E. (2013). Mosquito Taxonomic Inventory. http://mosquito-taxonomic-inventoryinfo/ *accessed on June 2013*.
Harris, C., Lambrechts, L., Rousset, F., Abate, L., Nsango, S.E., Fontenille, D., Morlais, I., and Cohuet, A. (2010a). Polymorphisms in Anopheles gambiae immune genes associated with natural resistance to Plasmodium falciparum. PLoS Pathog *6*.
Harris, C., Rousset, F., Morlais, I., Fontenille, D., and Cohuet, A. (2010b). Low linkage disequilibrium in wild Anopheles gambiae s.l. populations. BMC Genet *11*, 81.
Hartigan, J.A.W., M.A. (1979). Algorithm AS 136: A K-Means Clusering Algorithm. Journal of the Royal Statistical Society Series C (Applied Statistics) *28*, 100-108.
Hartl, D.L., Volkman, S.K., Nielsen, K.M., Barry, A.E., Day, K.P., Wirth, D.F., and Winzeler, E.A. (2002). The paradoxical population genetics of Plasmodium falciparum. Trends Parasitol *18*, 266-272.
Hattori, D., Chen, Y., Matthews, B.J., Salwinski, L., Sabatti, C., Grueber, W.B., and Zipursky, S.L. (2009). Robust discrimination between self and non-self neurites requires thousands of Dscam1 isoforms. Nature *461*, 644-648.
Hay, S.I., and Snow, R.W. (2006). The malaria Atlas Project: developing global maps of malaria risk. PLoS Med *3*, e473.
Hergarden, A.C., Tayler, T.D., and Anderson, D.J. (2012). Allatostatin-A neurons inhibit feeding behavior in adult Drosophila. Proc Natl Acad Sci U S A *109*, 3967-3972.
Hernandez, M.E., Martinez-Fong, D., Perez-Tapia, M., Estrada-Garcia, I., Estrada-Parra, S., and Pavon, L. (2010). Evaluation of the effect of selective serotonin-reuptake inhibitors on lymphocyte subsets in patients with a major depressive disorder. European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology *20*, 88-95.
Heutink, P., and Oostra, B.A. (2002). Gene finding in genetically isolated populations. Hum Mol Genet *11*, 2507-2515.
Hill, C.A., Fox, A.N., Pitts, R.J., Kent, L.B., Tan, P.L., Chrystal, M.A., Cravchik, A., Collins, F.H., Robertson, H.M., and Zwiebel, L.J. (2002). G protein-coupled receptors in Anopheles gambiae. Science *298*, 176-178.
Hino, T., Arakawa, T., Iwanari, H., Yurugi-Kobayashi, T., Ikeda-Suno, C., Nakada-Nakura, Y., Kusano-Arai, O., Weyand, S., Shimamura, T., Nomura, N*., et al.* (2012). G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. Nature *482*, 237-240.
Hoeppli, R. (1956). The knowledge of parasites and parasitic infections from ancient times to the 17th century. Exp Parasitol *5*, 398-419.
Hoffmann, A.A., Montgomery, B.L., Popovici, J., Iturbe-Ormaetxe, I., Johnson, P.H., Muzzi, F., Greenfield, M., Durkan, M., Leong, Y.S., Dong, Y*., et al.* (2011). Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. Nature *476*, 454-457.

Holm, I., Lavazec, C., Garnier, T., Mitri, C., Riehle, M.M., Bischoff, E., Brito-Fravallo, E., Takashima, E., Thiery, I., Zettor, A.*, et al.* (2012). Diverged Alleles of the Anopheles gambiae Leucine-Rich Repeat Gene APL1A Display Distinct Protective Profiles against Plasmodium falciparum. PLoS One *7*, e52684.
Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R.*, et al.* (2002). The genome sequence of the malaria mosquito Anopheles gambiae. Science *298*, 129-149.
Hooper, L.V., Littman, D.R., and Macpherson, A.J. (2012). Interactions between the microbiota and the immune system. Science *336*, 1268-1273.
Horton, A.A., Lee, Y., Coulibaly, C.A., Rashbrook, V.K., Cornel, A.J., Lanzaro, G.C., and Luckhart, S. (2010). Identification of three single nucleotide polymorphisms in Anopheles gambiae immune signaling genes that are associated with natural Plasmodium falciparum infection. Malar J *9*, 160.
Huber, A.H., Wang, Y.M., Bieber, A.J., and Bjorkman, P.J. (1994). Crystal structure of tandem type III fibronectin domains from Drosophila neuroglian at 2.0 A. Neuron *12*, 717-731.
Huff, C.G. (1929). The effects of selection upon susceptibility to bird malaria in Culex pipiens. Linn Ann Trop Med Parasitol *23*, 427-442.
Hughes, G.L., Pike, A.D., Xue, P., and Rasgon, J.L. (2012). Invasion of Wolbachia into Anopheles and Other Insect Germlines in an Ex vivo Organ Culture System. PLoS One *7*, e36277.
Hughes, G.L., Ren, X., Ramirez, J.L., Sakamoto, J.M., Bailey, J.A., Jedlicka, A.E., and Rasgon, J.L. (2011). Wolbachia infections in Anopheles gambiae cells: transcriptomic characterization of a novel host-symbiont interaction. PLoS Pathog *7*, e1001296.
Hughes, M.E., Bortnick, R., Tsubouchi, A., Baumer, P., Kondo, M., Uemura, T., and Schmucker, D. (2007). Homophilic Dscam interactions control complex dendrite morphogenesis. Neuron *54*, 417-427.
Hummel, T., Vasconcelos, M.L., Clemens, J.C., Fishilevich, Y., Vosshall, L.B., and Zipursky, S.L. (2003). Axonal targeting of olfactory receptor neurons in Drosophila is controlled by Dscam. Neuron *37*, 221-231.
Hunt, R.H., Coetzee, M., and Fettene, M. (1998). The Anopheles gambiae complex: a new species from Ethiopia. Trans R Soc Trop Med Hyg *92*, 231-235.
Hurst, M.R., Glare, T.R., Jackson, T.A., and Ronson, C.W. (2000). Plasmid-located pathogenicity determinants of Serratia entomophila, the causal agent of amber disease of grass grub, show similarity to the insecticidal toxins of Photorhabdus luminescens. J Bacteriol *182*, 5127-5138.
Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N., and Schuster, S.C. (2011). Integrative analysis of environmental sequences using MEGAN4. Genome Res *21*, 1552-1560.
Inomata, N., Goto, H., Itoh, M., and Isono, K. (2004). A single-amino-acid change of the gustatory receptor gene, Gr5a, has a major effect on trehalose sensitivity in a natural population of Drosophila melanogaster. Genetics *167*, 1749-1758.
Isaacs, A.T., Jasinskiene, N., Tretiakov, M., Thiery, I., Zettor, A., Bourgouin, C., and James, A.A. (2012). Transgenic Anopheles stephensi coexpressing single-chain antibodies resist Plasmodium falciparum development. Proc Natl Acad Sci U S A.
Isaacs, A.T., Li, F., Jasinskiene, N., Chen, X., Nirmala, X., Marinotti, O., Vinetz, J.M., and James, A.A. (2011). Engineered resistance to Plasmodium falciparum development in transgenic Anopheles stephensi. PLoS Pathog *7*, e1002017.
Ito, J., Ghosh, A., Moreira, L.A., Wimmer, E.A., and Jacobs-Lorena, M. (2002). Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. Nature *417*, 452-455.
Iwanaga, S., Kawabata, S., and Muta, T. (1998). New types of clotting factors and defense molecules found in horseshoe crab hemolymph: their structures and functions. J Biochem *123*, 1-15.
Ja, W.W., Carvalho, G.B., Mak, E.M., de la Rosa, N.N., Fang, A.Y., Liong, J.C., Brummel, T., and Benzer, S. (2007). Prandiology of Drosophila and the CAFE assay. Proc Natl Acad Sci U S A *104*, 8253-8256.
James, A.A. (2002). Engineering mosquito resistance to malaria parasites: the avian malaria model. Insect Biochem Mol Biol *32*, 1317-1323.
Jang, I.H., Chosa, N., Kim, S.H., Nam, H.J., Lemaitre, B., Ochiai, M., Kambris, Z., Brun, S., Hashimoto, C., Ashida, M.*, et al.* (2006). A Spatzle-processing enzyme required for toll signaling activation in Drosophila innate immunity. Dev Cell *10*, 45-55.
Jang, I.H., Nam, H.J., and Lee, W.J. (2008). CLIP-domain serine proteases in Drosophila innate immunity. BMB Rep *41*, 102-107.
Jasinskiene, N., Coates, C.J., Benedict, M.Q., Cornel, A.J., Rafferty, C.S., James, A.A., and Collins, F.H. (1998). Stable transformation of the yellow fever mosquito, Aedes aegypti, with the Hermes element from the housefly. Proc Natl Acad Sci U S A *95*, 3743-3747.
Jiang, H., and Kanost, M.R. (2000). The clip-domain family of serine proteinases in arthropods. Insect Biochem Mol Biol *30*, 95-105.
Jiang, H., Patel, P.H., Kohlmaier, A., Grenley, M.O., McEwen, D.G., and Edgar, B.A. (2009). Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the Drosophila midgut. Cell *137*, 1343-1355.
Jiang, H., Wang, Y., Yu, X.Q., Zhu, Y., and Kanost, M. (2003). Prophenoloxidase-activating proteinase-3 (PAP-3) from Manduca sexta hemolymph: a clip-domain serine proteinase regulated by serpin-1J and serine proteinase homologs. Insect Biochem Mol Biol *33*, 1049-1060.
Jiao, Y., Moon, S.J., and Montell, C. (2007). A Drosophila gustatory receptor required for the responses to sucrose, glucose, and maltose identified by mRNA tagging. Proc Natl Acad Sci U S A *104*, 14110-14115.
Johnson, A.W., Sidman, J.D., and Lin, J. (2013). Bioluminescent imaging of pneumococcal otitis media in chinchillas. The Annals of otology, rhinology, and laryngology *122*, 344-352.
Johnson, E.C., Kazgan, N., Bretz, C.A., Forsberg, L.J., Hector, C.E., Worthen, R.J., Onyenwoke, R., and Brenman, J.E. (2010). Altered metabolism and persistent starvation behaviors caused by reduced AMPK function in Drosophila. PLoS One *5*.
Johnson, O., Becnel, J., and Nichols, C.D. (2011). Serotonin receptor activity is necessary for olfactory learning and memory in Drosophila melanogaster. Neuroscience *192*, 372-381.
Jones, W.D., Cayirlioglu, P., Kadow, I.G., and Vosshall, L.B. (2007). Two chemosensory receptors together mediate carbon dioxide detection in Drosophila. Nature *445*, 86-90.
Joy, D.A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Krettli, A.U., Ho, M., Wang, A., White, N.J., Suh, E.*, et al.* (2003). Early origin and recent expansion of Plasmodium falciparum. Science *300*, 318-321.
Jung, S.H., and Jeong, J.H. (2003). Rank tests for clustered survival data. Lifetime data analysis *9*, 21-33.
Kacsoh, B.Z., Lynch, Z.R., Mortimer, N.T., and Schlenke, T.A. (2013). Fruit flies medicate offspring after seeing parasites. Science *339*, 947-950.
Kajla, M.K., Andreeva, O., Gilbreath, T.M., 3rd, and Paskewitz, S.M. (2010). Characterization of expression, activity and role in antibacterial immunity of Anopheles gambiae lysozyme c-1. Comp Biochem Physiol B Biochem Mol Biol *155*, 201-209.

Kajla, M.K., Shi, L., Li, B., Luckhart, S., Li, J., and Paskewitz, S.M. (2011). A New Role for an Old Antimicrobial: Lysozyme c-1 Can Function to Protect Malaria Parasites in Anopheles Mosquitoes. PLoS One *6*, e19649.
Kaltenpoth, M., Gottler, W., Herzner, G., and Strohm, E. (2005). Symbiotic bacteria protect wasp larvae from fungal infestation. Curr Biol *15*, 475-479.
Kamali, M., Xia, A., Tu, Z., and Sharakhov, I.V. (2012). A New Chromosomal Phylogeny Supports the Repeated Origin of Vectorial Capacity in Malaria Mosquitoes of the Anopheles gambiae Complex. PLoS Pathog *8*, e1002960.
Kambris, Z., Blagborough, A.M., Pinto, S.B., Blagrove, M.S., Godfray, H.C., Sinden, R.E., and Sinkins, S.P. (2010). Wolbachia stimulates immune gene expression and inhibits plasmodium development in Anopheles gambiae. PLoS Pathog *6*.
Kambris, Z., Brun, S., Jang, I.H., Nam, H.J., Romeo, Y., Takahashi, K., Lee, W.J., Ueda, R., and Lemaitre, B. (2006). Drosophila immunity: a large-scale in vivo RNAi screen identifies five serine proteases required for Toll activation. Curr Biol *16*, 808-813.
Kambris, Z., Cook, P.E., Phuc, H.K., and Sinkins, S.P. (2009). Immune activation by life-shortening Wolbachia and reduced filarial competence in mosquitoes. Science *326*, 134-136.
Kamhawi, S., Ramalho-Ortigao, M., Pham, V.M., Kumar, S., Lawyer, P.G., Turco, S.J., Barillas-Mury, C., Sacks, D.L., and Valenzuela, J.G. (2004). A role for insect galectins in parasite survival. Cell *119*, 329-341.
Kampfer, P., Lindh, J.M., Terenius, O., Haghdoost, S., Falsen, E., Busse, H.J., and Faye, I. (2006a). Thorsellia anophelis gen. nov., sp. nov., a new member of the Gammaproteobacteria. Int J Syst Evol Microbiol *56*, 335-338.
Kampfer, P., Mathews, H., Glaeser, S.P., Martin, K., Lodders, N., and Faye, I. (2010). Elizabethkingia anophelis sp. nov., isolated from the midgut of Anopheles gambiae. Int J Syst Evol Microbiol.
Kampfer, P., Terenius, O., Lindh, J.M., and Faye, I. (2006b). Janibacter anophelis sp. nov., isolated from the midgut of Anopheles arabiensis. Int J Syst Evol Microbiol *56*, 389-392.
Kaneko, T., Goldman, W.E., Mellroth, P., Steiner, H., Fukase, K., Kusumoto, S., Harley, W., Fox, A., Golenbock, D., and Silverman, N. (2004). Monomeric and polymeric gram-negative peptidoglycan but not purified LPS stimulate the Drosophila IMD pathway. Immunity *20*, 637-649.
Kaneko, T., Yano, T., Aggarwal, K., Lim, J.H., Ueda, K., Oshima, Y., Peach, C., Erturk-Hasdemir, D., Goldman, W.E., Oh, B.H.*, et al.* (2006). PGRP-LC and PGRP-LE have essential yet distinct functions in the drosophila immune response to monomeric DAP-type peptidoglycan. Nat Immunol *7*, 715-723.
Karamanou, M., Panayiotakopoulos, G., Tsoucalas, G., Kousoulis, A.A., and Androutsos, G. (2012). From miasmas to germs: a historical approach to theories of infectious disease transmission. Le infezioni in medicina : rivista periodica di eziologia, epidemiologia, diagnostica, clinica e terapia delle patologie infettive *20*, 58-62.
Katoh, N., Soga, F., Nara, T., Tamagawa-Mineoka, R., Nin, M., Kotani, H., Masuda, K., and Kishimoto, S. (2006). Effect of serotonin on the differentiation of human monocytes into dendritic cells. Clin Exp Immunol *146*, 354-361.
Katsura, K., Kawasaki, H., Potacharoen, W., Saono, S., Seki, T., Yamada, Y., Uchimura, T., and Komagata, K. (2001). Asaia siamensis sp. nov., an acetic acid bacterium in the alpha-proteobacteria. Int J Syst Evol Microbiol *51*, 559-563.
Kazura, J.W., Zou, Z., Souza-Neto, J., Xi, Z., Kokoza, V., Shin, S.W., Dimopoulos, G., and Raikhel, A. (2011). Transcriptome Analysis of Aedes aegypti Transgenic Mosquitoes with Altered Immunity. PLoS Pathogens *7*, e1002394.
Khan, W.I., and Ghia, J.E. (2010). Gut hormones: emerging role in immune activation and inflammation. Clin Exp Immunol *161*, 19-27.
Kidd, T., Brose, K., Mitchell, K.J., Fetter, R.D., Tessier-Lavigne, M., Goodman, C.S., and Tear, G. (1998). Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. Cell *92*, 205-215.
Kikuchi, Y., Hosokawa, T., and Fukatsu, T. (2007). Insect-microbe mutualism without vertical transmission: a stinkbug acquires a beneficial gut symbiont from the environment every generation. Appl Environ Microbiol *73*, 4308-4316.
Kikuchi, Y., Hosokawa, T., and Fukatsu, T. (2011a). An ancient but promiscuous host-symbiont association between Burkholderia gut symbionts and their heteropteran hosts. ISME J *5*, 446-460.
Kikuchi, Y., Hosokawa, T., and Fukatsu, T. (2011b). Specific developmental window for establishment of an insect-microbe gut symbiosis. Appl Environ Microbiol *77*, 4075-4081.
Kikuchi, Y., Meng, X.Y., and Fukatsu, T. (2005). Gut symbiotic bacteria of the genus Burkholderia in the broad-headed bugs Riptortus clavatus and Leptocorisa chinensis (Heteroptera: Alydidae). Appl Environ Microbiol *71*, 4035-4043.
Kim, W., Koo, H., Richman, A.M., Seeley, D., Vizioli, J., Klocko, A.D., and O'Brochta, D.A. (2004). Ectopic expression of a cecropin transgene in the human malaria vector mosquito Anopheles gambiae (Diptera: Culicidae): effects on susceptibility to Plasmodium. J Med Entomol *41*, 447-455.
Kinnamon, S.C. (2012). Taste receptor signalling - from tongues to lungs. Acta physiologica *204*, 158-168.
Kiselyov, V., Skladchikova, G., Hinsby, A., Jensen, P., Kulahin, N., Soroka, V., Pedersen, N., Tsetlin, V., Poulsen, F., and Berezin, V. (2003). Structural Basis for a Direct Interaction between FGFR1 and NCAM and Evidence for a Regulatory Role of ATP. Structure *11*, 691-701.
Klun, J.A., Kramer, M., and Debboun, M. (2013). Four simple stimuli that induce host-seeking and blood-feeding behaviors in two mosquito species, with a clue to DEET's mode of action. Journal of vector ecology : journal of the Society for Vector Ecology *38*, 143-153.
Knols, B.G., Bossin, H.C., Mukabana, W.R., and Robinson, A.S. (2007). Transgenic mosquitoes and the fight against malaria: managing technology push in a turbulent GMO world. Am J Trop Med Hyg *77*, 232-242.
Knols, B.G., Bukhari, T., and Farenhorst, M. (2010). Entomopathogenic fungi as the next-generation control agents against malaria mosquitoes. Future Microbiol *5*, 339-341.
Koch, C.M., Honemann-Capito, M., Egger-Adam, D., and Wodarz, A. (2009). Windei, the Drosophila homolog of mAM/MCAF1, is an essential cofactor of the H3K9 methyl transferase dSETDB1/Eggless in germ line development. PLoS Genet *5*, e1000644.
Kolodziej, P.A., Timpe, L.C., Mitchell, K.J., Fried, S.R., Goodman, C.S., Jan, L.Y., and Jan, Y.N. (1996). frazzled encodes a Drosophila member of the DCC immunoglobulin subfamily and is required for CNS and motor axon guidance. Cell *87*, 197-204.
Koonin, E.V. (2005). Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet *39*, 309-338.
Kounatidis, I., and Ligoxygakis, P. (2012). Drosophila as a model system to unravel the layers of innate immunity to infection. Open biology *2*, 120075.
Kousoulis, A.A., Chatzigeorgiou, K.S., Danis, K., Tsoucalas, G., Vakalis, N., Bonovas, S., and Tsiodras, S. (2012). Malaria in Laconia, Greece, then and now: a 2500-year-old pattern. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases.

Krafts, K., Hempelmann, E., and Skorska-Stania, A. (2012). From methylene blue to chloroquine: a brief review of the development of an antimalarial therapy. Parasitol Res *111*, 1-6.
Krashes, M.J., DasGupta, S., Vreede, A., White, B., Armstrong, J.D., and Waddell, S. (2009). A neural circuit mechanism integrating motivational state with memory expression in Drosophila. Cell *139*, 416-427.
Kruse, A.C., Hu, J., Pan, A.C., Arlow, D.H., Rosenbaum, D.M., Rosemond, E., Green, H.F., Liu, T., Chae, P.S., Dror, R.O., *et al.* (2012). Structure and dynamics of the M3 muscarinic acetylcholine receptor. Nature *482*, 552-556.
Kumar, S. (2003). The role of reactive oxygen species on Plasmodium melanotic encapsulation in Anopheles gambiae. Proceedings of the National Academy of Sciences *100*, 14139-14144.
Kumar, S., Molina-Cruz, A., Gupta, L., Rodrigues, J., and Barillas-Mury, C. (2010). A peroxidase/dual oxidase system modulates midgut epithelial immunity in Anopheles gambiae. Science *327*, 1644-1648.
Kuraishi, T., Binggeli, O., Opota, O., Buchon, N., and Lemaitre, B. (2011). Genetic evidence for a protective role of the peritrophic matrix against intestinal bacterial infection in Drosophila melanogaster. Proc Natl Acad Sci U S A *108*, 15966-15971.
Kurz, C.L., Chauvet, S., Andres, E., Aurouze, M., Vallet, I., Michel, G.P., Uh, M., Celli, J., Filloux, A., De Bentzmann, S., *et al.* (2003). Virulence factors of the human opportunistic pathogen Serratia marcescens identified by in vivo screening. EMBO J *22*, 1451-1460.
Kuss, S.K., Best, G.T., Etheredge, C.A., Pruijssers, A.J., Frierson, J.M., Hooper, L.V., Dermody, T.S., and Pfeiffer, J.K. (2011). Intestinal microbiota promote enteric virus replication and systemic pathogenesis. Science *334*, 249-252.
Lambrechts, L., Fellous, S., and Koella, J.C. (2006). Coevolutionary interactions between host and parasite genotypes. Trends Parasitol *22*, 12-16.
Lambrechts, L., Halbert, J., Durand, P., Gouagna, L.C., and Koella, J.C. (2005). Host genotype by parasite genotype interactions underlying the resistance of anopheline mosquitoes to Plasmodium falciparum. Malar J *4*, 3.
Lancet (2013). The Global Fund: "a historic opportunity". Lancet *381*, 1334.
Lapcharoen, P., Komalamisra, N., Rongsriyam, Y., Wangsuphachart, V., Dekumyoy, P., Prachumsri, J., Kajla, M.K., and Paskewitz, S.M. (2011). Investigations on the role of a lysozyme from the malaria vector Anopheles dirus during malaria parasite development. Dev Comp Immunol.
Lathrop, S.K., Bloom, S.M., Rao, S.M., Nutsch, K., Lio, C.W., Santacruz, N., Peterson, D.A., Stappenbeck, T.S., and Hsieh, C.S. (2011). Peripheral education of the immune system by colonic commensal microbiota. Nature *478*, 250-254.
Lavazec, C., and Bourgouin, C. (2008). Mosquito-based transmission blocking vaccines for interrupting Plasmodium development. Microbes Infect *10*, 845-849.
Lawniczak, M.K., Emrich, S.J., Holloway, A.K., Regier, A.P., Olson, M., White, B., Redmond, S., Fulton, L., Appelbaum, E., Godfrey, J., *et al.* (2010). Widespread divergence between incipient Anopheles gambiae species revealed by whole genome sequences. Science *330*, 512-514.
Lawson, D., Arensburger, P., Atkinson, P., Besansky, N.J., Bruggner, R.V., Butler, R., Campbell, K.S., Christophides, G.K., Christley, S., Dialynas, E., *et al.* (2009). VectorBase: a data resource for invertebrate vector genomics. Nucleic Acids Res *37*, D583-587.
Lazzaro, B.P., Sackton, T.B., and Clark, A.G. (2006). Genetic variation in Drosophila melanogaster resistance to infection: a comparison across bacteria. Genetics *174*, 1539-1554.
Lazzaro, B.P., Sceurman, B.K., and Clark, A.G. (2004). Genetic basis of natural variation in D. melanogaster antibacterial immunity. Science *303*, 1873-1876.
Le Prince, J.A. (1916). Mosquito control in Panama : the eradication of malaria and yellow fever in Cuba and Panama. (New York: G.P. Putnam's Sons).
Leahy, D.J., Aukhil, I., and Erickson, H.P. (1996). 2.0 A crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. Cell *84*, 155-164.
Lebon, G., Warne, T., Edwards, P.C., Bennett, K., Langmead, C.J., Leslie, A.G., and Tate, C.G. (2011). Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. Nature *474*, 521-525.
Lee, K.A., Kim, S.H., Kim, E.K., Ha, E.M., You, H., Kim, B., Kim, M.J., Kwon, Y., Ryu, J.H., and Lee, W.J. (2013a). Bacterial-derived uracil as a modulator of mucosal immunity and gut-microbe homeostasis in Drosophila. Cell *153*, 797-811.
Lee, W.J., and Brey, P.T. (2013). How Microbiomes Influence Metazoan Development: Insights from History and Drosophila Modeling of Gut-Microbe Interactions. Annual review of cell and developmental biology.
Lee, Y., Collier, T.C., Sanford, M.R., Marsden, C.D., Fofana, A., Cornel, A.J., and Lanzaro, G.C. (2013b). Chromosome Inversions, Genomic Differentiation and Speciation in the African Malaria Mosquito Anopheles gambiae. PLoS One *8*, e57887.
Lee, Y., Kang, M.J., Shim, J., Cheong, C.U., Moon, S.J., and Montell, C. (2012). Gustatory receptors required for avoiding the insecticide L-canavanine. J Neurosci *32*, 1429-1435.
Lee, Y., Kim, S.H., and Montell, C. (2010). Avoiding DEET through insect gustatory receptors. Neuron *67*, 555-561.
Lee, Y., Moon, S.J., and Montell, C. (2009). Multiple gustatory receptors required for the caffeine response in Drosophila. Proc Natl Acad Sci U S A *106*, 4495-4500.
Lee, Y.K., and Mazmanian, S.K. (2010). Has the microbiota played a critical role in the evolution of the adaptive immune system? Science *330*, 1768-1773.
Leevy, W.M., Serazin, N., and Smith, B.D. (2007). Optical Imaging of Bacterial Infection Models. Drug discovery today Disease models *4*, 91-97.
Lefevre, T., Oliver, L., Hunter, M.D., and De Roode, J.C. (2010). Evidence for trans-generational medication in nature. Ecology letters *13*, 1485-1493.
Lehane, A.M., McDevitt, C.A., Kirk, K., and Fidock, D.A. (2012). Degrees of chloroquine resistance in Plasmodium - is the redox system involved? International journal for parasitology, drugs and drug resistance *2*, 47-57.
Lemaitre, B. (2004). The road to Toll. Nature reviews Immunology *4*, 521-527.
Lemaitre, B., and Hoffmann, J. (2007). The host defense of Drosophila melanogaster. Annu Rev Immunol *25*, 697-743.
Lemaitre, B., Kromer-Metzger, E., Michaut, L., Nicolas, E., Meister, M., Georgel, P., Reichhart, J.M., and Hoffmann, J.A. (1995). A recessive mutation, immune deficiency (imd), defines two distinct control pathways in the Drosophila host defense. Proc Natl Acad Sci U S A *92*, 9465-9469.
Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.M., and Hoffmann, J.A. (1996). The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in Drosophila adults. Cell *86*, 973-983.

Lemaitre, B., Reichhart, J.M., and Hoffmann, J.A. (1997). Drosophila host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. Proc Natl Acad Sci U S A *94*, 14614-14619.
Leser, T.D., and Molbak, L. (2009). Better living through microbial action: the benefits of the mammalian gastrointestinal microbiota on the host. Environ Microbiol *11*, 2194-2206.
Lessie, T.G., Hendrickson, W., Manning, B.D., and Devereux, R. (1996). Genomic complexity and plasticity of Burkholderia cepacia. FEMS Microbiol Lett *144*, 117-128.
Levashina, E.A. (1999). Constitutive Activation of Toll-Mediated Antifungal Defense in Serpin-Deficient Drosophila. Science *285*, 1917-1919.
Levashina, E.A., Moita, L.F., Blandin, S., Vriend, G., Lagueux, M., and Kafatos, F.C. (2001). Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, Anopheles gambiae. Cell *104*, 709-718.
Ley, R.E., Peterson, D.A., and Gordon, J.I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell *124*, 837-848.
Lhocine, N., Ribeiro, P.S., Buchon, N., Wepf, A., Wilson, R., Tenev, T., Lemaitre, B., Gstaiger, M., Meier, P., and Leulier, F. (2008). PIMS modulates immune tolerance by negatively regulating Drosophila innate immune signaling. Cell Host Microbe *4*, 147-158.
Li, B., Calvo, E., Marinotti, O., James, A.A., and Paskewitz, S.M. (2005). Characterization of the c-type lysozyme gene family in Anopheles gambiae. Gene *360*, 131-139.
Li, B., and Paskewitz, S.M. (2006). A role for lysozyme in melanization of Sephadex beads in Anopheles gambiae. J Insect Physiol *52*, 936-942.
Liehl, P., Blight, M., Vodovar, N., Boccard, F., and Lemaitre, B. (2006). Prevalence of local immune response against oral infection in a Drosophila/Pseudomonas infection model. PLoS Pathog *2*, e56.
Liesch, J., Bellani, L.L., and Vosshall, L.B. (2013). Functional and Genetic Characterization of Neuropeptide Y-Like Receptors in. PLoS Negl Trop Dis *7*, e2486.
Ligoxygakis, P., Pelte, N., Hoffmann, J.A., and Reichhart, J.M. (2002). Activation of Drosophila Toll during fungal infection by a blood serine protease. Science *297*, 114-116.
Lim, J.H., Kim, M.S., Kim, H.E., Yano, T., Oshima, Y., Aggarwal, K., Goldman, W.E., Silverman, N., Kurata, S., and Oh, B.H. (2006). Structural basis for preferential recognition of diaminopimelic acid-type peptidoglycan by a subset of peptidoglycan recognition proteins. J Biol Chem *281*, 8286-8295.
Limmer, S., Haller, S., Drenkard, E., Lee, J., Yu, S., Kocks, C., Ausubel, F.M., and Ferrandon, D. (2011). Pseudomonas aeruginosa RhlR is required to neutralize the cellular immune response in a Drosophila melanogaster oral infection model. Proc Natl Acad Sci U S A *108*, 17378-17383.
Lin, H., Zhang, L., Luna, C., Ho, N.T., and Zheng, L. (2007). A splice variant of PGRP-LC required for expression of antimicrobial peptides in Anopheles gambiae. Insect Science *14*, 185-192.
Lindh, J.M., and Lehane, M.J. (2011). The tsetse fly Glossina fuscipes fuscipes (Diptera: Glossina) harbours a surprising diversity of bacteria other than symbionts. Antonie van Leeuwenhoek *99*, 711-720.
Lindh, J.M., Terenius, O., Eriksson-Gonzales, K., Knols, B.G., and Faye, I. (2006). Re-introducing bacteria in mosquitoes--a method for determination of mosquito feeding preferences based on coloured sugar solutions. Acta Trop *99*, 173-183.
Lindh, J.M., Terenius, O., and Faye, I. (2005). 16S rRNA gene-based identification of midgut bacteria from field-caught Anopheles gambiae sensu lato and A. funestus mosquitoes reveals new species related to known insect symbionts. Appl Environ Microbiol *71*, 7217-7223.
Littman, D.R., and Pamer, E.G. (2011). Role of the commensal microbiota in normal and pathogenic host immune responses. Cell Host Microbe *10*, 311-323.
Liu, G., Li, W., Wang, L., Kar, A., Guan, K.L., Rao, Y., and Wu, J.Y. (2009a). DSCAM functions as a netrin receptor in commissural axon pathfinding. Proceedings of the National Academy of Sciences *106*, 2951-2956.
Liu, J., Ward, A., Gao, J., Dong, Y., Nishio, N., Inada, H., Kang, L., Yu, Y., Ma, D., Xu, T.*, et al.* (2010a). C. elegans phototransduction requires a G protein-dependent cGMP pathway and a taste receptor homolog. Nat Neurosci *13*, 715-722.
Liu, L., Ishihara, K., Ichimura, T., Fujita, N., Hino, S., Tomita, S., Watanabe, S., Saitoh, N., Ito, T., and Nakao, M. (2009b). MCAF1/AM is involved in Sp1-mediated maintenance of cancer-associated telomerase activity. J Biol Chem *284*, 5165-5174.
Liu, W., Li, Y., Learn, G.H., Rudicell, R.S., Robertson, J.D., Keele, B.F., Ndjango, J.B., Sanz, C.M., Morgan, D.B., Locatelli, S.*, et al.* (2010b). Origin of the human malaria parasite Plasmodium falciparum in gorillas. Nature *467*, 420-425.
Livadas, G.A., and Georgopoulos, G. (1953). Development of resistance to DDT by Anopheles sacharovi in Greece. Bull World Health Organ *8*, 497-511.
Loaiza, J.R., Bermingham, E., Sanjur, O.I., Scott, M.E., Bickersmith, S.A., and Conn, J.E. (2012). Review of genetic diversity in malaria vectors (Culicidae: Anophelinae). Infect Genet Evol *12*, 1-12.
Logan, J.G., Birkett, M.A., Clark, S.J., Powers, S., Seal, N.J., Wadhams, L.J., Mordue Luntz, A.J., and Pickett, J.A. (2008). Identification of human-derived volatile chemicals that interfere with attraction of Aedes aegypti mosquitoes. Journal of chemical ecology *34*, 308-322.
Lu, Y., Wu, L.P., and Anderson, K.V. (2001). The antibacterial arm of the drosophila innate immune response requires an IkappaB kinase. Genes Dev *15*, 104-110.
Lubec, G., Afjehi-Sadat, L., Yang, J.W., and John, J.P. (2005). Searching for hypothetical proteins: theory and practice based upon original data and literature. Progress in neurobiology *77*, 90-127.
Lucas, K., and Raikhel, A.S. (2013). Insect microRNAs: biogenesis, expression profiling and biological functions. Insect Biochem Mol Biol *43*, 24-38.
Lucas, K.J., Myles, K.M., and Raikhel, A.S. (2013). Small RNAs: a new frontier in mosquito biology. Trends Parasitol *29*, 295-303.
Luckhart, S., and Riehle, M.A. (2007). The insulin signaling cascade from nematodes to mammals: insights into innate immunity of Anopheles mosquitoes to malaria parasite infection. Dev Comp Immunol *31*, 647-656.
Luckhart, S., Vodovotz, Y., Cui, L., and Rosenberg, R. (1998). The mosquito Anopheles stephensi limits malaria parasite development with inducible synthesis of nitric oxide. Proc Natl Acad Sci U S A *95*, 5700-5705.
Luz, C., Mnyone, L.L., and Russell, T.L. (2011). Survival of anopheline eggs and their susceptibility to infection with Metarhizium anisopliae and Beauveria bassiana under laboratory conditions. Parasitol Res.
Ly, A., Nikolaev, A., Suresh, G., Zheng, Y., Tessier-Lavigne, M., and Stein, E. (2008). DSCAM Is a Netrin Receptor that Collaborates with DCC in Mediating Turning Responses to Netrin-1. Cell *133*, 1241-1254.

Lycett, G., Blass, C., and Louis, C. (2001). Developmental variation in epidermal growth factor receptor size and localization in the malaria mosquito, Anopheles gambiae. Insect Mol Biol *10*, 619-628.
MacCallum, R.M., Redmond, S.N., and Christophides, G.K. (2011). An Expression Map for Anopheles gambiae. BMC Genomics *12*, 620.
MacDonald, S.J., Thomas, G.H., and Douglas, A.E. (2011). Genetic and metabolic determinants of nutritional phenotype in an insect-bacterial symbiosis. Mol Ecol *20*, 2073-2084.
Mackay, T.F., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S., Han, Y., Magwire, M.M., Cridland, J.M.*, et al.* (2012). The Drosophila melanogaster Genetic Reference Panel. Nature *482*, 173-178.
Mackenzie, J.S., Gubler, D.J., and Petersen, L.R. (2004). Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. Nat Med *10*, S98-109.
Madden, T. (2002). The BLAST Sequence Analysis Tool (Bethesda, MD, USA: National Center for Biotechnology Information (US)).
Maeshima, N., and Fernandez, R.C. (2013). Recognition of lipid A variants by the TLR4-MD-2 receptor complex. Frontiers in cellular and infection microbiology *3*, 3.
Magnusson, K., Mendes, A.M., Windbichler, N., Papathanos, P.A., Nolan, T., Dottorini, T., Rizzi, E., Christophides, G.K., and Crisanti, A. (2011). Transcription Regulation of Sex-Biased Genes during Ontogeny in the Malaria Vector Anopheles gambiae. PLoS One *6*, e21572.
Maguire, J.D., and Baird, J.K. (2010). The 'non-falciparum' malarias: the roles of epidemiology, parasite biology, clinical syndromes, complications and diagnostic rigour in guiding therapeutic strategies. Ann Trop Med Parasitol *104*, 283-301.
Magwire, M.M., Fabian, D.K., Schweyen, H., Cao, C., Longdon, B., Bayer, F., and Jiggins, F.M. (2012). Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in Drosophila melanogaster. PLoS Genet *8*, e1003057.
Maillet, F., Bischoff, V., Vignal, C., Hoffmann, J., and Royet, J. (2008). The Drosophila peptidoglycan recognition protein PGRP-LF blocks PGRP-LC and IMD/JNK pathway activation. Cell Host Microbe *3*, 293-303.
Main, A.L., Harvey, T.S., Baron, M., Boyd, J., and Campbell, I.D. (1992). The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. Cell *71*, 671-678.
Mallo, G.V., Kurz, C.L., Couillault, C., Pujol, N., Granjeaud, S., Kohara, Y., and Ewbank, J.J. (2002). Inducible antibacterial defense system in C. elegans. Curr Biol *12*, 1209-1214.
Manguin, S. (2008). Biodiversity of Malaria in the World (John Libbey Eurotext).
Manson, P. (2002). Experimental proof of the mosquito-malaria theory. 1900. The Yale journal of biology and medicine *75*, 107-112.
Marinotti, O., Calvo, E., Nguyen, Q.K., Dissanayake, S., Ribeiro, J.M., and James, A.A. (2006). Genome-wide analysis of gene expression in adult Anopheles gambiae. Insect Mol Biol *15*, 1-12.
Marinotti, O., Nguyen, Q.K., Calvo, E., James, A.A., and Ribeiro, J.M. (2005). Microarray analysis of genes showing variable expression following a blood meal in Anopheles gambiae. Insect Mol Biol *14*, 365-373.
Mariuzza, R.A., Velikovsky, C.A., Deng, L., Xu, G., and Pancer, Z. (2010). Structural insights into the evolution of the adaptive immune system: the variable lymphocyte receptors of jawless vertebrates. Biological chemistry *391*, 753-760.
Marrelli, M.T., Li, C., Rasgon, J.L., and Jacobs-Lorena, M. (2007). Transgenic malaria-resistant mosquitoes have a fitness advantage when feeding on Plasmodium-infected blood. Proc Natl Acad Sci U S A *104*, 5580-5583.
Marsden, C.D., Lee, Y., Nieman, C.C., Sanford, M.R., Dinis, J., Martins, C., Rodrigues, A., Cornel, A.J., and Lanzaro, G.C. (2011). Asymmetric introgression between the M and S forms of the malaria vector, Anopheles gambiae, maintains divergence despite extensive hybridization. Mol Ecol *20*, 4983-4994.
Martins, L.C., Rocha, N.P., Torres, K.C., Dos Santos, R.R., Franca, G.S., de Moraes, E.N., Mukhamedyarov, M.A., Zefirov, A.L., Rizvanov, A.A., Kiyasov, A.P.*, et al.* (2012). Disease-specific expression of the serotonin-receptor 5-HT(2C) in natural killer cells in Alzheimer's dementia. J Neuroimmunol *251*, 73-79.
Maslowski, K.M., Vieira, A.T., Ng, A., Kranich, J., Sierro, F., Yu, D., Schilter, H.C., Rolph, M.S., Mackay, F., Artis, D.*, et al.* (2009). Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. Nature *461*, 1282-1286.
Mathis, D., and Benoist, C. (2011). Microbiota and autoimmune disease: the hosted self. Cell Host Microbe *10*, 297-301.
Matsumoto, S., Fonagy, A., Kurihara, M., Uchiumi, K., Nagamine, T., Chijimatsu, M., and Mitsui, T. (1992). Isolation and primary structure of a novel pheromonotropic neuropeptide structurally related to leucopyrokinin from the armyworm larvae, Pseudaletia separata. Biochem Biophys Res Commun *182*, 534-539.
Matsushita, M. (2010). Ficolins: complement-activating lectins involved in innate immunity. Journal of innate immunity *2*, 24-32.
Matsushita, M. (2013). Ficolins in complement activation. Mol Immunol *55*, 22-26.
Matsushita, M., and Fujita, T. (2002). The role of ficolins in innate immunity. Immunobiology *205*, 490-497.
Matthews, B.J., and Grueber, W.B. (2011). Dscam1-mediated self-avoidance counters netrin-dependent targeting of dendrites in Drosophila. Curr Biol *21*, 1480-1487.
Mazier, D., Nitcheu, J., and Idrissa-Boubou, M. (2000). Cerebral malaria and immunogenetics. Parasite Immunol *22*, 613-623.
McCoy, D., Kembhavi, G., Patel, J., and Luintel, A. (2009). The Bill & Melinda Gates Foundation's grant-making programme for global health. Lancet *373*, 1645-1653.
McMeniman, C.J., Lane, R.V., Cass, B.N., Fong, A.W., Sidhu, M., Wang, Y.F., and O'Neill, S.L. (2009). Stable introduction of a life-shortening Wolbachia infection into the mosquito Aedes aegypti. Science *323*, 141-144.
Meijers, R., Puettmann-Holgado, R., Skiniotis, G., Liu, J.H., Walz, T., Wang, J.H., and Schmucker, D. (2007). Structural basis of Dscam isoform specificity. Nature *449*, 487-491.
Meister, S., Agianian, B., Turlure, F., Relogio, A., Morlais, I., Kafatos, F.C., and Christophides, G.K. (2009). Anopheles gambiae PGRPLC-mediated defense against bacteria modulates infections with malaria parasites. PLoS Pathog *5*, e1000542.
Meister, S., Kanzok, S.M., Zheng, X.L., Luna, C., Li, T.R., Hoa, N.T., Clayton, J.R., White, K.P., Kafatos, F.C., Christophides, G.K.*, et al.* (2005). Immune signaling pathways regulating bacterial and malaria parasite infection of the mosquito Anopheles gambiae. Proc Natl Acad Sci U S A *102*, 11420-11425.
Melcher, C., and Pankratz, M.J. (2005). Candidate gustatory interneurons modulating feeding behavior in the Drosophila brain. PLoS Biol *3*, e305.

Mendes, A.M., Awono-Ambene, P.H., Nsango, S.E., Cohuet, A., Fontenille, D., Kafatos, F.C., Christophides, G.K., Morlais, I., and Vlachou, D. (2011). Infection intensity dependent responses of Anopheles gambiae to African malaria parasites Plasmodium falciparum. Infect Immun.
Mendes, A.M., Schlegelmilch, T., Cohuet, A., Awono-Ambene, P., De Iorio, M., Fontenille, D., Morlais, I., Christophides, G.K., Kafatos, F.C., and Vlachou, D. (2008). Conserved mosquito/parasite interactions affect development of Plasmodium falciparum in Africa. PLoS Pathog *4*, e1000069.
Mendiratta, S.S., Sekulic, N., Hernandez-Guzman, F.G., Close, B.E., Lavie, A., and Colley, K.J. (2006). A novel alpha-helix in the first fibronectin type III repeat of the neural cell adhesion molecule is critical for N-glycan polysialylation. J Biol Chem *281*, 36052-36059.
Menge, D.M., Zhong, D., Guda, T., Gouagna, L., Githure, J., Beier, J., and Yan, G. (2006). Quantitative trait loci controlling refractoriness to Plasmodium falciparum in natural Anopheles gambiae mosquitoes from a malaria-endemic region in western Kenya. Genetics *173*, 235-241.
Meunier, N., Belgacem, Y.H., and Martin, J.R. (2007). Regulation of feeding behaviour and locomotor activity by takeout in Drosophila. J Exp Biol *210*, 1424-1434.
Meyer, J.M., Ejendal, K.F., Avramova, L.V., Garland-Kuntz, E.E., Giraldo-Calderon, G.I., Brust, T.F., Watts, V.J., and Hill, C.A. (2012). A "Genome-to-Lead" Approach for Insecticide Discovery: Pharmacological Characterization and Screening of Aedes aegypti D(1)-like Dopamine Receptors. PLoS Negl Trop Dis *6*, e1478.
Miguel-Aliaga, I. (2012). Nerveless and gutsy: intestinal nutrient sensing from invertebrates to humans. Seminars in cell & developmental biology *23*, 614-620.
Milan, N.F., Kacsoh, B.Z., and Schlenke, T.A. (2012). Alcohol Consumption as Self-Medication against Blood-Borne Parasites in the Fruit Fly. Curr Biol *22*, 488-493.
Millard, S.S., Flanagan, J.J., Pappu, K.S., Wu, W., and Zipursky, S.L. (2007). Dscam2 mediates axonal tiling in the Drosophila visual system. Nature *447*, 720-724.
Miller, L.H., Ackerman, H.C., Su, X.Z., and Wellems, T.E. (2013). Malaria biology and disease pathogenesis: insights for new treatments. Nat Med *19*, 156-167.
Miller, L.H., Baruch, D.I., Marsh, K., and Doumbo, O.K. (2002). The pathogenic basis of malaria. Nature *415*, 673-679.
Miller, L.H., and Su, X. (2011). Artemisinin: discovery from the Chinese herbal garden. Cell *146*, 855-858.
Miotto, O., Almagro-Garcia, J., Manske, M., Macinnis, B., Campino, S., Rockett, K.A., Amaratunga, C., Lim, P., Suon, S., Sreng, S.*, et al.* (2013). Multiple populations of artemisinin-resistant Plasmodium falciparum in Cambodia. Nat Genet *45*, 648-655.
Mita, T., and Tanabe, K. (2012). Evolution of Plasmodium falciparum drug resistance: implications for the development and containment of artemisinin resistance. Japanese journal of infectious diseases *65*, 465-475.
Mitraka, E., Stathopoulos, S., Siden-Kiamos, I., Christophides, G.K., and Louis, C. (2013). Asaia accelerates larval development of Anopheles gambiae. Pathogens and global health.
Mitri, C., Jacques, J.C., Thiery, I., Riehle, M.M., Xu, J., Bischoff, E., Morlais, I., Nsango, S.E., Vernick, K.D., and Bourgouin, C. (2009). Fine pathogen discrimination within the APL1 gene family protects Anopheles gambiae against human and rodent malaria species. PLoS Pathog *5*, e1000576.
Mitri, C., and Vernick, K.D. (2012). Anopheles gambiae pathogen susceptibility: the intersection of genetics, immunity and ecology. Curr Opin Microbiol *15*, 285-291.
Miyamoto, T., and Amrein, H. (2008). Suppression of male courtship by a Drosophila pheromone receptor. Nat Neurosci *11*, 874-876.
Miyamoto, T., Slone, J., Song, X., and Amrein, H. (2012). A Fructose Receptor Functions as a Nutrient Sensor in the Drosophila Brain. Cell *151*, 1113-1125.
Miyamoto, T., Wright, G., and Amrein, H. (2013). Nutrient sensors. Curr Biol *23*, R369-373.
Mlodzik, M., and Gehring, W.J. (1987). Expression of the caudal gene in the germ line of Drosophila: formation of an RNA and protein gradient during early embryogenesis. Cell *48*, 465-478.
Molina-Cruz, A., Garver, L.S., Alabaster, A., Bangiolo, L., Haile, A., Winikor, J., Ortega, C., van Schaijk, B.C., Sauerwein, R.W., Taylor-Salmon, E.*, et al.* (2013). The human malaria parasite Pfs47 gene mediates evasion of the mosquito immune system. Science *340*, 984-987.
Montell, C. (2013). Gustatory Receptors: Not Just for Good Taste. Curr Biol *23*, R929-R932.
Moon, S.J., Kottgen, M., Jiao, Y., Xu, H., and Montell, C. (2006). A taste receptor required for the caffeine response in vivo. Curr Biol *16*, 1812-1817.
Moon, S.J., Lee, Y., Jiao, Y., and Montell, C. (2009). A Drosophila gustatory receptor essential for aversive taste and inhibiting male-to-male courtship. Curr Biol *19*, 1623-1627.
Moreira, L.A., Ghosh, A.K., Abraham, E.G., and Jacobs-Lorena, M. (2002). Genetic transformation of mosquitoes: a quest for malaria control. Int J Parasitol *32*, 1599-1605.
Moreira, L.A., Iturbe-Ormaetxe, I., Jeffery, J.A., Lu, G., Pyke, A.T., Hedges, L.M., Rocha, B.C., Hall-Mendelin, S., Day, A., Riegler, M.*, et al.* (2009). A Wolbachia symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium. Cell *139*, 1268-1278.
Morlais, I., Poncon, N., Simard, F., Cohuet, A., and Fontenille, D. (2004). Intraspecific nucleotide variation in Anopheles gambiae: new insights into the biology of malaria vectors. Am J Trop Med Hyg *71*, 795-802.
Mu, J., Duan, J., Makova, K.D., Joy, D.A., Huynh, C.Q., Branch, O.H., Li, W.H., and Su, X.Z. (2002). Chromosome-wide SNPs reveal an ancient origin for Plasmodium falciparum. Nature *418*, 323-326.
Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., Gonzalez, A., Fontana, L., Henrissat, B., Knight, R., and Gordon, J.I. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science *332*, 970-974.
Mueller, K., Ash, C., Pennisi, E., and Smith, O. (2012). The gut microbiota. Introduction. Science *336*, 1245.
Mukabana, W.R., Mweresa, C.K., Otieno, B., Omusula, P., Smallegange, R.C., van Loon, J.J., and Takken, W. (2012). A novel synthetic odorant blend for trapping of malaria and other african mosquito species. Journal of chemical ecology *38*, 235-244.
Mukabana, W.R., Takken, W., Coe, R., and Knols, B.G. (2002). Host-specific cues cause differential attractiveness of Kenyan men to the African malaria vector Anopheles gambiae. Malar J *1*, 17.
Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P.*, et al.* (2003). The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res *31*, 315-318.
Mulder, N.J., Kersey, P., Pruess, M., and Apweiler, R. (2008). In silico characterization of proteins: UniProt, InterPro and Integr8. Molecular biotechnology *38*, 165-177.

Munde, E.O., Okeyo, W.A., Anyona, S.B., Raballah, E., Konah, S., Okumu, W., Ogonda, L., Vulule, J., and Ouma, C. (2012). Polymorphisms in Fc gamma receptor (FcgammaRIIIA-176 F/V) and Toll-like receptor (TLR9[-1237 T/C]) are associated with protection against severe malarial anemia and changes in circulating IFN-gamma. Infect Immun.
Naccarati, C., Audsley, N., Keen, J.N., Kim, J.H., Howell, G.J., Kim, Y.J., and Isaac, R.E. (2012). The host-seeking inhibitory peptide, Aea-HP-1, is made in the male accessory gland and transferred to the female during copulation. Peptides *34*, 150-157.
Najera, J.A., Gonzalez-Silva, M., and Alonso, P.L. (2011). Some lessons for the future from the Global Malaria Eradication Programme (1955-1969). PLoS Med *8*, e1000412.
Nakazawa, S., Culleton, R., and Maeno, Y. (2011). In vivo and in vitro gametocyte production of Plasmodium falciparum isolates from Northern Thailand. Int J Parasitol *41*, 317-323.
Ndiath, M.O., Brengues, C., Konate, L., Sokhna, C., Boudin, C., Trape, J.F., and Fontenille, D. (2008). Dynamics of transmission of Plasmodium falciparum by Anopheles arabiensis and the molecular forms M and S of Anopheles gambiae in Dielmo, Senegal. Malar J *7*, 136.
Ndiath, M.O., Cohuet, A., Gaye, A., Konate, L., Mazenot, C., Faye, O., Boudin, C., Sokhna, C., and Trape, J.F. (2011). Comparative susceptibility to Plasmodium falciparum of the molecular forms M and S of Anopheles gambiae and Anopheles arabiensis. Malar J *10*, 269.
Ndiath, M.O., Sougoufara, S., Gaye, A., Mazenot, C., Konate, L., Faye, O., Sokhna, C., and Trape, J.F. (2012). Resistance to DDT and pyrethroids and increased kdr mutation frequency in An. gambiae after the implementation of permethrin-treated nets in Senegal. PLoS One *7*, e31943.
Neafsey, D.E., Christophides, G.K., Collins, F.H., Emrich, S.J., Fontaine, M.C., Gelbart, W., Hahn, M.W., Howell, P.I., Kafatos, F.C., Lawson, D.*, et al.* (2013). The Evolution of the Anopheles 16 Genomes Project. G3.
Neafsey, D.E., Lawniczak, M.K., Park, D.J., Redmond, S.N., Coulibaly, M.B., Traore, S.F., Sagnon, N., Costantini, C., Johnson, C., Wiegand, R.C.*, et al.* (2010). SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. Science *330*, 514-517.
Neghina, R., Neghina, A.M., Marincu, I., and Iacobiciu, I. (2010). Malaria, a journey in time: in search of the lost myths and forgotten stories. The American journal of the medical sciences *340*, 492-498.
Nehme, N.T., Liegeois, S., Kele, B., Giammarinaro, P., Pradel, E., Hoffmann, J.A., Ewbank, J.J., and Ferrandon, D. (2007). A model of bacterial intestinal infections in Drosophila melanogaster. PLoS Pathog *3*, e173.
Nelson, K.E., Weinstock, G.M., Highlander, S.K., Worley, K.C., Creasy, H.H., Wortman, J.R., Rusch, D.B., Mitreva, M., Sodergren, E., Chinwalla, A.T.*, et al.* (2010). A catalog of reference genomes from the human microbiome. Science *328*, 994-999.
Nene, V., Wortman, J.R., Lawson, D., Haas, B., Kodira, C., Tu, Z.J., Loftus, B., Xi, Z., Megy, K., Grabherr, M.*, et al.* (2007). Genome sequence of Aedes aegypti, a major arbovirus vector. Science *316*, 1718-1723.
Nerlich, A.G., Schraut, B., Dittrich, S., Jelinek, T., and Zink, A.R. (2008). Plasmodium falciparum in ancient Egypt. Emerg Infect Dis *14*, 1317-1319.
Ni, L., Bronk, P., Chang, E.C., Lowell, A.M., Flam, J.O., Panzano, V.C., Theobald, D.L., Griffith, L.C., and Garrity, P.A. (2013). A gustatory receptor paralogue controls rapid warmth avoidance in Drosophila. Nature.
Niare, O., Markianos, K., Volz, J., Oduol, F., Toure, A., Bagayoko, M., Sangare, D., Traore, S.F., Wang, R., Blass, C.*, et al.* (2002). Genetic loci affecting resistance to human malaria parasites in a West African mosquito vector population. Science *298*, 213-216.
Nichols, D.E., and Nichols, C.D. (2008). Serotonin receptors. Chemical reviews *108*, 1614-1641.
Nirmala, X., and James, A.A. (2003). Engineering Plasmodium-refractory phenotypes in mosquitoes. Trends Parasitol *19*, 384-387.
Noriega, F.G. (2004). Nutritional regulation of JH synthesis: a mechanism to control reproductive maturation in mosquitoes? Insect Biochem Mol Biol *34*, 687-693.
Novais, R.C., and Thorstenson, Y.R. (2011). The evolution of Pyrosequencing(R) for microbiology: From genes to genomes. J Microbiol Methods *86*, 1-7.
Nowling, R.J., Abrudan, J.L., Shoue, D.A., Abdul-Wahid, B., Wadsworth, M., Stayback, G., Collins, F.H., McDowell, M.A., and Izaguirre, J.A. (2013). Identification of novel arthropod vector G protein-coupled receptors. Parasit Vectors *6*, 150.
Nunes, R.D., de Oliveira, R.L., and Braz, G.R. (2008). A novel method for measuring fructose ingestion by mosquitoes. Journal of vector ecology : journal of the Society for Vector Ecology *33*, 225-231.
Nwakanma, D.C., Neafsey, D.E., Jawara, M., Adiamoh, M., Lund, E., Rodrigues, A., Loua, K.M., Konate, L., Sy, N., Dia, I.*, et al.* (2013). Breakdown in the Process of Incipient Speciation in Anopheles gambiae. Genetics.
O'Brochta, D.A., Sethuraman, N., Wilson, R., Hice, R.H., Pinkerton, A.C., Levesque, C.S., Bideshi, D.K., Jasinskiene, N., Coates, C.J., James, A.A.*, et al.* (2003). Gene vector and transposable element behavior in mosquitoes. J Exp Biol *206*, 3823-3834.
O'Connell, P.J., Wang, X., Leon-Ponte, M., Griffiths, C., Pingle, S.C., and Ahern, G.P. (2006). A novel form of immune signaling revealed by transmission of the inflammatory mediator serotonin between dendritic cells and T cells. Blood *107*, 1010-1017.
O'Hara, A.M., and Shanahan, F. (2006). The gut flora as a forgotten organ. EMBO Rep *7*, 688-693.
Okombo, J., Ohuma, E., Picot, S., and Nzila, A. (2011). Update on genetic markers of quinine resistance in Plasmodium falciparum. Mol Biochem Parasitol *177*, 77-82.
Oliveira, E., Salgueiro, P., Palsson, K., Vicente, J.L., Arez, A.P., Jaenson, T.G., Caccone, A., and Pinto, J. (2008). High levels of hybridization between molecular forms of Anopheles gambiae from Guinea Bissau. J Med Entomol *45*, 1057-1063.
Oliveira, G.D., Lieberman, J., and Barillas-Mury, C. (2012). Epithelial Nitration by a Peroxidase/NOX5 System Mediates Mosquito Antiplasmodial Immunity. Science.
Ong'echa, J.M., Raballah, E.O., Kempaiah, P.M., Anyona, S.B., Were, T., Davenport, G.C., Konah, S., Vulule, J.M., Ouma, C., Hittner, J.B.*, et al.* (2011). Polymorphic variability in the 3' untranslated region (UTR) of IL12B is associated with susceptibility to severe anaemia in Kenyan children with acute Plasmodium falciparum malaria. BMC Genet *12*, 69.
Osei-Poku, J., Mbogo, C.M., Palmer, W.J., and Jiggins, F.M. (2012). Deep sequencing reveals extensive variation in the gut microbiota of wild mosquitoes from Kenya. Mol Ecol *21*, 5138-5150.
Osta, M.A., Christophides, G.K., and Kafatos, F.C. (2004a). Effects of mosquito genes on Plasmodium development. Science *303*, 2030-2032.
Osta, M.A., Christophides, G.K., Vlachou, D., and Kafatos, F.C. (2004b). Innate immunity in the malaria vector Anopheles gambiae: comparative and functional genomics. J Exp Biol *207*, 2551-2563.
Pace, K.E., and Baum, L.G. (2004). Insect galectins: roles in immunity and development. Glycoconj J *19*, 607-614.
Pace, K.E., Lebestky, T., Hummel, T., Arnoux, P., Kwan, K., and Baum, L.G. (2002). Characterization of a novel Drosophila melanogaster galectin. Expression in developing immune, neural, and muscle tissues. J Biol Chem *277*, 13091-13098.

Pacheco, R., Riquelme, E., and Kalergis, A.M. (2010). Emerging evidence for the role of neurotransmitters in the modulation of T cell responses to cognate ligands. Central nervous system agents in medicinal chemistry *10*, 65-83.
Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E.*, et al.* (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. Science *289*, 739-745.
Pan, X., Zhou, G., Wu, J., Bian, G., Lu, P., Raikhel, A.S., and Xi, Z. (2011). Wolbachia induces reactive oxygen species (ROS)-dependent activation of the Toll pathway to control dengue virus in the mosquito Aedes aegypti. Proc Natl Acad Sci U S A.
Paradis, S., Boissinot, M., Paquette, N., Belanger, S.D., Martel, E.A., Boudreau, D.K., Picard, F.J., Ouellette, M., Roy, P.H., and Bergeron, M.G. (2005). Phylogeny of the Enterobacteriaceae based on genes encoding elongation factor Tu and F-ATPase beta-subunit. Int J Syst Evol Microbiol *55*, 2013-2025.
Park, J.H., and Kwon, J.Y. (2011). Heterogeneous expression of Drosophila gustatory receptors in enteroendocrine cells. PLoS One *6*, e29022.
Paskewitz, S.M., Andreev, O., and Shi, L. (2006). Gene silencing of serine proteases affects melanization of Sephadex beads in Anopheles gambiae. Insect Biochem Mol Biol *36*, 701-711.
Pasvol, G. (2005). The treatment of complicated and severe malaria. British medical bulletin *75-76*, 29-47.
Pates, H., and Curtis, C. (2005). Mosquito behavior and vector control. Annu Rev Entomol *50*, 53-70.
Payne, D. (1987). Spread of chloroquine resistance in Plasmodium falciparum. Parasitol Today *3*, 241-246.
Pellegrino, M., Steinbach, N., Stensmyr, M.C., Hansson, B.S., and Vosshall, L.B. (2011). A natural polymorphism alters odour and DEET sensitivity in an insect odorant receptor. Nature.
Pelte, N., Robertson, A.S., Zou, Z., Belorgey, D., Dafforn, T.R., Jiang, H., Lomas, D., Reichhart, J.M., and Gubb, D. (2006). Immune challenge induces N-terminal cleavage of the Drosophila serpin Necrotic. Insect Biochem Mol Biol *36*, 37-46.
Pennisi, E. (2010). Body's hardworking microbes get some overdue respect. Science *330*, 1619.
Perez-Hedo, M., Rivera-Perez, C., and Noriega, F.G. (2013). The insulin/TOR signal transduction pathway is involved in the nutritional regulation of juvenile hormone synthesis in Aedes aegypti. Insect Biochem Mol Biol.
Perkins, D.J., Were, T., Davenport, G.C., Kempaiah, P., Hittner, J.B., and Ong'echa, J.M. (2011). Severe malarial anemia: innate immunity and pathogenesis. Int J Biol Sci *7*, 1427-1442.
Phyo, A.P., Nkhoma, S., Stepniewska, K., Ashley, E.A., Nair, S., McGready, R., ler Moo, C., Al-Saai, S., Dondorp, A.M., Lwin, K.M.*, et al.* (2012). Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. Lancet *379*, 1960-1966.
Piccin, A., Salameh, A., Benna, C., Sandrelli, F., Mazzotta, G., Zordan, M., Rosato, E., Kyriacou, C.P., and Costa, R. (2001). Efficient and heritable functional knock-out of an adult phenotype in Drosophila using a GAL4-driven hairpin RNA incorporating a heterologous spacer. Nucleic Acids Res *29*, E55-55.
Pitts, R.J., Rinker, D.C., Jones, P.L., Rokas, A., and Zwiebel, L.J. (2011). Transcriptome profiling of chemosensory appendages in the malaria vector Anopheles gambiae reveals tissue- and sex-specific signatures of odor coding. BMC Genomics *12*, 271.
Plottel, C.S., and Blaser, M.J. (2011). Microbiome and malignancy. Cell Host Microbe *10*, 324-335.
Povelones, M., Bhagavatula, L., Yassine, H., Tan, L.A., Upton, L.M., Osta, M.A., and Christophides, G.K. (2013). The CLIP-Domain Serine Protease Homolog SPCLIP1 Regulates Complement Recruitment to Microbial Surfaces in the Malaria Mosquito Anopheles gambiae. PLoS Pathog *9*, e1003623.
Povelones, M., Upton, L.M., Sala, K.A., and Christophides, G.K. (2011). Structure-Function Analysis of the Anopheles gambiae LRIM1/APL1C Complex and its Interaction with Complement C3-Like Protein TEP1. PLoS Pathog *7*, e1002023.
Povelones, M., Waterhouse, R.M., Kafatos, F.C., and Christophides, G.K. (2009). Leucine-rich repeat protein complex activates mosquito complement in defense against Plasmodium parasites. Science *324*, 258-261.
Pradel, E., Zhang, Y., Pujol, N., Matsuyama, T., Bargmann, C.I., and Ewbank, J.J. (2007). Detection and avoidance of a natural product from the pathogenic bacterium Serratia marcescens by Caenorhabditis elegans. Proc Natl Acad Sci U S A *104*, 2295-2300.
Proctor, L.M. (2011). The human microbiome project in 2011 and beyond. Cell Host Microbe *10*, 287-291.
Pujol, N., Link, E.M., Liu, L.X., Kurz, C.L., Alloing, G., Tan, M.W., Ray, K.P., Solari, R., Johnson, C.D., and Ewbank, J.J. (2001). A reverse genetic analysis of components of the Toll signaling pathway in Caenorhabditis elegans. Curr Biol *11*, 809-821.
Pumpuni, C.B., Beier, M.S., Nataro, J.P., Guers, L.D., and Davis, J.R. (1993). Plasmodium falciparum: inhibition of sporogonic development in Anopheles stephensi by gram-negative bacteria. Exp Parasitol *77*, 195-199.
Pumpuni, C.B., Demaio, J., Kent, M., Davis, J.R., and Beier, J.C. (1996). Bacterial population dynamics in three anopheline species: the impact on Plasmodium sporogonic development. Am J Trop Med Hyg *54*, 214-218.
Qu, Z., and Chaikof, E.L. (2010). Interface between hemostasis and adaptive immunity. Curr Opin Immunol *22*, 634-642.
Racloz, V., Ramsey, R., Tong, S., and Hu, W. (2012). Surveillance of dengue fever virus: a review of epidemiological models and early warning systems. PLoS Negl Trop Dis *6*, e1648.
Raju, T.N. (1999). The Nobel chronicles. 1948: Paul Hermann Muller (1899-1965). Lancet *353*, 1196.
Ramasamy, R. (1998). Molecular basis for evasion of host immunity and pathogenesis in malaria. Biochim Biophys Acta *1406*, 10-27.
Ramirez, J.L., and Dimopoulos, G. (2010). The Toll immune signaling pathway control conserved anti-dengue defenses across diverse Ae. aegypti strains and against multiple dengue virus serotypes. Dev Comp Immunol *34*, 625-629.
Ramirez, J.L., Souza-Neto, J., Torres Cosme, R., Rovira, J., Ortiz, A., Pascale, J.M., and Dimopoulos, G. (2012). Reciprocal Tripartite Interactions between the Aedes aegypti Midgut Microbiota, Innate Immune System and Dengue Virus Influences Vector Competence. PLoS Negl Trop Dis *6*, e1561.
Rani, A., Sharma, A., Rajagopal, R., Adak, T., and Bhatnagar, R.K. (2009). Bacterial diversity analysis of larvae and adult midgut microflora using culture-dependent and culture-independent methods in lab-reared and field-collected Anopheles stephensi-an Asian malarial vector. BMC Microbiol *9*, 96.
Ranson, H., N'Guessan, R., Lines, J., Moiroux, N., Nkuni, Z., and Corbel, V. (2011). Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? Trends Parasitol *27*, 91-98.
Rawlings, J.S., Rennebeck, G., Harrison, S.M., Xi, R., and Harrison, D.A. (2004). Two Drosophila suppressors of cytokine signaling (SOCS) differentially regulate JAK and EGFR pathway activities. BMC cell biology *5*, 38.

Reddy, K.C., Andersen, E.C., Kruglyak, L., and Kim, D.H. (2009). A polymorphism in npr-1 is a behavioral determinant of pathogen susceptibility in C. elegans. Science *323*, 382-384.
Reddy, M.R., Overgaard, H.J., Abaga, S., Reddy, V.P., Caccone, A., Kiszewski, A.E., and Slotman, M.A. (2011). Outdoor host seeking behaviour of Anopheles gambiae mosquitoes following initiation of malaria vector control on Bioko Island, Equatorial Guinea. Malar J *10*, 184.
Reidenbach, K.R., Cook, S., Bertone, M.A., Harbach, R.E., Wiegmann, B.M., and Besansky, N.J. (2009). Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology. BMC Evol Biol *9*, 298.
Reidenbach, K.R., Neafsey, D.E., Costantini, C., Sagnon, N., Simard, F., Ragland, G.J., Egan, S.P., Feder, J.L., Muskavitch, M.A., and Besansky, N.J. (2012). Patterns of genomic differentiation between ecologically differentiated M and S forms of Anopheles gambiae in West and Central Africa. Genome biology and evolution.
Reiter, P. (2000). From Shakespeare to Defoe: malaria in England in the Little Ice Age. Emerg Infect Dis *6*, 1-11.
Rey, F.E., Faith, J.J., Bain, J., Muehlbauer, M.J., Stevens, R.D., Newgard, C.B., and Gordon, J.I. (2010). Dissecting the in vivo metabolic potential of two human gut acetogens. J Biol Chem *285*, 22082-22090.
Richman, A.M., Dimopoulos, G., Seeley, D., and Kafatos, F.C. (1997). Plasmodium activates the innate immune response of Anopheles gambiae mosquitoes. EMBO J *16*, 6114-6119.
Riehle, M.A., Srinivasan, P., Moreira, C.K., and Jacobs-Lorena, M. (2003). Towards genetic manipulation of wild mosquito populations to combat malaria: advances and challenges. J Exp Biol *206*, 3809-3816.
Riehle, M.M., Guelbeogo, W.M., Gneme, A., Eiglmeier, K., Holm, I., Bischoff, E., Garnier, T., Snyder, G.M., Li, X., Markianos, K.*, et al.* (2011). A cryptic subgroup of Anopheles gambiae is highly susceptible to human malaria parasites. Science *331*, 596-598.
Riehle, M.M., Markianos, K., Niare, O., Xu, J., Li, J., Toure, A.M., Podiougou, B., Oduol, F., Diawara, S., Diallo, M.*, et al.* (2006). Natural malaria infection in Anopheles gambiae is regulated by a single genomic control region. Science *312*, 577-579.
Riehle, M.M., Xu, J., Lazzaro, B.P., Rottschaefer, S.M., Coulibaly, B., Sacko, M., Niare, O., Morlais, I., Traore, S.F., and Vernick, K.D. (2008). Anopheles gambiae APL1 is a family of variable LRR proteins required for Rel1-mediated protection from the malaria parasite, Plasmodium berghei. PLoS One *3*, e3672.
Riveron, J.M., Irving, H., Ndula, M., Barnes, K.G., Ibrahim, S.S., Paine, M.J., and Wondji, C.S. (2013). Directionally selected cytochrome P450 alleles are driving the spread of pyrethroid resistance in the major malaria vector Anopheles funestus. Proc Natl Acad Sci U S A *110*, 252-257.
Rocca, K.A., Gray, E.M., Costantini, C., and Besansky, N.J. (2009). 2La chromosomal inversion enhances thermal tolerance of Anopheles gambiae larvae. Malar J *8*, 147.
Rodrigues, J., Brayner, F.A., Alves, L.C., Dixit, R., and Barillas-Mury, C. (2010). Hemocyte differentiation mediates innate immune memory in Anopheles gambiae mosquitoes. Science *329*, 1353-1355.
Rodrigues, J., Oliveira, G.A., Kotsyfakis, M., Dixit, R., Molina-Cruz, A., Jochim, R., and Barillas-Mury, C. (2012). An Epithelial Serine Protease, AgESP, Is Required for Plasmodium Invasion in the Mosquito Anopheles gambiae. PLoS One *7*, e35210.
Rogers, C., Reale, V., Kim, K., Chatwin, H., Li, C., Evans, P., and de Bono, M. (2003). Inhibition of Caenorhabditis elegans social feeding by FMRFamide-related peptide activation of NPR-1. Nat Neurosci *6*, 1178-1185.
Rolff, J.R., SE (2009). Insect Infection and Immunity (New York: Oxford University Press Inc).
Rono, M.K., Whitten, M.M., Oulad-Abdelghani, M., Levashina, E.A., and Marois, E. (2010). The major yolk protein vitellogenin interferes with the anti-plasmodium response in the malaria mosquito Anopheles gambiae. PLoS Biol *8*, e1000434.
Rosner, D., and Markowitz, G. (2013). Persistent pollutants: a brief history of the discovery of the widespread toxicity of chlorinated hydrocarbons. Environmental research *120*, 126-133.
Rund, S.S., Hou, T.Y., Ward, S.M., Collins, F.H., and Duffield, G.E. (2011). Genome-wide profiling of diel and circadian gene expression in the malaria vector Anopheles gambiae. Proc Natl Acad Sci U S A.
Ruoslahti, E. (1984). Fibronectin in cell adhesion and invasion. Cancer metastasis reviews *3*, 43-51.
Rusten, T.E., Vaccari, T., Lindmo, K., Rodahl, L.M., Nezis, I.P., Sem-Jacobsen, C., Wendler, F., Vincent, J.P., Brech, A., Bilder, D.*, et al.* (2007). ESCRTs and Fab1 regulate distinct steps of autophagy. Curr Biol *17*, 1817-1825.
Ruwende, C., Khoo, S.C., Snow, R.W., Yates, S.N., Kwiatkowski, D., Gupta, S., Warn, P., Allsopp, C.E., Gilbert, S.C., Peschu, N.*, et al.* (1995). Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. Nature *376*, 246-249.
Ryu, J.H., Kim, S.H., Lee, H.Y., Bai, J.Y., Nam, Y.D., Bae, J.W., Lee, D.G., Shin, S.C., Ha, E.M., and Lee, W.J. (2008). Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in Drosophila. Science *319*, 777-782.
Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J.*, et al.* (2002). Detecting recent positive selection in the human genome from haplotype structure. Nature *419*, 832-837.
Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L.*, et al.* (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature *409*, 928-933.
Sakurai, M., Aoki, T., Yoshikawa, S., Santschi, L.A., Saito, H., Endo, K., Ishikawa, K., Kimura, K., Ito, K., Thomas, J.B.*, et al.* (2009). Differentially expressed Drl and Drl-2 play opposing roles in Wnt5 signaling during Drosophila olfactory system development. J Neurosci *29*, 4972-4980.
Sanchez-Arrones, L., Cardozo, M., Nieto-Lopez, F., and Bovolenta, P. (2012). Cdon and Boc: Two transmembrane proteins implicated in cell-cell communication. Int J Biochem Cell Biol *44*, 698-702.
Sangare, I., Michalakis, Y., Yameogo, B., Dabire, R., Morlais, I., and Cohuet, A. (2013). Studying fitness cost of Plasmodium falciparum infection in malaria vectors: validation of an appropriate negative control. Malar J *12*, 2.
Santiago-Sotelo, P., and Ramirez-Prado, J.H. (2012). prfectBLAST: a platform-independent portable front end for the command terminal BLAST+ stand-alone suite. Biotechniques *53*, 299-300.
Santini, A.C., Santos, H.R., Gross, E., and Correa, R.X. (2013). Genetic diversity of Burkholderia (Proteobacteria) species from the Caatinga and Atlantic rainforest biomes in Bahia, Brazil. Genetics and molecular research : GMR *12*, 655-664.
Santos, A.V., Dillon, R.J., Dillon, V.M., Reynolds, S.E., and Samuels, R.I. (2004). Ocurrence of the antibiotic producing bacterium Burkholderia sp. in colonies of the leaf-cutting ant Atta sexdens rubropilosa. FEMS Microbiol Lett *239*, 319-323.
Saper, C.B., Chou, T.C., and Elmquist, J.K. (2002). The need to feed: homeostatic and hedonic control of eating. Neuron *36*, 199-211.
Sato, K., Pellegrino, M., Nakagawa, T., Nakagawa, T., Vosshall, L.B., and Touhara, K. (2008). Insect olfactory receptors are heteromeric ligand-gated ion channels. Nature *452*, 1002-1006.

Sato, K., Tanaka, K., and Touhara, K. (2011). Sugar-regulated cation channel formed by an insect gustatory receptor. Proc Natl Acad Sci U S A *108*, 11680-11685.
Schlagenhauf, P., Adamcova, M., Regep, L., Schaerer, M.T., and Rhein, H.G. (2010). The position of mefloquine as a 21st century malaria chemoprophylaxis. Malar J *9*, 357.
Schlein, Y., Polacheck, I., and Yuval, B. (1985). Mycoses, bacterial infections and antibacterial activity in sandflies (Psychodidae) and their possible role in the transmission of leishmaniasis. Parasitology *90 (Pt 1)*, 57-66.
Schmucker, D. (2007). Molecular diversity of Dscam: recognition of molecular identity in neuronal wiring. Nature reviews Neuroscience *8*, 915-920.
Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., Dixon, J.E., and Zipursky, S.L. (2000). Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell *101*, 671-684.
Schnitger, A.K., Kafatos, F.C., and Osta, M.A. (2007). The melanization reaction is not required for survival of Anopheles gambiae mosquitoes after bacterial infections. J Biol Chem *282*, 21884-21888.
Schnitger, A.K., Yassine, H., Kafatos, F.C., and Osta, M.A. (2009). Two C-type lectins cooperate to defend Anopheles gambiae against Gram-negative bacteria. J Biol Chem *284*, 17616-17624.
Schuldt, K., Kretz, C.C., Timmann, C., Sievertsen, J., Ehmen, C., Esser, C., Loag, W., Ansong, D., Dering, C., Evans, J.*, et al.* (2011). A -436C>A polymorphism in the human FAS gene promoter associated with severe childhood malaria. PLoS Genet *7*, e1002066.
Schwartz, M.W., Sipols, A.J., Marks, J.L., Sanacora, G., White, J.D., Scheurink, A., Kahn, S.E., Baskin, D.G., Woods, S.C., Figlewicz, D.P.*, et al.* (1992). Inhibition of hypothalamic neuropeptide Y gene expression by insulin. Endocrinology *130*, 3608-3616.
Seitz, H.M., Maier, W.A., Rottok, M., and Becker-Feldmann, H. (1987). Concomitant infections of Anopheles stephensi with Plasmodium berghei and Serratia marcescens: additive detrimental effects. Zentralbl Bakteriol Mikrobiol Hyg A *266*, 155-166.
Serhan, C.N., Chiang, N., and Van Dyke, T.E. (2008). Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators. Nature reviews Immunology *8*, 349-361.
Sharakhov, I.V., White, B.J., Sharakhova, M.V., Kayondo, J., Lobo, N.F., Santolamazza, F., Della Torre, A., Simard, F., Collins, F.H., and Besansky, N.J. (2006). Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2La) in the Anopheles gambiae complex. Proc Natl Acad Sci U S A *103*, 6258-6262.
Sharma, A., Dhayal, D., Singh, O.P., Adak, T., and Bhatnagar, R.K. (2013). Gut microbes influence fitness and malaria transmission potential of Asian malaria vector Anopheles stephensi. Acta Trop.
Sharon, G., Segal, D., Ringo, J.M., Hefetz, A., Zilber-Rosenberg, I., and Rosenberg, E. (2010). Commensal bacteria play a role in mating preference of Drosophila melanogaster. Proc Natl Acad Sci U S A *107*, 20051-20056.
Shen, P., and Cai, H.N. (2001). Drosophila neuropeptide F mediates integration of chemosensory stimulation and conditioning of the nervous system by food. J Neurobiol *47*, 16-25.
Shin, S.C., Kim, S.H., You, H., Kim, B., Kim, A.C., Lee, K.A., Yoon, J.H., Ryu, J.H., and Lee, W.J. (2011). Drosophila microbiome modulates host developmental and metabolic homeostasis via insulin signaling. Science *334*, 670-674.
Shin, S.W., Kokoza, V., Bian, G., Cheon, H.M., Kim, Y.J., and Raikhel, A.S. (2005). REL1, a homologue of Drosophila dorsal, regulates toll antifungal immune pathway in the female mosquito Aedes aegypti. J Biol Chem *280*, 16499-16507.
Shirangi, T.R., Taylor, B.J., and McKeown, M. (2006). A double-switch system regulates male courtship behavior in male and female Drosophila melanogaster. Nat Genet *38*, 1435-1439.
Shohat-Ophir, G., Kaun, K.R., Azanchi, R., and Heberlein, U. (2012). Sexual deprivation increases ethanol intake in Drosophila. Science *335*, 1351-1355.
Shrivastava, S., McVey, J.H., and Dorling, A. (2007). The interface between coagulation and immunity. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons *7*, 499-506.
Siden-Kiamos, I., and Louis, C. (2004). Interactions between malaria parasites and their mosquito hosts in the midgut. Insect Biochem Mol Biol *34*, 679-685.
Sidhu, A.B., Verdier-Pinard, D., and Fidock, D.A. (2002). Chloroquine resistance in Plasmodium falciparum malaria parasites conferred by pfcrt mutations. Science *298*, 210-213.
Siebert, M., Banovic, D., Goellner, B., and Aberle, H. (2009). Drosophila motor axons recognize and follow a Sidestep-labeled substrate pathway to reach their target fields. Genes & Development *23*, 1052-1062.
Sigrist, C.J., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2013). New and continuing developments at PROSITE. Nucleic Acids Res *41*, D344-347.
Simard, F., Ayala, D., Kamdem, G.C., Pombi, M., Etouna, J., Ose, K., Fotsing, J.M., Fontenille, D., Besansky, N.J., and Costantini, C. (2009). Ecological niche partitioning between Anopheles gambiae molecular forms in Cameroon: the ecological side of speciation. BMC ecology *9*, 17.
Sinden, R.E. (2002). Molecular interactions between Plasmodium and its insect vectors. Cell Microbiol *4*, 713-724.
Sinden, R.E., Alavi, Y., and Raine, J.D. (2004). Mosquito--malaria interactions: a reappraisal of the concepts of susceptibility and refractoriness. Insect Biochem Mol Biol *34*, 625-629.
Singer, M.S., Mace, K.C., and Bernays, E.A. (2009). Self-medication as adaptive plasticity: increased ingestion of plant toxins by parasitized caterpillars. PLoS One *4*, e4796.
Sink, H., Rehm, E.J., Richstone, L., Bulls, Y.M., and Goodman, C.S. (2001). sidestep encodes a target-derived attractant essential for motor axon guidance in Drosophila. Cell *105*, 57-67.
Slone, J., Daniels, J., and Amrein, H. (2007). Sugar receptors in Drosophila. Curr Biol *17*, 1809-1816.
Smallegange, R.C., Qiu, Y.T., Bukovinszkine-Kiss, G., Van Loon, J.J., and Takken, W. (2009). The effect of aliphatic carboxylic acids on olfaction-based host-seeking of the malaria mosquito Anopheles gambiae sensu stricto. Journal of chemical ecology *35*, 933-943.
Smallegange, R.C., Qiu, Y.T., van Loon, J.J., and Takken, W. (2005). Synergism between ammonia, lactic acid and carboxylic acids as kairomones in the host-seeking behaviour of the malaria mosquito Anopheles gambiae sensu stricto (Diptera: Culicidae). Chem Senses *30*, 145-152.
Smallegange, R.C., van Gemert, G.J., van de Vegte-Bolmer, M., Gezan, S., Takken, W., Sauerwein, R.W., and Logan, J.G. (2013). Malaria infected mosquitoes express enhanced attraction to human odor. PLoS One *8*, e63602.

Smith, P.H., Mwangi, J.M., Afrane, Y.A., Yan, G., Obbard, D.J., Ranford-Cartwright, L.C., and Little, T.J. (2011). Alternative splicing of the Anopheles gambiae Dscam gene in diverse Plasmodium falciparum infections. Malar J *10*, 156.
Smits, A., Roelants, P., Van Bortel, W., and Coosemans, M. (1996). Enzyme polymorphisms in the Anopheles gambiae (Diptera:Culicidae) complex related to feeding and resting behavior in the Imbo Valley, Burundi. J Med Entomol *33*, 545-553.
Soukup, S.F., Culi, J., and Gubb, D. (2009). Uptake of the necrotic serpin in Drosophila melanogaster via the lipophorin receptor-1. PLoS Genet *5*, e1000532.
Souza-Neto, J.A., Sim, S., and Dimopoulos, G. (2009). An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense. Proc Natl Acad Sci U S A.
Spehr, M., and Munger, S.D. (2009). Olfactory receptors: G protein-coupled receptors and beyond. J Neurochem *109*, 1570-1583.
Spor, A., Koren, O., and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. Nat Rev Microbiol *9*, 279-290.
Stanek, D.M., Pohl, J., Crim, J.W., and Brown, M.R. (2002). Neuropeptide F and its expression in the yellow fever mosquito, Aedes aegypti. Peptides *23*, 1367-1378.
Stark, A., Lin, M.F., Kheradpour, P., Pedersen, J.S., Parts, L., Carlson, J.W., Crosby, M.A., Rasmussen, M.D., Roy, S., Deoras, A.N*., et al.* (2007). Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature *450*, 219-232.
Stecher, B., Robbiani, R., Walker, A.W., Westendorf, A.M., Barthel, M., Kremer, M., Chaffron, S., Macpherson, A.J., Buer, J., Parkhill, J*., et al.* (2007). Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota. PLoS Biol *5*, 2177-2189.
Stein, D., and Nusslein-Volhard, C. (1992). Multiple extracellular activities in Drosophila egg perivitelline fluid are required for establishment of embryonic dorsal-ventral polarity. Cell *68*, 429-440.
Stevenson, B.J., Pignatelli, P., Nikou, D., and Paine, M.J. (2012). Pinpointing P450s associated with pyrethroid metabolism in the dengue vector, Aedes aegypti: developing new tools to combat insecticide resistance. PLoS Negl Trop Dis *6*, e1595.
Stevenson, M.M., Ing, R., Berretta, F., and Miu, J. (2011). Regulating the adaptive immune response to blood-stage malaria: role of dendritic cells and CD4(+)Foxp3(+) regulatory T cells. Int J Biol Sci *7*, 1311-1322.
Stevenson, M.M., and Riley, E.M. (2004). Innate immunity to malaria. Nature reviews Immunology *4*, 169-180.
Storelli, G., Defaye, A., Erkosar, B., Hols, P., Royet, J., and Leulier, F. (2011). Lactobacillus plantarum Promotes Drosophila Systemic Growth by Modulating Hormonal Signals through TOR-Dependent Nutrient Sensing. Cell metabolism *14*, 403-414.
Stoven, S., Ando, I., Kadalayil, L., Engstrom, Y., and Hultmark, D. (2000). Activation of the Drosophila NF-kappaB factor Relish by rapid endoproteolytic cleavage. EMBO Rep *1*, 347-352.
Stump, A.D., Fitzpatrick, M.C., Lobo, N.F., Traore, S., Sagnon, N., Costantini, C., Collins, F.H., and Besansky, N.J. (2005). Centromere-proximal differentiation and speciation in Anopheles gambiae. Proc Natl Acad Sci U S A *102*, 15930-15935.
Styer, K.L., Singh, V., Macosko, E., Steele, S.E., Bargmann, C.I., and Aballay, A. (2008). Innate immunity in Caenorhabditis elegans is regulated by neurons expressing NPR-1/GPCR. Science *322*, 460-464.
Takken, F.L., and Goverse, A. (2012). How to build a pathogen detector: structural basis of NB-LRR function. Curr Opin Plant Biol *15*, 375-384.
Takken, W., and Knols, B.G. (1999). Odor-mediated behavior of Afrotropical malaria mosquitoes. Annu Rev Entomol *44*, 131-157.
Targett, G.A., and Greenwood, B.M. (2008). Malaria vaccines and their potential role in the elimination of malaria. Malar J *7 Suppl 1*, S10.
Taylor, C., Toure, Y.T., Carnahan, J., Norris, D.E., Dolo, G., Traore, S.F., Edillo, F.E., and Lanzaro, G.C. (2001). Gene flow among populations of the malaria vector, Anopheles gambiae, in Mali, West Africa. Genetics *157*, 743-750.
Terenius, O., de Oliveira, C.D., Pinheiro, W.D., Tadei, W.P., James, A.A., and Marinotti, O. (2008). 16S rRNA gene sequences from bacteria associated with adult Anopheles darlingi (Diptera: Culicidae) mosquitoes. J Med Entomol *45*, 172-175.
Terenius, O., Lindh, J.M., Eriksson-Gonzales, K., Bussiere, L., Laugen, A.T., Bergquist, H., Titanji, K., and Faye, I. (2012). Midgut bacterial dynamics in Aedes aegypti. FEMS Microbiol Ecol.
Thomas, D.D., Donnelly, C.A., Wood, R.J., and Alphey, L.S. (2000). Insect population control using a dominant, repressible, lethal genetic system. Science *287*, 2474-2476.
Thompson, A.A., Liu, W., Chun, E., Katritch, V., Wu, H., Vardy, E., Huang, X.P., Trapella, C., Guerrini, R., Calo, G*., et al.* (2012). Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. Nature *485*, 395-399.
Thurston, T.L., Wandel, M.P., von Muhlinen, N., Foeglein, A., and Randow, F. (2012). Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion. Nature *482*, 414-418.
Tian, A.G., Tamori, Y., Huang, Y.C., Melendez, N.T., and Deng, W.M. (2013). Efficient EGFR signaling and dorsal-ventral axis patterning requires syntaxin dependent Gurken trafficking. Dev Biol *373*, 349-358.
Tian, J., Lu, Y., Zhang, H., Chau, C.H., Dang, H.N., and Kaufman, D.L. (2004). Gamma-aminobutyric acid inhibits T cell autoimmunity and the development of inflammatory responses in a mouse type 1 diabetes model. J Immunol *173*, 5298-5304.
Tikar, S.N., Mendki, M.J., Sharma, A.K., Sukumaran, D., Veer, V., Prakash, S., and Parashar, B.D. (2011). Resistance Status of the Malaria Vector Mosquitoes, Anopheles stephensi and Anopheles subpictus Towards Adulticides and Larvicides in Arid and Semi-Arid Areas of India. J Insect Sci *11*, 85.
Timmann, C., Thye, T., Vens, M., Evans, J., May, J., Ehmen, C., Sievertsen, J., Muntau, B., Ruge, G., Loag, W*., et al.* (2012). Genome-wide association study indicates two novel resistance loci for severe malaria. Nature *489*, 443-446.
Tishkoff, S.A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J*., et al.* (2001). Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science *293*, 455-462.
Toure, Y.T., Petrarca, V., Traore, S.F., Coulibaly, A., Maiga, H.M., Sankare, O., Sow, M., Di Deco, M.A., and Coluzzi, M. (1994). Ecological genetic studies in the chromosomal form Mopti of Anopheles gambiae s.str. in Mali, west Africa. Genetica *94*, 213-223.
Toure, Y.T., Petrarca, V., Traore, S.F., Coulibaly, A., Maiga, H.M., Sankare, O., Sow, M., Di Deco, M.A., and Coluzzi, M. (1998). The distribution and inversion polymorphism of chromosomally recognized taxa of the Anopheles gambiae complex in Mali, West Africa. Parassitologia *40*, 477-511.
Tripet, F., Dolo, G., and Lanzaro, G.C. (2005). Multilevel analyses of genetic differentiation in Anopheles gambiae s.s. reveal patterns of gene flow important for malaria-fighting mosquito projects. Genetics *169*, 313-324.

Tripet, F., Toure, Y.T., Taylor, C.E., Norris, D.E., Dolo, G., and Lanzaro, G.C. (2001). DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of Anopheles gambiae. Mol Ecol *10*, 1725-1732.
Turner, T.L., Hahn, M.W., and Nuzhdin, S.V. (2005). Genomic islands of speciation in Anopheles gambiae. PLoS Biol *3*, e285.
Tzou, P., Ohresser, S., Ferrandon, D., Capovilla, M., Reichhart, J.M., Lemaitre, B., Hoffmann, J.A., and Imler, J.L. (2000). Tissue-specific inducible expression of antimicrobial peptide genes in Drosophila surface epithelia. Immunity *13*, 737-748.
Ueno, K., Ohta, M., Morita, H., Mikuni, Y., Nakajima, S., Yamamoto, K., and Isono, K. (2001). Trehalose sensitivity in Drosophila correlates with mutations in and expression of the gustatory receptor gene Gr5a. Curr Biol *11*, 1451-1455.
Vaishnava, S., Yamamoto, M., Severson, K.M., Ruhn, K.A., Yu, X., Koren, O., Ley, R., Wakeland, E.K., and Hooper, L.V. (2011). The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. Science *334*, 255-258.
Valanne, S., and Ramet, M. (2013). Uracil debases pathogenic but not commensal bacteria. Cell Host Microbe *13*, 505-506.
Valanne, S., Wang, J.H., and Ramet, M. (2011). The Drosophila Toll signaling pathway. J Immunol *186*, 649-656.
van de Leur, J.J., Vollaard, E.J., Janssen, A.J., and Dofferhoff, A.S. (1997). Influence of low dose ciprofloxacin on microbial colonization of the digestive tract in healthy volunteers during normal and during impaired colonization resistance. Scandinavian journal of infectious diseases *29*, 297-300.
Van Hiel, M.B., Van Loy, T., Poels, J., Vandersmissen, H.P., Verlinden, H., Badisco, L., and Vanden Broeck, J. (2010). Neuropeptide receptors as possible targets for development of insect pest control agents. Adv Exp Med Biol *692*, 211-226.
Vannini, L., Reed, T.W., and Willis, J.H. (2014). Temporal and spatial expression of cuticular proteins of Anopheles gambiae implicated in insecticide resistance or differentiation of M/S incipient species. Parasit Vectors *7*, 24.
Veenstra-VanderWeele, J., and Blakely, R.D. (2012). Networking in autism: leveraging genetic, biomarker and model system findings in the search for new treatments. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology *37*, 196-212.
Velavan, T.P., Buyukyazici, B., Kremsner, P.G., and Kun, J.F. (2012). Combined promoter haplotypes of the IL10R genes are associated with protection against severe malaria in Gabonese children. Immunogenetics *64*, 87-95.
Venkatakrishnan, A.J., Deupi, X., Lebon, G., Tate, C.G., Schertler, G.F., and Babu, M.M. (2013). Molecular signatures of G-protein-coupled receptors. Nature *494*, 185-194.
Verhulst, N.O., Andriessen, R., Groenhagen, U., Bukovinszkine Kiss, G., Schulz, S., Takken, W., van Loon, J.J., Schraa, G., and Smallegange, R.C. (2010). Differential attraction of malaria mosquitoes to volatile blends produced by human skin bacteria. PLoS One *5*, e15829.
Verhulst, N.O., Qiu, Y.T., Beijleveld, H., Maliepaard, C., Knights, D., Schulz, S., Berg-Lyons, D., Lauber, C.L., Verduijn, W., Haasnoot, G.W.*, et al.* (2011). Composition of human skin microbiota affects attractiveness to malaria mosquitoes. PLoS One *6*, e28991.
Vernick, K.D., Fujioka, H., Seeley, D.C., Tandler, B., Aikawa, M., and Miller, L.H. (1995). Plasmodium gallinaceum: a refractory mechanism of ookinete killing in the mosquito, Anopheles gambiae. Exp Parasitol *80*, 583-595.
Virgin, H.W., and Todd, J.A. (2011). Metagenomics and personalized medicine. Cell *147*, 44-56.
Visintin, A., Iliev, D.B., Monks, B.G., Halmen, K.A., and Golenbock, D.T. (2006). Md-2. Immunobiology *211*, 437-447.
Vizioli, J., Bulet, P., Charlet, M., Lowenberger, C., Blass, C., Muller, H.M., Dimopoulos, G., Hoffmann, J., Kafatos, F.C., and Richman, A. (2000). Cloning and analysis of a cecropin gene from the malaria vector mosquito, Anopheles gambiae. Insect Mol Biol *9*, 75-84.
Vizioli, J., Bulet, P., Hoffmann, J.A., Kafatos, F.C., Muller, H.M., and Dimopoulos, G. (2001). Gambicin: a novel immune responsive antimicrobial peptide from the malaria vector Anopheles gambiae. Proc Natl Acad Sci U S A *98*, 12630-12635.
Vlachou, D., Schlegelmilch, T., Christophides, G.K., and Kafatos, F.C. (2005). Functional genomic analysis of midgut epithelial responses in Anopheles during Plasmodium invasion. Curr Biol *15*, 1185-1195.
Vlachou, D., Schlegelmilch, T., Runn, E., Mendes, A., and Kafatos, F.C. (2006). The developmental migration of Plasmodium in mosquitoes. Curr Opin Genet Dev *16*, 384-391.
Vodovar, N., Vinals, M., Liehl, P., Basset, A., Degrouard, J., Spellman, P., Boccard, F., and Lemaitre, B. (2005). Drosophila host defense after oral infection by an entomopathogenic Pseudomonas species. Proc Natl Acad Sci U S A *102*, 11414-11419.
Vogel, G. (2012). Global health. Disappointing results blunt hopes for malaria vaccine. Science *338*, 871-872.
Vollaard, E.J., Clasener, H.A., and Janssen, A.J. (1992). Co-trimoxazole impairs colonization resistance in healthy volunteers. J Antimicrob Chemother *30*, 685-691.
Volz, J., Muller, H.M., Zdanowicz, A., Kafatos, F.C., and Osta, M.A. (2006). A genetic module regulates the melanization response of Anopheles to Plasmodium. Cell Microbiol *8*, 1392-1405.
Volz, J., Osta, M.A., Kafatos, F.C., and Muller, H.M. (2005). The roles of two clip domain serine proteases in innate immune responses of the malaria vector Anopheles gambiae. J Biol Chem *280*, 40161-40168.
Waldock, J., Olson, K.E., and Christophides, G.K. (2012). Anopheles gambiae Antiviral Immune Response to Systemic O'nyong-nyong Infection. PLoS Negl Trop Dis *6*, e1565.
Walsh, F.S., and Doherty, P. (1997). Neural cell adhesion molecules of the immunoglobulin superfamily: role in axon growth and guidance. Annual review of cell and developmental biology *13*, 425-456.
Wang, J., Weiss, B.L., and Aksoy, S. (2013). Tsetse fly microbiota: form and function. Frontiers in cellular and infection microbiology *3*, 69.
Wang, J., Zugates, C.T., Liang, I.H., Lee, C.H., and Lee, T. (2002). Drosophila Dscam is required for divergent segregation of sister branches and suppresses ectopic bifurcation of axons. Neuron *33*, 559-571.
Wang, L., Han, X., Mehren, J., Hiroi, M., Billeter, J.C., Miyamoto, T., Amrein, H., Levine, J.D., and Anderson, D.J. (2011a). Hierarchical chemosensory regulation of male-male social interactions in Drosophila. Nat Neurosci *14*, 757-762.
Wang, S., Ghosh, A.K., Bongio, N., Stebbings, K.A., Lampe, D.J., and Jacobs-Lorena, M. (2012). Fighting malaria with engineered symbiotic bacteria from vector mosquitoes. Proc Natl Acad Sci U S A.
Wang, S., and Jacobs-Lorena, M. (2013). Genetic approaches to interfere with malaria transmission by vector mosquitoes. Trends Biotechnol.
Wang, Y., Gilbreath, T.M., 3rd, Kukutla, P., Yan, G., and Xu, J. (2011b). Dynamic Gut Microbiome across Life History of the Malaria Mosquito Anopheles gambiae in Kenya. PLoS One *6*, e24767.
Wang, Y., Zou, Z., and Jiang, H. (2006). An expansion of the dual clip-domain serine proteinase family in Manduca sexta: gene organization, expression, and evolution of prophenoloxidase-activating proteinase-2, hemolymph proteinase 12, and other related proteinases. Genomics *87*, 399-409.

Warr, E., Das, S., Dong, Y., and Dimopoulos, G. (2008). The Gram-negative bacteria-binding protein gene family: its role in the innate immune system of anopheles gambiae and in anti-Plasmodium defence. Insect Mol Biol *17*, 39-51.
Watanabe, K., Toba, G., Koganezawa, M., and Yamamoto, D. (2011). Gr39a, a highly diversified gustatory receptor in Drosophila, has a role in sexual behavior. Behav Genet *41*, 746-753.
Waterhouse, R.M., Kriventseva, E.V., Meister, S., Xi, Z., Alvarez, K.S., Bartholomay, L.C., Barillas-Mury, C., Bian, G., Blandin, S., Christensen, B.M*., et al.* (2007). Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. Science *316*, 1738-1743.
Waterhouse, R.M., Povelones, M., and Christophides, G.K. (2010). Sequence-structure-function relations of the mosquito leucine-rich repeat immune proteins. BMC Genomics *11*, 531.
Waterhouse, R.M., Tegenfeldt, F., Li, J., Zdobnov, E.M., and Kriventseva, E.V. (2013). OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res *41*, D358-365.
Watson, F.L., Puttmann-Holgado, R., Thomas, F., Lamar, D.L., Hughes, M., Kondo, M., Rebel, V.I., and Schmucker, D. (2005). Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science *309*, 1874-1878.
Watthanasurorot, A., Jiravanichpaisal, P., Liu, H., Soderhall, I., and Soderhall, K. (2011). Bacteria-Induced Dscam Isoforms of the Crustacean, Pacifastacus leniusculus. PLoS Pathog *7*, e1002062.
Weber, A.L., Khan, G.F., Magwire, M.M., Tabor, C.L., Mackay, T.F., and Anholt, R.R. (2012). Genome-wide association analysis of oxidative stress resistance in Drosophila melanogaster. PLoS One *7*, e34745.
Weetman, D., Wilding, C.S., Steen, K., Morgan, J.C., Simard, F., and Donnelly, M.J. (2010). Association mapping of insecticide resistance in wild anopheles gambiae populations: major variants identified in a low-linkage disequilbrium genome. PLoS One *5*, e13140.
Weetman, D., Wilding, C.S., Steen, K., Pinto, J., and Donnelly, M.J. (2012). Gene flow-dependent genomic divergence between Anopheles gambiae M and S forms. Mol Biol Evol *29*, 279-291.
Weill, M., Chandre, F., Brengues, C., Manguin, S., Akogbeto, M., Pasteur, N., Guillet, P., and Raymond, M. (2000). The kdr mutation occurs in the Mopti form of Anopheles gambiae s.s. through introgression. Insect Mol Biol *9*, 451-455.
Weiss, B.L., Wang, J., and Aksoy, S. (2011a). Tsetse immune system maturation requires the presence of obligate symbionts in larvae. PLoS Biol *9*, e1000619.
Weiss, B.L., Wang, J., Maltz, M.A., Wu, Y., and Aksoy, S. (2013). Trypanosome infection establishment in the tsetse fly gut is influenced by microbiome-regulated host immune barriers. PLoS Pathog *9*, e1003318.
Weiss, L.A., Dahanukar, A., Kwon, J.Y., Banerjee, D., and Carlson, J.R. (2011b). The molecular and cellular basis of bitter taste in Drosophila. Neuron *69*, 258-272.
Welburn, S.C., and Maudlin, I. (1999). Tsetse-trypanosome interactions: rites of passage. Parasitol Today *15*, 399-403.
Wen, L., Ley, R.E., Volchkov, P.Y., Stranges, P.B., Avanesyan, L., Stonebraker, A.C., Hu, C., Wong, F.S., Szot, G.L., Bluestone, J.A*., et al.* (2008). Innate immunity and intestinal microbiota in the development of Type 1 diabetes. Nature *455*, 1109-1113.
Wen, T., Parrish, C.A., Xu, D., Wu, Q., and Shen, P. (2005). Drosophila neuropeptide F and its receptor, NPFR1, define a signaling pathway that acutely modulates alcohol sensitivity. Proc Natl Acad Sci U S A *102*, 2141-2146.
Werren, J.H., Baldo, L., and Clark, M.E. (2008). Wolbachia: master manipulators of invertebrate biology. Nat Rev Microbiol *6*, 741-751.
Whaley, K., and Schwaeble, W. (1997). Complement and complement deficiencies. Seminars in liver disease *17*, 297-310.
White, B.J., Cheng, C., Simard, F., Costantini, C., and Besansky, N.J. (2010). Genetic association of physically unlinked islands of genomic divergence in incipient species of Anopheles gambiae. Mol Ecol *19*, 925-939.
White, B.J., Lawniczak, M.K., Cheng, C., Coulibaly, M.B., Wilson, M.D., Sagnon, N., Costantini, C., Simard, F., Christophides, G.K., and Besansky, N.J. (2011). Adaptive divergence between incipient species of Anopheles gambiae increases resistance to Plasmodium. Proc Natl Acad Sci U S A *108*, 244-249.
White, B.J., Santolamazza, F., Kamau, L., Pombi, M., Grushko, O., Mouline, K., Brengues, C., Guelbeogo, W., Coulibaly, M., Kayondo, J.K*., et al.* (2007). Molecular karyotyping of the 2La inversion in Anopheles gambiae. Am J Trop Med Hyg *76*, 334-339.
White, J.F., Noinaj, N., Shibata, Y., Love, J., Kloss, B., Xu, F., Gvozdenovic-Jeremic, J., Shah, P., Shiloach, J., Tate, C.G*., et al.* (2012). Structure of the agonist-bound neurotensin receptor. Nature *490*, 508-513.
White, N. (2011). First Results of Phase 3 Trial of RTS,S/AS01 Malaria Vaccine in African Children. N Engl J Med.
WHO (1973). Handbook of resolutions and decisions of the World

Health Assembly and the Executive Board, Vol I; 1948-1972 (Geneva: WHO).
WHO (2012). World Malaria Report 2012 (Geneva: WHO Press).
Wilding, C.S., Weetman, D., Steen, K., and Donnelly, M.J. (2009). High, clustered, nucleotide diversity in the genome of Anopheles gambiae revealed through pooled-template sequencing: implications for high-throughput genotyping protocols. BMC Genomics *10*, 320.
Williams, T.N. (2006). Human red blood cell polymorphisms and malaria. Curr Opin Microbiol *9*, 388-394.
Williams, T.N., Mwangi, T.W., Roberts, D.J., Alexander, N.D., Weatherall, D.J., Wambua, S., Kortok, M., Snow, R.W., and Marsh, K. (2005a). An immune basis for malaria protection by the sickle cell trait. PLoS Med *2*, e128.
Williams, T.N., Mwangi, T.W., Wambua, S., Alexander, N.D., Kortok, M., Snow, R.W., and Marsh, K. (2005b). Sickle cell trait and the risk of Plasmodium falciparum malaria and other childhood diseases. J Infect Dis *192*, 178-186.
Windbichler, N., Menichelli, M., Papathanos, P.A., Thyme, S.B., Li, H., Ulge, U.Y., Hovde, B.T., Baker, D., Monnat, R.J., Burt, A*., et al.* (2011). A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature.
Wirth, D.F. (2002). Biological revelations. Nature *419*, 495-496.
Wojtowicz, W.M., Flanagan, J.J., Millard, S.S., Zipursky, S.L., and Clemens, J.C. (2004). Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. Cell *118*, 619-633.
Wojtowicz, W.M., Wu, W., Andre, I., Qian, B., Baker, D., and Zipursky, S.L. (2007). A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. Cell *130*, 1134-1145.
Wondji, C., Simard, F., and Fontenille, D. (2002). Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of Anopheles gambiae in an area of sympatry. Insect Mol Biol *11*, 11-19.

Wong, A.C., Chaston, J.M., and Douglas, A.E. (2013). The inconstant gut microbiota of Drosophila species revealed by 16S rRNA gene analysis. ISME J.
Wu, H.-J., Ivanov, I.I., Darce, J., Hattori, K., Shima, T., Umesaki, Y., Littman, D.R., Benoist, C., and Mathis, D. (2010). Gut-Residing Segmented Filamentous Bacteria Drive Autoimmune Arthritis via T Helper 17 Cells. Immunity *32*, 815-827.
Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G.W., Vardy, E., Liu, W., Thompson, A.A., Huang, X.P., Carroll, F.I.*, et al.* (2012). Structure of the human kappa-opioid receptor in complex with JDTic. Nature *485*, 327-332.
Wu, Q., Wen, T., Lee, G., Park, J.H., Cai, H.N., and Shen, P. (2003). Developmental Control of Foraging and Social Behavior by the Drosophila Neuropeptide Y-like System. Neuron *39*, 147-161.
Wu, Q., Zhang, Y., Xu, J., and Shen, P. (2005a). Regulation of hunger-driven behaviors by neural ribosomal S6 kinase in Drosophila. Proc Natl Acad Sci U S A *102*, 13289-13294.
Wu, Q., Zhao, Z., and Shen, P. (2005b). Regulation of aversion to noxious food by Drosophila neuropeptide Y- and insulin-like systems. Nat Neurosci *8*, 1350-1355.
Wu, Y., Ellis, R.D., Shaffer, D., Fontes, E., Malkin, E.M., Mahanty, S., Fay, M.P., Narum, D., Rausch, K., Miles, A.P.*, et al.* (2008). Phase 1 trial of malaria transmission blocking vaccine candidates Pfs25 and Pvs25 formulated with montanide ISA 51. PLoS One *3*, e2636.
Wyder, S., Kriventseva, E.V., Schroder, R., Kadowaki, T., and Zdobnov, E.M. (2007). Quantification of ortholog losses in insects and vertebrates. Genome Biol *8*, R242.
Xi, Z., Khoo, C.C., and Dobson, S.L. (2005). Wolbachia establishment and invasion in an Aedes aegypti laboratory population. Science *310*, 326-328.
Xi, Z., Ramirez, J.L., and Dimopoulos, G. (2008). The Aedes aegypti toll pathway controls dengue virus infection. PLoS Pathog *4*, e1000098.
Xiang, Y., Yuan, Q., Vogt, N., Looger, L.L., Jan, L.Y., and Jan, Y.N. (2010). Light-avoidance-mediating photoreceptors tile the Drosophila larval body wall. Nature *468*, 921-926.
Xin, N., Benchabane, H., Tian, A., Nguyen, K., Klofas, L., and Ahmed, Y. (2011). Erect Wing facilitates context-dependent Wnt/Wingless signaling by recruiting the cell-specific Armadillo-TCF adaptor Earthbound to chromatin. Development *138*, 4955-4967.
Xiong, B., and Jacobs-Lorena, M. (1995). Gut-specific transcriptional regulatory elements of the carboxypeptidase gene are conserved between black flies and Drosophila. Proc Natl Acad Sci U S A *92*, 9313-9317.
Xu, J., Li, M., and Shen, P. (2010). A G-protein-coupled neuropeptide Y-like receptor suppresses behavioral and sensory response to multiple stressful stimuli in Drosophila. J Neurosci *30*, 2504-2512.
Xu, J., Sheng, Z., and Palli, S.R. (2013). Juvenile hormone and insulin regulate trehalose homeostasis in the red flour beetle, Tribolium castaneum. PLoS Genet *9*, e1003535.
Yamada, Y., Katsura, K., Kawasaki, H., Widyastuti, Y., Saono, S., Seki, T., Uchimura, T., and Komagata, K. (2000). Asaia bogorensis gen. nov., sp. nov., an unusual acetic acid bacterium in the alpha-Proteobacteria. Int J Syst Evol Microbiol *50 Pt 2*, 823-829.
Yamakawa, K., Huot, Y.K., Haendelt, M.A., Hubert, R., Chen, X.N., Lyons, G.E., and Korenberg, J.R. (1998). DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system. Hum Mol Genet *7*, 227-237.
Yang, I.V., Chen, E., Hasseman, J.P., Liang, W., Frank, B.C., Wang, S., Sharov, V., Saeed, A.I., White, J., Li, J.*, et al.* (2002). Within the fold: assessing differential expression measures and reproducibility in microarray assays. Genome Biol *3*, research0062.
Yassine, H., Kamareddine, L., and Osta, M.A. (2012). The Mosquito Melanization Response Is Implicated in Defense against the Entomopathogenic Fungus Beauveria bassiana. PLoS Pathog *8*, e1003029.
Yassine, H., and Osta, M.A. (2010). Anopheles gambiae innate immunity. Cell Microbiol *12*, 1-9.
Ye, T., Tang, W., and Zhang, X. (2012). Involvement of Rab6 in the regulation of phagocytosis against virus infection in invertebrates. Journal of proteome research *11*, 4834-4846.
Yeka, A., Achan, J., D'Alessandro, U., and Talisuna, A.O. (2009). Quinine monotherapy for treating uncomplicated malaria in the era of artemisinin-based combination therapy: an appropriate public health policy? Lancet Infect Dis *9*, 448-452.
Yeka, A., Tibenderana, J., Achan, J., D'Alessandro, U., and Talisuna, A.O. (2013). Efficacy of quinine, artemether-lumefantrine and dihydroartemisinin-piperaquine as rescue treatment for uncomplicated malaria in Ugandan children. PLoS One *8*, e53772.
Yukphan, P., Potacharoen, W., Tanasupawat, S., Tanticharoen, M., and Yamada, Y. (2004). Asaia krungthepensis sp. nov., an acetic acid bacterium in the alpha-Proteobacteria. Int J Syst Evol Microbiol *54*, 313-316.
Zaidman-Remy, A., Herve, M., Poidevin, M., Pili-Floury, S., Kim, M.S., Blanot, D., Oh, B.H., Ueda, R., Mengin-Lecreulx, D., and Lemaitre, B. (2006). The Drosophila amidase PGRP-LB modulates the immune response to bacterial infection. Immunity *24*, 463-473.
Zdobnov, E.M., and Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics *17*, 847-848.
Zdobnov, E.M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R.R., Christophides, G.K., Thomasova, D., Holt, R.A., Subramanian, G.M.*, et al.* (2002). Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster. Science *298*, 149-159.
Zhan, X.-L., Clemens, J.C., Neves, G., Hattori, D., Flanagan, J.J., Hummel, T., Vasconcelos, M.L., Chess, A., and Zipursky, S.L. (2004). Analysis of Dscam Diversity in Regulating Axon Guidance in Drosophila Mushroom Bodies. Neuron *43*, 673-686.
Zhang, G., Lu, Z.Q., Jiang, H., and Asgari, S. (2004). Negative regulation of prophenoloxidase (proPO) activation by a clip-domain serine proteinase homolog (SPH) from endoparasitoid venom. Insect Biochem Mol Biol *34*, 477-483.
Zhang, H.J., Anderson, A.R., Trowell, S.C., Luo, A.R., Xiang, Z.H., and Xia, Q.Y. (2011). Topological and functional characterization of an insect gustatory receptor. PLoS One *6*, e24111.
Zhang, Y., Lu, H., and Bargmann, C.I. (2005). Pathogenic bacteria induce aversive olfactory learning in Caenorhabditis elegans. Nature *438*, 179-184.
Zheng, L., Benedict, M.Q., Cornel, A.J., Collins, F.H., and Kafatos, F.C. (1996). An integrated genetic map of the African human malaria vector mosquito, Anopheles gambiae. Genetics *143*, 941-952.
Zheng, L., Cornel, A.J., Wang, R., Erfle, H., Voss, H., Ansorge, W., Kafatos, F.C., and Collins, F.H. (1997). Quantitative trait loci for refractoriness of Anopheles gambiae to Plasmodium cynomolgi B. Science *276*, 425-428.

Zheng, L., Wang, S., Romans, P., Zhao, H., Luna, C., and Benedict, M.Q. (2003). Quantitative trait loci in Anopheles gambiae controlling the encapsulation response against Plasmodium cynomolgi Ceylon. BMC Genet *4*, 16.
Zhou, R., Hu, G., Liu, J., Gong, A.Y., Drescher, K.M., and Chen, X.M. (2009). NF-kappaB p65-dependent transactivation of miRNA genes following Cryptosporidium parvum infection stimulates epithelial cell immune responses. PLoS Pathog *5*, e1000681.
Zipursky, S.L., and Sanes, J.R. (2010). Chemoaffinity revisited: dscams, protocadherins, and neural circuit assembly. Cell *143*, 343-353.
Zou, Z., Shin, S.W., Alvarez, K.S., Kokoza, V., and Raikhel, A.S. (2010). Distinct Melanization Pathways in the Mosquito Aedes aegypti. Immunity *32*, 41-53.

Appendix

OPEN ACCESS Freely available online

PLOS | PATHOGENS

Genetic Dissection of *Anopheles gambiae* Gut Epithelial Responses to *Serratia marcescens*

Stavros Stathopoulos[1], Daniel E. Neafsey[2], Mara K. N. Lawniczak[1], Marc A. T. Muskavitch[3], George K. Christophides[1,4]*

1 Department of Life Sciences, Imperial College London, London, United Kingdom, **2** Broad Institute, Cambridge, Massachusetts, United States of America, **3** Boston College, Chestnut Hill, Massachusetts, United States of America, **4** The Cyprus Institute, Nicosia, Cyprus

Abstract

Genetic variation in the mosquito *Anopheles gambiae* profoundly influences its ability to transmit malaria. Mosquito gut bacteria are shown to influence the outcome of infections with *Plasmodium* parasites and are also thought to exert a strong drive on genetic variation through natural selection; however, a link between antibacterial effects and genetic variation is yet to emerge. Here, we combined SNP genotyping and expression profiling with phenotypic analyses of candidate genes by RNAi-mediated silencing and 454 pyrosequencing to investigate this intricate biological system. We identified 138 *An. gambiae* genes to be genetically associated with the outcome of *Serratia marcescens* infection, including the peptidoglycan recognition receptor *PGRPLC* that triggers activation of the antibacterial IMD/REL2 pathway and the epidermal growth factor receptor *EGFR*. Silencing of three genes encoding type III fibronectin domain proteins (*FN3Ds*) increased the *Serratia* load and altered the gut microbiota composition in favor of *Enterobacteriaceae*. These data suggest that natural genetic variation in immune-related genes can shape the bacterial population structure of the mosquito gut with high specificity. Importantly, *FN3D2* encodes a homolog of the hypervariable pattern recognition receptor Dscam, suggesting that pathogen-specific recognition may involve a broader family of immune factors. Additionally, we showed that silencing the gene encoding the gustatory receptor Gr9 that is also associated with the *Serratia* infection phenotype drastically increased *Serratia* levels. The Gr9 antibacterial activity appears to be related to mosquito feeding behavior and to mostly rely on changes of neuropeptide F expression, together suggesting a behavioral immune response following *Serratia* infection. Our findings reveal that the mosquito response to oral *Serratia* infection comprises both an epithelial and a behavioral immune component.

Citation: Stathopoulos S, Neafsey DE, Lawniczak MKN, Muskavitch MAT, Christophides GK (2014) Genetic Dissection of *Anopheles gambiae* Gut Epithelial Responses to *Serratia marcescens*. PLoS Pathog 10(3): e1003897. doi:10.1371/journal.ppat.1003897

Editor: David S. Schneider, Stanford University, United States of America

Received June 13, 2013; **Accepted** December 9, 2013; **Published** March 6, 2014

Funding: SS was supported by a Wellcome Trust PhD studentship (086723/Z/08/A). The work was additionally supported by the BBSRC grants BB/K009338/1 and BB/E002641/1. The generation of the DNA microarrays used here was financed by the EU-funded TransMalariaBloc project (HEALTH-F3-2008-223736). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: g.christophides@imperial.ac.uk

Introduction

Genetic variation within populations of the *An. gambiae* mosquito, especially with regard to genes encoding immune factors, is believed to play an important role in the mosquito susceptibility to infection by the malaria parasite *Plasmodium falciparum* [1–3]. Many immune factors exhibit both anti-*Plasmodium* and antibacterial activities, such as those involved in the IMD/REL2 pathway, which is triggered by bacteria through the peptidoglycan recognition receptor PGRPLC [4,5]. Bacterial infections can affect mosquito survival [6] and are thought to constitute a major evolutionary drive [7] as opposed to *Plasmodium* infections the impact of which on mosquito fitness is unclear [8]. An example is the segregation of *TEP1* alleles between the M and S molecular forms of *An. gambiae* in west Africa, which differentially affect *Plasmodium* infections, and is thought to be largely driven by bacterial pathogen pressure in larval habitats [2]. Therefore, genetic associations related to the outcome of bacterial infections may, directly or indirectly, influence mosquito vectorial capacity.

The adult mosquito gut harbors a wide spectrum of bacterial populations, mainly Gram-negative enterobacteria [9–11]. The broad variation in gut microbiota composition observed both at the individual and population levels is probably the result of an interplay between the environmental bacterial diversity and the mosquito genetic makeup [12–14]. Moreover, a precipitous bacterial increase after a blood meal, whose peak coincides with midgut invasion by *Plasmodium* [15], can affect the *Plasmodium* infection load both indirectly, by triggering PGRPLC-mediated mosquito immune responses [5,16] or through generation of immune memory [17], and directly, through the generation of reactive oxygen species by specific enterobacteria that compromise malaria parasites [18].

Epithelial responses against Gram-negative bacteria have been extensively studied in *Drosophila* [19]. They involve recognition of peptidoglycan [20,21] that triggers a finely-tuned immune response mainly through the Imd pathway, resulting in the expression of antimicrobial peptides that limit bacterial populations [22,23]. Production of reactive oxygen species, which target bacteria, through the Dual Oxidase (DUOX) pathway, has also been reported [24]. Gut stem cell proliferation and epithelial cell renewal following tissue damage due to bacterial infection are regulated by the EGFR and JAK/STAT pathways [25–27]. However, the mechanisms involved in achieving gut homeostasis

Author Summary

In malaria vector mosquitoes, the presence of bacteria and malaria parasites is tightly linked. Bacteria that are part of the mosquito gut ecosystem are critical modulators of the immune response elicited during infection with malaria parasites. Furthermore, responses against oral bacterial infections can affect malaria parasites. Here, we combined mosquito gut infections with the enterobacterium *Serratia marcescens* with genome-wide discovery and phenotypic analysis of genes involved in antibacterial responses to characterize molecular processes that control gut bacterial infections thus possibly affecting the mosquito susceptibility to infection by malaria parasites. Our data reveal complex genetic networks controlling the gut bacterial infection load and ecosystem homeostasis. These networks appear to exhibit much higher specificity toward specific classes of bacteria than previously thought and include behavioral response circuits involved in antibacterial immunity.

remain poorly understood. It has been suggested that regulation of Imd responses can influence the microbiota composition in *Drosophila* [28]. Further discrimination between commensal and pathogenic bacteria can be provided by recognition of pathogen-derived uracil, most likely by unidentified G protein-coupled receptors (GPCRs), which triggers the DUOX pathway [29], but the possibility of more specific responses that shape the gut microbiota remains open.

One unexplored aspect of antibacterial immunity is the behavioral immune responses that limit or disrupt the intake of pathogens, thus making an infection more controllable by the immune system. Feeding behavior in *Drosophila* is known to be finely regulated through an interplay between allatostatin A and neuropeptide F (NPF) [30] while feeding suppression is shown to occur following an immune challenge [31–33]. Gustatory receptors are shown to modulate feeding behavior by acting as nutrient sensors [34] and may also be involved in aversion circuits [35] or antibacterial responses through recognition of bacterial-derived metabolites as in mammalian chemoattractant receptors [36].

Here we set out to examine the genetic basis of bacterial infection in the mosquito gut using *An. gambiae* infections with the Gram-negative enterobacterium *Serratia marcescens* that is prevalent in both lab-reared and field collected mosquitoes and is shown to affect the *Plasmodium* infection load [10,12,37]. To achieve this, we used an Affymetrix 400 k SNP genotyping array to identify genetic variation associated with the outcome of oral *S. marcescens* infection in a recently established M form *An. gambiae* colony. The results identify 138 genes associated with the outcome of infection, including the gene encoding the major IMD/REL2 receptor PGRPLC and the epidermal growth factor receptor EGFR, and further suggest that epithelial immune responses against gut bacteria are more complex than previously thought. We identify a set of three type III fibronectins that modulate homeostasis of the gut microbiota with specificity mainly against *Enterobacteriaceae*. We also present evidence that behavioral responses following *S. marcescens* infection can modulate the bacterial load. These data could be further exploited in mosquito microbiota-based interventions aiming to limit malaria transmission.

Results

S. marcescens infection of the mosquito gut

An. gambiae female adults were treated with antibiotics to reduce their natural gut microbiota load (Figure S1) and subsequently fed with fluorescently labeled *S. marcescens* (Db11-GFP) added to the sugar meal. The bacterial levels in the gut of sugar-fed mosquitoes (henceforth referred to as infection) were monitored from day 2 to 6 post infection and showed considerable variation including highly and lowly infected mosquitoes as well as mosquitoes that despite ingesting bacteria-containing sugar showed no sign of fluorescence in their gut (Figure 1A). While the proportion of lowly infected mosquitoes remained rather constant at approximately 50% throughout the course of the experiment, the relative proportions of highly and non-infected mosquitoes changed between days 2 and 3 in favor of highly infected mosquitoes and remained stable thereafter until day 5 (Figure 1B). At day 6, highly infected mosquitoes decreased by ca. 15% with a parallel increase of non-infected mosquitoes.

Identification of SNP divergence associated with the outcome of *S. marcescens* infection

To investigate whether genetic variation could partly explain the observed *S. marcescens* infection phenotype, single nucleotide polymorphism (SNP) divergence between the highly and non-infected phenotypic pools was interrogated using a 400 k SNP genotyping array. Mosquitoes were orally infected with *S. marcescens*, and gut infection levels were determined at day 5 post infection. The results were similar to those obtained in the previous replicate experiments: 38.4% of mosquitoes could be classified as highly infected, 48.6% lowly infected and 13% non-infected (Figure 1C). Pools of equimolar amounts of genomic DNA (gDNA) prepared from carcasses of 15 highly infected and 15 non-infected mosquitoes out of 139 and 47 mosquitoes in each phenotypic group, respectively, were hybridized onto two Affymetrix SNP genotyping arrays. These SNP chips interrogate genetic variation at ~400,000 variable positions in the *An. gambiae* genome (Table S1) [38], and were previously shown to provide useful quantitative information regarding divergence between pooled mosquito samples [39].

Allele calls for each SNP locus were used to determine the minor allele frequency (MAF) differences between highly and non-infected gDNA pools. Two approaches were used to assess genotypic association with the *S. marcescens* infection phenotype. The first included MAF difference at a SNP locus between highly infected and non-infected pools >0.5, suggesting a preponderance of different genotypes between the two pools for the respective locus. The second involved a permutation analysis in which the average MAF difference of 10 adjacent SNP loci (SNPs) was compared with that of 10 random SNPs. Statistical significance was assessed for each of the ~40,000 non-overlapping 10-SNP windows (Table S2) and those showing a p-value$<10^{-5}$, following a Bonferroni correction for the number of tests conducted, were considered as being associated with the *S. marcescens* infection phenotype.

The two approaches detected 140 SNPs with MAF difference >0.5 and 44 10-SNP windows with significant p-values, respectively. As shown in Figure 2, these SNPs and 10-SNP windows together formed distinctive clusters along the *An. gambiae* genome that were designated as peaks so that they are discerned from each other, although assessed association was limited to genes within a 5 kb radius of highlighted SNPs or within genomic areas delineated by significant 10-SNP windows. Overall, 118 genes were found to reside within a 5 kb radius of highlighted SNPs (Table S3), while 27 genes fell within significant 10-SNP windows (Table S4). The two approaches combined detected 138 genes (Table S5), as there was an overlap of 7 genes between the two sets, including the highly relevant *CLIPE6* and *EGFR* as discussed below.

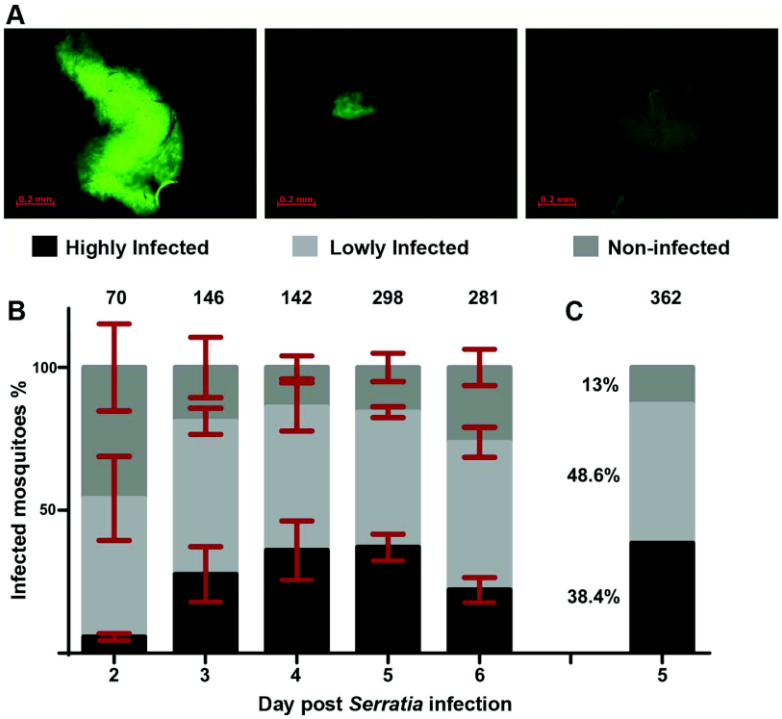


Figure 1. Gut infection with *S. marcescens* varies between individual *An. gambiae* mosquitoes. Mosquitoes were antibiotic treated for 5 days and subsequently fed on sugar containing the Db11-GFP strain of *S. marcescens*. Bacteria-fed mosquitoes were selected 2 days post infection and the prevalence of fluorescent bacteria in their gut was monitored from day 2 to 6 post infection. **1A:** The level of *S. marcescens* infection in the mosquito gut showed considerable variation: mosquitoes with intense fluorescence in most of the gut were characterized as highly infected (left panel), mosquitoes in which fluorescence was evident but confined to a part of the gut were characterized as lowly infected (middle panel) and mosquitoes with no sign of fluorescence were characterized as non-infected (right panel). **1B:** *S. marcescens* infected mosquitoes were dissected each day, from day 2 to 6 post infection, and the proportions of highly, lowly and non-infected mosquitoes were determined over 4 independent infections. The average percentage ±SEM for each level of infection is indicated for each day post infection, with the total number of mosquitoes dissected each day in all 4 infections shown over each bar. **1C:** In 2 independent infections used for SNP genotyping, mosquitoes were dissected 5 days post infection and the percentage of highly, lowly and non-infected mosquitoes, pooled from both infections, can be seen beside the respective part of the bar representing each level of infection.
doi:10.1371/journal.ppat.1003897.g001

In peak 2L-5 (chromosome 2L, peak 5), the gene encoding PGRPLC is found within a 5 kb radius of a highlighted SNP. PGRPLC recognizes peptidoglycan and activates the IMD/REL2 NF-kappaB signaling pathway, thus eliciting antibacterial responses [21,40,41]. This pathway is constitutively triggered by mosquito gut bacteria maintaining an elevated level of antimicrobial peptide production [5,42]. The association of *PGRPLC* with the *S. marcescens* infection phenotype suggests that genetic variation within the mosquito population may influence the ability to mount an antibacterial response via the IMD/REL2 pathway. Adjacent to *PGRPLC*, in peak 2L-5, is another peptidoglycan recognition protein encoding gene, *PGRPLA*.

Of the remaining genes, several exhibit homologies suggesting involvement in antibacterial immune responses, especially in recognition of pathogen or host derived signals as well as in signal transduction and regulation of immune responses (Table 1). The permutation analysis revealed 3 genes, out of a total of 27, encoding proteins with type III fibronectin domains (FN3D) in different peaks: *FN3D1* in peak 2L-4, which was also in the proximity of a highlighted SNP, *FN3D2* in 2L-14 and *FN3D3* in 2R-4. A total of 65 *An. gambiae* genes contain FN3 domains, including the hypervariable pattern recognition receptor *AgDscam*, the insulin receptor *INR* and the JAK/STAT receptor *DOME*. *FN3D2* and *FN3D3* additionally possess immunoglobulin and putative transmembrane domains, while *FN3D2* is an ortholog of *Drosophila Dscam4*. *Drosophila* Dscam is shown to bind bacteria and influence the efficiency of phagocytosis [43], while its *An. gambiae* ortholog, AgDscam, is also shown to bind bacteria and mediate antibacterial and anti-*Plasmodium* responses [44]. Importantly, *Dscam* genes in various organisms generate a diverse repertoire of isoforms, suggestive of challenge-specific pattern recognition through alternative splicing [43,45,46], with particular *AgDscam* isoforms specifically targeting *P. berghei*, *P. falciparum* or commensal bacteria [47,48].

Several putative transcription factors with homeobox-like or DNA-binding domains were found in the identified peaks. AGAP005096 in 2L-4 and AGAP005244 in 2L-5 (together with *PGRPLC* and *PGRPLA*) encode homeodomains. The homeobox gene, *Caudal*, has been previously implicated in the regulation of epithelial immune responses and shown to influence the gut bacterial population structure in *Drosophila* [28], while its mosquito homolog has been shown to regulate the IMD/REL2 pathway

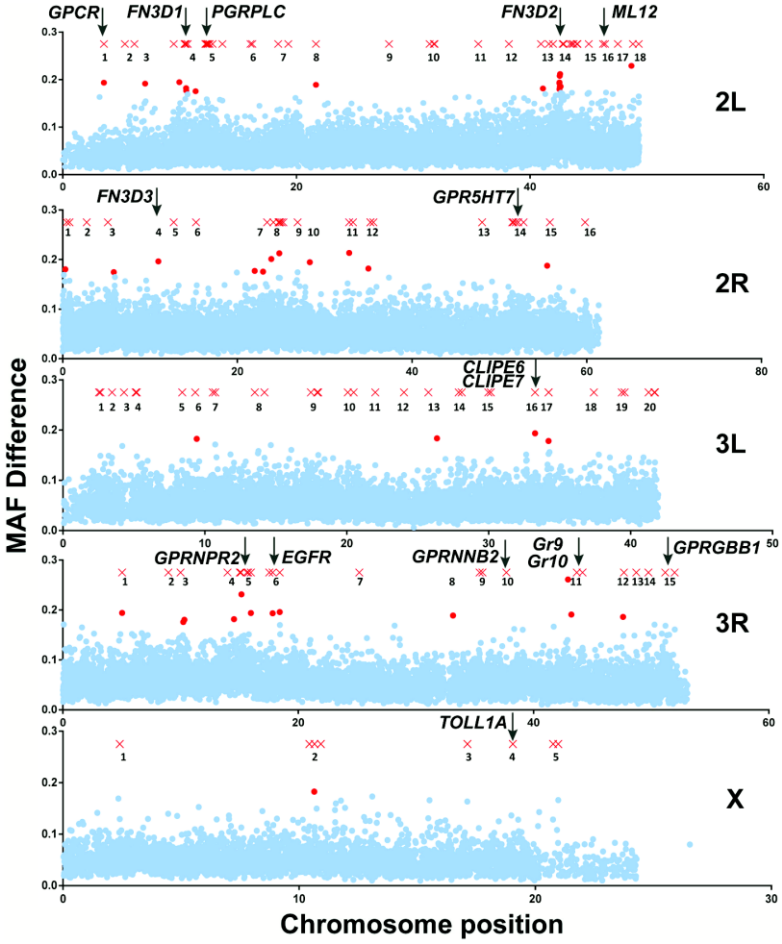


Figure 2. Mapping of *An. gambiae* genetic variation associated with the *S. marcescens* infection phenotype. SNPs with MAF difference >0.5 and 10-SNP windows with Bonferroni-corrected significance (p-value<10^{-5}) are shown in their respective chromosomal position as red X crosses and dots, respectively. Non-significant 10-SNP windows are shown as blue dots. Genomic areas with highlighted SNPs and/or significant 10-SNP windows in close proximity are referred to as peaks and are numbered. Each peak is referred to using the chromosomal arms it resides on and its respective assigned number. The genomic positions of genes of interest found within a 5 kb radius of highlighted SNPs or within genomic areas delineated by 10-SNP windows with a significant p-value are indicated by vertical arrows.
doi:10.1371/journal.ppat.1003897.g002

[49]. Thus, these putative transcription factors could play similar regulatory roles. AGAP002492 in peak 2R-7 encodes a DNA-binding domain, while its *Drosophila* ortholog *ewg* is involved in the Wnt/Wingless pathway [50]. AGAP005156, in peak 2L-4 encodes an ARID/BRIGHT DNA-binding domain, with its *Drosophila* ortholog, *retained*, is involved in behavioral modulations and repression of male courtship [51,52]. AGAP005661, in peak 2L-7, a putative ligand-regulated transcription factor, is an ortholog of the *Drosophila* nuclear receptor *FTZ-F1*, involved in juvenile hormone mediated gene expression [53].

Genes encoding alpha-glucosidase and alpha-mannosidase homologs were detected in peaks 2R-1 and 2R-13, respectively. These genes possess glycoside hydrolase domains that are also present in the conserved chitinase gene family [54], involved in bacterial clearance and host tolerance [55].

The gene encoding the epidermal growth factor receptor, EGFR, was identified in the prominent peak 3R-6 both by both the permutation and the individual SNP analysis. The *Drosophila* EGFR pathway has been implicated in gut remodeling following oral bacterial infection [27], suggesting that the EGFR pathway may influence the outcome of *S. marcescens* infection in *Anopheles*, possibly through synergistic functions in gut homeostasis.

CLIPE6 and CLIPE7, found in peak 3L-16, belong to the non-catalytic E sub-family of CLIP-type serine proteases, a family known to participate in proteolytic cascades in antibacterial and anti-*Plasmodium* responses [56,57], with SPCLIP1, another E sub-family member, involved in anti-*Plasmodium* responses by regulating complement recruitment [58,59]. Several leucine-rich repeat containing genes were also detected, including *LRIM15* (peak 2L-13), a transmembrane member of the LRIM family of immune

Table 1. Genes of interest associated with the *S. marcescens* infection phenotype.

Gene ID	Name/Description	SNP/Permutation analysis	Peak (Figure 2)
AGAP001111	alpha-glucosidase	Permutation	2R-1
AGAP004032	alpha-mannosidase	SNP	2R-13
AGAP011785	CLIPE6	SNP, Permutation	3L-16
AGAP011786	*CLIPE7*	SNP, Permutation	3L-16
AGAP002492	DNA-binding	Permutation	2R-7
AGAP005661	DNA-binding	SNP	2L-7
AGAP005156	DNA-binding	SNP	2L-4
AGAP008819	*EGFR*	SNP, Permutation	3R-6
AGAP005147	*FN3D1*	SNP, Permutation	2L-4
AGAP007092	*FN3D2*	Permutation	2L-14
AGAP001824	*FN3D3*	Permutation	2R-4
AGAP011277	*FREP6*	SNP	3L-10
AGAP011278	*GALE4*	SNP	3L-10
AGAP004223	*GPR5HT7*	SNP	2R-14
AGAP010281	*GPRGBB1*	SNP	3R-15
AGAP008702	*GPRNPR2*	SNP	3R-5
AGAP009804	*Gr10*	SNP	3R-11
AGAP009805	*Gr9*	SNP	3R-11
AGAP005096	Homeobox	Permutation	2L-4
AGAP005244	Homeodomain	SNP	2L-5
AGAP007291	*IAP4*	SNP	2L-15
AGAP007292	*IAP5*	SNP	2L-15
AGAP007045	*LRIM15*	SNP	2L-13
AGAP007415	*ML12*	SNP	2L-16
AGAP005205	*PGRPLA*	SNP	2L-5
AGAP005203	*PGRPLC*	SNP	2L-5
AGAP012252	*Protein C kinase 53E*	SNP	3L-19
AGAP004375	Ricin B lectin	Permutation	2R-15
AGAP001004	*TOLL1A*	SNP	X-4
AGAP001002	Toll-like receptor	SNP	X-4

The Gene ID is shown along with its assigned name, if any, or a homology description (Name/Description column). The SNP/Permutation analysis column indicates whether association is based on the presence of the gene within a 5 kb radius of a SNP with MAF difference >0.5 (SNP) or within a significant 10-SNP window (Permutation). The peak each gene is found corresponds to the designation shown in Figure 2.

doi:10.1371/journal.ppat.1003897.t001

proteins [60]. LRIMs have also been implicated in complement anti-*Plasmodium* responses [61–63].

Two Toll-like receptors, *TOLL1A* and a previously uncharacterized paralog of *TOLL5B*, were found in peak X-4. Little is known about the role of Toll-like receptors in *Anopheles* immunity, however, cross-talk between the REL1 and REL2 signaling pathways in the yellow fever mosquito *Aedes aegypti* [64] and synergistic interactions between the Toll and Imd pathway in *Drosophila* [65], leave open the possibility for involvement of Toll-like receptors in defenses against Gram-negative bacteria, also in *Anopheles* [66].

A gene encoding a protein with a ricin B lectin domain was found in peak 2R-15. Lectins bind oligosaccharides and have been shown to modulate mosquito immune responses [6,63], while mammalian lectins modulate host and gut microbiota interactions [67]. Genes belonging to other families of putative pattern recognition receptors were also found to be associated with the *S. marcescens* infection phenotype, including a fibrinogen-related protein (FBN or FREP) and a galectin in peak 3L-10 and an MD2-like receptor in 2L-16 [59,68,69].

Five annotated or putative GPCRs were found to be associated with the *S. marcescens* infection phenotype, including three putative neurotransmitter-triggered receptors: the serotonin receptor *GPR5HT7* in peak 2R-14, the GABA-B family receptor *GPRGBB1* in peak 3R-15 and the neuropeptide receptor *GPRNPR2* in 3R-5. GPCRs have been previously implicated in modulation of *P. falciparum* infection in *An. gambiae* [70], but the mechanism by which this is accomplished remains unclear. NPR-1, a neurotransmitter-triggered GPCR of *Caenorhabditis elegans*, has been shown to modulate antibacterial defenses in a behavior dependent or independent manner, and *NPR-1* genetic polymorphisms are suggested to be major determinants of bacterial susceptibility [71,72]. Serotonin is a major modulator of mammalian intestinal inflammation [73,74], in an interplay between the nervous and immune system [75]. The *Drosophila* ortholog of *GPR5HT7* is involved in various behavioral processes [76,77], including

aggressive behavior, a process also modulated by NPF [78]. Interestingly, the *Drosophila* ortholog of *GPRGBB1* has been implicated in behavioral responses to alcohol sensitivity [79], a process in which NPF is also a major modulator [80,81].

Two gustatory receptor genes, *Gr9* and *Gr10*, encoding 7-transmembrane chemoreceptor domains, were associated with the outcome of *S. marcescens* infection (Figure 2, peak 3R-11). *Gr9* and *Gr10* are paralogs and show co-orthologous relationships with the *Drosophila Gr32a*, *Gr39a* and *Gr68a* [82]. Gr32a and Gr68a act as pheromone receptors in modulating mating behavior [83,84], while Gr39a has been implicated, through 4 splice variants, in sustaining courtship behavior [85]. Gr32a is also involved in regulating aggressive behavior through recognition of small non-volatile hydrocarbons [86], or feeding suppression triggered by DEET or other antifeedants [87].

Gustatory receptor family members have also been implicated in aversive taste [35,88], CO_2 responses [89,90] and sugar recognition [91–94]. A *Drosophila* gustatory receptor, Gr43a, has been shown to recognize fructose and act as a nutrient sensor, promoting or suppressing feeding [34]. Since enhanced or suppressed feeding of bacteria-containing sugar can decisively influence the abundance of *S. marcescens* that the mosquito takes in and its immune system can handle, it is possible that *Gr9* or *Gr10* variants linked to altered mosquito feeding behavior can affect the outcome of infection. Furthermore, GPR43, a mammalian chemoattractant receptor, has been shown to recognize short-chain fatty acids of bacterial origin and participate in antibacterial responses [36], while other mammalian chemoattractant receptors regulate inflammatory responses by recognizing endogenous factors [95]. Recognition of bacterial-derived uracil has recently been shown to modulate *Drosophila* antibacterial responses through the DUOX pathway [29]. Therefore, another possibility is that *Gr9* or *Gr10* recognize bacterial-derived metabolites or infection-induced mosquito molecules and mediate antibacterial responses.

Several other genes with no known or unrelated to immune responses homologies were also associated with the *S. marcescens* infection phenotype such as AGAP013684 in peak 2R-8, encoding a putative miRNA. MiRNAs are known to modulate gene regulation in processes that include epithelial immunity [96,97]. AGAP006405 in peak 2L-10 encodes a tyrosine protein kinase, while its *Drosophila* ortholog, *derailed2*, is involved in Wnt5 signaling and establishment of olfactory circuits [98]. In peak 2L-15 the inhibitors of apoptosis *IAP4* and *IAP5* were found. The *Drosophila* IAP2 is known to regulate Imd signaling [99], suggesting that the *An. gambiae* IAP4 or IAP5 may also play similar roles. AGAP012252, in peak 3L-19, encodes the ortholog of *Drosophila* PKC53E, implicated in NPF-mediated alcohol sensitivity [100,101]. AGAP011363, in peak 3L-11, encodes the ortholog of *Drosophila* rab6, implicated in phagocytosis [102] but also trafficking of Grk, the EGFR ligand [103,104]. AGAP010503, in peak 3L-4, encodes the ortholog of the *Drosophila* SK channel, implicated in behavioral courtship memory [105]. AGAP005216, in peak 2L-5, encodes the ortholog of *Drosophila* fab1, involved in autophagy but also the lysosomal degradation of necrotic, a modulator of the Toll pathway [106–109].

Candidate gene prioritization for further phenotypic analysis was based on homologies with genes known to be involved in species-specific antibacterial responses, e.g. *FN3D2* and *Dscam* [43] or demonstrably regulating the response to gut microbiota in other systems, e.g. *Gr9* and the mammalian chemoattractant receptor *GPR43* [36], with the aim of the identification of novel functions of genes or gene families in antibacterial responses.

Serratia infection phenotypic analysis of *FN3D1-3*

The involvement of the three *FN3D* genes in shaping the outcome of *An. gambiae* gut infection with *S. marcescens* was investigated by RNAi-mediated gene silencing (Figure 3). Antibiotic treated mosquitoes were orally infected with *S. marcescens* following knockdown (kd) of each of the *FN3Ds* (Figure S2). The bacterial load in mosquito guts was determined 5 days post infection by quantitative RT-PCR (qRT-PCR), using both broad range bacterial 16S and *Serratia*-specific primers. Highly significant and robust increase of the *S. marcescens* load was observed after silencing any of the three genes compared to *dsLacZ*-treated controls: 21 to 53-fold in *FN3D1* (Figure 3A), 41 to 60-fold in *FN3D2* (Figure 3B) and 13 to 29-fold in *FN3D3* kd (Figure 3C).

We also assessed the role of *FN3Ds* in shaping the load of *Serratia* naturally found in the mosquito gut. Mosquitoes reared in

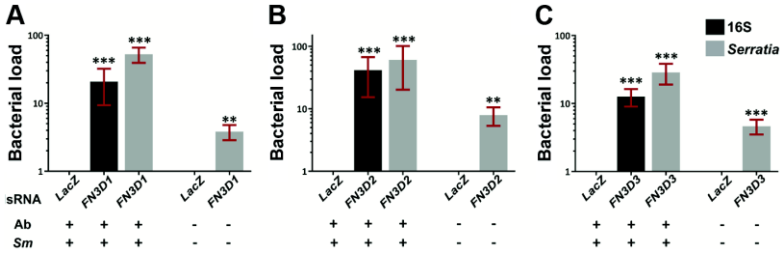


Figure 3. Silencing of *FN3D1–3* increases *Serratia* levels in orally infected mosquitoes or mosquitoes retaining their natural gut microbiota. Antibiotic treated and subsequently orally infected with *S. marcescens* (*Ab+Sm+*) or non-treated mosquitoes retaining their natural midgut microbiota (*Ab−Sm−*), were dsRNA treated to silence *FN3D1* (3A), *FN3D2* (3B) or *FN3D3* (3C) or treated with the *LacZ* dsRNA control. The bacterial load in the midguts of surface sterilized mosquitoes was determined 5 days post *S. marcescens* infection for *Ab+Sm+* mosquitoes, or 5 days post dsRNA treatment for *Ab−Sm−* mosquitoes. Bacterial load was determined using broad range bacterial 16S or *Serratia*-specific primers using qRT-PCR, in which relative to the endogenous *AgS7* control bacterial abundance was determined for each sample and then normalized to the relative abundance of the *dsLacZ* treated control. For *Ab+Sm+* mosquitoes, the average ±SEM of the fold-change in bacterial load is shown as determined over 7 independent infections for *FN3D1* (3A) and *FN3D2* (3B), or 8 independent infections for *FN3D3* (3C), with the qRT-PCR in each infection replicated at least twice. For *Ab−Sm−* mosquitoes, the average ±SEM of the fold-change in bacterial load is shown as determined over 4 independent assays for *FN3D1* and *FN3D3*, or 5 independent assays for *FN3D2*. Asterisks indicate significance in an one-sample t-test against zero using the log2-transformed fold-change values so that zero corresponds to no difference from *dsLacZ* treatment. Two asterisks indicate a p-value<0.005 while three asterisks indicate a p-value<0.0005.
doi:10.1371/journal.ppat.1003897.g003

standard conditions, without antibiotic treatment or infection with *S. marcescens*, were treated with dsRNA against each of *FN3D1–3* and the level of commensal *Serratia* was determined 5 days later (Figure 3 A–C, last bar in each panel). Silencing any of the three genes resulted in a significant 4 to 8-fold increase in the levels of commensal *Serratia* compared to *dsLacZ*-treated controls. These data indicate the involvement of FN3D1–3 in constitutive antibacterial effects that shape the load and composition of the mosquito natural gut microbiota.

FN3D1–3 kd alters the gut microbiota composition in favor of *Enterobacteriaceae*

When the effect of *FN3D1–3* kd was assessed on the total bacterial load in the gut of mosquitoes that retained their natural gut microbiota, a non-uniform effect was observed between 4 independent replicate assays (Figure S3). In some cases, *FN3D* silencing resulted in moderate increases of both *Serratia* and total bacterial load, while in other cases the total bacterial load showed no or marginal increase while *Serratia* showed a strong increase. This variability suggested that the FN3D effect on total bacteria may depend on the initial *Serratia* load and that FN3Ds may function in shaping the population structure of the gut microbiota by affecting a subset of bacteria inhabiting the mosquito gut, including *Serratia*.

To further investigate these hypotheses, we carried out a microbiome analysis using 454 pyrosequencing of samples from two of the replicate assays in which *FN3D1–3* kd increased *Serratia* but not total bacteria abundance (Figure 4A–B) and from a replicate assay in which *FN3D3* kd increased both *Serratia* and total bacterial load (Figure 4C). The resulting sequence reads were assigned to their respective bacterial family. Reads aligning to *Serratia* reference sequences were categorized separately from other *Enterobacteriaceae* (Table S6).

Considerable variation in bacterial composition was observed in control gut pools between the three assays (Figure 4). This variation is consistent with previously reported metagenomic analyses in lab-reared and field-collected mosquitoes, which revealed extensive gut microbiota diversity at both the individual and population levels [12–14]. Total *Enterobacteriaceae* (*Serratia* and other *Enterobacteriaceae*) were highly prevalent in all pools corresponding to 83.2%, 44.2% and 47.5% of total reads, respectively, while significant variation was observed in the specific representation of *Serratia* that corresponded to 1.9%, 24.5% and 9.5% of total sequence reads, respectively. This natural *Serratia* variation is consistent with the variation observed following oral infection with Db11-GFP *S. marcescens* (see Figure 1) and may be related to the underlying genetic variation. *Acetobacteriaceae* was a prominent family in all assays, while *Flavobacteriaceae* was the prevailing family in the second assay.

In the first assay, *FN3D1* or *FN3D2* kd increased the representation of total *Enterobacteriaceae* to 87.6% and 89.6%, respectively (Figure 4A). Remarkably, silencing *FN3D1* or *FN3D2* resulted in a dramatic increase in *Serratia* representation from 1.9% in the *dsLacZ*-treated control to 30% and 39.3% of total sequence reads, respectively, in agreement with the qRT-PCR analysis of the same samples (Figure S3). Similar results were obtained in the second assay whereby silencing *FN3D2* or *FN3D3* resulted in an increase in total *Enterobacteriaceae* representation, from 44.2% to 83.1% and 69.7%, respectively (Figure 4B). In both cases, *Serratia* representation showed a precipitous increase from an initial intermediate level of 24.5%, to almost all *Enterobacteriaceae* sequence reads aligning to *Serratia* reference sequences, again in consistence with the qRT-PCR analysis (Figure S3). Although non-*Enterobacteriaceae* representation decreased in both *FN3D2* and *FN3D3* kd, *Flavobacteriaceae* persisted following *FN3D2* kd but were completely eliminated following *FN3D3* kd, indicating a difference in the effect between the two *FN3Ds* related to non-*Enterobacteriaceae* strains.

Taken together, these data indicate that *FN3Ds* indeed play a major role in shaping the population structure of the mosquito gut microbiota, as silencing any of *FN3D1–3* led to increased *Serratia* abundance but also shifted the composition of the mosquito gut microbiota in favor of *Enterobacteriaceae*, mainly *Serratia* or strains that show similarity to *Serratia* reference sequences. This shift may be a result of a specific FN3D function against *Serratia* or a subset of gut bacteria. Alternatively, bacterial interactions or differential growth potential of different bacterial strains may account for the observed shift following a uniform FN3D antibacterial effect.

We tested this hypothesis by examining whether *FN3D1–3* silencing could affect the levels of gut infection with non-*Enterobacteriaceae*. Antibiotic treated *dsLacZ* treated controls and *FN3D1–3* kd *An. gambiae* mosquitoes were orally infected with bacteria of the genus *Asaia*, a member of the *Acetobacteriaceae* family, common in both field and laboratory-reared *An. gambiae* [9–11] and present in all of our sequenced samples. *FN3D1–3* silencing resulted in moderate, non-significant increases in bacterial load, compared to controls (Figure S4), distinguishably lower than following oral *S. marcescens* infection (Figure 3). These data suggest that the observed FN3D antibacterial effect is not uniform across all Gram-negative bacteria and may be specific to a subset of the gut bacterial population including *Enterobacteriaceae*.

The observed shift in favor of *Enterobacteriaceae* representation when both *Serratia* and total bacterial abundance increased following *FN3D3* kd was also confirmed by microbiome sequencing that showed an increase of *Serratia* from 9.5% to 33.7% of total sequence reads and of total *Enterobacteriaceae* from 47.5% to 66% (Figure 4C). Remarkably, *FN3D3* kd also increased the representation of bacteria of the genus *Burkholderia*, from an initial 0.71% to 15.1% of total reads (Figure 4C). *Burkholderia* were not traced in the *dsLacZ* treated control pool in which the effect of *FN3D3* kd was also assayed (Figure 4B). These data suggest that FN3D3 limits a subset of the mosquito gut bacterial community including *Enterobacteriaceae* but also bacteria of the genus *Burkholderia*.

Gr9 modulates *S. marcescens* infection levels

The genomic area encompassing genes encoding the gustatory receptors Gr9 and Gr10 was associated with the outcome of *S. marcescens* infection. As alternative splicing of *Gr9* has been previously suggested [82], with *Gr9* possessing 13 splice variants compared to one for the adjacent *Gr10*, we considered *Gr9* genetic variation more likely to influence the outcome of *S. marcescens* infection, leading to the observed SNP divergence. *Gr9* has shown significant upregulation compared to other tissues in the midgut of blood-fed adult mosquitoes [110] and also in the midgut of adult mosquito tissues [111]. The *Gr9* midgut expression was also confirmed here (Figure S2). Furthermore, comparison of transcription profiles between antennae or maxillary palps and whole body transcriptomes in female mosquitoes has previously shown a non-significant upregulation of *Gr9* in those two tissues (1.42 for antennae and 1.15 for maxillary palps) [112].

We carried out RNAi-mediated silencing of *Gr9* in adult mosquitoes and examined the outcome of oral *S. marcescens* Db11-GFP infection. *Gr9* knockdown resulted in a precipitous 36 to 48-fold increase in *S. marcescens* levels compared to *dsLacZ* treated controls, as determined using both broad range 16S and *Serratia*-specific primers (Figure 5A). These data suggest that Gr9 exerts an antibacterial effect that influences the outcome of *S. marcescens* infection.

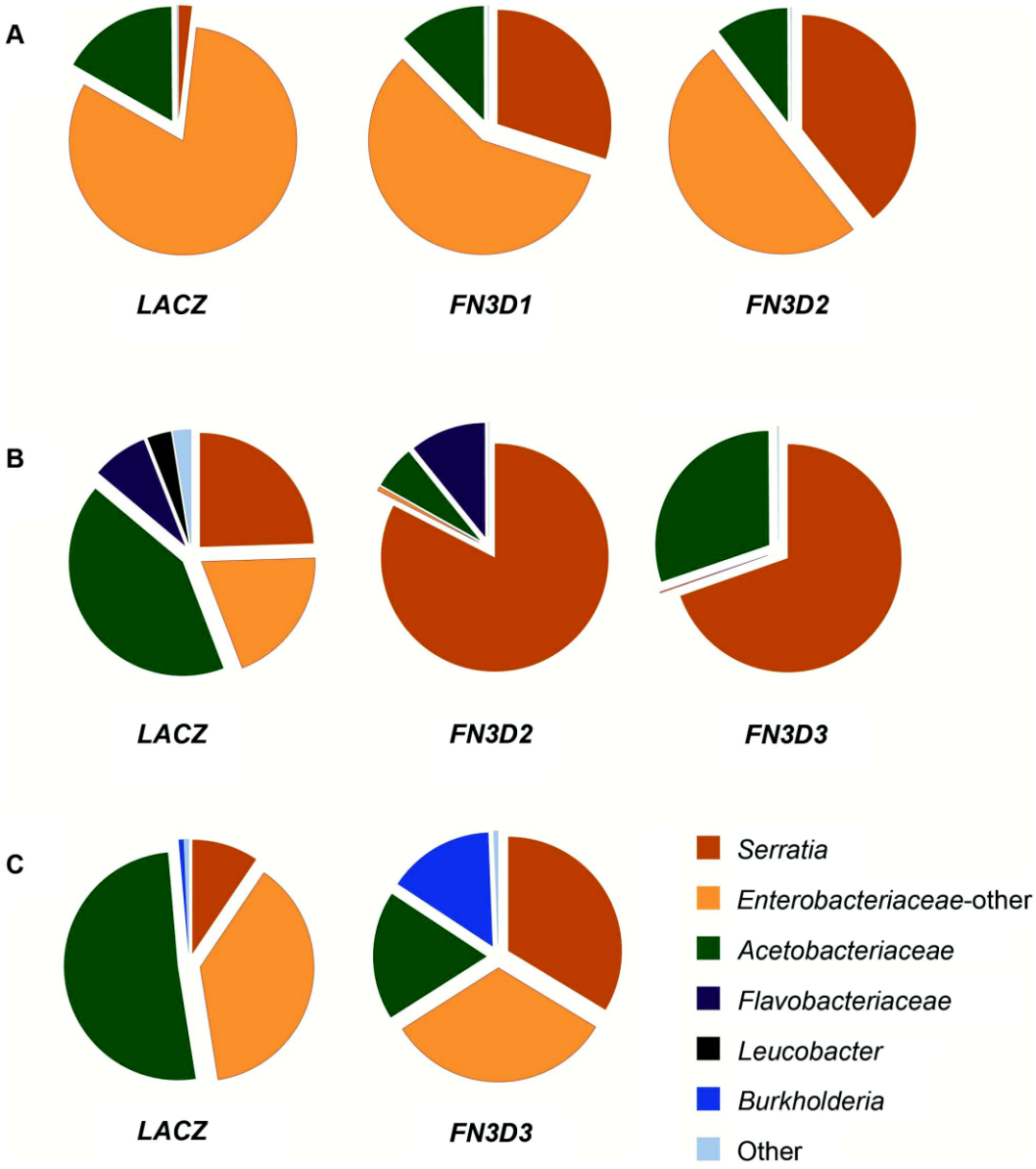


Figure 4. *FN3D1–3* silencing changes the composition of the mosquito gut microbiota in favor of *Enterobacteriaceae*. The 16S V4–V6 hypervariable regions of gut bacterial populations from mosquitoes retaining their natural gut microbiota without antibiotic treatment or *S. marcescens* infection (*Ab−Sm−*, Figure S3) were sequenced using 454 pyrosequencing (Table S6). cDNA pools from guts of *FN3D1–3* dsRNA treated mosquitoes or *dsLacZ* treated controls, surface sterilized and dissected 5 days post dsRNA treatment, were PCR amplified and sequenced over 3 independent assays (panels A to C). The gut microbiota composition of the *FN3D1–3* dsRNA treated pools or the *dsLacZ*-treated control in each independent assay can be seen in the respective pie charts, with the dsRNA treatment indicated below each pie chart. The color legend indicates the bacterial family corresponding to each pie chart color. Modulation of total bacteria or *Serratia* abundance can be seen for each sequenced pool in Figure S3, with *FN3D1* kd corresponding to replicate 1 in panel 4A, *FN3D2* kd corresponding to replicate 1 in panel 4A and 3 in panel 4B and *FN3D3* kd corresponding to replicate 2 in panel 4B and 4 in panel 4C.
doi:10.1371/journal.ppat.1003897.g004

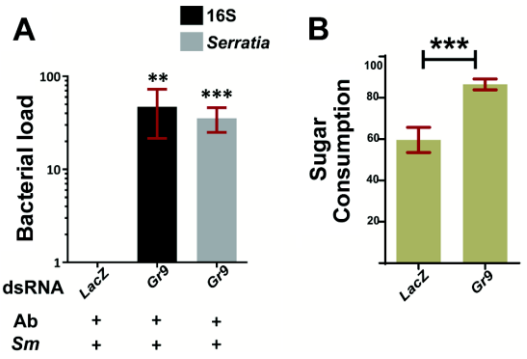


Figure 5. *Gr9* silencing increases *S. marcescens* levels in orally infected mosquitoes. 5A: The bacterial load of antibiotic treated mosquitoes orally infected with *S. marcescens* (*Ab+Sm+*), treated either with *Gr9* dsRNA or the *dsLacZ* control was determined at day 5 post infection either with broad range 16S or *Serratia*-specific primers. The average ±SEM of the bacterial fold-increase is shown, compared to *dsLacZ* treated mosquitoes over 5 independent infections, with the qRT-PCR performed at least twice for each infection. **5B:** Antibiotic treated mosquitoes were starved overnight and then offered a sugar meal through a 5 µl capillary. Sugar meal consumption was determined 16 hours later for 38 *LacZ* and 55 *Gr9* dsRNA treated mosquitoes. The average ±SEM percentage of sugar consumption through the capillary for each mosquito is shown for *LacZ* and *Gr9* dsRNA treatments. In panel 5A, asterisks indicate significance in a one-sample t-test against zero using the log2-transformed fold-change values while in panel 5B asterisks indicate significance in the non-parametric Mann-Whitney test. Two asterisks indicate a p-value<0.005 while three asterisks indicate a p-value<0.0005.
doi:10.1371/journal.ppat.1003897.g005

We next examined the possibility that the Gr9 antibacterial effect is related to changes in mosquito feeding behavior. Based on the *Gr9* many-to-many orthologous relationship with the *Drosophila Gr32a*, *Gr39a* and *Gr68a*, it would be more likely that any Gr9 effects on mosquito behavior would be exerted through the recognition of mosquito-induced or bacterial-derived molecules rather than nutrient sensing [83–85,87].Therefore we first examined the possibility that Gr9 mediates aversion to bacteria-containing sugar thus limiting sugar meal uptake upon oral *S. marcescens* infection. A two-choice preference assay, in which mosquitoes were offered to feed from a capillary that contained *S. marcescens* and another that contained only sugar, indicated that there was no significant difference due to *Gr9* silencing in consumption between the two capillaries (Figure S5).

Another possibility that could explain the Gr9 antibacterial effect is that Gr9 modulates meal size irrespective of the presence of bacteria. Antibiotic treated mosquitoes were starved and then offered a sugar meal. Consumption was determined 16 hours later in *LacZ* or *Gr9* dsRNA treated mosquitoes (Figure 5B). Indeed, *Gr9* silencing resulted in a significant 1.45-fold increase in meal size compared to *dsLacZ* treated mosquitoes. An increased meal size could result in higher *S. marcescens* uptake following oral infection, thus contributing to the precipitous increase in *S. marcescens* load following *Gr9* silencing. Our data suggest that Gr9 influences feeding behavior by triggers that do not rely on the presence of bacteria. As the presence of *S. marcescens* does not seem to affect sugar uptake following *Gr9* silencing, there is no reason to assume that the presence of *S. marcescens* influences the Gr9 effect on meal size, although Gr9-independent aversion circuits could conceivably taper overall consumption.

Transcriptional responses following *S. marcescens* infection

To examine the relationship between genes identified in the population genetics analysis to be associated with the *S. marcescens* infection outcome and infection-induced transcriptional responses, we used DNA microarrays to monitor the transcriptional profile of mosquito guts 3 days post infection with *S. marcescens* added to the sugar meal. Uninfected mosquitoes, which were also treated with antibiotics, were used as controls. Three independent replicate infections were performed. Overall, 55 and 44 transcripts were found to be up and down regulated by at least 1.75-fold, respectively, with 38 and 28 respective up or down regulated transcripts yielding a significant p-value in a t-test against zero, where zero corresponds to no transcriptional regulation (Figure 6A and Table S7). Functional classification of all 97 differentially regulated genes, accounting for multiple transcripts of the same gene, identified serine-type endopeptidases and protein/receptor binding as the most represented classes (Figure 6B). The protein/receptor binding functional class comprised 12 members, including several up or downregulated FREPs, zinc finger containing proteins, PGRPLC and the complement factor regulator LRIM1, which has been previously shown to be regulated by the IMD/REL2 pathway [113]. The oxidoreductase class comprised 7 members, including two P450 cytochromes, possibly involved in detoxification [114], the hydrolase class included a glycoside hydrolase and the nucleotide metabolic process class included 5 heat shock proteins, likely to be involved in stress responses [115]. The antimicrobial peptide *LYSC2*, showing the highest 3.44-fold upregulation of all genes, has been previously shown to be upregulated following a bacterial challenge [116].

A hypergeometric test followed by Benjamini-Hochberg correction was used to determine enriched GO terms in the set of 97 genes. The results identified 16 GO terms that were significantly overrepresented, most of which were related to just two functional classes: serine-type endopeptidases and chitin-binding genes (Figure S6 and Table S8). In total, 16 serine-type endopeptidase genes were differentially regulated, including *CLIPE6*, which was also associated with the outcome of infection, *CLIPB14* that has been implicated in defense against Gram-negative bacteria [57,117], *CLIPB17* and *CLIPB20*. The group of chitin-binding genes comprised 5 members, including the gene encoding the scavenger receptor SCRASP1, previously shown to be upregulated following bacterial infection and bind chitin [117,118] and two downregulated peritrophic matrix components identified by a previous proteomic analysis [119]. Chitin-binding genes are upregulated following oral bacterial infection in *Drosophila* [26], with one member participating in barrier formation that protects against oral *S. marcescens* infection [120], while their suggested role in mosquitoes is recognition of danger signals following tissue remodeling due to a bacterial infection [118].

Several transcriptionally regulated genes suggested a mosquito behavioral response following *S. marcescens* infection (Table 2). Among these genes, *NPF* was downregulated after infection. *NPF* is expressed in the midgut of *Drosophila* [121] and *Aedes aegypti* [122], and has been implicated in modulation of feeding behavior in *Drosophila* [30], aversion to noxious food [123] as well as in alcohol sensitivity [80] and regulation of reward systems [81]. It has been also linked to food signaling by integrating sugar gustatory stimuli [124] and behavioral immune responses against endoparasitoid wasps, by mediating oviposition behavior [125].

Additional behavior-related genes that were transcriptionally regulated following *S. marcescens* infection included the gustatory receptor *Gr13* with two downregulated transcripts, three upregulated

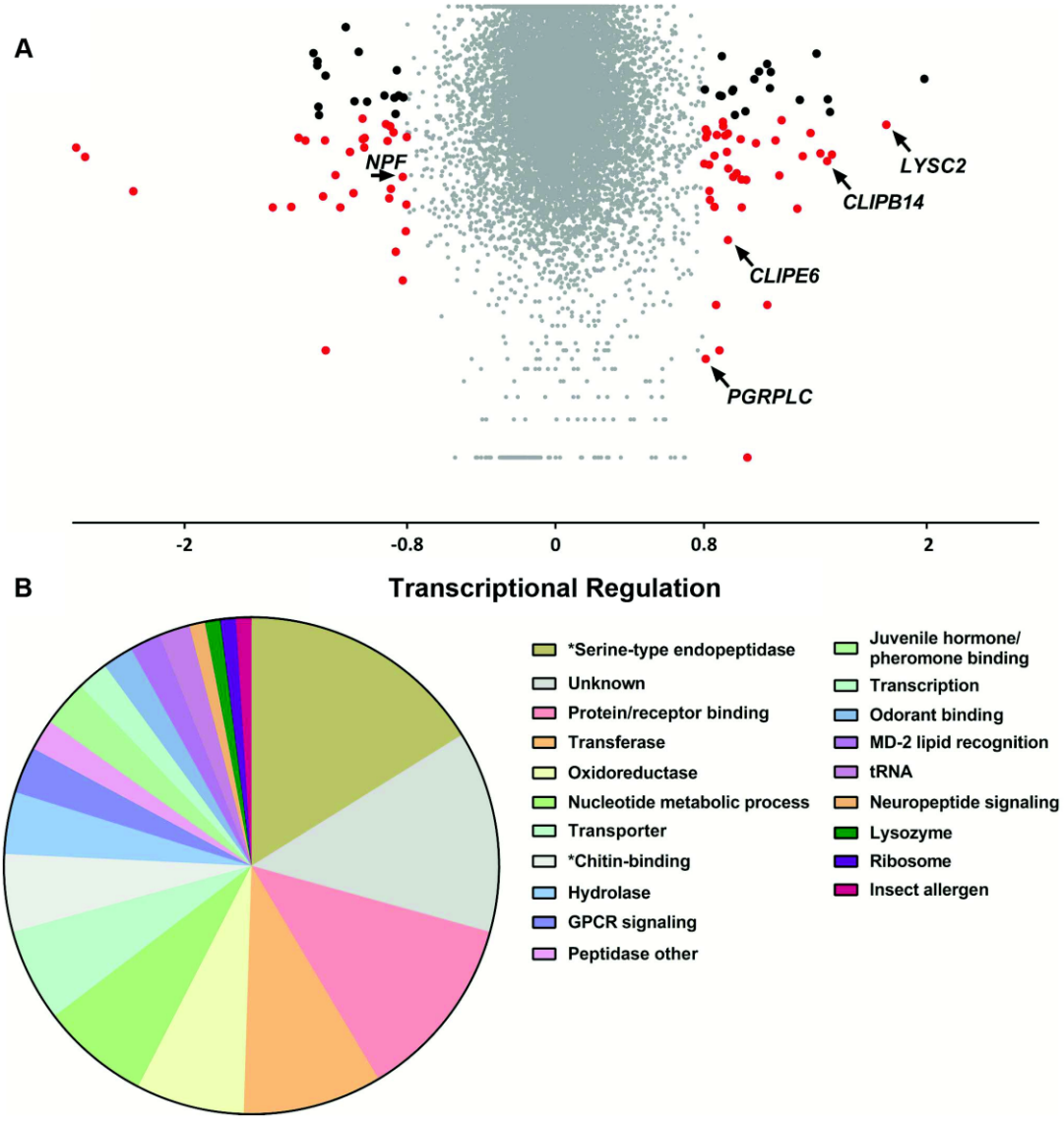


Figure 6. Transcriptional regulation following *S. marcescens* infection using DNA microarrays. Antibiotic treated mosquitoes were orally infected with *S. marcescens* and, 3 days post infection, transcriptional regulation in the gut of bacteria-fed mosquitoes was determined using DNA microarrays, compared to uninfected mosquitoes further antibiotic treated for 3 days. **6A:** Volcano plot of transcriptional regulation as determined over 3 independent infections. The log2-transformed fold-change values for each transcript, as determined by two probes for each of the three arrays, were used for a one-sample t-test against zero, where zero corresponds to no regulation. Transcripts with more than 1.75-fold regulation are indicated either by black dots if the p-value of the t-test is >0.05 or red dots if the p-value is <0.05. Transcripts corresponding to *LYSC2*, *PGRPLC*, *CLIPE6*, *CLIPB14* and *NPF* are indicated by arrows. **6B:** Functional classification of more than 1.75-fold regulated genes. The 97 genes with more than 1.75-fold regulation were assigned to a functional class based on assigned GO terms, *InterPro*-predicted domains or *Drosophila* orthologs. The pie chart shows the proportion of genes assigned to each functional class. Functional classes corresponding to significantly overrepresented GO terms are indicated by asterisks.
doi:10.1371/journal.ppat.1003897.g006

juvenile hormone-inducible kinases, two downregulated genes encoding a pheromone and a juvenile hormone binding protein and the downregulated odorant binding protein genes *OBP13* and *OBP54*. Juvenile hormone circuits are known to affect gustatory perception and feeding behavior in various organisms including *Ae. aegypti* [126–129], while pheromone and olfaction circuits are also

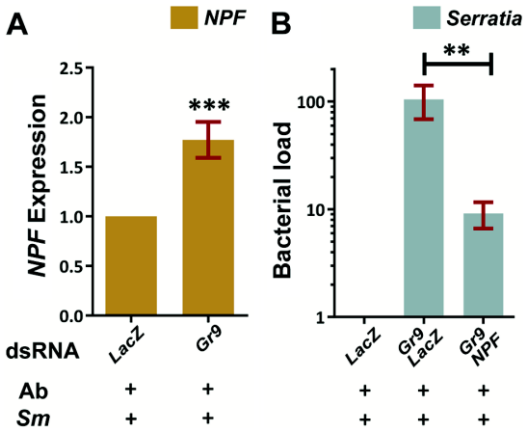


Figure 7. The Gr9 antibacterial effect mostly relies on changes in *NPF* expression. 7A: Antibiotic treated mosquitoes orally infected with *S. marcescens* (*Ab+Sm+*) were dissected 5 days post infection and the *NPF* levels of mosquitoes treated either with *LacZ* or *Gr9* dsRNA were determined in their guts. The average ±SEM of the *NPF* fold-increase as determined over 8 independent infections, with the qRT-PCR performed at least twice for each infection, can be seen. **7B**: The bacterial load of antibiotic treated mosquitoes orally infected with *S. marcescens* (*Ab+Sm+*) treated either with *LacZ* dsRNA or with a 50–50% mix of either *Gr9* and *LacZ* or *Gr9* and *NPF* dsRNA was determined and normalized to the levels of *dsLacZ* treated mosquitoes. The average ±SEM of the bacterial fold-increase as determined over 3 independent infections can be seen, with the qRT-PCR performed at least twice for each infection. In panel 7A asterisks indicate significance in a one-sample t-test against zero using the log2-transformed fold-change values while in panel 7B asterisks indicate significance in the non-parametric Mann-Whitney test. Two asterisks indicate a p-value<0.005 while three asterisks indicate a p-value<0.0005.
doi:10.1371/journal.ppat.1003897.g007

known to affect mosquito behavior [130,131]. The gene encoding the juvenile hormone binding protein TO2 (takeout2), was also upregulated following *S. marcescens* infection and may also participate in behavioral responses, as its *Drosophila* homolog, takeout, is known to regulate feeding behavior [132,133].

Gr9 modulates *S. marcescens* infection via NPF

We examined a possible link between the observed downregulation of the known modulator of feeding behavior, *NPF*, following *S. marcescens* infection and the role of Gr9 in modulating the *S. marcescens* infection outcome. The *Gr9* ortholog *Gr39a* has been previously shown to be expressed in the *Drosophila* midgut and co-localize with NPF in enteroendocrine cells [134], raising the possibility of a functional link between these two genes. In mosquitoes orally infected with *S. marcescens*, *NPF* expression showed a significant 1.8-fold increase following *Gr9* silencing compared to *dsLacZ* treated controls (Figure 7A). This modulation of *NPF* expression following *Gr9* silencing suggested that *NPF* expression may mediate the observed *Gr9* antibacterial effect. Therefore, we further examined whether silencing *NPF* can affect the increase of *S. marcescens* observed in *Gr9* kd mosquitoes. Indeed, concomitant silencing of *Gr9* and *NPF* resulted in a significant 10-fold decrease in infection load, compared to *Gr9* silencing alone (Figure 7B).

Taken together, our data suggest a behavioral immune response involving Gr9, mostly relying on changes of *NPF* expression. One hypothesis is that Gr9 activation in the midgut, most likely through a mosquito-induced cue, tapers the expression of *NPF*, resulting in feeding suppression that limits the mosquito meal size and thus the abundance of ingested *S. marcescens*. *Gr9* variants that influence the efficiency of this suppression may lead to enhanced feeding which, depending on the efficiency of the epithelial response to handle the infection, can influence the outcome of *S. marcescens* infection, thus explaining the observed *Gr9* association with the *S. marcescens* infection phenotype.

Discussion

The rapidly evolving and adapting mosquito species have become tractable systems for genetic association studies that could

Table 2. Transcripts related to mosquito behavior regulated more than 1.75-fold following *S. marcescens* oral infection.

Transcript ID	Name/Description	Functional class	Fold-change	Regulation
AGAP004642-RA	*NPF*	Neuropeptide signaling	1.77	down
AGAP002635-RB	*Gr13*	GPCR signaling	2.68	down
AGAP002635-RA	*Gr13*	GPCR signaling	2.13	down
AGAP008052-RA	Insect pheromone-binding protein	Juvenile hormone/pheromone binding	1.77	down
AGAP008182-RA	Juvenile hormone-binding	Juvenile hormone/pheromone binding	2.06	down
AGAP006080-RA	*OBP54*	Odorant binding	2.38	down
AGAP002905-RA	*OBP13*	Odorant binding	6.00	down
AGAP012703-RA	*TO2*	Juvenile hormone/pheromone binding	1.78	up
AGAP003759-RA	CHK kinase-like juvenile hormone-inducible	Transferase	1.81	up
AGAP003220-RA	CHK kinase-like juvenile hormone-inducible	Transferase	1.94	up
AGAP003762-RA	CHK kinase-like juvenile hormone-inducible	Transferase	2.04	up

The Transcript ID is shown along with its assigned name or homology description, the assigned functional class and the fold-change transcriptional regulation as determined over three independent infections.
doi:10.1371/journal.ppat.1003897.t002

yield important information about vector/parasite interactions leading to malaria transmission [135]. Previous studies have focused on the outcome of *Plasmodium* infections, using laboratory or field mosquitoes and genetic tools such as microsatellite markers and targeted SNP loci genotyping [1,3,136]. These studies have not considered the effect of gut bacteria on the outcome of *Plasmodium* infections, which has been revealed recently [5,16–18]. Furthermore, the influence of associated complement factors on natural *P. falciparum* infections remains questionable [137]. Indeed, the presence of *Enterobacteriaceae*, such as *S. marcescens*, a common member of the mosquito gut flora, has been correlated with *P. falciparum* susceptibility in field mosquito populations [12], while intraspecific variation within *S. marcescens* populations also is shown to affect the *Plasmodium* infection load [37]. Therefore, genome-wide studies to determine factors that modulate the levels of mosquito gut bacteria can provide novel insights into how midgut bacteria affect the outcome of *Plasmodium* infection and hence malaria transmission.

The unprecedented level of detail achieved in the population genetics analysis presented here in identifying SNPs associated with the outcome of *S. marcescens* infection is a result of the strong evolutionary drive exerted by gut bacteria on mosquito genetic variation, the use of a high-resolution SNP genotyping array and the use of a recently established laboratory colony of *An. gambiae* which retains genetic variation found in field populations but also shows elevated linkage disequilibrium due to colonization bottlenecks. This population homogeneity can facilitate gene discovery as shown in human genome-wide association studies in isolated populations [138,139].

A dual implication can be inferred for genes associated with the *S. marcescens* infection phenotype; they are putatively involved in shaping the infection outcome, while their level of involvement may also be affected by genetic variation within the mosquito population. It is possible that identified associations are the result of causal polymorphisms such as gain or loss of function mutations in coding or regulatory sequences or the result of allele combination in several genetic loci which shapes the outcome of infection through synergism, epistatic interactions or redundant function. In any of the latter cases, a reverse genetics approach may not be capable of capturing such interactions.

The involvement of the three *FN3Ds* in the outcome of *Serratia* infection reveals a novel function of this family in modulating the load and composition of the mosquito gut microbiota and opens new avenues in investigating the complexity of such responses and possible synergisms with known antibacterial pathways such as the IMD/REL2. The three *FN3D* genes identified here emerge as major modulators of the bacterial population structure in the mosquito gut, limiting the representation of *Enterobacteriaceae*, mainly *Serratia* or strains with similarity to *Serratia* reference sequences, but also, for *FN3D3*, bacteria of the genus *Burkholderia*. As shifts in gut microbiota population structure can elicit gut pathology [28,140], while *Serratia* can influence the outcome of *Plasmodium* infection [37], FN3Ds can play critical roles in gut homeostatic interactions and malaria transmission dynamics.

Further insights into the FN3D mode of action remain to be determined. Our data showing that the knockdown effects of *FN3Ds* may be limited to *Serratia* or to a fraction of the microbiota raise intriguing questions about the specificity of bacterial recognition in the mosquito gut. The homology of *FN3D2* with the hypervariable pattern recognition receptor *Dscam* opens the possibility that the specific pathogen recognition shown for *AgDscam* [47] concerns a broader family of *FN3Ds*, equipping mosquitoes with the capacity for specific recognition resembling that of animals possessing adaptive immune systems. The phylogenetically unrelated *FN3D2* and *FN3D3* share a similar domain architecture comprising immunoglobulin and FN3 domains, as is the case with *Dscam*. The identification of *FN3D2* and *FN3D3* as being both associated with the outcome of *S. marcescens* infection and exhibiting discrete but similar phenotypic characteristics in modulating the bacterial population structure in the mosquito gut, parallels the discrete but similar functions of the phylogenetically unrelated *Dscam*, *Frazzled* and *Roundabout* in *Drosophila* axon guidance, with all three receptors sharing immunoglobulin and FN3 domains [141–144]. *FN3D1* has a distinct domain architecture with an FN3 domain, while its orthologous relationship with *Drosophila windei* [145] and sequence similarity with the activating transcription factor 7- interacting protein [146], suggest a role in regulating gene expression.

The identification of *An. gambiae* genes involved in immune responses against bacteria and/or *Plasmodium* has been largely based to date on studies that combine bioinformatic identification of known immunity gene homologs and transcriptional profiling of genes following a pathogen challenge. This approach, however, has the limitation of the *a priori* assumption that genes of interest show significant change in transcriptional regulation, mostly induction, which is true for most effectors, but not all genes, for example pattern recognition receptors or transcription factors. In addition, it is possible that even strong changes in transcriptional regulation are the consequence of the infection rather than part of the response. Especially for quantitative traits within mosquito populations, such as *Plasmodium* infection intensity, different infection intensities can correlate with variable transcriptional responses [70], while the underlying genetic variation further complicates the observed transcriptional regulation.

The microarray approach adopted here has identified a limited set of 99 differentially regulated transcripts following oral *S. marcescens* infection. The number of regulated transcripts is consistent with that of a previous microarray-based comparison of antibiotic treated and untreated mosquitoes, which showed differential expression for 185 transcripts [16], attributing this limited transcriptional regulation to symbiotic relationships that have led to adaptation of commensal bacteria. A much broader set of differentially expressed genes has been identified following oral bacterial infections in *Drosophila* [26,32]. This is most likely due to differences in gene pool diversity between the genetically homogeneous fly lines and the recently established mosquito laboratory colony used here, which retains considerable genetic variation thus enabling the SNP genotyping analysis. The different levels of infection seen between mosquitoes (high, low and no infection), which are largely attributed to genetic variation within the colony population, are most likely linked to differences in the mosquito transcription profiles that are averaged out in our study design. Therefore, our analysis identifies transcripts with the most pronounced and consistent differential expression, comprising the core response to *S. marcescens* infection. Future studies investigating the transcription profile of highly, lowly or non-infected mosquitoes are most likely to reveal components of transcriptional regulation that lead to the respective outcome of infection. Indeed, genes identified to show prominent differential expression after bacterial challenge in previous studies also showed transcriptional regulation following oral *S. marcescens* infection, including *CLIPB14* [117], *LRIM1* [63,113], *LYSC2* [116] and *SCRASP1* [117,118].

The identification of diverse transcriptional responses to different bacteria in *Drosophila* [32] along with the specificity of mosquito responses to a subset of bacteria, as suggested by the SNP genotyping analysis presented here, may explain the surprisingly little overlap between differentially expressed genes following *S. marcescens* infection and antibiotic treated vs. untreated

mosquitoes [16]. Remarkably, however, consistency is seen in gene families present in both datasets, including CLIPs, chitin-binding genes, homeobox genes, *PGRPs* and *FREPs*, suggesting that similar defense strategies are employed, which are customized for each type of infection through utilization of different gene family members.

The approach we adopted here to identify genes involved in mosquito gut infection with *S. marcescens* combines transcriptional profiling of infected guts with the identification of SNPs segregating between phenotypic pools, whereby an association implies contribution to the outcome of infection, while the study design incorporates variation that leads to different observed phenotypes. This approach addresses some of the aforementioned shortcomings but introduces others, as it cannot capture genes with redundant functions, genes with additional housekeeping functions or a role during development, of which variants are eliminated from the population, or genes with rare variants that are not in the variation pool of our colony. Furthermore, an association may be the result of a selective sweep in the proximity of the gene that creates linkage disequilibrium and leads to SNP divergence between the phenotypic pools. Therefore, although each of the approaches cannot provide by itself a complete picture, the combination of the two can provide novel insights into the mosquito gut responses to *S. marcescens*.

The comparison between the datasets of transcriptionally regulated genes and genes associated with the outcome of *S. marcescens* infection shows limited overlap, with only *PGRPLC* and *CLIPE6* found in both datasets. Again, considerable overlap is detected in identified gene families, which are represented by different members in each dataset. These include acyl-transferase, glycoside hydrolase, kinase, GPCR, LRIM, homeobox, zinc-finger, PGRP, peptidase, FREP, MD2-like and chitin-binding genes. Interestingly, a previous study investigating differential expression following a bacterial challenge in mosquito immunoglobulin-containing genes failed to identify significant regulation for *FN3D2* or *FN3D3* [147], strengthening the case for the complementarity of the SNP genotyping and expression analysis approaches. The specific role of gene family members, especially those showing considerable expansion in *Anopheles*, e.g. FREPs [68,148], remains unclear. Therefore, SNP genotyping reveals a different set of candidate genes involved in antibacterial immunity while at the same time it is intriguing to postulate whether this divergence between associated and differentially expressed genes within each gene family constitutes a functional divergence between them.

A novel finding stemming from this combinatorial approach is a mosquito behavioral response to *S. marcescens* infection that involves *Gr9* signaling and is mediated by changes of *NPF* expression. Although *Gr9* orthologs in *Drosophila* recognize chemosensory cues and mediate aversive behaviors [83,84,87], surprisingly, Gr9 appears to suppress feeding irrespective of the presence of bacteria. One explanation is that the Gr9 antibacterial effect relies on its expression in the midgut rather than external sensory organs, where the role of its *Drosophila* counterparts has been studied. The role of gustatory receptor midgut expression [134] remains poorly understood and could involve detection of nutrients or host-derived molecules that triggers downstream responses. The role of *NPF* midgut expression [121,122] also remains poorly understood. *NPF* downregulation following *S. marcescens* infection implies its involvement in an aversion circuit triggered by the presence of *S. marcescens*, with a possible NPF role in integrating aversion and satiation signals that lead to feeding suppression. Such NPF involvement remains to be further investigated, in conjunction with the involvement of other genes related to mosquito behavior which were either associated with the outcome of *S. marcescens* infection or were transcriptionally regulated following infection. These include *Gr13*, downregulated following *S. marcescens* infection but also three neurotransmitter-triggered GPCRs, associated with the outcome of infection, pointing to complex behavioral circuits involved in antibacterial responses, which are yet to be revealed.

The identification of *FN3Ds* as well as *Gr9* and *NPF* in responses affecting the outcome of *S. marcescens* infection, in addition to known responses including the IMD/REL2 and DUOX pathways, suggests that the response to gut infection is the result of a complex molecular interplay. Both the SNP genotyping and expression analysis suggest that the mosquito response to oral *S. marcescens* infection involves two discrete but inextricably linked modes of defense, a behavioral and an epithelial immune response. A behavioral immune response involving Gr9 and NPF can limit or disrupt pathogen intake, a defense conceptually similar to barrier responses that inhibit pathogen contact with triggers of epithelial or systemic immune responses. An impaired behavioral response, e.g. due to *Gr9* variants that affect feeding behavior, can decisively influence the efficacy of the epithelial response and thus the infection outcome. This implies a threshold after which epithelial immunity cannot efficiently handle the pathogen load, an aspect of immunity that remains poorly understood. Nevertheless, in mosquito infections with *Plasmodium* parasites, the intensity of infection has been correlated with the efficacy of different components of the IMD/REL2 pathway, suggesting that different effectors may be deployed in low, mid or high intensity parasite infections [4].

As pathogen abundance most likely relies on feeding behavior, the interplay between behavioral and epithelial immunity can shape both responses. Our implementation of a model of natural bacterial infections through the oral route integrated both behavioral and epithelial responses and not only revealed the previously unknown behavioral component but also allowed the study of aspects of epithelial immunity that, by being infection intensity dependent, possibly rely on the behavioral component. This integrative approach to behavioral and epithelial immunity can be further employed to reveal aspects of this interplay that may involve regulation of behavioral responses by host-derived factors induced by the epithelial component. This implies that the study of behavioral immunity alone may be insufficient to uncover some aspects of its biological consequences.

In *Drosophila*, a balance between immune response and tolerance, achieved by various Imd regulators, largely shapes the gut microbiota population structure, although the only known elicitor of such responses is DAP-type peptidoglycan, common to all Gram-negative bacteria [28,149,150]. A similar mechanism has been suggested for mosquitoes through alternative splicing of the modular IMD/REL2 pathway receptor *PGRPLC* that leads to production of positive and negative pathway modulators [5,42]. Indeed, utilization of alternative splicing as a mechanism to derive new immune functions and increase the specificity of pathogen recognition by the mosquito innate immune system has been described for the *FN3D2* homolog, *Dscam* [43,44,47]. Whether the *Enterobacteriaceae*-specific effect of *FN3D2* knockdown is due to specific recognition and activation of highly specialized or targeted effector reactions remains to be investigated. Furthermore, the significance of alternative splicing suggested for *Gr9* [82] remains to be determined along with the cue that triggers its antibacterial effect, and could also involve recognition of, most likely, host-derived signals. In addition, recognition of differentially produced metabolites after infection as shown for the DUOX pathway [29] could further increase the specificity in antibacterial responses.

Whether PAMPs (bacterial-derived) or DAMPs (host-derived), such metabolites can be recognized by gustatory receptors triggering specific antibacterial responses, which together with *FN3Ds* and the rather generalist response of the IMD/REL2 pathway can shape the load and composition of the mosquito gut microbiota. In conclusion, our findings suggest that mosquitoes can mount a much more complex and specific antibacterial response than previously thought, which not only contributes to fending off intestinal bacterial infections but also to achieving homeostasis of the complex gut ecosystem.

Materials and Methods

Ethics statement

This study was carried out in strict accordance with the United Kingdom Animals (Scientific Procedures) Act 1986. The protocols for maintenance of mosquitoes by blood feeding were approved and carried out under the UK Home Office License PPL70/7170. The procedures are of mild to moderate severity and the numbers of animals used are minimized by incorporation of the most economical protocols. Opportunities for reduction, refinement and replacement of animal experiments are constantly monitored and new protocols are implemented following approval by the Imperial College Ethical Review Committee.

Mosquito rearing and maintenance

The *N'gousso* strain of *An. gambiae* was used. This is an M form strain colonized in 2006 [1] and kept in large numbers to retain genetic variation. Rearing and maintenance of the strain was performed as described previously [151]. Mosquitoes were collected after emergence and kept on a cocktail of 25 μg/ml gentamicin, 10 μg/ml penicillin and 10 units/ml streptomycin, diluted in 10% D-(-)-Fructose (Sigma). This antibiotic treatment regime was used for 5 days, with the antibiotic solution refreshed every 24 hours. At day 5 post emergence, the antibiotic solution was replaced by dH_2O and mosquitoes were starved overnight prior to oral bacterial infection.

Mosquito oral infection with *S. marcescens* or *Asaia*

We used the *S. marcescens* Db11-GFP strain, modified to be GFP-fluorescent and resistant to tetracycline and carbenicillin [152]. *S. marcescens* glycerol stock was grown in 5 ml LB cultures containing 50 μg/ml tetracycline and carbenicillin (Sigma) at 37°C. Following overnight incubation, the cultures were expanded to 100 ml and further incubated overnight at 37°C. OD_{600} and GFP fluorescence (excitation/emission at 485/520 nm) were then measured to ensure cultures maintained GFP fluorescence, using the Fluostar Omega spectrophotometer (BMG Labtech). Bacterial pellets following centrifugation at 2500 rpm for 5 minutes were washed twice with PBS and resuspended in such volume of 10% D-(-)-Fructose, so that 1 ml of the bacteria-containing sugar solution corresponded to $OD_{600} = 0.1$ of the initial 100 ml culture. The sugar solution was further diluted 1:12 in a 10% D-(-)-Fructose solution that contained tetracycline and carbenicillin at 50 μg/ml and 5% v/v of scarlet dye (Langdale). Mosquitoes were fed with this solution for 2 days. Subsequently, mosquitoes fed with bacteria-containing sugar were separated based on the presence of the dye in their gut and kept on 10% D-(-)-Fructose containing tetracycline and carbenicillin at 50 μg/ml.

Oral infections with *Asaia* were conducted in a similar manner. The *Asaia* SF2.1 (GFP) strain was used, grown as previously described [153] and maintained in 50 μg/ml kanamycin (Sigma).

Fluorescence microscopy

The levels of *S. marcescens* infection were determined by microscopic observation of dissected midguts immersed in *Vectashield* mounting medium (Vecta), immediately after dissection. The Zeiss Axiophot fluorescence microscope was used, equipped with light and GFP filters while photos of observed midguts were taken with the Axiocam HRc and Axiovision software (Zeiss).

SNP genotyping arrays

All carcasses corresponding to midguts of *S. marcescens* infected mosquitoes were kept numbered in 96-well plates immersed in 75% ethanol at −80°C. Carcasses from selected midguts were used for gDNA extraction using the *QIAquick Blood and Tissue* kit (QIAGEN). Subsequently, gDNA concentrations were determined using the *Picogreen dsDNA* kit (Invitrogen) and equimolar gDNA quantities from each mosquito were pooled. The design and validation of the SNP genotyping array used along with the treatment of gDNA pools, hybridization, calling of SNP genotypes and measurement of differentiation in each pooled hybridization between allele A and B have been described previously [39,154]. The frequency of designated allele A was considered as the minor allele frequency and was used to measure the difference between pooled hybridizations. The permutation analysis used has been described previously [39], with a modified length of non-overlapping 10-SNP windows. Determination of genes residing in identified genomic areas and homology analysis was performed using Biomart 0.7 and the AgamP3.7 *An. gambiae* gene annotation [155]. The SNP genotyping array datasets have been deposited to ArrayExpress under the experiment name *Serratia_SNP1* and accession number E-MEXP-3951.

DNA microarrays

Total RNA was extracted from midguts using the *Trizol* reagent (Invitrogen), and treated with *Turbo DNAse I* (Ambion). Samples were further purified using the *RNeasy* kit (QIAGEN). Quantification was performed using the *Nanodrop 1000* spectrophotometer (Thermo Scientific) and RNA integrity was assessed using the *RNA 6000 Pico Chip* kit (Agilent). Labeling and hybridization were performed using the *Low Input Quick Amp Labeling* kit for two-color microarray based expression analysis (Agilent). We used Agilent custom 4×44 k gene expression microarrays. The microarray design *Pfalcip_Agamb2009* (A-MEXP-2324) comprises oligonucleotide probes encompassing all *An. gambiae* annotated transcripts of the AgamP3.6 release along with *P. falciparum* probes, with each probe represented in duplicate. Slides were scanned using the *Genepix 4000B* scanner equipped with the *Genepix Pro 6.1* software (Axon instruments).

All dataset files were normalized using the *Genespring 11.0 GX* software (Agilent). The Lowess normalization method was used while the threshold of raw signals was set to 5, which was sufficient to eliminate background regulation of *P. falciparum* probes. Further analysis of transcriptionally regulated genes and GO analysis was performed using the Genespring 11.0 GX software. For GO analysis, GO accession numbers for all *An. gambiae* transcripts were obtained using Biomart 0.7 and a hypergeometric test with Benjamini-Hochberg correction was performed on the set of more than 1.75-fold regulated genes. The corrected p-value for testing multiple GO accession numbers for their significance was set to 0.1. The log2-transformed transcriptional regulation for each transcript was extracted from the normalized datasets for each of the two probes corresponding to each transcript and the obtained values from all three independent infections were used in a t-test against zero, with a p-value cut-off of 0.05. The DNA microarray datasets

have been deposited to ArrayExpress under the experiment name *Serratia_infections* and accession number E-MEXP-3952.

Bacterial load following RNAi-mediated silencing

Mosquitoes were treated with the respective dsRNA at the day of emergence, as described previously [156]. For each targeted gene or the *dsLacZ* control, dsRNA was synthesized using the *T7 Megascript* kit (Invitrogen) and further purified using the *RNeasy* kit (QIAGEN) to a concentration of 3 μg/μl. For each *T7 Megascript* reaction, 1 μg of purified PCR product was added, derived using the T7 primer sets shown in Table S9, using *An. gambiae* cDNA as template.

Mosquitoes were surface sterilized by immersing them in 70% ethanol for 30 seconds and washing them twice in PBS and midguts were dissected in *RNA later* (Invitrogen). Total RNA from mosquito midguts was extracted after homogenization with a pestle motor in *RNA later* using the *RNeasy* kit (QIAGEN). cDNA was synthesized from total RNA using the *QuantiTect Reverse Transcription* kit (QIAGEN).

Quantification of bacterial load or the efficiency of RNAi-mediated silencing was performed using qRT-PCR with the respective primers shown in Table S9. In a 20 μl reaction of *Fast SYBR Green Master Mix* (Applied Biosystems), 1 μl of cDNA template and 2 μl of each respective primer at a 0.5 to 9 μM concentration, optimized for each primer set, were added. The 7500 Real-Time PCR System (Applied Biosystems) was used with its respective software to perform the reaction and any further analysis. The relative abundance of each sample was determined using the standard curve method as described in User Bulletin #2 for the ABI Prism 7700 Sequence Detection system (Applied Biosystems) in which the housekeeping *AgS7* gene was used as an endogenous control.

454 pyrosequencing

cDNA pools were amplified with the *GO Taq* DNA polymerase (Promega) using the 16S V4–V6 primers shown in Table S9 and suitable barcode sequences and purified using *PCR purification* and *Gel Extraction* kits (QIAGEN). PCR products were sequenced by Beckman Genomics (Grenoble, France) using the Roche 454 GS FLX+ and standard procedures. The resulting FASTA files were filtered to a minimum read length of 250 bp using *Galaxy* [157] and blasted against the NCBI *16SMicrobial* database using *BLAST+* and *prfectBLAST* [158] with standard *blastn* algorithm settings and 10 maximum target sequences. Further analysis was performed using *MEGAN4* [159].

Meal size and two-choice preference assays

Sugar meal size was determined through a modified capillary feeder assay [160]. Mosquitoes treated with *LacZ* or *Gr9* dsRNA were antibiotic treated for 5 days, starved overnight and, subsequently, individual mosquitoes were fed on a 5 μl glass capillary (VWR) containing 10% D-(-)-Fructose and 5% v/v scarlet dye. For alive mosquitoes, sugar consumption was determined 16 hours later through the reduction of sugar solution in each capillary. The two-choice preference assay was also conducted based on a previously described capillary feeder assay [160]. Mosquitoes treated with *LacZ* or *Gr9* dsRNA were antibiotic treated for 5 days, starved overnight and placed in pools of 8–11 mosquitoes. Mosquitoes were offered to feed from two capillaries, one containing a sugar solution as above and one also containing *S. marcescens*, prepared as described above for oral infection. Water-containing cotton pads were also used and pools with mosquito mortality were disregarded. 16 hours later, consumption for each capillary was determined based on the reduction of the sugar solution.

Supporting Information

Figure S1 Efficacy of antibiotic treatment in reducing the presence of gut bacteria in treated mosquitoes. *An. gambiae* mosquitoes were antibiotic treated for 5 days with a cocktail of gentamicin, penicillin and streptomycin. Subsequently, 10–15 mosquitoes were surface sterilized and their guts were dissected, homogenized and total RNA was extracted and further used for cDNA synthesis. Mosquitoes kept untreated were processed in the same way. Subsequently, cDNA from antibiotic treated or untreated mosquito pools was used in a qRT-PCR using broad range bacterial 16S primers while *AgS7* primers were used as controls. The bacterial load ±SEM in 3 independent assays, with the qRT-PCR performed at least twice for each assay, can be seen. Asterisks indicate significance, with a p-value<0.0005, in a Mann-Whitney non-parametric test between the bacterial load of untreated and antibiotic treated mosquitoes.
(TIF)

Figure S2 Efficacy of RNAi-mediated silencing of *FN3D1–3*, *Gr9* and *NPF*. Mosquitoes were treated with dsRNA targeting the *FN3D1*, *FN3D2*, *FN3D3*, *Gr9* or *NPF* transcripts or with the *dsLacZ* control. The *Gr9* dsRNA targets the 5′ exon 1, which mainly limits the RA and RG splice variants. The relative expression of each transcript in the mosquito gut was determined 5 to 6 days post dsRNA treatment, normalized to the endogenous *AgS7* control, in mosquitoes treated with the respective dsRNA by qRT-PCR and primers targeting the respective transcript in a region not targeted by the respective dsRNA. Silencing efficiency was determined by further normalizing the relative expression of each transcript in mosquitoes treated with the respective dsRNA to the relative expression in the *dsLacZ* treated control. The average ±SEM of relative expression is shown for at least 3 independent assays, with the qRT-PCR performed at least twice for each assay.
(TIF)

Figure S3 *FN3D1–3* silencing modulates total bacteria and *Serratia* in a non-uniform way in mosquitoes retaining their natural gut microbiota. Mosquitoes retaining their natural gut microbiota (*Ab−Sm−*) were treated with *FN3D1* (S3A), *FN3D2* (S3B) or *FN3D3* (S3C) dsRNA and bacterial load was normalized to the respective *dsLacZ* treated control. Bacterial load using broad range 16S or *Serratia*-specific primers is shown for 4 independent assays (Replicates 1–4 as indicated above each pair of bars).
(TIF)

Figure S4 Oral infection with *Asaia* following *FN3D1–3* silencing. Antibiotic treated mosquitoes were orally infected with bacteria of the genus *Asaia* following treatment with *FN3D1–3* dsRNA or the *dsLacZ* control. Bacterial load was determined in the guts of surface sterilized mosquitoes dissected 5 days post infection, using qRT-PCR with 16S broad range bacterial primers and the *AgS7* control. The bacterial load ±SEM in 3 independent assays, with the qRT-PCR reaction performed at least twice for each assay, can be seen.
(TIF)

Figure S5 Two-choice preference assay between sugar solutions containing or not *S. marcescens*. Antibiotic treated mosquitoes treated either with *LacZ* or *Gr9* dsRNA were starved overnight and, subsequently, pools of 8–11 mosquitoes were offered a choice of two meals in separate 5 μl capillaries, one containing a sugar solution containing a dye used to measure

consumption and ensure uptake from the mosquitoes and one also containing *S. marcescens*, at the same concentration as used for oral infection. 16 hours later, consumption was measured for each capillary and, for each mosquito pool, the ratio of the % consumption in the *Serratia*-containing capillary to the % consumption in the sugar-only capillary was determined. Overall, the consumption percentage ratio of *Serratia*-containing versus sugar-only capillaries was determined for 6 *LacZ* and 11 *Gr9* dsRNA treated mosquito pools. The average ±SEM percentage ratio for each dsRNA treatment can be seen. Significant differences were assessed using the non-parametric Mann-Whitney test resulting in a p-value of 0.7395, indicated as non-significant (NS). A percentage ratio of <100% would indicate aversion to the *Serratia*-containing solution while a percentage ratio of >100% would indicate attraction to the *Serratia*-containing solution for mosquitoes treated with the respective dsRNA, as indicated.
(TIF)

Figure S6 Significantly overrepresented GO terms in the set of more than 1.75-fold regulated genes following *S. marcescens* infection. A hypergeometric test with Benjamini-Hochberg correction was used to compare the representation of genes corresponding to the same GO term in the set of 97 more than 1.75-fold regulated genes following *S. marcescens* infection to the respective representation in the *An. gambiae* genome, as annotated in the *Pfalciip_Agamb2009* microarray design. 16 GO terms corresponded to significantly overrepresented groups of genes, shown in orange type in the GO directed acyclic graph, with GO terms in the same path that did not meet the p-value cut-off shown in black type. For significantly overrepresented GO terms, the number of corresponding genes in the set of more than 1.75-fold regulated genes is shown in parenthesis. A table with the regulated transcripts corresponding to GO terms at each final leaf node is shown, including, for each transcript, the Transcript ID, an assigned name or description based on *Interpro*-predicted domains or homologies with *Drosophila* counterparts and the observed transcriptional regulation following *S. marcescens* infection.
(TIF)

Table S1 Complete list of SNP loci assayed for association with the *S. marcescens* infection phenotype. SNP loci interrogated in a 400 k Affymetrix SNP genotyping array. The MAF for the highly infected and non-infected phenotypic pools and the calculated MAF difference can be seen for each interrogated SNP locus.
(RAR)

Table S2 Complete list of 10-SNP loci windows assayed for association with the *S. marcescens* infection phenotype. The 40,007 sliding 10-SNP windows assayed for association with the *S. marcescens* infection phenotype are shown along with the genomic position of each 10-SNP window, the calculated average MAF difference and the assigned p-value of the permutation analysis conducted.
(XLSX)

Table S3 Individual SNP loci associated with the *S. marcescens* infection phenotype. A total of 140 out of 400,071 SNP loci that showed MAF difference >0.5 between the highly infected and non-infected *S. marcescens* phenotypic pools are shown. Each column contains for each SNP locus: the chromosome it resides, the MAF for the highly infected and non-infected phenotypic pools and the calculated MAF difference, the genomic position of the SNP locus and the calculated 5 kb radius (position −5 kb and position +5 kb columns), the residing peak for each SNP locus as assigned in Figure 2 and genes found within a 5 kb radius of this SNP locus based on the AgamP3.7 annotation of the *An. gambiae* genome.
(XLSX)

Table S4 10-SNP loci windows associated with the *S. marcescens* infection phenotype. A total of 44 out of 40,007 10-SNP loci windows showed a significant p-value of $<10^{-5}$ in a permutation analysis following Bonferroni correction, when assayed for association with the *S. marcescens* infection phenotype. Each column contains for each 10-SNP locus window: The chromosome it resides, the genomic position (Center-Start-Stop columns), the average MAF difference calculated, the assigned p-value of the permutation analysis, the peak it resides as assigned in Figure 2 and the genes residing within the 10-SNP locus window based on the AgamP3.7 annotation of the *An. gambiae* genome.
(XLSX)

Table S5 Complete list of *An. gambiae* genes associated with the outcome *of S. marcescens* infection. 138 *An. gambiae* genes were associated with the outcome of *S. marcescens* infection, either by residing within a 5 kb radius of SNP loci with MAF difference >0.5 (SNP) or by residing within 10-SNP loci windows with a $<10^{-5}$ p-value in a permutation analysis (Permutation). For genes identified by both approaches, both SNP and Permutation are noted. For each gene, the Description column includes the assigned name, if any, or a description based on homologs or *Interpro*-predicted domains.
(XLSX)

Table S6 16S metagenomic profiling in *FN3D1–3* or *LacZ* dsRNA treated mosquitoes retaining their natural gut microbiota. cDNA pools from *Ab−Sm−* mosquitoes following *FN3D1–3* or *LacZ* dsRNA treatment were sequenced using 454 pyrosequencing targeting the V4–V6 region of the bacterial 16S rRNA transcript. Results are shown for each cDNA pool corresponding to a dsRNA treatment shown in Figure 4. In each case, taxonomic identification and alignment of reference sequences for each identified taxon is shown along with the number of sequence reads assigned. For each bacterial genus, the reference sequence with most alignment hits (sequence reads) is shown. For *Serratia*, the reference sequence with most alignment hits is highlighted and was used for categorization between *Serratia* and other *Enterobacteriaceae*.
(XLSX)

Table S7 *An. gambiae* transcripts with more than 1.75-fold regulation following *S. marcescens* infection. 99 *An. gambiae* transcripts showing more than 1.75-fold up or down regulation following *S. marcescens* infection. For each transcript, the Transcript ID is shown along with its designated name or functional description, the assigned functional class, the fold-change regulation and the p-values obtained by a t-test against zero using the log2-transformed values for each interrogated probe.
(XLSX)

Table S8 GO terms significantly overrepresented in the set of more than 1.75-fold regulated genes following *S. marcescens* infection. 16 enriched GO terms as determined by a hypergeometric test followed by Benjamini-Hochberg correction using a 0.1 p-value cut-off are shown. The GO ID, accession number and description of each GO term is shown along with the results of the test for each GO term.
(XLSX)

Table S9 Sequences of primers used for qRT-PCR, dsRNA synthesis or 454 pyrosequencing. For each forward

or reverse primer, its use is indicated along with the respective sequence.
(XLSX)

Acknowledgments

We thank Dina Vlachou and Amanda Jackson for help and advice in carrying out the DNA microarray experiments, Katarzyna Sala for helping with mosquito rearing, Dominique Ferrandon for sharing the *S. marcescens* stock and Guido Favia for sharing the *Asaia* stock. The authors are grateful to Fotis C. Kafatos for his multifaceted and invaluable support.

Author Contributions

Conceived and designed the experiments: GKC SS. Performed the experiments: SS. Analyzed the data: SS. Contributed reagents/materials/analysis tools: GKC MKNL DEN MATM. Wrote the paper: SS GKC.

References

1. Harris C, Lambrechts L, Rousset F, Abate L, Nsango SE, et al. (2010) Polymorphisms in Anopheles gambiae immune genes associated with natural resistance to Plasmodium falciparum. PLoS Pathog 6: e1001112.
2. White BJ, Lawniczak MK, Cheng C, Coulibaly MB, Wilson MD, et al. (2011) Adaptive divergence between incipient species of Anopheles gambiae increases resistance to Plasmodium. Proc Natl Acad Sci U S A 108: 244–249.
3. Riehle MM, Markianos K, Niare O, Xu J, Li J, et al. (2006) Natural malaria infection in Anopheles gambiae is regulated by a single genomic control region. Science 312: 577–579.
4. Garver LS, Bahia AC, Das S, Souza-Neto JA, Shiao J, et al. (2012) Anopheles imd pathway factors and effectors in infection intensity-dependent anti-Plasmodium action. PLoS Pathog 8: e1002737.
5. Meister S, Agianian B, Turlure F, Relogio A, Morlais I, et al. (2009) Anopheles gambiae PGRPLC-mediated defense against bacteria modulates infections with malaria parasites. PLoS Pathog 5: e1000542.
6. Schnitger AK, Yassine H, Kafatos FC, Osta MA (2009) Two C-type lectins cooperate to defend Anopheles gambiae against Gram-negative bacteria. J Biol Chem 284: 17616–17624.
7. Grossman SR, Andersen KG, Shlyakhter I, Tabrizi S, Winnicki S, et al. (2013) Identifying recent adaptations in large-scale genomic data. Cell 152: 703–713.
8. Sangare I, Michalakis Y, Yameogo B, Dabire R, Morlais I, et al. (2013) Studying fitness cost of Plasmodium falciparum infection in malaria vectors: validation of an appropriate negative control. Malar J 12: 2.
9. Rani A, Sharma A, Rajagopal R, Adak T, Bhatnagar RK (2009) Bacterial diversity analysis of larvae and adult midgut microflora using culture-dependent and culture-independent methods in lab-reared and field-collected Anopheles stephensi-an Asian malarial vector. BMC Microbiol 9: 96.
10. Gonzalez-Ceron L, Santillan F, Rodriguez MH, Mendez D, Hernandez-Avila JE (2003) Bacteria in midguts of field-collected Anopheles albimanus block Plasmodium vivax sporogonic development. J Med Entomol 40: 371–374.
11. Gendrin M, Christophides GK (2013) The Anopheles Mosquito Microbiota and Their Impact on Pathogen Transmission. In Manguin S, Ed. Anopheles mosquitoes - New insights into malaria vectors
12. Boissière A, Tchioffo MT, Bachar D, Abate L, Marie A, et al. (2012) Midgut Microbiota of the Malaria Mosquito Vector Anopheles gambiae and Interactions with Plasmodium falciparum Infection. PLoS Pathogens 8: e1002742.
13. Wang Y, Gilbreath TM, 3rd, Kukutla P, Yan G, Xu J (2011) Dynamic Gut Microbiome across Life History of the Malaria Mosquito Anopheles gambiae in Kenya. PLoS One 6: e24767.
14. Osei-Poku J, Mbogo CM, Palmer WJ, Jiggins FM (2012) Deep sequencing reveals extensive variation in the gut microbiota of wild mosquitoes from Kenya. Mol Ecol 21: 5138–5150.
15. Kumar S, Molina-Cruz A, Gupta L, Rodrigues J, Barillas-Mury C (2010) A peroxidase/dual oxidase system modulates midgut epithelial immunity in Anopheles gambiae. Science 327: 1644–1648.
16. Dong Y, Manfredini F, Dimopoulos G (2009) Implication of the mosquito midgut microbiota in the defense against malaria parasites. PLoS Pathog 5: e1000423.
17. Rodrigues J, Brayner FA, Alves LC, Dixit R, Barillas-Mury C (2010) Hemocyte differentiation mediates innate immune memory in Anopheles gambiae mosquitoes. Science 329: 1353–1355.
18. Cirimotich CM, Dong Y, Clayton AM, Sandiford SL, Souza-Neto JA, et al. (2011) Natural microbe-mediated refractoriness to Plasmodium infection in Anopheles gambiae. Science 332: 855–858.
19. Lemaitre B, Hoffmann J (2007) The host defense of Drosophila melanogaster. Annu Rev Immunol 25: 697–743.
20. Kaneko T, Goldman WE, Mellroth P, Steiner H, Fukase K, et al. (2004) Monomeric and polymeric gram-negative peptidoglycan but not purified LPS stimulate the Drosophila IMD pathway. Immunity 20: 637–649.
21. Choe KM, Lee H, Anderson KV (2005) Drosophila peptidoglycan recognition protein LC (PGRP-LC) acts as a signal-transducing innate immune receptor. Proc Natl Acad Sci U S A 102: 1122–1126.
22. Maillet F, Bischoff V, Vignal C, Hoffmann J, Royet J (2008) The Drosophila peptidoglycan recognition protein PGRP-LF blocks PGRP-LC and IMD/JNK pathway activation. Cell Host Microbe 3: 293–303.
23. Lhocine N, Ribeiro PS, Buchon N, Wepf A, Wilson R, et al. (2008) PIMS modulates immune tolerance by negatively regulating Drosophila innate immune signaling. Cell Host Microbe 4: 147–158.
24. Ha EM, Oh CT, Bae YS, Lee WJ (2005) A direct role for dual oxidase in Drosophila gut immunity. Science 310: 847–850.
25. Cronin SJ, Nehme NT, Limmer S, Liegeois S, Pospisilik JA, et al. (2009) Genome-wide RNAi screen identifies genes involved in intestinal pathogenic bacterial infection. Science 325: 340–343.
26. Buchon N, Broderick NA, Poidevin M, Pradervand S, Lemaitre B (2009) Drosophila intestinal response to bacterial infection: activation of host defense and stem cell proliferation. Cell Host Microbe 5: 200–211.
27. Buchon N, Broderick NA, Kuraishi T, Lemaitre B (2010) Drosophila EGFR pathway coordinates stem cell proliferation and gut remodeling following infection. BMC Biol 8: 152.
28. Ryu JH, Kim SH, Lee HY, Bai JY, Nam YD, et al. (2008) Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in Drosophila. Science 319: 777–782.
29. Lee KA, Kim SH, Kim EK, Ha EM, You H, et al. (2013) Bacterial-derived uracil as a modulator of mucosal immunity and gut-microbe homeostasis in Drosophila. Cell 153: 797–811.
30. Hergarden AC, Tayler TD, Anderson DJ (2012) Allatostatin-A neurons inhibit feeding behavior in adult Drosophila. Proc Natl Acad Sci U S A 109: 3967–3972.
31. Adamo SA (2005) Parasitic suppression of feeding in the tobacco hornworm, Manduca sexta: parallels with feeding depression after an immune challenge. Arch Insect Biochem Physiol 60: 185–197.
32. Chakrabarti S, Liehl P, Buchon N, Lemaitre B (2012) Infection-induced host translational blockage inhibits immune responses and epithelial renewal in the Drosophila gut. Cell Host Microbe 12: 60–70.
33. Liehl P, Blight M, Vodovar N, Boccard F, Lemaitre B (2006) Prevalence of local immune response against oral infection in a Drosophila/Pseudomonas infection model. PLoS Pathog 2: e56.
34. Miyamoto T, Slone J, Song X, Amrein H (2012) A Fructose Receptor Functions as a Nutrient Sensor in the Drosophila Brain. Cell 151: 1113–1125.
35. Moon SJ, Lee Y, Jiao Y, Montell C (2009) A Drosophila gustatory receptor essential for aversive taste and inhibiting male-to-male courtship. Curr Biol 19: 1623–1627.
36. Maslowski KM, Vieira AT, Ng A, Kranich J, Sierro F, et al. (2009) Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. Nature 461: 1282–1286.
37. Bando H, Okado K, Guelbeogo WM, Badolo A, Aonuma H, et al. (2013) Intra-specific diversity of Serratia marcescens in Anopheles mosquito midgut defines Plasmodium transmission capacity. Sci Rep 3: 1641.
38. Lawniczak MK, Emrich SJ, Holloway AK, Regier AP, Olson M, et al. (2010) Widespread divergence between incipient Anopheles gambiae species revealed by whole genome sequences. Science 330: 512–514.
39. Neafsey DE, Lawniczak MK, Park DJ, Redmond SN, Coulibaly MB, et al. (2010) SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes. Science 330: 514–517.
40. Gottar M, Gobert V, Michel T, Belvin M, Duyk G, et al. (2002) The Drosophila immune response against Gram-negative bacteria is mediated by a peptidoglycan recognition protein. Nature 416: 640–644.
41. Choe KM, Werner T, Stoven S, Hultmark D, Anderson KV (2002) Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in Drosophila. Science 296: 359–362.
42. Lin H, Zhang L, Luna C, Ho NT, Zheng L (2007) A splice variant of PGRP-LC required for expression of antimicrobial peptides in Anopheles gambiae. Insect Science 14: 185–192.
43. Watson FL, Puttmann-Holgado R, Thomas F, Lamar DL, Hughes M, et al. (2005) Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science 309: 1874–1878.
44. Dong Y, Taylor HE, Dimopoulos G (2006) AgDscam, a hypervariable immunoglobulin domain-containing receptor of the Anopheles gambiae innate immune system. PLoS Biol 4: e229.
45. Watthanasurorot A, Jiravanichpaisal P, Liu H, Soderhall I, Soderhall K (2011) Bacteria-Induced Dscam Isoforms of the Crustacean, Pacifastacus leniusculus. PLoS Pathog 7: e1002062.
46. Wojtowicz WM, Wu W, Andre I, Qian B, Baker D, et al. (2007) A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. Cell 130: 1134–1145.
47. Dong Y, Cirimotich CM, Pike A, Chandra R, Dimopoulos G (2012) Anopheles NF-kappaB-Regulated Splicing Factors Direct Pathogen-Specific Repertoires

of the Hypervariable Pattern Recognition Receptor AgDscam. Cell Host Microbe 12: 521–530.
48. Smith PH, Mwangi JM, Afrane YA, Yan G, Obbard DJ, et al. (2011) Alternative splicing of the Anopheles gambiae Dscam gene in diverse Plasmodium falciparum infections. Malar J 10: 156.
49. Clayton AM, Cirimotich CM, Dong Y, Dimopoulos G (2012) Caudal is a negative regulator of the Anopheles IMD Pathway that controls resistance to P. falciparum infection. Dev Comp Immunol 39(4):323–32.
50. Xin N, Benchabane H, Tian A, Nguyen K, Klofas L, et al. (2011) Erect Wing facilitates context-dependent Wnt/Wingless signaling by recruiting the cell-specific Armadillo-TCF adaptor Earthbound to chromatin. Development 138: 4955–4967.
51. Ditch LM, Shirangi T, Pitman JL, Latham KL, Finley KD, et al. (2005) Drosophila retained/dead ringer is necessary for neuronal pathfinding, female receptivity and repression of fruitless independent male courtship behaviors. Development 132: 155–164.
52. Shirangi TR, Taylor BJ, McKeown M (2006) A double-switch system regulates male courtship behavior in male and female Drosophila melanogaster. Nat Genet 38: 1435–1439.
53. Dubrovsky EB, Dubrovskaya VA, Bernardo T, Otte V, DiFilippo R, et al. (2011) The Drosophila FTZ-F1 nuclear receptor mediates juvenile hormone activation of E75A gene expression through an intracellular pathway. J Biol Chem 286: 33689–33700.
54. Funkhouser JD, Aronson NN, Jr. (2007) Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. BMC Evol Biol 7: 96.
55. Dela Cruz CS, Liu W, He CH, Jacoby A, Gornitzky A, et al. (2012) Chitinase 3-like-1 Promotes Streptococcus pneumoniae Killing and Augments Host Tolerance to Lung Antibacterial Responses. Cell Host Microbe 12: 34–46.
56. Volz J, Muller HM, Zdanowicz A, Kafatos FC, Osta MA (2006) A genetic module regulates the melanization response of Anopheles to Plasmodium. Cell Microbiol 8: 1392–1405.
57. Volz J, Osta MA, Kafatos FC, Muller HM (2005) The roles of two clip domain serine proteases in innate immune responses of the malaria vector Anopheles gambiae. J Biol Chem 280: 40161–40168.
58. Povelones M, Bhagavatula L, Yassine H, Tan LA, Upton LM, et al. (2013) The CLIP-Domain Serine Protease Homolog SPCLIP1 Regulates Complement Recruitment to Microbial Surfaces in the Malaria Mosquito Anopheles gambiae. PLoS Pathog 9: e1003623.
59. Dong Y, Aguilar R, Xi Z, Warr E, Mongin E, et al. (2006) Anopheles gambiae immune responses to human and rodent Plasmodium parasite species. PLoS Pathog 2: e52.
60. Waterhouse RM, Povelones M, Christophides GK (2010) Sequence-structure-function relations of the mosquito leucine-rich repeat immune proteins. BMC Genomics 11: 531.
61. Povelones M, Upton LM, Sala KA, Christophides GK (2011) Structure-Function Analysis of the Anopheles gambiae LRIM1/APL1C Complex and its Interaction with Complement C3-Like Protein TEP1. PLoS Pathog 7: e1002023.
62. Povelones M, Waterhouse RM, Kafatos FC, Christophides GK (2009) Leucine-rich repeat protein complex activates mosquito complement in defense against Plasmodium parasites. Science 324: 258–261.
63. Osta MA, Christophides GK, Kafatos FC (2004) Effects of mosquito genes on Plasmodium development. Science 303: 2030–2032.
64. Kazura JW, Zou Z, Souza-Neto J, Xi Z, Kokoza V, et al. (2011) Transcriptome Analysis of Aedes aegypti Transgenic Mosquitoes with Altered Immunity. PLoS Pathogens 7: e1002394.
65. Valanne S, Wang JH, Ramet M (2011) The Drosophila Toll signaling pathway. J Immunol 186: 649–656.
66. Garver LS, Dong Y, Dimopoulos G (2009) Caspar controls resistance to Plasmodium falciparum in diverse anopheline species. PLoS Pathog 5: e1000335.
67. Vaishnava S, Yamamoto M, Severson KM, Ruhn KA, Yu X, et al. (2011) The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. Science 334: 255–258.
68. Dong Y, Dimopoulos G (2009) Anopheles fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. J Biol Chem 284: 9835–9844.
69. Pace KE, Baum LG (2004) Insect galectins: roles in immunity and development. Glycoconj J 19: 607–614.
70. Mendes AM, Awono-Ambene PH, Nsango SE, Cohuet A, Fontenille D, et al. (2011) Infection intensity dependent responses of Anopheles gambiae to African malaria parasites Plasmodium falciparum. Infect Immun 79(11):4708–15.
71. Styer KL, Singh V, Macosko E, Steele SE, Bargmann CI, et al. (2008) Innate immunity in Caenorhabditis elegans is regulated by neurons expressing NPR-1/GPCR. Science 322: 460–464.
72. Reddy KC, Andersen EC, Kruglyak L, Kim DH (2009) A polymorphism in npr-1 is a behavioral determinant of pathogen susceptibility in C. elegans. Science 323: 382–384.
73. Cirillo C, Vanden Berghe P, Tack J (2011) Role of serotonin in gastrointestinal physiology and pathology. Minerva Endocrinol 36: 311–324.
74. Khan WI, Ghia JE (2010) Gut hormones: emerging role in immune activation and inflammation. Clin Exp Immunol 161: 19–27.
75. Baganz NL, Blakely RD (2013) A dialogue between the immune system and brain, spoken in the language of serotonin. ACS Chem Neurosci 4: 48–63.
76. Becnel J, Johnson O, Luo J, Nassel DR, Nichols CD (2011) The serotonin 5-HT7Dro receptor is expressed in the brain of Drosophila, and is essential for normal courtship and mating. PLoS One 6: e20800.
77. Johnson O, Becnel J, Nichols CD (2011) Serotonin receptor activity is necessary for olfactory learning and memory in Drosophila melanogaster. Neuroscience 192: 372–381.
78. Dierick HA, Greenspan RJ (2007) Serotonin and neuropeptide F have opposite modulatory effects on fly aggression. Nat Genet 39: 678–682.
79. Dzitoyeva S, Dimitrijevic N, Manev H (2003) Gamma-aminobutyric acid B receptor 1 mediates behavior-impairing actions of alcohol in Drosophila: adult RNA interference and pharmacological evidence. Proc Natl Acad Sci U S A 100: 5485–5490.
80. Wen T, Parrish CA, Xu D, Wu Q, Shen P (2005) Drosophila neuropeptide F and its receptor, NPFR1, define a signaling pathway that acutely modulates alcohol sensitivity. Proc Natl Acad Sci U S A 102: 2141–2146.
81. Shohat-Ophir G, Kaun KR, Azanchi R, Heberlein U (2012) Sexual deprivation increases ethanol intake in Drosophila. Science 335: 1351–1355.
82. Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, et al. (2002) G protein-coupled receptors in Anopheles gambiae. Science 298: 176–178.
83. Bray S, Amrein H (2003) A putative Drosophila pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. Neuron 39: 1019–1029.
84. Miyamoto T, Amrein H (2008) Suppression of male courtship by a Drosophila pheromone receptor. Nat Neurosci 11: 874–876.
85. Watanabe K, Toba G, Koganezawa M, Yamamoto D (2011) Gr39a, a highly diversified gustatory receptor in Drosophila, has a role in sexual behavior. Behav Genet 41: 746–753.
86. Wang L, Han X, Mehren J, Hiroi M, Billeter JC, et al. (2011) Hierarchical chemosensory regulation of male-male social interactions in Drosophila. Nat Neurosci 14: 757–762.
87. Lee Y, Kim SH, Montell C (2010) Avoiding DEET through insect gustatory receptors. Neuron 67: 555–561.
88. Moon SJ, Kottgen M, Jiao Y, Xu H, Montell C (2006) A taste receptor required for the caffeine response in vivo. Curr Biol 16: 1812–1817.
89. Jones WD, Cayirlioglu P, Kadow IG, Vosshall LB (2007) Two chemosensory receptors together mediate carbon dioxide detection in Drosophila. Nature 445: 86–90.
90. Erdelyan CN, Mahood TH, Bader TS, Whyard S (2011) Functional validation of the carbon dioxide receptor genes in Aedes aegypti mosquitoes using RNA interference. Insect Mol Biol.
91. Chyb S, Dahanukar A, Wickens A, Carlson JR (2003) Drosophila Gr5a encodes a taste receptor tuned to trehalose. Proc Natl Acad Sci U S A 100 Suppl 2: 14526–14530.
92. Ueno K, Ohta M, Morita H, Mikuni Y, Nakajima S, et al. (2001) Trehalose sensitivity in Drosophila correlates with mutations in and expression of the gustatory receptor gene Gr5a. Curr Biol 11: 1451–1455.
93. Dahanukar A, Foster K, van der Goes van Naters WM, Carlson JR (2001) A Gr receptor is required for response to the sugar trehalose in taste neurons of Drosophila. Nat Neurosci 4: 1182–1186.
94. Jiao Y, Moon SJ, Montell C (2007) A Drosophila gustatory receptor required for the responses to sucrose, glucose, and maltose identified by mRNA tagging. Proc Natl Acad Sci U S A 104: 14110–14115.
95. Serhan CN, Chiang N, Van Dyke TE (2008) Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators. Nat Rev Immunol 8: 349–361.
96. Lucas KJ, Myles KM, Raikhel AS (2013) Small RNAs: a new frontier in mosquito biology. Trends Parasitol 29: 295–303.
97. Zhou R, Hu G, Liu J, Gong AY, Drescher KM, et al. (2009) NF-kappaB p65-dependent transactivation of miRNA genes following Cryptosporidium parvum infection stimulates epithelial cell immune responses. PLoS Pathog 5: e1000681.
98. Sakurai M, Aoki T, Yoshikawa S, Santschi LA, Saito H, et al. (2009) Differentially expressed Drl and Drl-2 play opposing roles in Wnt5 signaling during Drosophila olfactory system development. J Neurosci 29: 4972–4980.
99. Gesellchen V, Kuttenkeuler D, Steckel M, Pelte N, Boutros M (2005) An RNA interference screen identifies Inhibitor of Apoptosis Protein 2 as a regulator of innate immune signalling in Drosophila. EMBO Rep 6: 979–984.
100. Chen J, Zhang Y, Shen P (2008) A protein kinase C activity localized to neuropeptide Y-like neurons mediates ethanol intoxication in Drosophila melanogaster. Neuroscience 156: 42–47.
101. Chen J, Zhang Y, Shen P (2010) Protein kinase C deficiency-induced alcohol insensitivity and underlying cellular targets in Drosophila. Neuroscience 166: 34–39.
102. Ye T, Tang W, Zhang X (2012) Involvement of Rab6 in the regulation of phagocytosis against virus infection in invertebrates. J Proteome Res 11: 4834–4846.
103. Coutelis JB, Ephrussi A (2007) Rab6 mediates membrane organization and determinant localization during Drosophila oogenesis. Development 134: 1419–1430.
104. Tian AG, Tamori Y, Huang YC, Melendez NT, Deng WM (2013) Efficient EGFR signaling and dorsal-ventral axis patterning requires syntaxin dependent Gurken trafficking. Dev Biol 373: 349–358.

105. Abou Tayoun AN, Pikielny C, Dolph PJ (2012) Roles of the Drosophila SK channel (dSK) in courtship memory. PLoS One 7: e34665.
106. Rusten TE, Vaccari T, Lindmo K, Rodahl LM, Nezis IP, et al. (2007) ESCRTs and Fab1 regulate distinct steps of autophagy. Curr Biol 17: 1817–1825.
107. Soukup SF, Culi J, Gubb D (2009) Uptake of the necrotic serpin in Drosophila melanogaster via the lipophorin receptor-1. PLoS Genet 5: e1000532.
108. Levashina EA (1999) Constitutive Activation of Toll-Mediated Antifungal Defense in Serpin-Deficient Drosophila. Science 285: 1917–1919.
109. Pelte N, Robertson AS, Zou Z, Belorgey D, Dafforn TR, et al. (2006) Immune challenge induces N-terminal cleavage of the Drosophila serpin Necrotic. Insect Biochem Mol Biol 36: 37–46.
110. Marinotti O, Calvo E, Nguyen QK, Dissanayake S, Ribeiro JM, et al. (2006) Genome-wide analysis of gene expression in adult Anopheles gambiae. Insect Mol Biol 15: 1–12.
111. Baker DA, Nolan T, Fischer B, Pinder A, Crisanti A, et al. (2011) A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, Anopheles gambiae. BMC Genomics 12: 296.
112. Pitts RJ, Rinker DC, Jones PL, Rokas A, Zwiebel LJ (2011) Transcriptome profiling of chemosensory appendages in the malaria vector Anopheles gambiae reveals tissue- and sex-specific signatures of odor coding. BMC Genomics 12: 271.
113. Meister S, Kanzok SM, Zheng XL, Luna C, Li TR, et al. (2005) Immune signaling pathways regulating bacterial and malaria parasite infection of the mosquito Anopheles gambiae. Proc Natl Acad Sci U S A 102: 11420–11425.
114. Felix RC, Silveira H (2011) The Interplay between Tubulins and P450 Cytochromes during Plasmodium berghei Invasion of Anopheles gambiae Midgut. PLoS One 6: e24181.
115. Benoit JB, Lopez-Martinez G, Patrick KR, Phillips ZP, Krause TB, et al. (2011) From the Cover: Drinking a hot blood meal elicits a protective heat shock response in mosquitoes. Proc Natl Acad Sci U S A 108: 8026–8029.
116. Li B, Calvo E, Marinotti O, James AA, Paskewitz SM (2005) Characterization of the c-type lysozyme gene family in Anopheles gambiae. Gene 360: 131–139.
117. Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, et al. (2002) Immunity-related genes and gene families in Anopheles gambiae. Science 298: 159–165.
118. Danielli A, Loukeris TG, Lagueux M, Muller HM, Richman A, et al. (2000) A modular chitin-binding protease associated with hemocytes and hemolymph in the mosquito Anopheles gambiae. Proc Natl Acad Sci U S A 97: 7136–7141.
119. Dinglasan RR, Devenport M, Florens L, Johnson JR, McHugh CA, et al. (2009) The Anopheles gambiae adult midgut peritrophic matrix proteome. Insect Biochem Mol Biol 39: 125–134.
120. Kuraishi T, Binggeli O, Opota O, Buchon N, Lemaitre B (2011) Genetic evidence for a protective role of the peritrophic matrix against intestinal bacterial infection in Drosophila melanogaster. Proc Natl Acad Sci U S A 108: 15966–15971.
121. Brown MR, Crim JW, Arata RC, Cai HN, Chun C, et al. (1999) Identification of a Drosophila brain-gut peptide related to the neuropeptide Y family. Peptides 20: 1035–1042.
122. Stanek DM, Pohl J, Crim JW, Brown MR (2002) Neuropeptide F and its expression in the yellow fever mosquito, Aedes aegypti. Peptides 23: 1367–1378.
123. Wu Q, Zhao Z, Shen P (2005) Regulation of aversion to noxious food by Drosophila neuropeptide Y- and insulin-like systems. Nat Neurosci 8: 1350–1355.
124. Shen P, Cai HN (2001) Drosophila neuropeptide F mediates integration of chemosensory stimulation and conditioning of the nervous system by food. J Neurobiol 47: 16–25.
125. Kacsoh BZ, Lynch ZR, Mortimer NT, Schlenke TA (2013) Fruit flies medicate offspring after seeing parasites. Science 339: 947–950.
126. Wang Y, Brent CS, Fennern E, Amdam GV (2012) Gustatory perception and fat body energy metabolism are jointly affected by vitellogenin and juvenile hormone in honey bees. PLoS Genet 8: e1002779.
127. Xu J, Sheng Z, Palli SR (2013) Juvenile hormone and insulin regulate trehalose homeostasis in the red flour beetle, Tribolium castaneum. PLoS Genet 9: e1003535.
128. Noriega FG (2004) Nutritional regulation of JH synthesis: a mechanism to control reproductive maturation in mosquitoes? Insect Biochem Mol Biol 34: 687–693.
129. Perez-Hedo M, Rivera-Perez C, Noriega FG (2013) The insulin/TOR signal transduction pathway is involved in the nutritional regulation of juvenile hormone synthesis in Aedes aegypti. Insect Biochem Mol Biol 43(6):495–500.
130. Sharon G, Segal D, Ringo JM, Hefetz A, Zilber-Rosenberg I, et al. (2010) Commensal bacteria play a role in mating preference of Drosophila melanogaster. Proc Natl Acad Sci U S A 107: 20051–20056.
131. Carey AF, Wang G, Su CY, Zwiebel LJ, Carlson JR (2010) Odorant reception in the malaria mosquito Anopheles gambiae. Nature 464: 66–71.
132. Meunier N, Belgacem YH, Martin JR (2007) Regulation of feeding behaviour and locomotor activity by takeout in Drosophila. J Exp Biol 210: 1424–1434.
133. Vanaphan N, Dauwalder B, Zufall RA (2012) Diversification of takeout, a male-biased gene family in Drosophila. Gene 491: 142–148.
134. Park JH, Kwon JY (2011) Heterogeneous expression of Drosophila gustatory receptors in enteroendocrine cells. PLoS One 6: e29022.
135. Niare O, Markianos K, Volz J, Oduol F, Toure A, et al. (2002) Genetic loci affecting resistance to human malaria parasites in a West African mosquito vector population. Science 298: 213–216.
136. Blandin SA, Wang-Sattler R, Lamacchia M, Gagneur J, Lycett G, et al. (2009) Dissecting the genetic basis of resistance to malaria parasites in Anopheles gambiae. Science 326: 147–150.
137. Molina-Cruz A, Garver LS, Alabaster A, Bangiolo L, Haile A, et al. (2013) The human malaria parasite Pfs47 gene mediates evasion of the mosquito immune system. Science 340: 984–987.
138. Heutink P, Oostra BA (2002) Gene finding in genetically isolated populations. Hum Mol Genet 11: 2507–2515.
139. Gudmundsson J, Sulem P, Gudbjartsson DF, Masson G, Agnarsson BA, et al. (2012) A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. Nat Genet 44: 1326–1329.
140. Lee K-A, Kim S-H, Kim E-K, Ha E-M, You H, et al. (2013) Bacterial-Derived Uracil as a Modulator of Mucosal Immunity and Gut-Microbe Homeostasis in Drosophila. Cell 153: 797–811.
141. Meijers R, Puettmann-Holgado R, Skiniotis G, Liu JH, Walz T, et al. (2007) Structural basis of Dscam isoform specificity. Nature 449: 487–491.
142. Kolodziej PA, Timpe LC, Mitchell KJ, Fried SR, Goodman CS, et al. (1996) frazzled encodes a Drosophila member of the DCC immunoglobulin subfamily and is required for CNS and motor axon guidance. Cell 87: 197–204.
143. Kidd T, Brose K, Mitchell KJ, Fetter RD, Tessier-Lavigne M, et al. (1998) Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. Cell 92: 205–215.
144. Evans TA, Bashaw GJ (2010) Axon guidance at the midline: of mice and flies. Curr Opin Neurobiol 20: 79–85.
145. Koch CM, Honemann-Capito M, Egger-Adam D, Wodarz A (2009) Windei, the Drosophila homolog of mAM/MCAF1, is an essential cofactor of the H3K9 methyl transferase dSETDB1/Eggless in germ line development. PLoS Genet 5: e1000644.
146. Liu L, Ishihara K, Ichimura T, Fujita N, Hino S, et al. (2009) MCAF1/AM is involved in Sp1-mediated maintenance of cancer-associated telomerase activity. J Biol Chem 284: 5165–5174.
147. Garver LS, Xi Z, Dimopoulos G (2008) Immunoglobulin superfamily members play an important role in the mosquito immune system. Dev Comp Immunol 32: 519–531.
148. Zdobnov EM, von Mering C, Letunic I, Torrents D, Suyama M, et al. (2002) Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster. Science 298: 149–159.
149. Bosco-Drayon V, Poidevin M, Boneca IG, Narbonne-Reveau K, Royet J, et al. (2012) Peptidoglycan Sensing by the Receptor PGRP-LE in the Drosophila Gut Induces Immune Responses to Infectious Bacteria and Tolerance to Microbiota. Cell Host Microbe 12: 153–165.
150. Paredes JC, Welchman DP, Poidevin M, Lemaitre B (2011) Negative Regulation by Amidase PGRPs Shapes the Drosophila Antibacterial Response and Protects the Fly from Innocuous Infection. Immunity 35: 770–779.
151. Crampton JM, Beard CB, Louis C, World Health Organization. (1997) The molecular biology of insect disease vectors : a methods manual. London: Chapman & Hall. xxv, 578 p.
152. Nehme NT, Liegeois S, Kele B, Giammarinaro P, Pradel E, et al. (2007) A model of bacterial intestinal infections in Drosophila melanogaster. PLoS Pathog 3: e173.
153. Favia G, Ricci I, Damiani C, Raddadi N, Crotti E, et al. (2007) Bacteria of the genus Asaia stably associate with Anopheles stephensi, an Asian malarial mosquito vector. Proc Natl Acad Sci U S A 104: 9047–9051.
154. Reidenbach KR, Neafsey DE, Costantini C, Sagnon N, Simard F, et al. (2012) Patterns of genomic differentiation between ecologically differentiated M and S forms of Anopheles gambiae in West and Central Africa. Genome Biol Evol 4(12):1202–12.
155. Lawson D, Arensburger P, Atkinson P, Besansky NJ, Bruggner RV, et al. (2009) VectorBase: a data resource for invertebrate vector genomics. Nucleic Acids Res 37: D583–587.
156. Blandin S, Moita LF, Kocher T, Wilm M, Kafatos FC, et al. (2002) Reverse genetics in the mosquito Anopheles gambiae: targeted disruption of the Defensin gene. EMBO Rep 3: 852–856.
157. Goecks J, Nekrutenko A, Taylor J, Galaxy T (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11: R86.
158. Santiago-Sotelo P, Ramirez-Prado JH (2012) prfectBLAST: a platform-independent portable front end for the command terminal BLAST+ stand-alone suite. Biotechniques 53: 299–300.
159. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. Genome Res 21: 1552–1560.
160. Ja WW, Carvalho GB, Mak EM, de la Rosa NN, Fang AY, et al. (2007) Prandiology of Drosophila and the CAFE assay. Proc Natl Acad Sci U S A 104: 8253–8256.

The research paper “Genetic Dissection of *Anopheles gambiae* Gut Epithelial Responses to *Serratia marcescens*”, based on work presented in this thesis. (*PLoS Pathogens* 10(3): e1003897. doi:10.1371/journal.ppat.1003897).

Primer name	T7 Reaction - qRT-PCR or 454 sequencing	Sequence
16S-0515F	454	TGYCAGCMGCCGCGGTA
16S-1061R	454	TCACGRCACGAGCTGACG
16S F	qRT-PCR	TCCTACGGGAGGCAGCAGT
16S R	qRT-PCR	GGACTACCAGGGTATCTAATCCTGTT
AGAP006991 F	qRT-PCR	CTCTCGCGCTGTTCGTCGCA
AGAP006991 R	qRT-PCR	CCACAGGCGGGCCGTCATTT
AgS7 FOR	qRT-PCR	GTGCGCGAGTTGGAGAAGA
AgS7 REV	qRT-PCR	ATCGGTTTGGGCAGAATGC
CEC1 F	qRT-PCR	TCATCTTTGTCGTGCTGGCA
CEC1 R	qRT-PCR	TCTTCAGCCGTCCCGCT
CLIPB14 F	qRT-PCR	GAGCCGAACTGTGGGCCGTC
CLIPB14 R	qRT-PCR	GCCTGAAACCGCAGCAACGC
CLIPB17 F	qRT-PCR	ACTCGGGCGGACCGTACCAA
CLIPB17 R	qRT-PCR	GTACCCCACCAGCACACCGC
DEF1 F	qRT-PCR	CGGCTACCCTGCTGAACGGC
DEF1 R	qRT-PCR	GTTCTCCAGCGCGGCATGGT
Ect4 F	qRT-PCR	TGACGACGCACAACCCGTCG
Ect4 R	qRT-PCR	GAGCTCAGCACCGGCACCAG
FN3D1 F	qRT-PCR	GGCCGCCGGTTTCATCCACA
FN3D1 R	qRT-PCR	TTGCGGTCCAGGTGGTGGGA
FN3D2 F	qRT-PCR	GGTGTCGCTGACGGTGACGG
FN3D2 R	qRT-PCR	TGCGCCGGAAAGCCGGAAAT
FN3D3 F	qRT-PCR	CGTGACGGCCAACGTGACGA
FN3D3 R	qRT-PCR	GGTCGCGAACCCACCGACTG
Gr9 F	qRT-PCR	CGCTTGTTCTGCTGCATTGT
Gr9 R	qRT-PCR	AACGGCCACAGAATGTTTGC
LYSC2 F	qRT-PCR	TGCCTGACATCCGCGAGTGC
LYSC2 R	qRT-PCR	GGCTGGTCATTGCCGCATGG
NPF F	qRT-PCR	TCGACGATGGCGTCAGGCAC
NPF R	qRT-PCR	CGGCAGCTACAGAGGCAGCG
PGRPLC F	qRT-PCR	AGAACGGAATGGCTGGCGCA
PGRPLC R	qRT-PCR	TGCAACCTTCGGTGGCAGTGT
Serratia F	qRT-PCR	ACGTTCATCAATTGACGTTACTCGCA
Serratia R	qRT-PCR	AACCGCCTGCGTGCGCTTTA
AGAP002492 F	T7	taatacgactcactatagggCGGCAAACCCAACAACAT
AGAP002492 R	T7	taatacgactcactatagggGGTATACCGTCGATGATGAGC
AGAP004375 F	T7	taatacgactcactatagggAGCAGCATCCTGCAAAAGAC
AGAP004375 R	T7	taatacgactcactatagggGTTCCAGCCATCCTCGATT
AGAP006991 F	T7	taatacgactcactatagggAGCTTGCTACAACCCGTCAC
AGAP006991 R	T7	taatacgactcactatagggGGTGGGGTCCCAGTTGTTAC
AGAP010012 F	T7	taatacgactcactatagggGAGCAACTGCGGCAAATTAT
AGAP010012 R	T7	taatacgactcactatagggCGCAGCACAGAAGATTCTCA
CLIPE6 F	T7	taatacgactcactatagggTGCACTGAAGTTCCAGCAGA
CLIPE6 R	T7	taatacgactcactatagggGGGACGAACCGTTTACCC
dsLacZ F	T7	taatacgactcactatagggAGAATCCGACGGGTTGTTACT
dsLacZ R	T7	taatacgactcactatagggCACCACGCTCATCGATAATTT
Ect4 F	T7	taatacgactcactatagggGCGACACGCTGCGCAATACG
Ect4 R	T7	taatacgactcactatagggGCACGTGTCCTCGCTGTGCT
FN3D1 F	T7	taatacgactcactatagggGATGGACGTGGATCAGCC
FN3D1 R	T7	taatacgactcactatagggTGGATCGTCCTCATCACTGT
FN3D2 F	T7	taatacgactcactatagggACGGCCGTTTTAAAGTGTCA
FN3D2 R	T7	taatacgactcactatagggCCCAGCATTGTTGATGTACAGA
FN3D3 F	T7	taatacgactcactatagggAATCATTTCCTTTCGCATTCC
FN3D3 R	T7	taatacgactcactatagggACATTGTCCTTGTACCACACCA
Gr9 F	T7	taatacgactcactatagggAGCACCCGGCATGCGACATC
Gr9 R	T7	taatacgactcactatagggAGCTCAGCTCGTTTTGGCGCA
LRIM15 F	T7	taatacgactcactatagggGCCACCGTCGGCTGAGCTTT
LRIM15 R	T7	taatacgactcactatagggCCCACAGCAGGAAGCCCACG

NPF F	T7	taatacgactcactatagggACCTCCAGGAGCTTGAAACT
NPF R	T7	taatacgactcactatagggAAACATGATGCTGCGGGTTG
PGRPLC F	T7	taatacgactcactatagggTCAGTCAACACCGGATACCA
PGRPLC R	T7	taatacgactcactatagggTCAGTCAACACCGGATACCA

Table 3.1: Primers used for dsRNA synthesis, qRT-PCR or 454 sequencing. The use of each primer pair is indicated as used for dsRNA synthesis (T7 reaction), qRT-PCR or 454 sequencing. T7 flanking sequences are included in lower case letters.

SNP	Chromosome	Position	MAF Diff	Peak	Gene hits						
Ag_2L_003518656	2L	3518656	0.54	2L-1	AGAP004783	AGAP004784					
Ag_2L_005323700	2L	5323700	0.61	2L-2							
Ag_2L_006129603	2L	6129603	0.56		AGAP004931						
Ag_2L_009499844	2L	9499844	0.52	2L-4							
Ag_2L_010421601	2L	10421601	0.54								
Ag_2L_010482763	2L	10482763	0.60		AGAP005143						
Ag_2L_010562614	2L	10562614	0.61		AGAP005147						
Ag_2L_010693205	2L	10693205	0.51		AGAP005156						
Ag_2L_012210700	2L	12210700	0.52	2L-5	AGAP005203	AGAP005205					
Ag_2L_012228038	2L	12228038	0.68		AGAP005206						
Ag_2L_012239158	2L	12239158	0.79								
Ag_2L_012327390	2L	12327390	0.62								
Ag_2L_012416410	2L	12416410	0.69		AGAP005216	AGAP005217					
Ag_2L_012450546	2L	12450546	0.58								
Ag_2L_012569927	2L	12569927	0.54								
Ag_2L_012823420	2L	12823420	0.58		AGAP005244						
Ag_2L_013668860	2L	13668860	0.55		AGAP005287						
Ag_2L_016074637	2L	16074637	0.51	2L-6	AGAP005473	AGAP005474	AGAP005475	AGAP005476	AGAP005477	AGAP005478	AGAP005479
Ag_2L_016227362	2L	16227362	0.50		AGAP005497						
Ag_2L_018440454	2L	18440454	0.54	2L-7	AGAP005661						
Ag_2L_019303208	2L	19303208	0.51								
Ag_2L_021674505	2L	21674505	0.51	2L-8							
Ag_2L_027924370	2L	27924370	0.60	2L-9	AGAP006186						
Ag_2L_031407343	2L	31407343	0.57	2L-10	AGAP006405						
Ag_2L_031764861	2L	31764861	0.52		AGAP006428						
Ag_2L_031831396	2L	31831396	0.51								
Ag_2L_035563670	2L	35563670	0.53	2L-11	AGAP006643	AGAP006644					
Ag_2L_038192521	2L	38192521	0.51	2L-12	AGAP006784	AGAP006785					
Ag_2L_040933889	2L	40933889	0.56	2L-13	AGAP007012						
Ag_2L_041612176	2L	41612176	0.68		AGAP007045						
Ag_2L_041971478	2L	41971478	0.51								
Ag_2L_042790121	2L	42790121	0.55	2L-14	AGAP007104						
Ag_2L_042879419	2L	42879419	0.50								
Ag_2L_043431017	2L	43431017	0.74								
Ag_2L_043585853	2L	43585853	0.53		AGAP007148	AGAP007149	AGAP007150	AGAP007151	AGAP007152	AGAP007153	AGAP007154
Ag_2L_043726092	2L	43726092	0.50								
Ag_2L_044010910	2L	44010910	0.54								
Ag_2L_044057845	2L	44057845	0.50		AGAP007181						
Ag_2L_045022088	2L	45022088	0.54	2L-15	AGAP007291	AGAP007292					
Ag_2L_046250290	2L	46250290	0.53	2L-16	AGAP007387	AGAP007388					
Ag_2L_046421090	2L	46421090	0.55		AGAP007415						
Ag_2L_047521026	2L	47521026	0.51	2L-17	AGAP007550						
Ag_2L_048801371	2L	48801371	0.55	2L-18	AGAP007657						
Ag_2L_049299743	2L	49299743	0.67		AGAP007723						
Ag_2R_000465473	2R	465473	0.57	2R-1							
Ag_2R_000775042	2R	775042	0.52		AGAP001179						
Ag_2R_002747832	2R	2747832	0.50	2R-2							
Ag_2R_005170891	2R	5170891	0.50	2R-3	AGAP001446						
Ag_2R_012714381	2R	12714381	0.52	2R-5	AGAP001944						
Ag_2R_015263254	2R	15263254	0.71	2R-6	AGAP002094						
Ag_2R_023401663	2R	23401663	0.52	2R-8							
Ag_2R_024127613	2R	24127613	0.56								
Ag_2R_024713048	2R	24713048	0.68								
Ag_2R_024728111	2R	24728111	0.52								
Ag_2R_024803451	2R	24803451	0.65								
Ag_2R_024804136	2R	24804136	0.57								
Ag_2R_024873492	2R	24873492	0.52		AGAP013110	AGAP013353	AGAP013322	AGAP013225	AGAP013456	AGAP013086	
Ag_2R_025058493	2R	25058493	0.55								
Ag_2R_025258455	2R	25258455	0.51		AGAP002665	AGAP002667	AGAP002666	AGAP002664			
Ag_2R_026906125	2R	26906125	0.78	2R-9							
Ag_2R_032804506	2R	32804506	0.54	2R-11							
Ag_2R_033238552	2R	33238552	0.52								
Ag_2R_035225086	2R	35225086	0.51	2R-12							
Ag_2R_035238077	2R	35238077	0.55								
Ag_2R_035578739	2R	35578739	0.57								
Ag_2R_048020283	2R	48020283	0.59	2R-13	AGAP004032						
Ag_2R_051472055	2R	51472055	0.54	2R-14							
Ag_2R_051612164	2R	51612164	0.52								
Ag_2R_051875990	2R	51875990	0.51		AGAP004208						
Ag_2R_052153947	2R	52153947	0.55		AGAP004223						
Ag_2R_052758256	2R	52758256	0.57								
Ag_2R_055777531	2R	55777531	0.53	2R-15	AGAP004405						
Ag_2R_059805200	2R	59805200	0.51	2R-16							
Ag_3L_002543806	3L	2543806	0.56	3L-1							
Ag_3L_002637624	3L	2637624	0.53		AGAP010394						

Ag_3L_003457739	3L	3457739	0.70	3L-2							
Ag_3L_004302915	3L	4302915	0.53	3L-3							
Ag_3L_005119760	3L	5119760	0.56	3L-4	AGAP010503						
Ag_3L_005203143	3L	5203143	0.53								
Ag_3L_008405079	3L	8405079	0.53	3L-5							
Ag_3L_009319745	3L	9319745	0.53	3L-6							
Ag_3L_010558698	3L	10558698	0.51	3L-7	AGAP010798						
Ag_3L_010732196	3L	10732196	0.50								
Ag_3L_013522657	3L	13522657	0.57	3L-8	AGAP010940	AGAP010941					
Ag_3L_014194976	3L	14194976	0.51								
Ag_3L_017475128	3L	17475128	0.54	3L-9							
Ag_3L_017910149	3L	17910149	0.50								
Ag_3L_017918407	3L	17918407	0.71								
Ag_3L_017984337	3L	17984337	0.51		AGAP011147	AGAP011148	AGAP011149	AGAP011150	AGAP011151		
Ag_3L_020065464	3L	20065464	0.54	3L-10							
Ag_3L_020485207	3L	20485207	0.68		AGAP011277	AGAP011278					
Ag_3L_022008938	3L	22008938	0.56	3L-11	AGAP011363						
Ag_3L_024035717	3L	24035717	0.52	3L-12	AGAP011415						
Ag_3L_025747899	3L	25747899	0.68	3L-13	AGAP011506						
Ag_3L_027890931	3L	27890931	0.51	3L-14	AGAP011566						
Ag_3L_028103457	3L	28103457	0.52								
Ag_3L_029973214	3L	29973214	0.57	3L-15	AGAP011618						
Ag_3L_030153784	3L	30153784	0.73								
Ag_3L_033262927	3L	33262927	0.53	3L-16	AGAP011784	AGAP011785	AGAP011786				
Ag_3L_034218156	3L	34218156	0.66	3L-17							
Ag_3L_034222963	3L	34222963	0.75								
Ag_3L_034223030	3L	34223030	0.66								
Ag_3L_037418174	3L	37418174	0.52	3L-18	AGAP012075						
Ag_3L_039384389	3L	39384389	0.58	3L-19	AGAP012252						
Ag_3L_039575688	3L	39575688	0.55								
Ag_3L_041270221	3L	41270221	0.59	3L-20							
Ag_3L_041678356	3L	41678356	0.57								
Ag_3L_041718066	3L	41718066	0.54		AGAP012395	AGAP012396	AGAP012397	AGAP012398			
Ag_3R_005034029	3R	5034029	0.56	3R-1	AGAP008068						
Ag_3R_005034078	3R	5034078	0.55		AGAP008068						
Ag_3R_008966312	3R	8966312	0.74	3R-2	AGAP008318	AGAP008319					
Ag_3R_009999019	3R	9999019	0.52	3R-3							
Ag_3R_013989461	3R	13989461	0.52	3R-4	AGAP008648						
Ag_3R_015038909	3R	15038909	0.54	3R-5							
Ag_3R_015129848	3R	15129848	0.58		AGAP028214						
Ag_3R_015637096	3R	15637096	0.55		AGAP008702						
Ag_3R_015754624	3R	15754624	0.61								
Ag_3R_015978046	3R	15978046	0.62		AGAP008712						
Ag_3R_017548264	3R	17548264	0.50	3R-6							
Ag_3R_017824446	3R	17824446	0.53								
Ag_3R_018435599	3R	18435599	0.62		AGAP008819						
Ag_3R_025188436	3R	25188436	0.53	3R-7							
Ag_3R_035422418	3R	35422418	0.55	3R-9	AGAP009531	AGAP009532					
Ag_3R_035658134	3R	35658134	0.52								
Ag_3R_037687542	3R	37687542	0.53	3R-10	AGAP009661						
Ag_3R_043677054	3R	43677054	0.55	3R-11	AGAP009804	AGAP009805					
Ag_3R_044167320	3R	44167320	0.54		AGAP009848						
Ag_3R_047682861	3R	47682861	0.56	3R-12	AGAP010015						
Ag_3R_048730013	3R	48730013	0.62	3R-13							
Ag_3R_049768031	3R	49768031	0.51	3R-14							
Ag_3R_051193999	3R	51193999	0.53	3R-15							
Ag_3R_052001389	3R	52001389	0.53		AGAP010281						
Ag_X_002405425	X	2405425	0.53	X-1	AGAP000150						
Ag_X_010434384	X	10434384	0.63	X-2	AGAP000585						
Ag_X_010651239	X	10651239	0.64		AGAP000596	AGAP000595					
Ag_X_010922663	X	10922663	0.73								
Ag_X_017126779	X	17126779	0.55	X-3	AGAP000907	AGAP000908					
Ag_X_019053317	X	19053317	0.52	X-4	AGAP000998	AGAP001002	AGAP001004				
Ag_X_020760306	X	20760306	0.60	X-5							
Ag_X_020977056	X	20977056	0.57		AGAP001052						

Table 5.1: *An. gambiae* SNP loci associated with the outcome of *S. marcescens* infection. 140 SNPs with MAF difference >0.5 between the *S. marcescens* infection phenotypic pools were considered associated with the outcome of *S. marcescens* infection. The SNP column shows the ID of the respective SNP as designated in the *AgSNP01* genotyping dataset (Neafsey et al., 2010) along with the chromosomal arm it resides and its position (Chromosome-Position columns). The MAF Difference (MAF Diff) column shows the difference between the measured MAFs of the SNP genotyping arrays corresponding to the highly and non-infected phenotypic pools for the respective SNP. The peak each SNP resides corresponds to the designation in Figure 5.3. Gene hits show the Gene IDs of *An. gambiae* annotated genes, based on the *AgamP3.7* annotation that reside within a 5 kb radius of the respective SNP.

Chromosome	Center	Start	Stop	Average MAF Difference	P-value	Peak	Gene hits	
2L	3519452	3518391	3520513	0.19	1.00E-06	2L-1	AGAP004783	AGAP004784
2L	7047259	7045403	7049115	0.19	1.20E-06	2L-3		
2L	9983550	9981909	9985191	0.19	1.00E-06	2L-4	AGAP005096	
2L	10562133	10560684	10563583	0.18	3.70E-06			
2L	10573387	10572072	10574703	0.18	7.00E-06		AGAP005147	
2L	11376893	11375261	11378526	0.18	8.40E-06		AGAP005174	AGAP005175
2L	21675710	21674481	21676940	0.19	1.30E-06	2L-8		
2L	41101131	41098434	41103829	0.18	4.00E-06	2L-13		
2L	42507399	42505794	42509004	0.19	1.10E-06	2L-14		
2L	42520218	42518241	42522196	0.19	1.00E-06			
2L	42523530	42522314	42524746	0.18	4.50E-06			
2L	42542456	42539544	42545369	0.21	5.00E-07		AGAP007092	
2L	42552900	42551223	42554578	0.18	2.80E-06		AGAP007093	
2L	42570829	42567723	42573935	0.21	1.98E-07		AGAP007095	
2L	42615484	42613716	42617252	0.19	2.50E-06		AGAP007099	
2L	48678293	48675337	48681250	0.23	1.00E-07	2L-18		
2R	304340	303431	305250	0.18	4.20E-06	2R-1	AGAP001111	
2R	5857012	5856205	5857819	0.17	9.80E-06	2R-3	AGAP013478	AGAP001528
2R	10947314	10945724	10948904	0.20	9.00E-07	2R-4	AGAP001824	
2R	21981715	21980967	21982464	0.18	6.90E-06	2R-7	AGAP002492	
2R	22942308	22940286	22944330	0.18	8.50E-06			
2R	23888710	23885420	23892001	0.20	8.00E-07	2R-8	AGAP013684	
2R	24803393	24801383	24805404	0.21	1.98E-07			
2R	28287819	28286484	28289155	0.19	1.00E-06	2R-10		
2R	32803261	32800760	32805762	0.21	1.00E-07	2R-11		
2R	35004447	35000752	35008143	0.18	3.70E-06	2R-12		
2R	55469472	55468530	55470415	0.19	1.80E-06	2R-15	AGAP004375	
3L	9424852	9420247	9429458	0.18	3.60E-06	3L-6		
3L	26357281	26355918	26358645	0.18	3.50E-06	3L-13		
3L	33263629	33262613	33264646	0.19	1.00E-06	3L-16	AGAP011785	AGAP011786
3L	34221906	34219843	34223969	0.18	5.80E-06	3L-17		
3R	5032142	5028219	5036066	0.19	1.00E-06	3R-1		
3R	10229439	10227872	10231006	0.18	8.20E-06	3R-3	AGAP008384	
3R	10329068	10327571	10330566	0.18	4.20E-06		AGAP008399	
3R	14547533	14538128	14556939	0.18	3.80E-06	3R-4		
3R	15173263	15170968	15175559	0.23	1.00E-07	3R-5		
3R	15979321	15972534	15986108	0.19	1.00E-06		AGAP008712	
3R	17828600	17824124	17833076	0.19	1.00E-06	3R-6		
3R	18435501	18432543	18438459	0.20	9.00E-07		AGAP008819	
3R	33161118	33159170	33163067	0.19	1.30E-06	3R-8		
3R	42927106	42924488	42929725	0.26	1.00E-07	3R-11		
3R	43212443	43210733	43214154	0.19	1.20E-06		AGAP009777	AGAP009778
3R	47609353	47608641	47610066	0.19	2.40E-06	3R-12	AGAP010012	
X	10643881	10642989	10644773	0.18	3.60E-06	X-2	AGAP000594	

Table 5.2: *An. gambiae* 10-SNP windows associated with the outcome of *S. marcescens* infection. The measured MAF differences between the SNP genotyping arrays corresponding of highly and non-infected phenotypic pools following *S. marcescens* infection were used in a permutation analysis in which the average MAF difference of 10 adjacent SNPs was compared to a null distribution of the average MAF difference of 10 random SNPs. Association was assessed for 40,007 sliding 10-SNP windows. 44 windows with significant p-values were considered associated with the outcome of *S. marcescens* infection. The genomic area delineated by each 10-SNP window is shown (Chromosome-Center-Start-Stop columns), the average MAF difference of the 10 included SNPs and the Bonferroni-corrected p-value of the permutation analysis is shown in the respective columns. The peak each 10-SNP window resides corresponds to the designation in Figure 5.3. Gene hits include the Gene IDs of annotated *An. gambiae* genes based on the *AgamP3.7* annotation that reside within the genomic area corresponding to the respective 10-SNP window.

Gene ID	Peak	SNP/Permutation analysis	Description
AGAP000150	X-1	SNP	Unknown
AGAP000585	X-2	SNP	Zinc finger
AGAP000594	X-2	Permutation	Unknown
AGAP000595	X-2	SNP	Phospholipid/glycerol acyltransferase
AGAP000596	X-2	SNP	Phospholipid/glycerol acyltransferase
AGAP000907	X-3	SNP	AMP dependent ligase
AGAP000908	X-3	SNP	Dynactin p62 subunit
AGAP000998	X-4	SNP	Mannose 6-phosphate/insulin receptor
AGAP001002	X-4	SNP	Toll-like receptor
AGAP001004	X-4	SNP	TOLL1A
AGAP001052	X-5	SNP	Ubiquitin specific protease
AGAP001111	2R-1	Permutation	Alpha-glucosidase
AGAP001179	2R-1	SNP	Nucleoporin
AGAP001446	2R-3	SNP	AMP-activated protein kinase,
AGAP001528	2R-3	Permutation	Unknown
AGAP001824	2R-4	Permutation	FN3D3
AGAP001944	2R-5	SNP	G-protein beta WD-40 repeat
AGAP002094	2R-6	SNP	Cyclic nucleotide-binding domain
AGAP002492	2R-7	Permutation	DNA binding protein
AGAP002664	2R-8	SNP	Mediator of RNA polymerase II transcription subunit 27
AGAP002665	2R-8	SNP	Vacuolar membrane protein
AGAP002666	2R-8	SNP	Dolichyl pyrophosphate phosphatase
AGAP002667	2R-8	SNP	Translationally-controlled tumor protein homolog
AGAP004032	2R-13	SNP	Alpha mannosidase
AGAP004208	2R-14	SNP	Kinectin
AGAP004223	2R-14	SNP	GPR5HT7 putative serotonin 5HT-7 receptor
AGAP004375	2R-15	Permutation	Ricin B lectin
AGAP004405	2R-15	SNP	Leucine-rich repeat
AGAP004783	2L-1	SNP, Permutation	GPCR Rhodopsin 7TM
AGAP004784	2L-1	SNP, Permutation	Fatty acyl-CoA reductase
AGAP004931	2L-2	SNP	NTKL-Binding
AGAP005096	2L-4	Permutation	Homeobox
AGAP005143	2L-4	SNP	Ubiquitin conjugating enzyme 7 interacting protein
AGAP005147	2L-4	SNP, Permutation	FN3D1
AGAP005156	2L-4	SNP	ARID/BRIGHT DNA-binding domain
AGAP005174	2L-4	Permutation	Nucleoporin
AGAP005175	2L-4	Permutation	Acetyl-coA carboxylase
AGAP005203	2L-5	SNP	PGRPLC
AGAP005205	2L-5	SNP	PGRPLA
AGAP005206	2L-5	SNP	Unknown
AGAP005216	2L-5	SNP	Phosphatidylinositol-4-phosphate 5-kinase
AGAP005217	2L-5	SNP	Unknown
AGAP005244	2L-5	SNP	DNA-binding HTH Homeodomain-like
AGAP005287	2L-5	SNP	Zinc finger
AGAP005473	2L-6	SNP	Zinc finger
AGAP005474	2L-6	SNP	Unknown
AGAP005475	2L-6	SNP	Unknown
AGAP005476	2L-6	SNP	Unknown
AGAP005477	2L-6	SNP	Unknown
AGAP005478	2L-6	SNP	Unknown
AGAP005479	2L-6	SNP	Zinc finger
AGAP005497	2L-6	SNP	Unknown
AGAP005661	2L-7	SNP	Betha Ftz transcription factor isoform A (nuclear receptor)
AGAP006186	2L-9	SNP	Calcium-transporting ATPase sarcoplasmic/endoplasmic reticulum type
AGAP006405	2L-10	SNP	Tyrosine-protein kinase
AGAP006428	2L-10	SNP	Methyl-CpG DNA binding
AGAP006643	2L-11	SNP	Leucine-rich transmembrane protein
AGAP006644	2L-11	SNP	Leucine-rich repeat
AGAP006784	2L-12	SNP	F-box protein FBX9, putative

AGAP006785	2L-12	SNP	BTB/POZ domain-containing protein
AGAP007012	2L-13	SNP	Na,K-ATPase Interacting
AGAP007045	2L-13	SNP	LRIM15
AGAP007092	2L-14	Permutation	FN3D2
AGAP007093	2L-14	Permutation	Goodpasture antigen-binding protein
AGAP007095	2L-14	Permutation	Nhl repeat-containing protein, putative
AGAP007099	2L-14	Permutation	Shavenoid family protein
AGAP007104	2L-14	SNP	Farnesyl pyrophosphate synthetase
AGAP007148	2L-14	SNP	Cuticular protein unclassified
AGAP007149	2L-14	SNP	Unknown
AGAP007150	2L-14	SNP	Unknown
AGAP007151	2L-14	SNP	Unknown
AGAP007152	2L-14	SNP	Unknown
AGAP007153	2L-14	SNP	Unknown
AGAP007154	2L-14	SNP	Unknown
AGAP007181	2L-14	SNP	Supervillin
AGAP007291	2L-15	SNP	Inhibitor of Apoptosis 4
AGAP007292	2L-15	SNP	Inhibitor of Apoptosis 3
AGAP007387	2L-16	SNP	DNA methyltransferase 1-associated protein 1
AGAP007388	2L-16	SNP	AAA+ ATPase domain
AGAP007415	2L-16	SNP	ML12
AGAP007550	2L-17	SNP	Alpha carbonic anhydrase
AGAP007657	2L-18	SNP	Syndecan
AGAP007723	2L-18	SNP	Pleckstrin homology domain
AGAP008068	3R-1	SNP	tRNA-Thr
AGAP008318	3R-2	SNP	Sperm-tail PG-rich repeat
AGAP008319	3R-2	SNP	Ser/Thr kinase
AGAP008384	3R-3	Permutation	PDZ domain
AGAP008399	3R-3	Permutation	Unknown
AGAP008648	3R-4	SNP	Unknown
AGAP008702	3R-5	SNP	GPRNPR2
AGAP008712	3R-5	SNP, Permutation	Organic anion transporting polypeptide 26F
AGAP008819	3R-6	SNP, Permutation	EGFR
AGAP009531	3R-9	SNP	Amino acid transporter
AGAP009532	3R-9	SNP	Unknown
AGAP009661	3R-10	SNP	GPRNNB2
AGAP009777	3R-11	Permutation	SMAD domain, Dwarfin-type
AGAP009778	3R-11	Permutation	Ankyrin repeat
AGAP009804	3R-11	SNP	Gr10
AGAP009805	3R-11	SNP	Gr9
AGAP009848	3R-11	SNP	Pleckstrin homology domain
AGAP010012	3R-12	Permutation	Leucine-rich repeat
AGAP010015	3R-12	SNP	Peptidase S1A, chymotrypsin-type
AGAP010281	3R-15	SNP	GPRGBB1
AGAP010394	3L-1	SNP	Patched domain-containing protein
AGAP010503	3L-4	SNP	Potassium channel, calcium-activated
AGAP010798	3L-7	SNP	Peptidase S1, trypsin family, active site
AGAP010940	3L-8	SNP	EGF-like, conserved site
AGAP010941	3L-8	SNP	Zinc finger
AGAP011147	3L-9	SNP	DEAD box ATP-dependent RNA helicase
AGAP011148	3L-9	SNP	Mitochondrial ribosomal protein, L14, putative
AGAP011149	3L-9	SNP	ATP-dependent RNA helicase
AGAP011150	3L-9	SNP	GrpE nucleotide exchange factor
AGAP011151	3L-9	SNP	Src homology 2 (SH2) domain
AGAP011277	3L-10	SNP	FREP6
AGAP011278	3L-10	SNP	GALE4
AGAP011363	3L-11	SNP	Small GTPase superfamily, Rab type
AGAP011415	3L-12	SNP	Chitin binding domain
AGAP011506	3L-13	SNP	Insect cuticle protein
AGAP011566	3L-14	SNP	Zinc finger
AGAP011618	3L-15	SNP	Dopamine beta hydroxylase
AGAP011784	3L-16	SNP	Unknown

AGAP011785	3L-16	SNP, Permutation	CLIPE6
AGAP011786	3L-16	SNP, Permutation	CLIPE7
AGAP012075	3L-18	SNP	Unknown
AGAP012252	3L-19	SNP	Protein C kinase 53E
AGAP012395	3L-20	SNP	Peptide methionine sulfoxide reductase, putative
AGAP012396	3L-20	SNP	Unknown
AGAP012397	3L-20	SNP	DNA-directed RNA polymerase subunit rpb8
AGAP012398	3L-20	SNP	COP9 signalosome complex subunit 7
AGAP013086	2R-8	SNP	Unknown
AGAP013110	2R-8	SNP	Unknown
AGAP013225	2R-8	SNP	Unknown
AGAP013322	2R-8	SNP	Unknown
AGAP013353	2R-8	SNP	Unknown
AGAP013456	2R-8	SNP	Unknown
AGAP013478	2R-3	Permutation	Glyoxylate/hydroxypyruvate reductase
AGAP013684	2R-8	Permutation	miRNA
AGAP028214	3R-5	SNP	Unknown

Table 5.3: *An. gambiae* genes associated with the outcome of *S. marcescens* infection. 138 *An. gambiae* genes were associated with the outcome of *S. marcescens* infection, either by residing within a 5 kb radius of individual SNPs with MAF difference >0.5 (SNP) or by residing within genomic areas of 10-SNP windows with a p-value of $<10^{-5}$ in the conducted permutation analysis (Permutation). For genes identified by both approaches, both SNP and Permutation are noted. Peaks refer to genomic areas numbered in Figure 5.3, along with the chromosomal arm they reside. The Description column includes the assigned name for the gene, if any, or a description based on *Drosophila* homologues and *Interpro* domains.

Transcript ID	Name/Description	*SmA2* Regulation	*SmA3i* Regulation
AGAP006795-RA	Aper1	-0.94	-1.09
AGAP006434-RA	Aper57	-3.13	-0.21
AGAP006793-RA	Chitin-binding	-1.16	-0.68
AGAP001203-RA	Chitin-binding	-1.20	-1.92
AGAP010897-RA	FN3	-1.01	-0.78
AGAP010759-RA	FREP32	-1.15	-0.62
AGAP006414-RA	Glycoside hydrolase	-0.52	-0.92
AGAP006423-RA	Glycoside hydrolase	-0.62	-0.85
AGAP002635-RB	Gr13	-2.07	-1.49
AGAP004727-RA	Gr25	-0.75	-0.84
AGAP001137-RA	Gr59	-0.09	-1.29
AGAP003233-RA	Haem peroxidase	-1.50	-0.91
AGAP011167-RA	Haemolymph juvenile hormone binding	-0.22	-1.73
AGAP008182-RA	Haemolymph juvenile hormone binding	-0.96	-2.20
AGAP009383-RA	Haemolymph juvenile hormone binding	-0.61	-1.40
AGAP009386-RA	Haemolymph juvenile hormone binding	-0.72	-1.45
AGAP005041-RA	Homeobox_metazoa	-1.34	-1.75
AGAP009500-RA	Homeodomain-like	-1.25	-1.10
AGAP007291-RA	IAP4	-0.28	-0.96
AGAP004879-RA	Immunoglobulin E-set	-1.47	-0.67
AGAP007345-RA	LYSC3	-1.34	-1.27
AGAP001956-RA	ML28	-1.15	-1.58
AGAP002847-RA	ML3	-1.20	-1.04
AGAP002804-RA	ML4	-2.37	-1.48
AGAP002850-RA	ML5	-0.32	-1.16
AGAP002848-RA	ML9	-1.25	-0.49
AGAP002905-RA	OBP13	-3.55	-4.20
AGAP006080-RA	OBP54	-2.08	-1.27
AGAP011368-RA	OBP57	-0.85	-1.32
AGAP007684-RA	Peptidase	-0.32	-1.55
AGAP011442-RA	Peptidase	-1.77	-0.87
AGAP008290-RA	Peptidase	-2.72	-0.61
AGAP008291-RA	Peptidase	-2.47	-0.46
AGAP004770-RA	Peptidase	-0.94	-0.81
AGAP005196-RA	Peptidase	-0.25	-1.31
AGAP005310-RA	Peptidase	-1.56	-0.46
AGAP011984-RA	Ricin B lectin	-1.19	-0.98
AGAP010132-RA	SCRBQ1	-0.63	-0.95
AGAP005246-RD	SRPN10	-0.47	-0.77
AGAP009212-RA	SRPN6	-0.73	-1.01
AGAP006974-RA	TOLL9	-0.91	-0.61
AGAP010290-RA	Wnt5	-0.43	-0.77

Table 6.1: Genes of interest found in a cluster of mostly downregulated transcripts between the *SmA2*-*SmA3i* datasets, stemming from midguts of *S. marcescens* infected and uninfected mosquitoes. Genes of interest with a putative immune function found in cluster 7 (Figure 6.5) of co-regulated transcripts are shown. The Transcript ID corresponding to the identified transcript is shown along with the assigned name or description of the homologies of the corresponding gene, based on *Interpro*-predicted domains, assigned GO terms or *Drosophila* homologues and the determined log2-transformed fold-change regulation in the respective dataset

Transcript ID	Name/Description	*SmA2* Regulation	*SmA3i* Regulation
AGAP007033-RA	APL1C	0.24	1.17
AGAP007503-RA	ARID/BRIGHT DNA binding	0.44	1.04
AGAP012252-RA	classical protein kinase C sensory perception	0.32	0.98
AGAP011791-RA	CLIPA1	0.33	1.16
AGAP011790-RB	CLIPA2	0.30	1.31
AGAP010731-RA	CLIPA8	0.16	1.70
AGAP009214-RA	CLIPB11	0.07	1.63
AGAP010833-RA	CLIPB14	1.18	2.79
AGAP001648-RA	CLIPB17	0.51	2.20
AGAP009215-RA	CLIPB18	-0.05	1.86
AGAP012037-RA	CLIPB20	0.48	1.78
AGAP013184-RA	CLIPB36	0.35	1.98
AGAP004811-RA	CTL1	1.32	0.97
AGAP002625-RB	CTL9	0.47	0.82
AGAP002417-RA	CYP4AR1	1.24	0.94
AGAP000877-RA	CYP4G17	0.28	0.94
AGAP006048-RA	CYP4J5	0.23	0.92
AGAP008213-RA	CYP6M3	1.46	0.81
AGAP008210-RA	CYP6N1	2.31	1.14
AGAP002865-RA	CYP6P3	2.18	2.11
AGAP012296-RA	CYP9J5	1.13	1.04
AGAP005901-RA	Ect4	0.02	1.42
AGAP012395-RA	oxide reductase	0.27	1.04
AGAP011228-RA	FREP24	0.94	1.77
AGAP005848-RA	FREP44	0.21	1.98
AGAP006426-RA	Glycoside hydrolase	0.17	2.44
AGAP013120-RA	Glycoside hydrolase	0.51	1.07
AGAP009049-RA	Glycoside hydrolase	0.38	1.10
AGAP002318-RA	Glycoside hydrolase	0.18	1.02
AGAP002317-RA	Glycoside hydrolase	0.06	1.14
AGAP013260-RA	Glycoside hydrolase	0.13	1.01
AGAP013149-RA	GPROP1	0.01	1.24
AGAP001170-RA	Gr48	0.00	1.33
AGAP013533-RA	Haemolymph juvenile hormone binding	1.29	1.16
AGAP001352-RA	Haemolymph juvenile hormone binding	0.46	1.08
AGAP002341-RA	Haemolymph juvenile hormone binding	0.29	1.16
AGAP000190-RA	Homeobox	0.30	0.85
AGAP002372-RA	Homeobox	0.60	0.79
AGAP004662-RA	Homeobox	0.48	0.73
AGAP013287-RA	Immunoglobulin E-set	0.27	0.99
AGAP001350-RA	Immunoglobulin I-set	0.87	0.93
AGAP001826-RA	Lipophorin (Lp)	0.48	1.13
AGAP006348-RA	LRIM1	0.95	2.11
AGAP007455-RA	LRIM10	0.82	0.83

Transcript ID	Name/Description	*SmA2* Regulation	*SmA3i* Regulation
AGAP007037-RA	LRIM3	0.65	0.88
AGAP007039-RA	LRIM4	0.41	1.75
AGAP007457-RA	LRIM7	0.30	0.92
AGAP007347-RA	LYSC1	0.30	1.64
AGAP007343-RA	LYSC2	1.74	3.61
AGAP008648-RA	Methyltransferase	0.83	0.95
AGAP002852-RA	ML8	0.73	0.81
AGAP008255-RA	NOS	0.57	1.90
AGAP001409-RA	OBP3	0.23	1.34
AGAP001960-RA	Peptidase	0.45	0.98
AGAP006670-RA	Peptidase	0.93	1.28
AGAP005728-RA	Peptidase	0.30	0.84
AGAP004860-RA	Peptidase	0.01	1.08
AGAP011737-RA	Peptidase	0.35	0.92
AGAP001026-RA	Peptidase	0.61	0.99
AGAP007904-RA	Peptidase	0.61	1.22
AGAP007904-RA	Peptidase	0.61	1.22
AGAP011654-RA	Peptidase	1.12	1.83
AGAP005669-RA	Peptidase	0.39	2.51
AGAP003689-RA	Peptidase	0.53	1.59
AGAP011477-RA	Peptidase	0.93	1.15
AGAP013221-RA	Peptidase	0.90	3.40
AGAP005690-RA	Peptidase	0.82	3.46
AGAP005689-RA	Peptidase	0.70	3.15
AGAP001251-RA	Peptidase	1.12	1.11
AGAP009849-RA	Peptidase	0.17	2.06
AGAP009849-RA	Peptidase	0.17	2.06
AGAP005686-RA	Peptidase	0.00	1.71
AGAP013252-RA	Peptidase	0.45	1.09
AGAP005195-RA	Peptidase	0.57	0.80
AGAP010545-RA	Peptidase	0.41	0.89
AGAP007043-RA	Peptidase	0.32	0.93
AGAP005194-RA	Peptidase	0.24	0.87
AGAP010730-RA	Peptidase	0.43	1.09
AGAP011325-RA	Peptidase	0.00	3.20
AGAP005203-RA	PGRPLC	0.86	0.89
AGAP008782-RA	Ricin B lectin	0.22	1.14
AGAP005625-RA	SCRASP1	0.26	1.08
AGAP009213-RA	SRPN16	0.17	1.27
AGAP006910-RA	SRPN3	0.37	0.97
AGAP010815-RA	TEP1	0.19	1.37
AGAP008654-RA	TEP12	1.00	1.30
AGAP010816-RA	TEP3	0.36	1.27
AGAP010812-RA	TEP4	0.08	1.40
AGAP004262-RA	TO3	0.96	1.13
AGAP010669-RA	TOLL5B	0.82	0.85

Table 6.2: Genes of interest found in the cluster *SmA*, corresponding to mostly upregulated transcripts between the *SmA2*-*SmA3i* datasets, following *S. marcescens* infection. Genes of interest with a putative immune function found in cluster *SmA* of co-regulated transcripts are shown. The Transcript ID corresponding to the identified transcript is shown along with the assigned name or description of the homologies of the corresponding gene, based on *Interpro*-predicted domains, assigned GO terms or *Drosophila* homologues and the determined log2-transformed fold-change regulation in the respective dataset.

Transcript ID	Name/Description	*SmB2i* Regulation	*SmB3* Regulation
AGAP001446-RA	AMP-activated protein kinase starvation-induced	0.03	1.19
AGAP007036-RA	APL1A	0.42	0.69
AGAP006785-RA	BTB POZ	0.82	1.89
AGAP000359-RA	Chitin binding	0.60	0.74
AGAP012591-RA	CLIPA3	0.57	0.68
AGAP002813-RA	CLIPD6	1.31	1.58
AGAP011785-RA	CLIPE6	0.94	1.17
AGAP005335-RA	CTL4	0.38	0.87
AGAP010196-RA	CTLGA1	0.49	1.44
AGAP007411-RA	CTLMA1	0.46	1.19
AGAP008022-RA	CYP12F1	-0.11	0.56
AGAP010077-RA	CYP303A1	0.84	0.67
AGAP001076-RA	CYP4G16	0.72	0.43
AGAP000088-RA	CYP4H19	0.49	0.73
AGAP013490-RA	CYP4H24	0.00	0.00
AGAP009375-RA	CYP9M2	0.63	0.95
AGAP011294-RA	DEF1	1.92	1.87
AGAP010884-RA	Dscam	0.51	0.89
AGAP001944-RA	F-box and WD-40 domain behaviour defective	1.54	1.79
AGAP001633-RA	FN3	0.32	0.79
AGAP002154-RB	FN3	0.80	1.25
AGAP006721-RA	FN3	0.33	0.82
AGAP001824-RA	FN3D3	0.40	1.19
AGAP011225-RA	FREP28	0.47	1.23
AGAP000143-RA	Glycoside hydrolase	0.28	1.05
AGAP001111-RC	Glycoside hydrolase	0.86	0.23
AGAP005381-RC	Glycoside hydrolase	1.33	1.90
AGAP006371-RA	Glycoside hydrolase	0.38	0.96
AGAP009022-RA	Glycoside hydrolase	0.71	0.54
AGAP009621-RA	Glycoside hydrolase	0.44	0.68
AGAP012453-RA	Glycoside hydrolase	0.38	0.90
AGAP003658-RB	GPRALS1	0.61	0.76
AGAP004222-RB	GPRNNA20	0.38	0.81
AGAP008347-RA	GPRRK	0.55	1.19
AGAP002519-RA	GPRTYR	0.59	0.90
AGAP003254-RA	Gr16	1.06	0.58
AGAP001117-RG	Gr37	0.52	1.73
AGAP007756-RA	Gr46	0.40	1.02
AGAP006143-RE	Gr56	0.68	0.79
AGAP009805-RE	Gr9	0.87	0.76
AGAP004944-RA	Heat shock	0.60	1.14
AGAP006958-RA	Heat shock	1.03	0.09
AGAP010876-RA	Heat shock	0.34	0.87
AGAP003671-RA	Homeobox	0.55	1.11
AGAP005281-RA	Homeobox	0.48	0.60
AGAP005311-RA	Homeobox	0.53	0.53
AGAP005878-RA	Homeobox	0.41	1.31
AGAP000067-RA	Homeodomain-like	-0.01	1.63
AGAP003871-RA	Homeodomain-like	1.27	1.91
AGAP010700-RA	Homeodomain-like	0.61	0.87
AGAP011902-RA	Homeodomain-like	0.64	0.43
AGAP010735-RA	HPX12	0.40	0.87
AGAP000051-RA	HPX5	0.69	0.34
AGAP004038-RA	HPX8	0.75	0.29
AGAP008538-RA	Immunoglobulin	1.27	0.23
AGAP010746-RA	Immunoglobulin	0.63	0.89
AGAP002228-RA	Immunoglobulin E-set	0.00	0.46
AGAP005767-RA	Immunoglobulin E-set	0.44	0.87
AGAP011489-RA	Immunoglobulin I-set	0.00	0.34
AGAP009270-RA	Immunoglobulin V-set	0.96	0.59
AGAP002007-RA	Leucine rich repeat	0.00	0.82
AGAP003994-RA	Leucine rich repeat	0.00	0.58
AGAP006183-RA	Leucine rich repeat	0.00	0.39
AGAP007462-RA	Leucine rich repeat	0.00	0.50

Transcript ID	Name/Description	*SmB2i* Regulation	*SmB3* Regulation
AGAP009594-RA	Leucine rich repeat	0.00	0.53
AGAP012385-RA	Leucine rich repeat	0.00	0.32
AGAP007388-RA	midasin ATPase	0.46	0.73
AGAP002851-RA	ML6	0.63	0.53
AGAP011150-RA	molecular chaperone GrpE	0.98	0.90
AGAP000038-RA	gamma-aminobutyric acid receptor subunit beta	0.52	0.77
AGAP002974-RA	nicotinic acetylcholine receptor subunit alpha 1	0.17	1.06
AGAP000819-RA	nuclear receptor subfamily 2 (Tailless)	0.93	0.29
AGAP009413-RA	Or20	1.21	1.64
AGAP010505-RA	Or44	0.41	1.15
AGAP009706-RA	Or70	0.55	0.53
AGAP000829-RA	Peptidase	0.72	0.66
AGAP000994-RA	Peptidase	0.31	1.22
AGAP001190-RA	Peptidase	1.01	1.21
AGAP001241-RA	Peptidase	0.88	1.08
AGAP001245-RA	Peptidase	0.16	1.36
AGAP001330-RA	Peptidase	0.54	0.59
AGAP001330-RA	Peptidase	0.54	0.59
AGAP003475-RB	Peptidase	0.39	0.73
AGAP003643-RA	Peptidase	0.61	0.94
AGAP005043-RB	Peptidase	0.47	0.72
AGAP005301-RA	Peptidase	0.59	0.80
AGAP005591-RA	Peptidase	0.57	0.83
AGAP005591-RA	Peptidase	-0.22	0.57
AGAP005642-RA	Peptidase	0.50	0.94
AGAP005671-RA	Peptidase	1.18	-0.10
AGAP005691-RA	Peptidase	0.32	1.38
AGAP006616-RA	Peptidase	0.45	0.83
AGAP007252-RA	Peptidase	0.71	1.12
AGAP007600-RA	Peptidase	0.20	1.23
AGAP008022-RA	Peptidase	0.56	0.53
AGAP009252-RA	Peptidase	0.03	0.37
AGAP010617-RA	Peptidase	0.82	0.33
AGAP011040-RA	Peptidase	0.05	1.18
AGAP011430-RA	Peptidase	0.93	0.92
AGAP011653-RA	Peptidase	0.42	0.69
AGAP011917-RA	Peptidase	0.37	0.81
AGAP011920-RA	Peptidase	-0.18	0.65
AGAP012269-RA	Peptidase	0.11	1.45
AGAP012270-RA	Peptidase	0.24	1.17
AGAP012473-RA	Peptidase	0.67	0.45
AGAP012492-RA	Peptidase	0.74	0.90
AGAP012671-RA	Peptidase	0.33	1.13
AGAP012778-RA	Peptidase	0.00	0.82
AGAP013089-RA	Peptidase	-0.07	0.40
AGAP013181-RA	Peptidase	0.86	0.99
AGAP005203-RC	PGRPLC	0.58	0.69
AGAP005552-RB	PGRPLD	0.78	0.53
AGAP010078-RA	Ricin B lectin	0.04	1.32
AGAP004118-RA	SCRAL1	0.05	0.81
AGAP004847-RA	SCRB7	0.28	0.88
AGAP004847-RA	SCRB7	0.81	0.28
AGAP008318-RA	SHIPPO repeat sensory perception	1.50	1.81
AGAP007177-RA	SPZ5	1.47	1.70
AGAP012703-RA	TO2	0.73	1.57
AGAP001002-RA	Toll protein	0.57	1.24
AGAP005478-RA	Unknown	0.36	0.77
AGAP007149-RA	Unknown	0.20	0.98
AGAP013086-RA	Unknown	0.31	1.18
AGAP008369-RA	Vitellinogen	0.71	0.42
AGAP005287-RA	Zinc finger	1.36	0.75

Table 6.3: Genes of interest found in cluster *SmB*, corresponding to mostly upregulated transcripts between the *SmB2-SmB3i* datasets, stemming from midguts of *S. marcescens* infected and uninfected mosquitoes. Genes of interest with a putative immune function found in cluster *SmB* (Figure 6.5) of co-regulated transcripts are shown. The Transcript ID corresponding to the identified transcript is shown along with the assigned name or description of the homologies of the corresponding gene, based on *Interpro*-predicted domains, assigned GO terms or *Drosophila* homologues and the determined log2-transformed fold-change regulation in the respective d

Transcript ID	Name/Description	Functional class	Fold-change	Regulation	t-test p-value
AGAP005625-RB	SCRASP1	Chitin binding	1.76	up	0.04
AGAP001203-RA	Chitin binding	Chitin binding	2.05	down	0.03
AGAP006991-RA	Chitin binding	Chitin binding	2.77	up	0.07
AGAP006433-RA	AgAper34	Chitin binding	1.83	down	0.07
AGAP006434-RA	AgAper57	Chitin binding	2.36	down	0.10
AGAP002471-RA	ADP ribosylation factor	GPCR signaling	2.21	up	0.00
AGAP002635-RB	Gr13	GPCR signaling	2.13	down	0.01
AGAP002635-RA	Gr13	GPCR signaling	2.68	down	0.01
AGAP011046-RA	Deoxyribonuclease I	Hydrolase	1.75	up	0.03
AGAP011698-RA	Deoxyribonuclease I	Hydrolase	2.23	down	0.01
AGAP006426-RA	Glycoside hydrolase	Hydrolase	1.86	up	0.15
AGAP005381-RC	Beta-hexosaminidase	Hydrolase	1.94	up	0.08
AGAP006187-RA	Insect allergen G12	Insect allergen	2.27	up	0.03
AGAP012703-RA	takeout2	Juvenile hormone/pheromone binding	1.78	up	0.01
AGAP008052-RA	Insect pheromone-binding protein	Juvenile hormone/pheromone binding	1.77	down	0.00
AGAP008182-RA	Juvenile hormone-binding	Juvenile hormone/pheromone binding	2.06	down	0.05
AGAP007343-RA	LYSC2	Lysozyme	3.44	up	0.04
AGAP001956-RA	ML28	MD2 lipid recognition	2.36	down	0.00
AGAP002804-RA	ML4	MD2 lipid recognition	2.55	down	0.03
AGAP004642-RA	Neuropeptide F	Neuropeptide signaling	1.77	down	0.02
AGAP008898-RA	Mitochondrial ATPase inhibitor	Nucleotide metabolic process	1.81	down	0.11
AGAP012891-RA	Heat shock protein 70	Nucleotide metabolic process	2.02	down	0.06
AGAP007790-RA	Sodium/potassium-dependent ATPase	Nucleotide metabolic process	2.09	down	0.16
AGAP004583-RA	Heat shock protein 70	Nucleotide metabolic process	2.43	down	0.13
AGAP004582-RA	Heat shock protein 70	Nucleotide metabolic process	2.44	down	0.12
AGAP004581-RA	Heat shock protein 70	Nucleotide metabolic process	2.47	down	0.16
AGAP006080-RA	OBP54	Odorant binding	2.38	down	0.01
AGAP002905-RA	OBP13	Odorant binding	6.00	down	0.03
AGAP000946-RA	Sterol desaturase	Oxidoreductase	2.00	up	0.02
AGAP003581-RA	Alcohol dehydrogenase	Oxidoreductase	2.05	up	0.00
AGAP008210-RA	Cytochrome P450 CYP6N1	Oxidoreductase	2.11	up	0.03
AGAP002865-RA	Cytochrome P450 CYP6P3	Oxidoreductase	2.69	up	0.03
AGAP005501-RA	Oxidoreductase	Oxidoreductase	2.81	up	0.02
AGAP007621-RC	Cytochrome-c oxidase	Oxidoreductase	2.19	down	0.25
AGAP003209-RA	Sterol desaturase	Oxidoreductase	2.43	down	0.06
AGAP011654-RA	Microsomal dipeptidase	Peptidase other	2.00	up	0.03
AGAP011442-RA	Serine carboxypeptidase	Peptidase other	1.87	down	0.03
AGAP005203-RA	PGRPLC	Protein/receptor binding	1.75	up	0.00
AGAP001350-RA	Immunoglobulin-like	Protein/receptor binding	1.78	up	0.02
AGAP011228-RA	FREP24	Protein/receptor binding	1.87	up	0.04
AGAP006348-RA	LRIM1	Protein/receptor binding	1.88	up	0.03
AGAP011765-RA	F-spondin	Protein/receptor binding	1.90	up	0.04
AGAP005848-RA	FREP44	Protein/receptor binding	2.33	up	0.05
AGAP011482-RA	Serine proteinase inhibitor	Protein/receptor binding	2.52	up	0.02
AGAP006689-RA	Protein binding BTB/POZ-like	Protein/receptor binding	1.83	down	0.04
AGAP010762-RA	FREP4	Protein/receptor binding	1.86	down	0.01
AGAP010760-RA	FREP69	Protein/receptor binding	2.04	down	0.03
AGAP006367-RA	Zinc finger	Protein/receptor binding	2.16	down	0.03
AGAP000668-RA	Zinc finger	Protein/receptor binding	2.03	up	0.05

Transcript ID	Name/Description	Functional class	Fold-change	Regulation	t-test p-value
AGAP003025-RC	60S acidic ribosomal protein LP2	Ribosome	1.80	down	0.07
AGAP012037-RA	CLIPB20	Serine-type endopeptidase	1.75	up	0.04
AGAP007142-RA	Serine-type endopeptidase	Serine-type endopeptidase	1.78	up	0.01
AGAP011785-RA	CLIPE6	Serine-type endopeptidase	1.90	up	0.01
AGAP013221-RA	Trypsin putative	Serine-type endopeptidase	2.59	up	0.04
AGAP010833-RA	CLIPB14	Serine-type endopeptidase	2.76	up	0.02
AGAP004316-RA	Serine-type endopeptidase Disulphide knot CLIP	Serine-type endopeptidase	1.85	up	0.07
AGAP001648-RA	CLIPB17	Serine-type endopeptidase	1.86	up	0.07
AGAP005669-RA	Serine collagenase	Serine-type endopeptidase	1.94	up	0.08
AGAP012966-RA	Serine-type endopeptidase	Serine-type endopeptidase	1.96	up	0.05
AGAP006707-RA	Serine-type endopeptidase	Serine-type endopeptidase	2.10	up	0.10
AGAP011325-RA	Coagulation factor	Serine-type endopeptidase	2.20	up	0.13
AGAP005689-RA	Serine collagenase	Serine-type endopeptidase	2.49	up	0.07
AGAP005690-RA	Serine collagenase	Serine-type endopeptidase	2.79	up	0.05
AGAP008291-RA	Trypsin-5	Serine-type endopeptidase	1.82	down	0.05
AGAP001198-RA	Trypsin putative	Serine-type endopeptidase	2.12	down	0.06
AGAP008290-RA	Trypsin-6	Serine-type endopeptidase	2.42	down	0.05
AGAP004344-RA	Hairless Transcription corepressor activity	Transcription	1.88	down	0.04
AGAP005041-RA	Homeobox	Transcription	2.04	down	0.03
AGAP004382-RA	Glutathione S-transferase delta class GSTD3	Transferase	1.81	up	0.01
AGAP003759-RA	CHK kinase-like Juvenile hormone-inducible	Transferase	1.81	up	0.02
AGAP003490-RA	Serine-pyruvate aminotransferase	Transferase	1.82	up	0.00
AGAP003220-RA	CHK kinase-like Juvenile hormone-inducible	Transferase	1.94	up	0.02
AGAP003762-RA	CHK kinase-like Juvenile hormone-inducible	Transferase	2.04	up	0.02
AGAP004164-RD	Glutathione S-transferase delta class GSTD1	Transferase	2.31	up	0.02
AGAP003757-RA	CHK kinase-like	Transferase	2.47	up	0.01
AGAP006788-RA	Acyl CoA acyltransferase	Transferase	2.61	down	0.03
AGAP012672-RA	Sulfotransferase	Transferase	2.14	up	0.11
AGAP006222-RA	UDP glucosyl transferase	Transferase	2.23	up	0.08
AGAP002293-RA	Intraflaggelar transport	Transporter	1.84	up	0.00
AGAP005045-RA	Glucose transporter	Transporter	1.87	up	0.04
AGAP013063-RA	Synaptic vesicle protein	Transporter	1.90	up	0.03
AGAP002826-RA	Synaptic vesicle protein	Transporter	1.97	up	0.02
AGAP002043-RA	Sugar transporter	Transporter	2.00	up	0.01
AGAP013318-RA	Apolipoprotein D	Transporter	1.77	down	0.07
AGAP007267-RA	tRNA-Gly	tRNA	1.86	down	0.04
AGAP012794-RA	tRNA-Arg	tRNA	2.37	down	0.03
AGAP013453-RA	Unknown	Unknown	1.83	up	0.04
AGAP006200-RA	Unknown	Unknown	1.91	up	0.02
AGAP009309-RA	Unknown	Unknown	1.75	down	0.01
AGAP000570-RA	Unknown	Unknown	1.82	down	0.00
AGAP003281-RA	Unknown	Unknown	1.85	down	0.01
AGAP013493-RA	Unknown	Unknown	2.28	down	0.02
AGAP009146-RA	Unknown	Unknown	2.88	down	0.01
AGAP010502-RA	Unknown	Unknown	4.85	down	0.01
AGAP010502-RB	Unknown	Unknown	5.80	down	0.02
AGAP013005-RA	Unknown	Unknown	2.23	up	0.11
AGAP006404-RA	Unknown	Unknown	2.65	up	0.15
AGAP001508-RA	Unknown	Unknown	3.97	up	0.10
AGAP012552-RA	Unknown	Unknown	1.90	down	0.07

Table 6.4: List of more than 1.75-fold regulated transcripts following oral *S. marcescens* infection. 99 *An. gambiae* transcripts were identified as more than 1.75-fold regulated following *S. marcescens* infection, over 3 independent infections. For each differentially expressed transcript, the Transcript ID is shown along with the assigned name or description of the homologies of the corresponding gene, based on *Interpro*-predicted domains, assigned GO terms or *Drosophila* homologues. Furthermore, the functional class that the gene corresponding to each transcript has been assigned is shown, along with the transcriptional regulation determined over 3 independent infections, fold-change regulation and the computed p-value in a t-test against zero using the log2-transformed fold-change values for each probe corresponding to the respective transcript, over the 3 independent infections. Significance was assessed using a p-value cut-off of 0.05.

Transcript ID	Name/Description	*AsaA1ii* Regulation	*AsaA2i* Regulation
AGAP007033-RA	APL1C	1.38	1.13
AGAP005620-RA	Attacin C	0.53	0.78
AGAP007938-RA	Cactus	0.47	0.54
AGAP000987-RA	Chitin-binding	0.87	1.71
AGAP009405-RA	Chitin-binding	0.28	0.75
AGAP011791-RA	CLIPA1	0.93	0.97
AGAP003245-RA	CLIPA19	1.66	1.66
AGAP011790-RB	CLIPA2	0.8	0.82
AGAP011780-RA	CLIPA4	0.85	0.59
AGAP011787-RA	CLIPA5	0.56	0.56
AGAP010731-RA	CLIPA8	1.54	1.27
AGAP010968-RA	CLIPA9	0.85	0.93
AGAP003251-RA	CLIPB1	0.45	0.59
AGAP004855-RA	CLIPB13	0.44	0.42
AGAP009844-RA	CLIPB15	0.87	0.31
AGAP003247-RA	CLIPB19	1.15	1.24
AGAP003246-RA	CLIPB2	1.37	1.68
AGAP003057-RA	CLIPB8	0.34	0.66
AGAP002422-RA	CLIPD1	0.96	1.75
AGAP010530-RA	CLIPE4	1.38	1.07
AGAP010547-RA	CLIPE5	1.56	0.95
AGAP002625-RB	CTL9	0.3	0.65
AGAP006430-RB	CTLGA2	0.56	1.05
AGAP009200-RA	C-type lectin	0.74	0.23
AGAP000284-RA	CYP315A1	0.42	0.86
AGAP009241-RA	CYP4C36	1.35	0.88
AGAP001076-RB	CYP4G16	0.82	1.17
AGAP002416-RA	CYP4K2	0.66	0.18
AGAP008213-RA	CYP6M3	1.65	0.2
AGAP008210-RA	CYP6N1	0.64	1.28
AGAP008204-RA	CYP6S1	0.56	0.37
AGAP008208-RA	CYP6Y1	0.45	0.41
AGAP008350-RA	DBLOX Double Oxidase	0.77	0.4
AGAP007556-RA	FN3	0.4	0.66
AGAP011197-RA	FREP13	1.24	0.28
AGAP010531-RA	FREP2	2.31	0.64
AGAP011228-RA	FREP24	-0.25	2.13
AGAP005848-RA	FREP44	1.16	1.57
AGAP009556-RA	FREP50	0.84	1.98
AGAP011231-RA	FREP59	0.79	0.16
AGAP011239-RA	FREP60	1.09	0.18
AGAP009184-RA	FREP7	0.6	0.69
AGAP006191-RA	Glycoside hydrolase	1.15	1.96
AGAP002457-RA	Glycoside hydrolase	0.71	0.9
AGAP002156-RA	GPRGNR1	0.66	1.25
AGAP012427-RA	GPRNNA16	0.49	0.98
AGAP002076-RA	Heat shock	0.53	0.32
AGAP011065-RA	Homeobox	0.15	1.09
AGAP005281-RA	Homeobox	0.44	0.9
AGAP003669-RA	Homeodomain	0.45	0.81
AGAP004647-RA	Homeodomain	0.74	0.09
AGAP004036-RA	HPX7	0.27	0.96
AGAP009166-RA	IKK1/ird5 (IMD pathway)	0.67	0.54
AGAP002268-RA	Immunoglobulin	1.18	0.42
AGAP002737-RA	Immunoglobulin	0.7	0.24
AGAP001504-RA	Immunoglobulin	0.35	0.68
AGAP009098-RA	Immunoglobulin E-set	1.15	0.66
AGAP010221-RA	Immunoglobulin I-set	0.33	0.68
AGAP010600-RA	Insulin-like peptide 2	0.27	0.72
AGAP005055-RA	ITP Crust neurhormone	0.67	0.24
AGAP005870-RA	Leucine-rich repeat	0.8	0.32
AGAP005870-RA	Leucine-rich repeat	0.8	0.32
AGAP003878-RA	Leucine-rich repeat	1.19	1.2
AGAP011141-RA	Leucine-rich repeat	0.45	0.51
AGAP010012-RA	Leucine-rich repeat	0.5	0.55
AGAP001826-RA	Lipophorin (Lp)	1.16	1.33
AGAP006348-RA	LRIM1	1.37	0.83
AGAP007455-RA	LRIM10	1.81	1.08
AGAP005693-RA	LRIM17	1.17	1.6
AGAP007039-RA	LRIM4	1.19	1.54
AGAP007457-RA	LRIM7	1.11	0.53
AGAP007454-RA	LRIM8A	1.67	2.16
AGAP007456-RA	LRIM8B	1.22	2.84
AGAP007453-RA	LRIM9	1.09	1.9
AGAP007347-RA	LYSC1	-0.24	2.12
AGAP012352-RA	ML1	1.27	1.26
AGAP010059-RA	Neurotrans-gated channel	0.59	0.86
AGAP008382-RA	nuclear receptor subfamily 2 group C	0	1.29
AGAP004693-RA	nuclear receptor subfamily 6 group A	0.71	0.87
AGAP012113-RA	Peptidase	0.33	0.9
AGAP009593-RA	Peptidase	1.51	0.85
AGAP010351-RA	Peptidase	0.49	0.54
AGAP006610-RA	Peptidase	1.22	-0.03
AGAP004700-RA	Peptidase	1.13	0.55
AGAP007043-RA	Peptidase	1.34	1.14
AGAP013487-RA	Peptidase	0.19	0.5
AGAP003689-RA	Peptidase	0.46	1.21
AGAP003691-RA	Peptidase	0.65	0.3
AGAP004638-RA	Peptidase	0.32	0.7
AGAP010415-RA	Peptidase	1.24	0.51
AGAP010730-RA	Peptidase	1.05	0.85
AGAP012034-RA	Peptidase	1.05	0.6
AGAP005689-RA	Peptidase	0.63	0.58
AGAP005690-RA	Peptidase	1.3	0.41
AGAP006385-RA	Peptidase	1.31	-0.1
AGAP003248-RA	Peptidase	1.72	0.63
AGAP001365-RA	Peptidase	0.28	0.74
AGAP005434-RA	Peptidase	1.21	1.09
AGAP010545-RA	Peptidase	1.18	-0.01
AGAP012328-RA	Peptidase	0.48	0.61
AGAP009849-RA	Peptidase	1.46	1.01
AGAP010078-RA	Ricin B lectin	0.49	0.4
AGAP004844-RA	Socs44A (JAK/STAT pathway)	0.2	0.87
AGAP001377-RA	SRPN11	0.52	0.7
AGAP007692-RA	SRPN14	1	0.28
AGAP009213-RA	SRPN16	1.08	0.17
AGAP001376-RA	SRPN17	0.82	0.21
AGAP007691-RB	SRPN18	0.73	0.68
AGAP006911-RA	SRPN2	0.82	0.71
AGAP006910-RA	SRPN3	0.83	0.76
AGAP003194-RA	SRPN8	0.84	0.51
AGAP010815-RA	TEP1	0.94	0.96
AGAP008654-RA	TEP12	2.11	1.24
AGAP008368-RA	TEP14	1.31	0.69
AGAP010816-RA	TEP3	1.15	0.51
AGAP010812-RA	TEP4	0.44	0.81
AGAP004262-RA	TO3	0.83	0.66
AGAP010669-RA	TOLL5B	0.97	1.76
AGAP003062-RA	TUBE (Toll pathway)	0.39	0.71
AGAP008369-RA	Vitellinogen	0.94	1.92
AGAP008810-RA	Vitellinogen	0.47	0.85

Table 6.5: Genes of interest found in a cluster of mostly upregulated transcripts between the *AsaA1ii-AsaA2i* datasets, stemming from midguts of *Asaia* infected and uninfected mosquitoes. Genes of interest with a putative immune function found in cluster *AsaA* (Figure 6.9) of co-regulated transcripts are shown. The Transcript ID corresponding to the identified transcript is shown along with the assigned name or description of the homologies of the corresponding gene, based on *Interpro*-predicted domains, assigned GO terms or *Drosophila* homologues and the determined log2-transformed fold-change regulation in the respective dataset

Transcript ID	Name/Description	*AsaB1* Regulation	*AsaB2* Regulation
AGAP002629-RA	Chitin-binding	-0.44	-0.88
AGAP010364-RA	Chitin-binding	-0.58	-1.03
AGAP011617-RA	Chitin-binding	-1.31	-1.96
AGAP009217-RA	CLIPB12	-0.67	-0.65
AGAP002270-RB	CLIPB7	-0.33	-0.89
AGAP011782-RA	CLIPE2	-1.39	-1.97
AGAP000443-RA	CTL5	-0.50	-0.46
AGAP002912-RA	C-type lectin	-0.55	-0.39
AGAP012294-RA	CYP9L2	-0.17	-0.90
AGAP009363-RA	CYP9M1	-0.40	-0.96
AGAP006028-RC	GABA-gated chloride channel subunit	-0.64	-0.25
AGAP004934-RA	GALE3	-2.03	-1.14
AGAP000708-RA	Glycoside hydrolase	-0.42	-0.62
AGAP008965-RA	Glycoside hydrolase	-0.42	-1.10
AGAP010486-RA	GPRALS3	-0.25	-1.07
AGAP009661-RA	GPRNNB2	-0.36	-0.78
AGAP001592-RB	GPRTAK2	-0.87	-0.83
AGAP009256-RD	Gr44	-0.40	-1.76
AGAP005096-RA	Homeobox	-0.25	-0.79
AGAP004696-RA	Homeobox	-0.37	-0.68
AGAP004659-RA	Homeobox	-0.89	-1.30
AGAP005878-RB	Homeodomain	-0.49	-0.42
AGAP010735-RA	HPX12	-1.15	-0.14
AGAP01081C-RA	HPX14	-0.70	-0.46
AGAP011326-RA	IAP2	-0.46	-1.26
AGAP010604-RA	Insulin-like peptide 3 precursor	-0.38	-0.66
AGAP010605-RA	Insulin-like peptide 3 precursor	-0.71	-0.68
AGAP011947-RA	Leucine-rich repeat	-0.35	-0.69
AGAP007385-RA	LYSC4	-1.14	-1.32
AGAP007346-RA	LYSC5	-0.18	-1.20
AGAP002804-RA	ML4	-0.94	-0.11
AGAP006988-RA	Neurospecific receptor kinase	-0.51	-0.63
AGAP000966-RA	Nicotinic acetylcholine receptor	-0.80	-0.21
AGAP010057-RA	Nicotinic acetylcholine receptor	-1.43	-1.31
AGAP008255-RA	NOS	-1.24	-1.38
AGAP009519-RA	Or2	-0.54	-0.98
AGAP010504-RA	Or43	-0.40	-0.71
AGAP011813-RA	Or54	-1.36	0.09
AGAP000884-RA	Peptidase	-1.48	-0.35
AGAP010834-RA	Peptidase	-0.72	-0.31
AGAP012068-RA	Peptidase	-1.04	-0.94
AGAP001026-RB	Peptidase	-0.79	-0.61
AGAP011434-RA	Peptidase	-1.37	-1.13
AGAP004563-RA	Peptidase	-0.45	-0.71
AGAP004775-RB	Peptidase	-0.24	-1.33
AGAP001931-RA	Peptidase	-0.55	-0.79
AGAP01187C-RA	Peptidase	-0.72	-0.65
AGAP005686-RA	Peptidase	-0.68	-0.26
AGAP001199-RA	Peptidase	-1.17	-0.76
AGAP004639-RA	Peptidase	-0.37	-0.67
AGAP011914-RA	Peptidase	-0.63	-0.79
AGAP008861-RA	Peptidase	-0.69	-1.86
AGAP008911-RA	Peptidase	-0.49	-0.43
AGAP007262-RA	Peptidase	-0.82	-0.89
AGAP002842-RA	Peptidase	-0.22	-0.88
AGAP006485-RA	Peptidase	0.02	-1.43
AGAP001212-RB	PGRPLB	-0.54	-1.14
AGAP009212-RA	SRPN6	0.01	-1.38
AGAP009731-RA	Wnt10	-0.15	-1.17

Table 6.6: Genes of interest found in a cluster of mostly downregulated transcripts between the *AsaB1-AsaB2* datasets, stemming from midguts of *Asaia* infected and uninfected mosquitoes. Genes of interest with a putative immune function found in cluster 8 (Figure 6.9) of co-regulated transcripts are shown. The Transcript ID corresponding to the identified transcript is shown along with the assigned name or description of the homologies of the corresponding gene, based on *Interpro*-predicted domains, assigned GO terms or *Drosophila* homologues and the determined log2-transformed fold-change regulation in the respective dataset.

Transcript ID	Name/Description	*AsaB1* Regulation	*AsaB2* Regulation
AGAP005541-RA	ARID/BRIGHT DNA-binding	0.74	0.36
AGAP010828-RA	CASPS9	0.48	0.57
AGAP011416-RA	Chitin-binding	1.44	0.83
AGAP012133-RA	Chitin-binding	1.11	0.24
AGAP011781-RA	CLIPA12	0.68	0.04
AGAP011788-RA	CLIPA14	1.21	0.56
AGAP012591-RA	CLIPA3	0.43	0.75
AGAP002813-RA	CLIPD6	1.99	1.26
AGAP000871-RA	C-type lectin	0.90	0.42
AGAP003066-RA	CYP304B1	0.57	0.12
AGAP001443-RA	CYP325J1	0.58	0.41
AGAP013511-RC	CYP6AG2	0.31	0.60
AGAP008205-RA	CYP6R1	0.66	0.57
AGAP011294-RA	DEF1	1.25	1.98
AGAP011224-RA	FREP57	1.16	0.55
AGAP011225-RA	FREP28	0.64	0.69
AGAP005147-RB	FN3D1	0.52	0.52
AGAP002342-RA	FN3	0.58	0.34
AGAP004246-RA	FN3 Receptor protein-tyrosine phosphatase	1.02	0.78
AGAP001633-RA	FN3	0.95	0.62
AGAP001674-RA	sidestep FN3D3 paralogue	0.70	0.29
AGAP010184-RA	FN3 Echinoid-(EGFR pathway)	0.68	0.12
AGAP005381-RA	Glycoside hydrolase	0.96	0.60
AGAP006422-RA	Glycoside hydrolase	0.82	0.50
AGAP000143-RA	Glycoside hydrolase	1.19	0.51
AGAP008562-RA	Glycoside hydrolase	0.41	0.35
AGAP009621-RA	Glycoside hydrolase	0.79	1.14
AGAP010089-RA	GPROP9	1.45	0.26
AGAP006126-RB	GPROP8	0.63	0.43
AGAP002519-RA	GPRTYR	1.77	0.88
AGAP003254-RA	Gr16	0.96	1.24
AGAP006876-RA	Gr31	0.78	0.23
AGAP001117-RA	Gr37	0.43	0.72
AGAP006143-RE	Gr56	0.57	0.93
AGAP001981-RB	Haemolymph juvenile hormone binding	0.82	0.39
AGAP004944-RA	Heat shock	0.78	0.66
AGAP006958-RA	Heat shock	0.75	0.71
AGAP001187-RA	Homeobox	0.45	0.54
AGAP004664-RA	Homeobox	0.80	0.24
AGAP003871-RA	Homeodomain	2.05	1.38
AGAP001495-RA	Homeodomain	0.48	0.43
AGAP004649-RA	Homeodomain	1.00	0.53
AGAP011902-RA	Homeodomain	0.57	0.43
AGAP010734-RA	HPX1	0.62	0.05
AGAP003714-RA	HPX3	0.62	0.00
AGAP000051-RA	HPX5	0.68	0.56
AGAP013034-RA	IAP8	0.89	0.01
AGAP004979-RA	Immunoglobulin E-set	0.47	0.64
AGAP007361-RA	Immunoglobulin E-set	0.54	0.56
AGAP008935-RA	Immunoglobulin E-set	0.43	0.47
AGAP004915-RA	Immunoglobulin I-set	0.79	0.67
AGAP009247-RA	Immunoglobulin	1.18	0.41
AGAP011094-RA	Immunoglobulin	0.40	0.33
AGAP007500-RA	Immunoglobulin V-set	0.35	0.48
AGAP004416-RA	Leucine-rich repeat	0.42	0.51
AGAP010847-RA	Leucine-rich repeat	0.70	-0.01
AGAP007846-RA	Leucine-rich repeat	0.66	0.19
AGAP009594-RA	Leucine-rich repeat	1.22	0.96
AGAP007470-RA	Leucine-rich repeat	0.73	0.72
AGAP008257-RA	nitric-oxide synthase, invertebrate	0.56	0.07
AGAP002188-RA	OBP12	0.41	0.33
AGAP006077-RB	OBP51	0.38	0.38
AGAP009413-RA	Or20	1.97	1.35
AGAP003053-RA	Or45	1.18	0.46
AGAP009393-RA	Or47	0.58	0.13
AGAP011979-RA	Or60	0.44	0.45
AGAP002957-RA	Peptidase	0.46	0.51
AGAP003643-RA	Peptidase	0.67	0.55
AGAP006208-RA	Peptidase	0.41	0.63
AGAP001814-RA	Peptidase	0.91	0.49
AGAP008371-RA	Peptidase	0.60	0.70
AGAP007619-RA	Peptidase	0.98	0.17
AGAP005790-RA	Peptidase	0.75	0.36
AGAP006487-RA	Peptidase	1.28	0.27
AGAP012022-RA	Peptidase	0.48	0.51
AGAP010659-RA	Peptidase	1.41	0.01
AGAP011429-RA	Peptidase	0.43	0.29
AGAP012270-RA	Peptidase	0.50	0.80
AGAP012671-RA	Peptidase	0.70	0.36
AGAP005688-RA	Peptidase	1.68	0.67
AGAP006672-RA	Peptidase	0.55	0.96
AGAP001241-RA	Peptidase	0.71	0.75
AGAP011040-RA	Peptidase	0.80	0.96
AGAP011430-RA	Peptidase	1.29	0.93
AGAP012269-RA	Peptidase	0.54	0.48
AGAP005671-RA	Peptidase	2.03	1.21
AGAP008276-RA	Peptidase	0.78	1.30
AGAP002351-RA	Peptidase	1.04	-0.01
AGAP005203-RD	PGRPLC	0.66	0.72
AGAP009515-RA	REL1	0.55	0.07
AGAP008782-RA	Ricin B lectin	0.63	0.52
AGAP002451-RA	SCRB1	0.82	0.52
AGAP005725-RA	SCRB3	0.82	0.59
AGAP004847-RA	SCRB7	0.88	0.89
AGAP006483-RA	SPZ2	0.36	0.47
AGAP007177-RA	SPZ5	2.06	1.67
AGAP005126-RA	SPZ6	1.25	0.62
AGAP000099-RA	STAT2	0.86	0.05
AGAP010831-RA	TEP8	0.73	0.12

Table 6.7: Genes of interest found in a cluster of mostly upregulated transcripts between the *AsaB1-AsaB2* datasets, stemming from midguts of *Asaia* infected and uninfected mosquitoes. Genes of interest with a putative immune function found in cluster *AsaB* (Figure 6.9) of co-regulated transcripts are shown. The Transcript ID corresponding to the identified transcript is shown along with the assigned name or description of the homologies of the corresponding gene, based on *Interpro*-predicted domains, assigned GO terms or *Drosophila* homologues and the determined log2-transformed fold-change regulation in the respective dataset.

Name	Name/Description	Functional class	Fold-change	Regulation	t-test p-value
AGAP011442-RA	serine carboxypeptidase 1	Peptidase	2.16	down	0.03
AGAP001249-RA	Eupolytin	Peptidase	2.06	down	0.16
AGAP008292-RA	Trypsin-4	Peptidase	1.93	down	0.10
AGAP008861-RA	female reproductive tract protease GLEANR_896	Peptidase	1.89	down	0.14
AGAP007684-RB	Peptidase Wnt signalling	Peptidase	1.89	down	0.06
AGAP007684-RA	Peptidase Wnt signalling	Peptidase	1.83	down	0.03
AGAP008293-RA	Trypsin-7	Peptidase	1.81	down	0.03
AGAP001198-RA	chymotrypsin	Peptidase	1.80	down	0.03
AGAP013078-RA	Plasma glutamate carboxypeptidase precursor	Peptidase	1.78	down	0.13
AGAP010243-RA	Female reproductive tract protease GLEANR 897	Peptidase	1.76	down	0.14
AGAP008468-RA	fatty acid synthase	Oxidoreductase	1.78	up	0.02
AGAP003209-RA	C-4 methylsterol oxidase	Oxidoreductase	1.86	down	0.02
AGAP003231-RA	ADH dehydrogenase (ubiquinone) iron-sulfur protein 7, mitochondrial	Oxidoreductase	1.84	down	0.02
AGAP004880-RC	L-lactate dehydrogenase	Oxidoreductase	1.76	down	0.03
AGAP013185-RA	Putative glycine-rich cell wall structural protei	Other	2.24	up	0.07
AGAP008441-RA	CUB domain complement/neurotransmission	Other	2.20	down	0.00
AGAP000570-RA	Putative salivary protein	Other	2.03	down	0.00
AGAP010534-RA	Putative secreted salivary gland peptide	Other	1.83	down	0.04
AGAP006464-RA	Putative secreted mucin	Other	1.77	down	0.03
AGAP002905-RA	OBP13	Odorant binding	1.82	down	0.04
AGAP005201-RA	Nucleic acid-binding, OB-fold	Nucleic acid binding	1.78	up	0.01
AGAP002847-RA	ML3	MD2 lipid recognition	1.98	down	0.05
AGAP006197-RA	Methyltransferase	Juvenile hormone-related	2.29	up	0.16
AGAP012703-RA	TO2 Haemolymph juvenile hormone-binding	Juvenile hormone-related	1.80	up	0.02
AGAP003749-RA	Pancreatic triacylglycerol lipase	Hydrolase	2.75	down	0.01
AGAP002102-RA	Glycoside hydrolase	Hydrolase	2.31	down	0.07
AGAP013161-RA	Fatty-acid amide hydrolase 2	Hydrolase	1.96	down	0.04
AGAP006086-RA	Glycoside hydrolase	Hydrolase	1.87	down	0.31
AGAP011305-RA	Alkaline_phosphatase_AS	Hydrolase	1.83	down	0.01
AGAP009049-RA	Glycoside hydrolase	Hydrolase	1.81	down	0.06
AGAP002635-RA	Gr13	Gustatory receptor	2.08	down	0.03
AGAP002635-RB	Gr13	Gustatory receptor	1.98	down	0.00
AGAP008703-RA	GPCR	GPCR signalling	1.82	down	0.12
AGAP006521-RA	Adenylate cyclase-stimulating G alpha protein	GPCR signalling	1.77	down	0.03
AGAP011642-RA	methuselah-like 14	GPCR signalling	1.75	down	0.05
AGAP001975-RA	Protein SYS1 homolog	Golgi apparatus	2.04	down	0.05
AGAP007583-RA	Golgi-associated plant pathogenesis-related protein 1	Golgi apparatus	1.97	down	0.01
AGAP007583-RB	Golgi-associated plant pathogenesis-related protein 1	Golgi apparatus	1.93	down	0.04
AGAP001477-RA	Innexin	Gap junction channel	1.75	down	0.07
AGAP011228-RA	FREP24	Fibrinogen-related protein	2.32	up	0.03
AGAP010759-RA	FREP32	Fibrinogen-related protein	2.11	down	0.01
AGAP010762-RA	FREP4	Fibrinogen-related protein	1.83	down	0.04
AGAP010760-RA	FREP69	Fibrinogen-related protein	1.79	down	0.03
AGAP011238-RA	Zinc finger	DNA binding	1.98	up	0.03
AGAP012027-RA	Zinc finger	DNA binding	1.80	up	0.04
AGAP010220-RA	proliferating cell nuclear antigen	DNA binding	1.77	up	0.06
AGAP010642-RA	FH2 actin-binding	Cytoskeletal protein binding	1.97	down	0.03
AGAP002867-RA	CYP6P4	Cytochrome P450	1.83	down	0.10
AGAP001864-RA	CYP4H15	Cytochrome P450	1.81	down	0.04
AGAP005656-RA	CYP305A1	Cytochrome P450	1.77	down	0.03
AGAP000382-RA	CPF4 Cuticle	Cuticle	1.82	up	0.28
AGAP001203-RA	Chitin-binding	Chitin-binding	2.19	down	0.03
AGAP009830-RA	Chitin-binding	Chitin-binding	1.91	down	0.02
AGAP006068-RC	Alkaline phosphatase core	Catalytic activity	1.76	up	0.05
AGAP010071-RA	acyl-CoA synthetase	Catalytic activity	1.92	down	0.03
AGAP004376-RA	aldose 1-epimerase	Carbohydrate binding	1.81	down	0.01
AGAP004403-RA	myosin I	ATP binding	1.81	up	0.00

Name	Name/Description	Functional class	Fold-change	Regulation	t-test p-value
AGAP001718-RA	Unknown	Unknown	1.77	up	0.02
AGAP009146-RA	Unknown	Unknown	7.54	down	0.02
AGAP013493-RA	Unknown	Unknown	2.93	down	0.01
AGAP006507-RA	Unknown	Unknown	2.70	down	0.15
AGAP012552-RA	Unknown	Unknown	2.28	down	0.05
AGAP005379-RA	DUF457	Unknown	2.28	down	0.25
AGAP005424-RA	Unknown	Unknown	2.04	down	0.02
AGAP006442-RA	Unknown	Unknown	1.95	down	0.02
AGAP011958-RA	Unknown	Unknown	1.93	down	0.08
AGAP004076-RA	Unknown	Unknown	1.89	down	0.12
AGAP011037-RA	Unknown	Unknown	1.85	down	0.08
AGAP005614-RA	Unknown	Unknown	1.79	down	0.00
AGAP006664-RA	Unknown	Unknown	1.77	down	0.02
AGAP006506-RA	Unknown	Unknown	1.75	down	0.03
AGAP001989-RA	Unknown	Unknown	1.75	down	0.08
AGAP007267-RA	tRNA-Gly	tRNA	2.11	down	0.04
AGAP012794-RA	tRNA-Arg	tRNA	1.92	down	0.01
AGAP005367-RA	tRNA-Gly	tRNA	1.78	down	0.07
AGAP012644-RA	Cyt b561	Transport	1.82	up	0.02
AGAP005251-RA	potassium voltage-gated channel	Transport	2.00	down	0.04
AGAP005563-RC	facilitated trehalose transporter Tret1	Transport	1.96	down	0.01
AGAP000732-RA	Synaptic vesicle protein	Transport	1.85	down	0.03
AGAP005988-RA	pyrimidine nucleoside transport protein	Transport	1.80	down	0.06
AGAP009672-RA	protein kinase autophagic cell death	Transferase	1.84	up	0.00
AGAP013329-RA	Acyl CoA acyltransferase	Transferase	2.90	down	0.01
AGAP010111-RA	Acyl CoA acyltransferase	Transferase	1.99	down	0.04
AGAP010125-RA	Acyl CoA acyltransferase	Transferase	1.88	down	0.03
AGAP011984-RA	Ricin B lectin	Transferase	1.78	down	0.05
AGAP010118-RA	Acyl CoA acyltransferase	Transferase	1.76	down	0.02
AGAP006788-RA	glycerol kinase	Tranferase	1.99	down	0.17
AGAP007158-RA	Alpha-crystallin B chain HSP2C-like_chaperone	Stress response	4.84	down	0.04
AGAP007159-RA	Alpha-crystallin B chain HSP2C-like_chaperone	Stress response	4.70	down	0.02
AGAP004582-RA	Heat shock protein 70kDa	Stress response	3.67	down	0.04
AGAP004583-RA	Heat shock protein 70kDa	Stress response	3.58	down	0.04
AGAP004581-RA	Heat shock protein 70kDa	Stress response	3.53	down	0.05
AGAP012891-RA	heat shock 70kDa protein	Stress response	2.92	down	0.02
AGAP011707-RA	RhoGEF4	Signal transduction	1.99	down	0.33
AGAP006438-RA	ribosomal biogenesis protein LAS1	Ribosome	2.34	down	0.11
AGAP005942-RB	Thyroglobulin 1	Proteinase inhibitor	1.77	down	0.11
AGAP010265-RA	Delta, transcript variant A external compound sense organ	Protein/receptor binding	1.89	up	0.01
AGAP002310-RA	protocadherin-15	Protein/receptor binding	1.84	up	0.02
AGAP000783-RA	ARM-type_fold	Protein/receptor binding	1.83	up	0.03
AGAP001177-RA	Ca2+-triggered coelenterazine-binding protein 1	Protein/receptor binding	2.33	down	0.02
AGAP005093-RA	Cadherin	Protein/receptor binding	2.06	down	0.06
AGAP006689-RA	BTB (POZ) domain containing 9	Protein/receptor binding	1.82	down	0.01
AGAP007611-RA	EGF-like_CS	Protein/receptor binding	1.81	down	0.03
AGAP011326-RA	IAP2	Protein/receptor binding	1.76	down	0.10
AGAP012317-RA	Leucine-rich repeat	Protein/receptor binding	1.76	down	0.16
AGAP006343-RA	PGRPS2	Peptidoglycan recognition	1.94	down	0.03
AGAP009288-RA	Derlin-1 ER proteolytic degradation	Peptidase	2.20	up	0.14
AGAP001697-RA	20S proteasome subunit beta 3	Peptidase	1.84	up	0.05
AGAP008290-RA	Trypsin-6	Peptidase	3.06	down	0.04
AGAP008291-RA	Trypsin-5	Peptidase	2.91	down	0.01
AGAP007165-RA	Late trypsin	Peptidase	2.31	down	0.07

Table 6.8: List of more than 1.75-fold regulated transcripts following oral *Asaia* infection. 111 *An. gambiae* transcripts were identified as more than 1.75-fold regulated following *Asaia* infection, over 3 independent infections. For each differentially expressed transcript, the Transcript ID is shown along with the assigned name or description of the homologies of the corresponding gene, based on *Interpro*-predicted domains, assigned GO terms or *Drosophila* homologues. The functional class that the gene corresponding to each transcript has been assigned is shown, along with the transcriptional regulation determined over 3 independent infections, the fold-change regulation and the computed p-value in a t-test against zero using the log2-transformed fold-change values for each probe corresponding to the respective transcript, over the 3 independent infections. Significance was assessed with a p-value cut-off of 0.05.

Dataset	Transcript ID	Microarrays regulation	qRT-PCR Regulation
SmA2	AGAP007343-RA	1.74	0.43
SmA2	AGAP006991-RA	3.41	4.18
SmA2	AGAP001648-RA	0.51	-1.30
SmA2	AGAP010833-RA	1.18	1.03
SmA2	AGAP005901-RA	0.02	-0.48
SmA3i	AGAP007343-RA	3.61	4.05
SmA3i	AGAP006991-RA	0.33	0.54
SmA3i	AGAP001648-RA	2.20	0.74
SmA3i	AGAP010833-RA	2.79	3.46
SmA3i	AGAP005901-RA	1.42	0.38
SmB2i	AGAP007343-RA	0.00	-0.30
SmB2i	AGAP006991-RA	0.13	-1.06
SmB2i	AGAP001648-RA	0.16	0.67
SmB2i	AGAP010833-RA	0.06	0.43
SmB2i	AGAP005901-RA	-0.23	-0.46
SmB2ii	AGAP007343-RA	0.00	0.04
SmB2ii	AGAP006991-RA	0.04	-0.74
SmB2ii	AGAP001648-RA	0.27	2.41
SmB2ii	AGAP010833-RA	0.15	1.22
SmB2ii	AGAP005901-RA	-0.20	0.16
SmB3	AGAP007343-RA	0.00	-2.77
SmB3	AGAP006991-RA	0.67	0.35
SmB3	AGAP001648-RA	-0.02	-2.37
SmB3	AGAP010833-RA	0.43	-2.94
SmB3	AGAP005901-RA	-0.17	0.25

Table 6.9: Dataset-transcript combinations used for validating the DNA microarray expression data using qRT-PCR. The fold-change differential expression of 5 transcripts corresponding to LYSC2 (AGAP007343), a chitin-binding protein encoded by AGAP006991-RA, CLIPB17 (AGAP001648)-RA, CLIPB14 (AGAP010833-RA) and Ect4 (AGAP005901-RA) was determined by qRT-PCR. Total RNA stemming from midguts of *S. marcescens* infected mosquitoes or uninfected controls used in hybridizations corresponding to the datasets *SmA2, SmA3i, SmB2i, SmB2ii* and *SmB3*, was used for synthesis of cDNA further used as a template in the respective qRT-PCR reaction. The obtained microarray and qRT-PCR derived values are shown for each transcript-dataset combination.

A

dsRNA	Figure			dsRNA	Figure			dsRNA	Figure		
lacZ	7.11A			*FN3D1*	7.11A			*FN3D2*	7.11A		
Node	Reads	Total Reads	%	Node	Reads	Total Reads	% prevalence	Node	Reads	Total Reads	% prevalence
Enterobacteriaceae, 3047	3047	3661	83.23	*Enterobacteriaceae*, 4690	4690	5352	87.63	*Enterobacteriaceae*, 4242	4242	4735	89.59
Acetobacteraceae, 610	610	3661	16.66	*Acetobacteraceae*, 660	660	5352	12.33	*Acetobacteraceae*, 490	490	4735	10.35
Proteobacteria, 4	4	3661	0.11	*Proteobacteria*, 1	1	5352	0.02	*Proteobacteria*, 2	2	4735	0.04
				Gammaproteobacteria, 1	1	5352	0.02	*Bacteria*, 1	1	4735	0.02
Reference sequence	Reads	Total Reads	%	*Not assigned*, 1	1	5352	0.02				
Enterobacteriaceae, 3047								Reference sequence	Reads	Total Reads	% prevalence
Erwinia	2916	3661	79.65	Reference sequence	Reads	Total Reads	% prevalence	*Enterobacteriaceae*, 4242			
Citrobacter	2907	3661	79.40	*Enterobacteriaceae*, 4690				*Citrobacter*	2935	4735	61.99
Kluyvera	1586	3661	43.32	*Citrobacter*	3461	5352	64.67	*Raoultella*	2225	4735	46.99
Raoultella	1533	3661	41.87	*Klebsiella*	2324	5352	43.42	*Klebsiella*	2014	4735	42.53
Klebsiella	817	3661	22.32	*Raoultella*	2215	5352	41.39	*Enterobacter*	1890	4735	39.92
Enterobacter	542	3661	14.80	*Enterobacter*	1914	5352	35.76	*Serratia*	1862	4735	39.32
Pectobacterium	409	3661	11.17	*Serratia*	1603	5352	29.95	*Erwinia*	1139	4735	24.05
Serratia	69	3661	1.88	*Erwinia*	1298	5352	24.25	*Kluyvera*	889	4735	18.78
Escherichia	50	3661	1.37	*Kluyvera*	865	5352	16.16	*Cedecea*	830	4735	17.53
Pantoea	24	3661	0.66	*Pectobacterium*	398	5352	7.44	*Pectobacterium*	466	4735	9.84
Cedecea	19	3661	0.52	*Cedecea*	302	5352	5.64	*Buttiauxella*	195	4735	4.12
				Escherichia	280	5352	5.23	*Shigella*	193	4735	4.08
Acetobacteraceae, 610				*Shigella*	137	5352	2.56	*Escherichia*	132	4735	2.79
Swaminathania	610	3661	16.66	*Buttiauxella*	121	5352	2.26	*Salmonella*	83	4735	1.75
Asaia	600	3661	16.39	*Salmonella*	120	5352	2.24	*Hafnia*	59	4735	1.25
Gluconoacetobacter	549	3661	15.00	*Yersinia*	74	5352	1.38	*Cronobacter*	30	4735	0.63
Gluconobacter	485	3661	13.25	*Hafnia*	55	5352	1.03	*Yersinia*	27	4735	0.57
Kozakia	15	3661	0.41	*Samsonia*	48	5352	0.90	*Samsonia*	24	4735	0.51
				Rahnella	41	5352	0.77	*Rahnella*	17	4735	0.36
Proteobacteria, 4				*Glossina*	22	5352	0.41				
Burkholderia	2	3661	0.05	*Pragia*	21	5352	0.39	*Acetobacteraceae*, 490			
Erwinia	2	3661	0.05	*Cronobacter*	12	5352	0.22	*Swaminathania*	490	4735	10.35
Asaia	1	3661	0.03	*Pantoea*	11	5352	0.21	*Asaia*	487	4735	10.29
Kozakia	1	3661	0.03					*Gluconoacetobacter*	417	4735	8.81
Aquabacterium	1	3661	0.03	*Acetobacteraceae*, 660				*Gluconobacter*	356	4735	7.52
Alkalispirillum	1	3661	0.03	*Swaminathania*	660	5352	12.33	*Kozakia*	27	4735	0.57
Swaminathania	1	3661	0.03	*Asaia*	642	5352	12.00				
Schlegelella	1	3661	0.03	*Gluconoacetobacter*	571	5352	10.67	*Proteobacteria*, 2			
Gluconacetobacter	1	3661	0.03	*Gluconobacter*	493	5352	9.21	*Asaia*	1	4735	0.02
Saccharospirillum	1	3661	0.03	*Kozakia*	34	5352	0.64	*Burkholderia*	1	4735	0.02
Gluconacetobacter	1	3661	0.03					*Swaminathania*	1	4735	0.02
Azohydromona	1	3661	0.03	*Gammaproteobacteria*, 1				*Gluconoacetobacter*	1	4735	0.02
Methylibium	1	3661	0.03	*Pseudomonas*	1	5352	0.02	*Gluconobacter*	1	4735	0.02
								Erwinia	1	4735	0.02
				Proteobacteria, 1							
				Asaia	1	5352	0.02	*Bacteria*, 1			
				Swaminathania	1	5352	0.02	*Chryseobacterium*	1	4735	0.02
				Gluconoacetobacter	1	5352	0.02	*Elizabethkingia*	1	4735	0.02
				Gluconobacter	1	5352	0.02	*Riemerella*	1	4735	0.02
				Klebsiella	1	5352	0.02				
				Erwinia	1	5352	0.02				

B

dsRNA	Figure			dsRNA	Figure			dsRNA	Figure		
lacZ	7.11B			*FN3D2*	7.11B			*FN3D3*	7.11B		
Node	Reads	Total Reads	% prevalence	Node	Reads	Total Reads	% prevalence	Node	Reads	Total Reads	% prevalence
Enterobacteriaceae, 1405	1405	3182	44.15	*Enterobacteriaceae*,	4039	4860	83.11	*Enterobacteriaceae*, 2782	2782	3987	69.78
Acetobacteraceae, 1335	1335	3182	41.95	*Flavobacteriaceae*, 478	478	4860	9.84	*Acetobacteraceae*, 1201	1201	3987	30.12
Flavobacteriaceae, 234	234	3182	7.35	*Acetobacteraceae*, 296	296	4860	6.09	*Proteobacteria*, 3	3	3987	0.08
Leucobacter, 109	109	3182	3.43	*Elizabethkingia*, 46	46	4860	0.95	*Bacteria*, 1	1	3987	0.03
Microbacteriaceae, 64	64	3182	2.01	*Proteobacteria*, 1	1	4860	0.02				
Elizabethkingia, 19	19	3182	0.60					Reference sequence	Reads	Total Reads	% prevalence
Micrococcineae, 6	6	3182	0.19	Reference sequence	Reads	Total Reads	% prevalence	*Enterobacteriaceae*, 2782			
Bacteria, 5	5	3182	0.16	*Enterobacteriaceae*,				*Serratia*	2777	3987	69.65
Proteobacteria, 3	3	3182	0.09	*Serratia*	4011	4860	82.53	*Klebsiella*	2591	3987	64.99
Actinomycetales, 2	2	3182	0.06	*Klebsiella*	3681	4860	75.74	*Enterobacter*	2447	3987	61.37
				Enterobacter	3419	4860	70.35	*Citrobacter*	2299	3987	57.66
Reference sequence	Reads	Total Reads	% prevalence	*Citrobacter*	3344	4860	68.81	*Pectobacterium*	909	3987	22.80
Enterobacteriaceae, 1405				*Pectobacterium*	1561	4860	32.12	*Raoultella*	417	3987	10.46
Enterobacter	1373	3182	43.15	*Raoultella*	765	4860	15.74	*Dickeya*	373	3987	9.36
Cedecea	1296	3182	40.73	*Dickeya*	635	4860	13.07	*Cedecea*	11	3987	0.28
Raoultella	1119	3182	35.17	*Cedecea*	62	4860	1.28	*Erwinia*	11	3987	0.28
Citrobacter	847	3182	26.62	*Buttiauxella*	14	4860	0.29				
Serratia	780	3182	24.51					*Acetobacteraceae*, 1201			
Klebsiella	626	3182	19.67	*Acetobacteraceae*, 296				*Swaminathania*	1201	3987	30.12
Buttiauxella	554	3182	17.41	*Swaminathania*	296	4860	6.09	*Asaia*	1200	3987	30.10
Pectobacterium	82	3182	2.58	*Asaia*	295	4860	6.07	*Gluconoacetobacter*	1059	3987	26.56
Kluyvera	21	3182	0.66	*Gluconoacetobacter*	253	4860	5.21	*Gluconobacter*	1036	3987	25.98
Dickeya	14	3182	0.44	*Gluconobacter*	234	4860	4.81	*Kozakia*	79	3987	1.98
Erwinia	11	3182	0.35	*Kozakia*	26	4860	0.53	*Acetobacter*	25	3987	0.63
Acetobacteraceae, 1335				*Elizabethkingia*, 46				*Proteobacteria*, 3			
Swaminathania	1335	3182	41.95	*Elizabethkingia*	46	4860	0.95	*Burkholderia*	2	3987	0.05
Asaia	1297	3182	40.76	*Chryseobacterium*	46	4860	0.95	*Asaia*	1	3987	0.03
Gluconoacetobacter	1224	3182	38.47	*Riemerella*	39	4860	0.80	*Swaminathania*	1	3987	0.03
Gluconobacter	1219	3182	38.31					*Gluconoacetobacter*	1	3987	0.03
Kozakia	31	3182	0.97	*Flavobacteriaceae*, 478				*Serratia*	1	3987	0.03
Acetobacter	15	3182	0.47	*Chryseobacterium*	478	4860	9.84	*Gluconobacter*	1	3987	0.03
				Elizabethkingia	478	4860	9.84				
Elizabethkingia, 19				*Chryseobacterium*	478	4860	9.84	*Bacteria*, 1			
Elizabethkingia	19	3182	0.60	*Riemerella*	469	4860	9.65	*Corynebacterium*	1	3987	0.03
Chryseobacterium	19	3182	0.60								
Riemerella	17	3182	0.53	*Proteobacteria*, 1							
				Burkholderia	1	4860	0.02				
Flavobacteriaceae, 234											
Chryseobacterium	234	3182	7.35								
Elizabethkingia	234	3182	7.35								
Riemerella	234	3182	7.35								
Flavobacteriaceae bacterium	11	3182	0.35								
Leucobacter, 109											
Leucobacter	109	3182	3.43								
Curtobacterium	108	3182	3.39								
Microbacterium	108	3182	3.39								
Arthrobacter	13	3182	0.41								
Microbacteriaceae, 64											
Leucobacter	64	3182	2.01								
Microbacterium	64	3182	2.01								
Curtobacterium	51	3182	1.60								
Arthrobacter	12	3182	0.38								
Micrococcineae, 6											
Arthrobacter	3	3182	0.09								
Microbacterium	3	3182	0.09								
Leucobacter	3	3182	0.09								
Micrococcus	3	3182	0.09								
Curtobacterium	1	3182	0.03								
Rathayibacter	1	3182	0.03								
Actinomycetales, 2											
Corynebacterium	2	3182	0.06								
Gordonia	1	3182	0.03								
Mycobacterium	1	3182	0.03								

Proteobacteria, 3											
Burkholderia	2	3182	0.06								
Asaia	1	3182	0.03								
Swaminathania	1	3182	0.03								
Serratia	1	3182	0.03								
Bacteria, 5											
Elizabethkingia	5	3182	0.16								
Enterobacter	5	3182	0.16								
Raoultella	4	3182	0.13								
Cedecea	4	3182	0.13								
Kluyvera	4	3182	0.13								
Chryseobacterium	4	3182	0.13								
Buttiauxella	1	3182	0.03								

C

dsRNA	Figure			dsRNA	Figure		
lacZ	7.11C			*FN3D1*	7.11C		
Node	Reads	Total Reads	% prevalence	Node	Reads	Total Reads	% prevalence
Enterobacteriaceae, 4075	4075	5546	73.48	*Enterobacteriaceae*, 2041	2041	2389	85.43
Flavobacteriaceae, 1149	1149	5546	20.72	*Flavobacteriaceae*, 225	225	2389	9.42
Acetobacteraceae, 168	168	5546	3.03	*Acetobacteraceae*, 102	102	2389	4.27
Elizabethkingia, 90	90	5546	1.62	*Elizabethkingia*, 18	18	2389	0.75
Burkholderia cepacia complex, 19	19	5546	0.34	*Proteobacteria*, 2	2	2389	0.08
Leucobacter, 16	16	5546	0.29	*Bacteria*, 1	1	2389	0.04
Microbacteriaceae, 14	14	5546	0.25				
Bacteria, 5	5	5546	0.09	Reference sequence	Reads	Total Reads	
Burkholderia, 5	5	5546	0.09	*Enterobacteriaceae*, 2041			
Micrococcineae, 2	2	5546	0.04	*Enterobacter*	1576	2389	65.97
Gammaproteobacteria, 2	2	5546	0.04	*Citrobacter*	1559	2389	65.26
Proteobacteria, 1	1	5546	0.02	*Serratia*	1453	2389	60.82
				Klebsiella	1155	2389	48.35
				Raoultella	802	2389	33.57
Reference sequence	Reads	Total Reads	% prevalence	*Cedecea*	345	2389	14.44
Enterobacteriaceae, 4075				*Pectobacterium*	318	2389	13.31
Serratia	3680	5546	66.35	*Salmonella*	164	2389	6.86
Enterobacter	3439	5546	62.01	*Buttiauxella*	142	2389	5.94
Citrobacter	3173	5546	57.21	*Escherichia*	130	2389	5.44
Klebsiella	2998	5546	54.06	*Shigella*	116	2389	4.86
Raoultella	1317	5546	23.75	*Yersinia*	94	2389	3.93
Pectobacterium	1231	5546	22.20	*Dickeya*	88	2389	3.68
Cedecea	591	5546	10.66	*Rahnella*	67	2389	2.80
Dickeya	388	5546	7.00	*Pragia*	31	2389	1.30
Buttiauxella	165	5546	2.98	*Kluyvera*	24	2389	1.00
Kluyvera	63	5546	1.14	*Erwinia*	16	2389	0.67
Yersinia	47	5546	0.85				
Erwinia	43	5546	0.78	*Acetobacteraceae*, 102			
Samsonia	35	5546	0.63	*Swaminathania*	102	2389	4.27
Escherichia	32	5546	0.58	*Gluconoacetobacter*	102	2389	4.27
Shigella	31	5546	0.56	*Asaia*	100	2389	4.19
Salmonella	28	5546	0.50	*Gluconobacter*	90	2389	3.77
Rahnella	20	5546	0.36	*Kozakia*	12	2389	0.50
Pragia	15	5546	0.27				
				Elizabethkingia, 18			
Burkholderia cepacia complex, 19				*Elizabethkingia*	18	2389	0.75
Burkholderia	19	5546	0.34	*Chryseobacterium*	18	2389	0.75
				Riemerella	17	2389	0.71
Acetobacteraceae, 168							
Swaminathania	168	5546	3.03	*Flavobacteriaceae*, 225			
Gluconoacetobacter	167	5546	3.01	*Elizabethkingia*	225	2389	9.42
Asaia	165	5546	2.98	*Chryseobacterium*	225	2389	9.42
Gluconobacter	144	5546	2.60	*Riemerella*	224	2389	9.38
Kozakia	18	5546	0.32				
				Proteobacteria, 2			
Flavobacteriaceae, 1149				*Burkholderia*	2	2389	0.08
Chryseobacterium	1149	5546	20.72				
Elizabethkingia	1149	5546	20.72	*Bacteria*, 1			
Riemerella	1124	5546	20.27	*Riemerella*	1	2389	0.04
				Samsonia	1	2389	0.04
Leucobacter, 16				*Kaistella*	1	2389	0.04
Leucobacter	16	5546	0.29	*Chryseobacterium*	1	2389	0.04
Curtobacterium	16	5546	0.29	*Elizabethkingia*	1	2389	0.04
Microbacterium	16	5546	0.29				
Elizabethkingia, 90							
Elizabethkingia	90	5546	1.62				
Chryseobacterium	90	5546	1.62				
Riemerella	81	5546	1.46				
Microbacteriaceae, 14							
Leucobacter	14	5546	0.25				
Microbacterium	14	5546	0.25				
Burkholderia, 5							
Burkholderia	5	5546	0.09				
Gammaproteobacteria, 2							

Pseudomonas	2	5546	0.04				
Stenotrophomonas	2	5546	0.04				
Micrococcineae, 2							
Arthrobacter	2	5546	0.04				
Microbacterium	2	5546	0.04				
Leucobacter	2	5546	0.04				
Knoellia	1	5546	0.02				
Proteobacteria, 1							
Asaia	1	5546	0.02				
Kozakia	1	5546	0.02				
Swaminathania	1	5546	0.02				
Serratia	1	5546	0.02				
Bacteria, 5							
Elizabethkingia	5	5546	0.09				
Chryseobacterium	4	5546	0.07				
Serratia	3	5546	0.05				
Enterobacter	1	5546	0.02				
Streptococcus	1	5546	0.02				
Raoultella	1	5546	0.02				
Wautersiella	1	5546	0.02				

D

dsRNA	Figure			dsRNA	Figure		
lacZ	7.11D			*FN3D3*	7.11D		
Node	Reads	Total Reads	% prevalence	Node	Reads	Total Reads	% prevalence
Acetobacteraceae, 1800	1800	3518	51.17	*Enterobacteriaceae*, 3349	3349	5074	66.00
Enterobacteriaceae, 1670	1670	3518	47.47	*Acetobacteraceae*, 926	926	5074	18.25
Burkholderia cepacia complex, 19	19	3518	0.54	*Burkholderia cepacia complex*,	647	5074	12.75
Mycobacterium, 15	15	3518	0.43	*Burkholderia*, 121	121	5074	2.38
Burkholderia, 6	6	3518	0.17	*Aeromonas*, 25	25	5074	0.49
Bacteria, 4	4	3518	0.11	*Proteobacteria*, 3	3	5074	0.06
Proteobacteria, 3	3	3518	0.09	*Gammaproteobacteria*, 2	2	5074	0.04
Actinomycetales, 1	1	3518	0.03	*Bacteria*, 1	1	5074	0.02
Reference sequence	Reads	Total Reads	% prevalence	Reference sequence	Reads	Total Reads	% prevalence
Enterobacteriaceae, 1670				*Enterobacteriaceae*, 3349			
Enterobacter	955	3518	27.15	*Enterobacter*	3185	5074	62.77
Shigella	750	3518	21.32	*Cedecea*	2525	5074	49.76
Cedecea	712	3518	20.24	*Raoultella*	2521	5074	49.68
Raoultella	675	3518	19.19	*Serratia*	1708	5074	33.66
Escherichia	644	3518	18.31	*Citrobacter*	1472	5074	29.01
Citrobacter	555	3518	15.78	*Buttiauxella*	1324	5074	26.09
Salmonella	516	3518	14.67	*Klebsiella*	1184	5074	23.33
Klebsiella	469	3518	13.33	*Pectobacterium*	375	5074	7.39
Serratia	333	3518	9.47	*Kluyvera*	50	5074	0.99
Buttiauxella	310	3518	8.81	*Hafnia*	20	5074	0.39
Erwinia	50	3518	1.42				
Pectobacterium	47	3518	1.34	*Aeromonas*, 25			
Kluyvera	15	3518	0.43	*Aeromonas*	25	5074	0.49
Burkholderia cepacia complex, 19				*Burkholderia cepacia complex*,			
Burkholderia	19	3518	0.54	*Burkholderia*	647	5074	12.75
Burkholderia, 6				*Burkholderia*, 121			
Burkholderia	6	3518	0.17	*Burkholderia*	121	5074	2.38
Acetobacteraceae, 1800				*Acetobacteraceae*, 926			
Swaminathania	1798	3518	51.11	*Swaminathania*	925	5074	18.23
Asaia	1778	3518	50.54	*Asaia*	924	5074	18.21
Gluconoacetobacter	1670	3518	47.47	*Gluconoacetobacter*	811	5074	15.98
Gluconobacter	1643	3518	46.70	*Gluconobacter*	804	5074	15.85
Acetobacter	64	3518	1.82	*Kozakia*	51	5074	1.01
Kozakia	33	3518	0.94	*Acetobacter*	21	5074	0.41
Tanticharoenia	15	3518	0.43				
				Gammaproteobacteria, 2			
Mycobacterium, 15				*Enterobacter*	1	5074	0.02
Mycobacterium	14	3518	0.40	*Raoultella*	1	5074	0.02
				Shewanella	1	5074	0.02
Actinomycetales, 1				*Oleiphilus*	1	5074	0.02
Rothia	1	3518	0.03	*Kluyvera*	1	5074	0.02
Kocuria	1	3518	0.03	*Salmonella*	1	5074	0.02
Nesterenkonia	1	3518	0.03	*Aeromonas*	1	5074	0.02
Arthrobacter	1	3518	0.03				
Micrococcus	1	3518	0.03	*Proteobacteria*, 3			
				Asaia	2	5074	0.04
Proteobacteria, 3				*Swaminathania*	2	5074	0.04
Asaia	2	3518	0.06	*Serratia*	1	5074	0.02
Citrobacter	2	3518	0.06	*Lysobacter*	1	5074	0.02
Swaminathania	2	3518	0.06	*Citrobacter*	1	5074	0.02
Salmonella	2	3518	0.06	*Gluconobacter*	1	5074	0.02
Klebsiella	1	3518	0.03	*Burkholderia*	1	5074	0.02
Erwinia	1	3518	0.03	*Rhodanobacter*	1	5074	0.02
Xenorhabdus	1	3518	0.03	*Pectobacterium*	1	5074	0.02
Enterobacter	1	3518	0.03	*Enterobacter*	1	5074	0.02
Pectobacterium	1	3518	0.03	*Erwinia*	1	5074	0.02
Proteus	1	3518	0.03				
				Bacteria, 1			
Bacteria, 4				*Riemerella*	1	5074	0.02
Streptococcus	4	3518	0.11	*Chryseobacterium*	1	5074	0.02
Asaia	1	3518	0.03	*Elizabethkingia*	1	5074	0.02
Swaminathania	1	3518	0.03				
Gluconoacetobacter	1	3518	0.03				

Table 7.1: Proportions of assigned bacterial families and alignment of sequence reads to respective reference sequences as determined by 454 pyrosequencing of *FN3D1-3* or *LacZ* dsRNA treated mosquitoes retaining their natural gut microbiota. For each sequenced pool, the dsRNA treatment and the panel it corresponds in Figure 7.8 are shown. Sequenced pools from the same independent assay are shown adjacent to each other, as in Figure 7.8A-D (panels A-D). In each case, the sequence reads were blasted against the NCBI *16SMicrobial* database and were assigned to their respective bacterial family. The sequence reads assigned to each family, the total sequence reads of the respective pool and the calculated prevalence of the bacterial family are indicated. Sequence reads assigned to each bacterial family aligned with reference sequences corresponding to genera belonging to the respective family. For each genus, the reference sequence with the highest number of aligned reference sequences is shown, along with the corresponding number of sequence reads, the total number of sequence reads in the pool and the calculated prevalence. *Serratia* reference sequences with the highest number of aligned sequence reads were used to separately assign the prevalence of *Serratia* or strains with similarity to *Serratia* reference sequences.