University of Groningen

Evolutionary genomics of the immune response against parasitoids in Drosophila

Salazar Jaramillo, Laura

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2014

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Salazar Jaramillo, L. (2014). Evolutionary genomics of the immune response against parasitoids in Drosophila [S.l.]: [S.n.]

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 11-02-2018

Evolutionary genomics of the immune response against parasitoids in *Drosophila*

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with the decision by the College of Deans.

This thesis will be defended in public on

Monday 22 December 2014 at 12.45 hours

by

Laura Salazar Jaramillo
born on 11 July 1980
in Medellín, Colombia

Promotors:
Prof. Dr. B. Wertheim
Prof. Dr. L.W. Beukeboom

Co-promotor:
Dr. L. van de Zande

Assessment committee:
Prof. Dr. J.L. Olsen
Prof. Dr. L.E.M. Vet
Dr. F.M. Jiggins

This work was financed by a Rosalind Franklin Fellowship of the University of Groningen to Bregje Wertheim and by the Netherlands Organisation for Scientific Research (NWO) [VIDI grant 864.08.008 to BW].

ISBN: 978-90-5335-991-4
Cover design: Laura Salazar Jaramillo
Printed by: Ridderprint

Contents

Chapter 1 General Introduction **3**
1.1 Functional Genomics 4
1.2 Evolutionary Genomics 5
1.3 *Drosophila* as model organism 7
1.4 Immune system 7
1.5 The *Drosophila* immune response against parasitoids 9
1.6 This thesis 11

Chapter 2 Evolution of a cellular immune response in *Drosophila*: a phenotypic and genomic comparative analysis **13**
2.1 Introduction 14
2.2 Materials and Methods 16
2.3 Results 19
2.4 Discussion 29
2.5 Acknowledgements 34
2.6 Supplementary Material Chapter 2 36

Chapter 3 Genetic variation of the immune receptor *Tep1* among natural populations of *D. melanogaster* **47**
3.1 Introduction 47
3.2 Materials and Methods 50
3.3 Results 53
3.4 Discussion 55
3.5 Acknowledgements 57
3.6 Supplementary Material Chapter 3 58

Chapter 4 Inter- and intra-species variation in the genome-wide gene expression of *Drosophila* in response to parasitoid wasp attack **69**
4.1 Introduction 70
4.2 Materials and Methods 72
4.3 Results 74
4.4 Discussion 80
4.5 Acknowledgements 85
4.6 Supplementary Material Chapter 4 86

Chapter 5 From genomes to natural history: does *Drosophila sechellia* escape parasitoid attack by feeding on a "toxic" resource? **95**

5.1 Introduction . . . 95
5.2 Materials and Methods . . . 96
5.3 Results . . . 97
5.4 Discussion . . . 101
5.5 Acknowledgements . . . 102

Chapter 6 General Discussion **105**
6.1 Overview . . . 105
6.2 The encapsulation ability in *Drosophila* species . . . 105
6.3 Evolving encapsulation ability: a trait involving new genes . . . 106
6.4 Fast evolving genes . . . 108
6.5 Encapsulation ability among lines of *D. melanogaster* . . . 110
6.6 Contrasting long- and short-term evolution . . . 111
6.7 Loss of resistance in *D. sechellia* . . . 112
6.8 Future work . . . 113
6.9 Conclusions . . . 114

References **115**

Summary **131**

Samenvatting **135**

Resumen **139**

Curriculum vitae **143**

List of Publications **145**

Acknowledgments **147**

Little Fly
Thy summer's play,
My thoughtless hand
Has brush'd away.

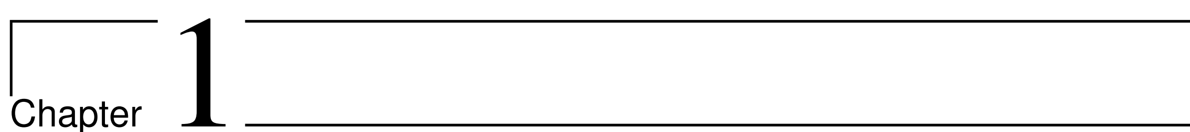
Am not I
A fly like thee?
Or art not thou
A man like me?

For I dance
And drink and sing;
Till some blind hand
Shall brush my wing.

If thought is life
And strength and breath;
And the want
Of thought is death;

Then am I
A happy fly,
If I live,
Or if I die.

William Blake



General Introduction

Ecological interactions can exert strong selection pressures on organisms. Organisms need to cope with abiotic conditions as well as with biotic interactions. Abiotic factors include variable environments through seasonal change and different environmental conditions during distinct life stages. Biotic factors involve, for example, predator-prey or host-parasite relationships. All these ecological interactions may severely impact the fitness of organisms. The variation in the ability to survive and reproduce under a particular set of ecological interactions affect the contribution of genotypes to the next generation (Pelletier *et al.*, 2009). In return, the evolution of traits influences population dynamics, community structure, and ecosystem function, creating so-called eco-evolutionary dynamics (Fussmann *et al.*, 2010).

How organisms respond to their ecological interactions can drastically affect their fitness. It, thus, seems evident that mechanisms to cope with these interactions, such as abilities to, for instance, avoid predators, defend against parasites or survive adverse conditions confer organisms with a selective advantage. Less obvious is under what circumstances such responses evolve and how they are encoded in the genome. Moreover, such mechanisms can also bear costs, which determine the extent to which they are maintained or lost during evolution (Flatt & Heyland, 2011).

Among the strongest selective forces in nature are host-parasite interactions due to the antagonistic fitness relationships. Hosts often suffer costs from infections in terms of morbidity, fertility and survival. To reduce these costs, the host may try to resist the parasite, thereby inflicting costs to the parasites in terms of its development, survival and/or propagation (Zuk & Stoehr, 2002; McKean *et al.*, 2008). The strong selection pressures that the antagonists impose on each other may lead to co-evolutionary arms races between the hosts and parasites (Schmid-Hempel, 2005). Organisms are exposed to a variety of parasites (e.g., bacteria, viruses, fungi, nematodes, parasitoids) and often possess different defence mechanisms against these various types of parasites. These interactions with parasites are considered to have great impact on an organism's life history traits, and as a consequence, on its genome (Christophides *et al.*, 2002; Nielsen *et al.*, 2005; Sackton *et al.*, 2007).

In this thesis I study the evolutionary genomics of host-parasite interactions, using *Drosophila* fruit flies and their parasitoids as model system. The central aim of my thesis is to understand the genomic basis of evolving a response in an ecological interaction. I focus on the defence mechanisms used by *Drosophila* species to fight parasitoid wasps

that lay their eggs in *Drosophila* larvae. The main defence mechanism consists of an immune response, called melanotic encapsulation (Lavine & Strand, 2002). To address the evolution and genomic basis of this immune response, I investigate the molecular mechanisms of encapsulation, its effects on the host's fitness, and the evolution of these mechanisms, through a combination of comparative genomics, parasitization assays, population genetics, field observations, functional genomics and gene expression experiments across different *Drosophila* populations and species.

1.1 Functional Genomics

Genomes are composed of protein-coding and regulatory non-protein coding DNA, as well as structural DNA. Many of these elements may not have direct consequences for an organism's fitness, remaining to a large extent "invisible" to selection (Eddy, 2012). To understand the molecular mechanisms underlying evolutionary processes it is essential to identify the genomic regions that affect an organism's fitness. The aim of functional genomics is to identify these regions by combining traditional genetic and novel sequencing techniques in order to establish how the genotype leads to the phenotype, the so-called genotype-phenotype map. It studies how genetic variation affects molecular, cellular and organismal function and how these processes translate into organismal fitness (Feder & Mitchell-Olds, 2003).

A classic technique to infer the link between genotype and phenotype is linkage mapping, which consists of measuring the correlation between one or multiple phenotypic traits to genetic variation (estimated more recently through, for example, microsatellite markers or single nucleotide polymorphism) (Brown, 2002). This type of "forward genetics" technique, focusing on the display of a particular phenotype and aiming to associate it to its genetic basis, has made a great contribution in finding genetic variants associated with particular phenotypes. For example, it has been successful in identifying disease resistance genes in plants (Stahl *et al.*, 1999) or coat colour genes of pocket mice (Nachman *et al.*, 2003). A serious caveat of using genotype mapping is that it is limited to the mere correlation of variation. In order to gain insights into causal relationships, it is necessary to not only correlate phenotypes to genotypes, but also to measure the phenotypic effects of gene disruption. Measuring the phenotypic effects of induced mutations, in particular DNA sequences or known genes, is the alternative traditional approach, the so-called "reverse genetics" approach. Disruption of gene function can be induced at the DNA level through targeted or site-directed mutagenesis, transgenetic manipulation (knock-out), or by silencing the gene at the level of mRNA (knock-down). The last technique is referred to as RNA interference (RNAi), because it consists in introducing RNA complementary to the target gene, which induces degradation of the endogenous transcript (Wilson & Doudna, 2013). This has become a very important tool in molecular genetics as it provides a controlled way of targeting genes, while phenotypes can be directly compared between the genetic wild type and null mutants. One disadvantage of this technique, however, is that it is primary designed to test one or few genes.

The recent advances in genomics and sequencing have, to a certain extent, overcome the limitations of studying individual genes. High-throughput sequencing allows us to explore the co-variance of multiple genes in a relatively unbiased way, without focusing

only on genes that are already implicated for involvement in a trait or genes with strong effects on the phenotype. Interestingly, massive screens of RNAi along the genome of several model organisms have yielded the insight that a large fraction of the genes in a genome are non-essential or do not induce obvious differences in the phenotype (Carroll & Potts, 2006). This has been interpreted as: 1) redundancy in gene function 2) several genes with very small effects and 3) ecology-dependent function (i.e the ecological condition determines the function of the gene) (Carroll & Potts, 2006). Further study of these alternative scenarios is required to unravel the role of these genes in phenotypic traits and evolutionary processes. Moreover, it emphasizes a need to shift research focus from one gene in one experimental condition to multiple genes in multiple conditions.

Understanding which genes are activated under different conditions and their pattern of co-regulation would allow us to study the intermediate levels of information transfer between the genome and the phenotypic trait. Such intermediate levels encompass messenger RNA and proteins, giving us insights into the complexity of the information transmission. Gene products interact in networks that can, for instance "buffer" the effect of genetic variants or amplify them during transcription, translation, post-translational regulation and activation of signalling pathways. Understanding these intermediate levels is fundamental to characterizing the function of the genome (Barabási & Oltvai, 2004; Schadt *et al.*, 2009).

1.2 Evolutionary Genomics

To understand the history of the selective forces behind the existing mechanisms we need an evolutionary approach. The essence of comparative genomics is to elucidate evolutionary processes by comparing sequences across different clades. To compare sequences, we first need to ensure that those sequences are homologous, that is, that they have a shared ancestry. Sequences can be homologous and belong to different species (orthologs) or have duplicated and diverged within one species (paralogs). From these two clear categories, more complex relationships can be defined, such as a duplication preceding a speciation event (outparalogs) or a duplication that occurred in a specific lineage (in-paralogs) (Koonin, 2005).

Among homologous genes, it is possible to measure the rates of sequence change in order to estimate the force of selection. Conserved sequences are assumed to be under constrained selection because they are functionally important. For example, among the most conserved sequences are functional RNAs (tRNA and rRNA), which are ancient molecules that mediate the most basic cellular processes (Isenbarger *et al.*, 2008). The opposite, however, is not true. Functionally important sequences can diverge fast and be under positive selection (i.e., selection for change), for instance under co-evolution or genetic conflict (Alföldi & Lindblad-Toh, 2013). This has been found particularly in immune genes or genes involved in sexual traits (Wagner, 2007). "New" sequences can also appear in some lineages, which can sometimes be indicative of innovations. These new sequences can appear either *de novo* (e.g from retroposon insertions) (Wang *et al.*, 2002), or, the most common case, because of duplication accompanied by fast divergence. When duplicated copies diverge so fast that their homologs are beyond recognition and hence undetectable, these sequences are classified as lineage restricted. Duplications are a fundamental source of raw material for evolution (Zhang, 2003), because in the

short-term they may provide an advantage on protein dosage regulation, and in the long-term they may enable the acquisition of new protein functions (Kondrashov *et al.*, 2002).

One of the most challenging problems in evolutionary biology is to infer whether a trait has evolved through adaptation, random drift or as side effect of other adaptations (a "spandrel") (Gould & Lewontin, 1979; Barrett & Hoekstra, 2011). In the genomics approach, this problem has been partially solved by defining the neutral mutation rate as the rate of synonymous substitutions (dS), i.e. nucleotide substitutions that do not change the encoded amino acid (Yang & Nielsen, 2000) (synonymous codons may, however, be used in unequal frequencies, referred to as "codon bias" (Plotkin & Kudla, 2011)). Non-synonymous substitutions (dN) are nucleotide substitutions that change the encoded amino acid. Since natural selection operates mainly at the protein level, synonymous and non-synonymous substitutions are fixed at different rates (Yang, 2006). Thus, a common approach to identify deviation from neutrality is by comparing the rate of non-synonymous (dN) to synonymous substitutions given the number of silent substitutions per synonymous site ($\omega = dN/dS$). The rate ω has been defined to describe the type of selection, such that a similar rate of non-synonymous to synonymous substitutions ($\omega \approx 1$) indicates neutrality, a rate where dS exceeds dN ($\omega < 1$) indicates stabilizing selection and a dN greater than dS ($\omega > 1$) indicates directional selection (Yang, 2006).

Unfortunately, the rate ω can only be calculated for coding sequences (i.e sequences that code for a protein), while non-coding DNA and RNA are also fundamental in the regulation of gene expression and thus in the production of phenotypes. In fact, regulatory regions have been shown to play an important role in the evolution of innovations and in the differentiation of major lineages (Wittkopp & Kalay, 2012). This is reflected in their evolutionary rates, which tend to be generally faster for enhancers and transcription factors binding sites than coding regions, both in prokaryotes and eukaryotes (Cordero & Hogeweg, 2007; Otterloo *et al.*, 2013; Rubinstein & de Souza, 2013). Methods to assess the signatures of selection on non-coding regions are also becoming available (Kohn *et al.*, 2004; Andolfatto, 2005; Haddrill *et al.*, 2008).

The availability of genomes from different lineages, together with statistical methods, have contributed to the rapid development of the field "comparative genomics". Different evolutionary questions can be addressed by comparing genomes with different phylogenetic distances (Hardison, 2003). For example, questions regarding the size of protein families in different genomes and how many genes are shared among distantly related clades could be addressed with the availability of full genomes. The proteome size of yeast, worms, and flies revealed that these were of a comparable size, and that much of the genomes of flies and worms consisted of duplicated genes (Rubin *et al.*, 2000). The first comparative genomics study within a genus of multicellular eukaryotes was carried out in *Drosophila*. This study came with the publication of the *D. pseudoobscura* genome (Richards *et al.*, 2005) after that of *D. melanogaster* genome in 2000 (Myers *et al.*, 2000). The comparison of closely related species provided the opportunity to investigate questions related to the evolution of gene content, gene regulation, gene order, chromosome structure, and genome architecture (Singh *et al.*, 2009). The pioneering studies that have been carried out with this model system for over a century are now being continued with a great contribution of *Drosophila* to the comparative

genomics field (Rubin & Lewis, 2000; Singh *et al.*, 2009).

1.3 *Drosophila* as model organism

Drosophila melanogaster has been extensively studied as a model organism in diverse fields of Biology. In particular, research on genetics, developmental biology and more recently immunity has contributed greatly to our understanding of the molecular and developmental mechanisms underlying phenotypic traits (Letsou & Bohmann, 2005; Lemaitre & Hoffman, 2007). Many of the findings in *Drosophila* have been shown to be largely representative for both invertebrates and vertebrates. For example, the Toll pathway was first described for *D. melanogaster* in both development and immunity, and it was thereafter found to play a fundamental role in immunity from invertebrates to vertebrates (Kimbrell & Bruce, 2001).

The genus *Drosophila* comprises approximately 1500 species. Its biogeographical origin is believed to be in the tropics of south Asia, from where it radiated around 80-120 million of years ago (late Cretaceous). A major split occurred in the Old World tropics, giving rise to the two subgroups, *Sophophora* and *Drosophila* 50 million years ago. These two groups in turn split again into Old World and New World groups. *Sophophora* gave rise to the Old World *melanogaster* and New World *willistoni* and *saltans*, while *obscura* evolved in the African tropics (Powell, 1997). The subgroup *Drosophila* gave rise to the Old World *virilis* and the New World *repleta*, while the origin of the *hawaiian* clade is not clear. Most of the *Drosophila* species breed on rotting plant or fungal material and use microorganisms as main nutritional source (Powell, 1997).

At the start of this project, 12 species of the genus *Drosophila* were fully sequenced and annotated, coming from different geographical, phylogenetic and ecological ranges (*Drosophila* 12 Genomes Consortium, 2007) (Figure 1.1). Since then other genomes have been added to the public databases. The 12 sequenced species comprised cosmopolitan species such as *D. melanogaster* and *D. simulans*, species with large geographical ranges such as *D. annanassae* (Asia and Pacific), *D. yakuba* (Africa), *D. virilis* (Holartic), *D. pseudoobscura*, *D. persimilis* and *D. willistoni* (America), and some species with (very) limited distributions such as *D. erecta* (west Africa), *D. mojavensis* (Mojave desert), *D. grimshawi* (Hawaii) and *D. sechellia* (Seychelles Islands) (Singh *et al.*, 2009).

1.4 Immune system

The immune response of an organism enables it to cope with the diversity of parasitic organisms to which it is exposed. As these parasites can impose direct and severe fitness effects, the immune response is under strong selection pressure. This is largely reflected in the great diversity of immune defences and the rapid evolutionary rates found among certain immune genes (Kimbrell & Bruce, 2001; Christophides *et al.*, 2002; Nielsen *et al.*, 2005; Sackton *et al.*, 2007; Obbard *et al.*, 2009b). At the same time, immune systems share important attributes across distant phylogenetic taxa (e.g., both insects and humans), which suggests that they also rely largely on ancient basic blocks

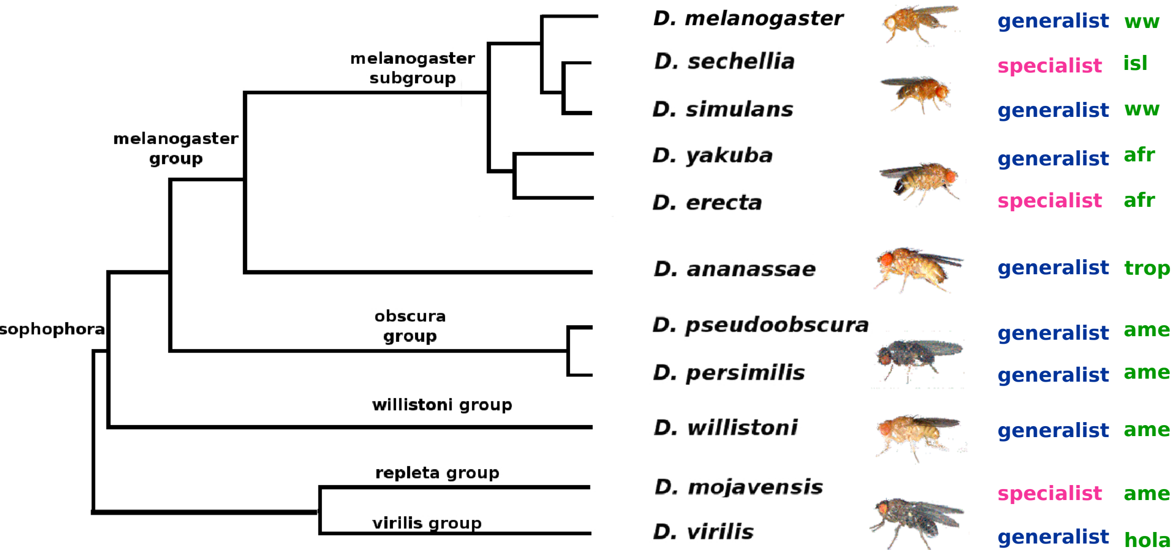


Figure 1.1: **11 sequenced *Drosophila* species used in this thesis.** Complete genomes were publicly available in 2007 for 12 species. For 11 of the species (i.e., all except *D. grimshawi*), the strains that were sequenced were available for experiments. Species show different distributions (ww: world wide, isl:island, afr: Africa, ame: America, trop: tropics, holar: holartic) and feeding ecology

(Williams, 2007; Fauverque & Williams, 2011). The immune system is usually classified into adaptive and innate immunity. The adaptive arm (the antibody-based immunity) is a derived monophyletic defence response present in jawed vertebrates, which was built atop the innate immune system (Klein & Nikolaidis, 2005). The innate immunity includes numerous independent systems that compose the first line of defence present in all multicellular organisms (Kimbrell & Bruce, 2001).

Information about the mechanisms and evolution of innate immunity comes to a large extent from studies in *Drosophila melanogaster*. From these studies we know that after an immune challenge certain signal transduction pathways are activated (Toll, IMD, JAK/STAT and JNK) that can trigger the release of antimicrobial peptides and other effector proteins (referred to as humoral response), and the proliferation and differentiation of specialized cell-groups (referred to as cellular response) (Lemaitre & Hoffman, 2007). Both the humoral and cellular components act together. However, their activation and regulation can differ substantially as the process of producing and releasing (humoral) molecules is fundamentally different from the process of differentiating and proliferating specialized cells.

The large diversity of parasites has lead to both variety and specificity in the immune responses. The triggered immune response needs to adjust to the type of parasite it is directed to. Microparasites (e.g bacteria and fungi) can be neutralized by effector molecules and phagocytized, while macro-parasites (e.g parasitic wasps) are too big to

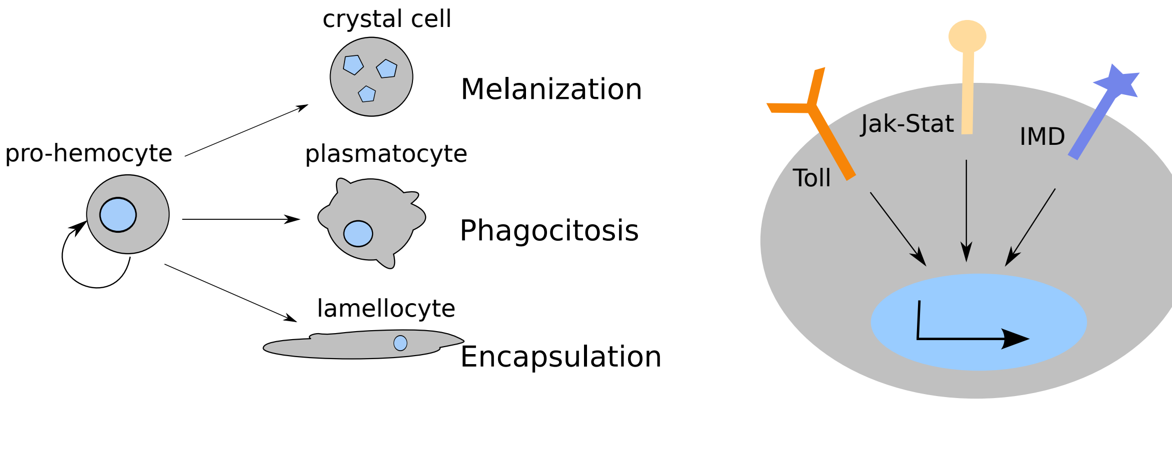


Figure 1.2: **Innate Immunity.** Left: the cellular component consists of epithelial barriers and controls the proliferation and differentiation of specialized cells. Right: the humoral component triggers the production of cellular-free molecules that directly attack the pathogens.

be phagocytized and need to be sequestered by a multicellular layer of specialized cells (Figure 1.2). Besides the size, parasites also differ in their biochemical composition and phylogenetic distance to the host, which may be important for their recognition. Most importantly, parasites also differ in their effect on the host. Some parasites are relatively innocuous, others can be detrimental to the survival and or reproduction and some are always lethal. Among the last type are parasitoids, which are insects that use other arthropods as hosts and kill them often before reproduction, thus exerting an extremely strong selection pressure (Godfray, 1994).

1.5 The *Drosophila* immune response against parasitoids

Parasitoids are a common and important factor in controlling insect populations (Hawkings & Sheehan, 1994). Parasitoid females inject their eggs into a host, where they develop into larvae that consume the host. It is thought that radiation of insects during the Cretaceous was enabled by the emergence of flowering plants that created new conditions for mutual divergence (Ehrlich & Raven, 1964). In this scenario, insects rapidly adapted to new environments. They themselves became new niches for insect parasitoids (ancestral Diptera and Hymenoptera), which probably gave rise to the great diversity in parasitoids found today (Godfray, 1994).

In some host species, parasitoid attack triggers an immune response to encapsulate the parasitoid egg, resulting in the death of the parasitoid and survival of the host. If the response is not complete or fast enough, the parasitoid develops by killing the host. Encapsulation ability has evolved in a wide range of insect orders, such as Hemiptera (Suma *et al.*, 2012), Orthoptera (Lackie *et al.*, 1985), Lepidoptera (Blumberg & Ferkovich, 1994; Strand & Pech, 1995), Dictyoptera (Lackie *et al.*, 1985), Coleoptera (Reed

et al., 2007) and Diptera (Nappi & Stoffolano, 1971; Eslin *et al.*, 2009). The underlying mechanisms, including the type of differentiated blood cells vary greatly among the different orders (Strand & Pech, 1995), and it remains unclear whether there is a shared evolutionary relationship. Similarly, parasitoids have evolved a variety of strategies to counter-attack the immune response of the hosts. These include, for example, the injection of immuno-suppressive virulence genes coming from DNA viruses (Bitra *et al.*, 2012) and the production of RhoGAP toxins by the parasitoids that induce changes in morphology and adhesion properties of host hemocytes (Colinet *et al.*, 2007).

Most species of *Drosophila* are natural hosts to parasitoid wasps, with members of the genera *Asobara* (Hymenoptera: Braconidae) and *Leptopilina* (Hymenoptera: Figitidae) being the most widespread (Carton *et al.*, 1986; Fleury *et al.*, 2009). When a parasitoid female lays an egg in the host larva, this egg needs to be detected by the host as a "foreign body", surrounded with multiple layers of hemocytes, and fully melanized for the host to survive the parasitoid attack. The differentiation and mobilization of hemocytes (blood cells) is a critical step in this process (Fauverque & Williams, 2011). In *D. melanogaster* three types of differentiated blood cells have been described: 1) plasmatocytes, which perform phagocytosis of bacteria and other small pathogens and are also recruited in the cellular capsules around parasitoid eggs, 2) lamellocytes, which are large, adhesive and flat cells that are important for the formation of the cellular layers around the foreign bodies (e.g., parasitoid eggs); and 3) crystal cells, which store the precursors of the melanin that is deposited on invading pathogens (Pech & Strand, 1996; Williams, 2007).

Within the *Drosophila* genus there is large variation in the encapsulation ability against parasitoids, from completely absent in some species to being strong in others. An important aspect of this variation has been associated with characteristics of the hemocytic response after parasitoid attack (Eslin & Prévost, 1998). Some species in the obscura group are deficient for encapsulation ability and this was associated to the lack of lamellocytes (Eslin & Doury, 2006). In the melanogaster subgroup the encapsulation success rate varies among species, and this was shown to correlate with either the constitutive or induced hemocyte load after parasitoid attack (Eslin & Prévost, 1998). Substantial variation has not only been found among species but also among populations of *D. melanogaster*, with a strong geographic component (Kraaijeveld & Godfray, 1999; Gerritsma *et al.*, 2013). Strains of *D. melanogaster* can also be experimentally selected to double their resistance in a few generations (Kraaijeveld & Godfray, 1997; Jalvingh *et al.*, 2014). Strikingly, lines selected for higher resistance show a significant increase in the constitutive level of hemocytes (Kraaijeveld *et al.*, 2001; Wertheim *et al.*, 2011), while in natural populations, the level of constitutive and induced hemocyte types varies greatly and does not necessarily correlate with the level of resistance (Gerritsma *et al.*, 2013).

The main parasitoid species that are used in this thesis to study the immune response of *Drosophila* hosts are solitary parasitoid wasps from the genus *Asobara*, in particular *A. tabida* (Figure 1.3). The distribution of *A. tabida* is holartic, but representatives of the genus are also found in Africa (e.g., *A. citri*) (Prevost *et al.*, 2005), Asia (e.g., *A. pleuralis, A. japonica*) (Mitsui *et al.*, 2007) and Australia (e.g., *A. persimilis*) (Papp, 1977). In Europe, *A. tabida* is known to use various *Drosophila* species as natural hosts. These host species differ in their abilities to raise a successful im-

mune response, ranging from high (e.g., *D. simulans*), to various intermediate levels (*D. melanogaster*) to completely absent (*D. subobscura*) (Kraaijeveld & Godfray, 2009). Thus, co-occurrence with the parasitoid is insufficient to explain the great variability found in defence mechanisms, which partly motivated the study presented here.

Figure 1.3: ***Asobara tabida* parasitizing *Drosophila* larvae** The female wasp injects an egg inside the fly larvae, where it can develop by consuming the host

1.6 This thesis

In this thesis I use *Drosophila* fruit flies and their immune response to parasitoids to study the evolution of a response to an ecological interaction. I combine comparative genomics, parasitization assays, population genetics, field observations, functional genomics and gene expression experiments to compare the immune response against parasitoids in *Drosophila* species and lines. I use these approaches to address different aspects of the mechanism and genomic basis of this immune response. These aspects are finally integrated to propose specific hypotheses on the evolution of the immune response against parasitoids in *Drosophila*. More generally, the insights gained from this work are used to motivate a general discussion on the evolution of a defence mechanism shaped by the interplay between ecology and genomes.

Chapter 2 consists of a genomic and phenotypic characterization of the cellular immune response against parasitoids across 11 sequenced *Drosophila* species. First, I quantify the level of resistance and report the production of different types of hemocytes in the different *Drosophila* species after exposure to *A. tabida*. Then I use a comparative genomics approach on a list of candidate genes coming from a previous microarray study (Wertheim *et al.*, 2005) and from genes with hemopoietic functional annotation in *D. melanogaster*, to quantify the presence-absence and the rates of amino acid substitution of these genes across all species. For a subset of genes, the level of expression is also measured and associated to the level of resistance. Combined data from phenotypic

characterization, genomics and expression patterns, lead us to propose a scenario for the evolution of the resistance against parasitoids, highlighting a subset of genes that may be tightly associated with the evolution of this response.

While one strain per species was used for the characterization in **chapter 2**, it is well established that large intra-species variation exists in the resistance against parasitoids trait. In **chapter 3**, I shift focus to 8 field lines of *D. melanogaster* collected across Europe, which differ in resistance against *A. tabida*. For these field lines, we zoom in onto one of the immune receptors, *Tep1*. In the comparative approach across species, this gene was shown to be evolving fast. Furthermore, the expression of *Tep1* correlated with the level of resistance of a subset of species. In this chapter, I quantify the sequence variation among the field lines in *Tep1* and four other immune receptors, and measure *Tep1* expression after parasitization. To establish the function of *Tep1* in the immune response against parasitoids, we use a RNAi approach to knock down *Tep1* expression.

The candidate genes that were used in the genomic characterization in **chapter 2** were based on genome-wide expression studies in of *D. melanogaster*. No genomic information was available on the parasitoid response of closely related *Drosophila* species. In **chapter 4** I report on a RNAseq approach to characterize and compare the genome-wide expression of *D. melanogaster*, *D. simulans*, *D. sechellia* and *D. yakuba* when exposed to *A. tabida*. Moreover, the expression profile of two *D. melanogaster* lines selected for higher resistance against parasitoids and two non-selected (control) lines is compared to characterize the changes in gene expression after short-term evolution of higher resistance. This combination enabled a comparison of the inter- and intra- species variation in gene expression in response to parasitoid attack, while also presenting insights into the short- and long-term evolutionary patterns of this immune response.

The genomic and phenotypic characterization in **chapter 2** and the expression profiles found in **chapter 4** led to the formulation of specific hypotheses about the implications of *D. sechellia*'s ecological specialization and its inability to resist against parasitoids. In **chapter 5** I report on a laboratory experiment to test whether the breeding substrate of *D. sechellia* may be toxic for parasitoids. I also describe a field study conducted to investigate the host-parasite interactions of this species in its natural habitat. I collected biological material in the natural reserve of the Seychelles, Cousin, in order to characterize the fruit fly and wasp community. Based on my findings, I re-formulate our hypothesis on the implications of the ecological specialization of *D. sechellia* for parasitoid resistance. This also provides new insights into the role of the ecological context in which genomes evolve.

Finally, in **chapter 6**, I integrate the results from my whole study on the evolution of the *Drosophila* immune response to parasitoids. I discuss the insights gained from the different approaches and how they contributed to understanding the genomic changes and ecological conditions that enabled the evolution of an ecological response, and the factors that contribute to its variation.

Evolution of a cellular immune response in *Drosophila*: a phenotypic and genomic comparative analysis

Laura Salazar-Jaramillo, Angeliki Paspati, Louis van de Zande, Cornelis Joseph Vermeulen, Tanja Schwander, Bregje Wertheim.

Published in: *Genome Biology and Evolution 2014, 6: 273-289*

Abstract

Understanding the genomic basis of evolutionary adaptation requires insight into the molecular basis underlying phenotypic variation. Yet, even changes in molecular pathways associated with extreme variation, gains and losses of specific phenotypes, remain largely uncharacterized. Here we investigate the large interspecific differences in the ability to survive infection by parasitoids across 11 *Drosophila* species and identify genomic changes associated with gains and losses of parasitoid resistance. We show that a cellular immune defense, encapsulation, and the production of a specialized blood cell, lamellocytes, are restricted to a sublineage of *Drosophila*, but that encapsulation is absent in one species of this sublineage, *D. sechellia*. Our comparative analyses of hemopoiesis pathway genes and of genes differentially expressed during the encapsulation response revealed that hemopoiesis-associated genes are highly conserved and present in all species independently of their resistance. In contrast, 11 genes that are differentially expressed during the response to parasitoids are novel genes, specific to the *Drosophila* sublineage capable of lamellocyte-mediated encapsulation. These novel genes, which are predominantly expressed in hemocytes, arose via duplications, whereby five of them also showed signatures of positive selection, as expected if they were recruited for new functions. Three of these novel genes further showed large-scale and presumably loss-of-function sequence changes in *D. sechellia*, consistent with the loss of resistance in this species. In combination, these convergent lines of evidence suggest that co-option of duplicated genes in existing pathways and subsequent neofunctionalization are likely to have contributed to the evolution of the lamellocyte-mediated encapsulation in *Drosophila*.

2.1 Introduction

The evolution of immune systems is driven by the large diversity of parasites that organisms are exposed to. The ongoing selection pressure is at the root of the extensive variation underlying many of the genes in the immune defense pathways (Christophides *et al.*, 2002; Nielsen *et al.*, 2005; Sackton *et al.*, 2007; Obbard *et al.*, 2009b). Yet immune defense pathways also comprise elements that are highly conserved across multi-cellular organisms, such as Toll receptors that function in innate immunity of both vertebrates and invertebrates (Kimbrell & Bruce, 2001). Conservation may be expected for genes involved in multiple processes or genes that occupy key positions in interaction networks, because increased connectivity can generate greater constraints on protein structure (Fraser *et al.*, 2002); but see also (Kopp & McIntyre, 2010). With the availability of genome sequences of related species and the tools to investigate genome changes, tackling the complexity of the evolutionary history of the immune systems has become possible.

Insects are ideal for studying the evolution of the immune response, because their immune system is relatively simple compared to vertebrates yet potent and multi-facetted. Like all invertebrates they rely solely on innate immunity (Lemaitre & Hoffman, 2007). This innate immunity system consists of two interacting components, a humoral component, involving the release of molecules such as antimicrobial peptides, and a cellular component, involving the differentiation of several specialized cell-groups. Both humoral and cellular components are activated after an immune challenge, but the reaction cascades induced by microparasites (e.g bacteria and fungi) and macroparasites (e.g parasitic wasps) result in substantially different defensive responses, because micro- and macro-parasites differ in size and biochemical composition (important for recognition), and they require different mechanisms to be eliminated or disarmed (Lemaitre & Hoffman, 2007).

During the humoral response, surface proteins of pathogens are detected by pattern recognition proteins of the host, which activate two primary signal transduction pathways, the Toll and IMD pathways. A third immunity pathway, RNAi, is directed against viruses. These pathways trigger the transcription and release of antimicrobial peptides and other effector proteins, which directly attack parasites (Lemaitre & Hoffman, 2007). Comparative genomic studies in the genus *Drosophila* revealed divergent evolutionary patterns for different groups of humoral immune genes. Most genes in the signal transduction pathways occur as single orthologous copies in each species' genome and are highly conserved, while genes encoding pattern recognition and effector proteins have diversified rapidly across species (Sackton *et al.*, 2007). This diversification has been interpreted as the result of a co-evolutionary process with the parasites interacting with the hosts' immune response (Obbard *et al.*, 2009b). Genes encoding recognition proteins diversified mainly by accumulating coding mutations, whereas genes encoding effector proteins diversified primarily through duplication (Sackton *et al.*, 2007; Waterhouse *et al.*, 2007).

The cellular response involves epithelial barriers, as well as specialized blood cells. Different types of blood cells (collectively called hemocytes) mediate defensive processes, whereby the hemocytes can change in morphology and abundance after infection (Gillespie & Kanost, 1997; Krzemien *et al.*, 2010). In *D. melanogaster*, the three

most common blood cell types are plasmatocytes, lamellocytes and crystal cells. Plasmatocytes perform phagocytosis of bacteria and other small pathogens; lamellocytes form a layer around large foreign bodies; and crystal cells store the precursors of the melanin that is deposited on invading pathogens (Fauverque & Williams, 2011). In unchallenged *Drosophila* larvae, lamellocytes are typically absent or detectable only in very low densities among the circulating hemocytes, while parasitization by macroparasites can (strongly) induce the proliferation and differentiation of lamellocytes from both the lymph glands (the hematopoietic organ in *Drosophila*) and from circulating undifferentiated hemocytes (for simplicity, we may refer to this induced proliferation and differentiation as the "production" of lamellocytes). The main cellular immunity pathways are the Toll, JAK/STAT and JNK pathways (Meister, 2004), but it is not clear whether selection pressures imposed by parasites may have driven diversification patterns in these pathways similar to those found in the humoral pathways.

There are at least two reasons why the evolutionary patterns found for the humoral response may not be representative of the cellular response. Firstly, the process of producing and releasing (humoral) molecules is fundamentally different from the process of differentiating and proliferating specialized cells. Secondly, expression experiments indicated that the genes differentially expressed after microbial infection differ considerably from those differentially expressed under parasitoid attack, and the humoral pathways RNAi and IMD do not show up-regulation under wasp attack (Wertheim *et al.*, 2005; Schlenke *et al.*, 2007). These substantial differences may be the consequence of different evolutionary dynamics for the humoral and cellular innate immune responses.

In this study, we investigate the genomic changes associated with the evolution of cellular immunity in the *Drosophila* genus, specifically the encapsulation response against parasitoids. Parasitoids are insects that lay eggs in or on other insects, and kill their host during development (Godfray, 1994). To neutralize a parasitoid egg by encapsulation, the host has to detect the egg, surround it with multiple layers of hemocytes, and fully melanize it (from hereon this process is referred to as "encapsulation ability"). When the melanotic encapsulation response is not fast or strong enough, the developing wasp kills the host (Strand & Pech, 1995). Within the *Drosophila* genus, there is large variation in encapsulation ability, from completely absent in some species to high in others. The hemocyte load of the host (constitutive or induced) was shown to correlate with encapsulation success rates in species of the melanogaster subgroup (Eslin & Prévost, 1998). The ability to encapsulate does not (only) depend on the natural exposure to parasitoids, since some species in the obscura group are natural hosts of parasitoid wasps, but completely deficient for encapsulation ability (Eslin & Doury, 2006; Havard *et al.*, 2009). In order to investigate the genomic basis of the ability to encapsulate, we conducted parasitization experiments and a genomic characterization across a broad taxonomic range of 11 sequenced *Drosophila* species (Figure 2.1) (*Drosophila* 12 Genomes Consortium, 2007). Focusing on genes that have been shown to be involved in hemopoiesis (Zettervall *et al.*, 2004; Williams, 2007; Stofanko *et al.*, 2010; Avet-Rochex *et al.*, 2010) and on genes differentially expressed after parasitoid attack in *D.melanogaster* (Wertheim *et al.*, 2005; Schlenke *et al.*, 2007), we identified orthologs in all 11 species, and studied the divergence in terms of both (i) presence - absence and (ii) sequence variation of protein coding genes.

2.2 Materials and Methods

Species strains The 11 *Drosophila* strains used in this study were all genome project strains from the Drosophila Stock Center (San Diego University) (Drosophila 12 Genomes Consortium 2007) (Table S2.1). Flies were reared at 20 ℃ under a dark:light regime of 12:12 and 50 % relative humidity in quarter-pint bottles containing 30 mL standard medium (26 g dried yeast, 54 g sugar, 17 g agar and 13 mL nipagine solution per litre), supplemented with a small piece of banana. The *A. tabida* strain was originally collected in Sospel, France and has been maintained on *D. subobscura* at 20 ℃ under a dark:light regime of 12:12. It has a moderately to high virulence and produces so-called "sticky eggs" that can adhere to host tissue to evade full encapsulation. The *A. citri* strain was collected in Ivory Coast and has been maintained on *D. melanogaster* at 25 ℃ under a dark:light regime of 12:12.

Encapsulation assay We tested the encapsulation ability of the 11 *Drosophila* species (Drosophila 12 Genomes Consortium 2007) against two different parasitoid wasp species from the *Asobara* genus, *A. tabida* and *A. citri*. Fifty second-instar larvae (approximately 48 hours after egg laying at 25 ℃) were exposed to either two wasp females of *A. tabida* or one female of *A. citri*. We used 2 wasps for *A. tabida* to increase parasitization rates, while for *A. citri*, single females achieved high parasitization rates. All infections were carried out at 20 ℃ on a petri dish of 70 mm diameter filled with standard medium. Typically 8 petri dishes with 50 larvae were examined, while for some species only 4 (due to culturing difficulties). Wasps were removed 3 days later and 5 larvae per petri dish were dissected to confirm parasitization by the wasp (except for the *D. mojavensis*, for which dissections were not carried out because the amount of eggs laid and larvae developed was too small). We recorded super-parasitism in our dissection assays, which was occasionally found but did not differ substantially among host species, nor affected the results qualitatively (data not shown). The rest of the larvae were allowed to complete development and the number of emerging flies with capsule and wasps was recorded for each petri dish. Capsules in adult flies were recorded by squashing the adult between two glass slides under a stereo microscope. Each petri dish was considered an independent replicate. We used a Generalized Linear Model (glm) implemented in R 2.15.1 (R Development Core Team, 2008) to analyze the number of wasps and flies with capsule (ratio) that emerged (binding the variables in a matrix) and considering fly species (FlySp) as explanatory variable. We used Binomial error distribution (logit-link function), and a quasibinomial distribution in order to correct for overdispersion.

$$glm(ratio \sim FlySp, fam = quasibinomial) \tag{2.1}$$

To test the contribution of the explanatory variable to the model, we used an analysis of deviance for generalized model fit using F-tests.

Lamellocyte identification To assess lamellocyte production, we exposed 50 second-instar larvae to *A. tabida* and observed the oviposition behavior of the wasps. We collected only larvae for which parasitization was recorded (the wasp spent at least

10 seconds ovipositing). We also collected larvae that were not exposed to wasps as control. At 96 hours after parasitization, i.e when the larvae were in the third-instar stage, we pricked five larvae with a fine needle and collected their pooled hemolymph. We diluted 1 μl of the pooled hemolymph into 7 μl of Ringer's solution (13.6 g KCl, 2.7 g NaCl, 0.33 g CaCl2 and 1.21 g tris solution per liter) to fill a hemocytometer slide Neubauer Improved ®(0.1mm depth). We repeated this at least five times per species. We observed the samples at 40x objective magnification under a phase-contrast microscope. Lamellocytes can be recognized by their flat shape compared to other blood cells (Figure S2.1). Pictures were made with a Moticam 2000 (2 M pixel) camera.

Melanization In order to test the ability to melanize after injury, five second-instar larvae were pricked with a fine needle and scored for the presence of a black spot after 2 and 4 hours.

Candidate genes The set of candidate genes we analyzed was composed of 144 protein-coding genes, 35 with a GeneOntology annotation of "hemocyte differentiation", "hemocyte proliferation" or "regulation of hemocyte differentiation" in Flybase (version FB2012 04) (McQuilton *et al.*, 2012), and 109 protein-coding-genes based on the studies by Wertheim *et al.* (2005); Schlenke *et al.* (2007), and compiled in Kraaijeveld & Wertheim (2009). Both studies are genome-wide expression data from microarray experiments of *Drosophila* larvae parasitized by wasps: the first study by *A. tabida* and the second by *Leptopilina heterotoma* and *L. boulardi.*

Orthologous groups and homology categories Orthologs to the *Drosophila* candidate genes in the remaining 10 *Drosophila* species were found using OrthoMCL (Li *et al.*, 2003). This algorithms uses BLAST similarity score to find best reciprocal hits between complete genomes (we used the default cut-off value, i.e 10^{-5}), clustered into within-species best reciprocal hits (inparalogs) and between-species best reciprocal hits (outparalogs). Outparalogs are those proteins that share orthologs inside and across species and represent ancient duplicates (predating speciation). The distinction between in- and out- paralogs allows the differentiation of recent from ancient paralogs. We used these clusters of orthologs to detect the pattern of gene presence-absence. We used three general homology classes: "single copy ortholog" (SCO) for genes that have exactly one copy in each species, "paralog" (PAR) for genes with multiple orthologs in more than 2 species and "lineage restricted" (LR) for those genes present in a (monophyletic) subset of the lineage. Recent paralogs were included in the lineage restricted class, since they constitute lineage-specific expansions. The clusters of orthologous groups were aligned using ClustalW 2.0.10 (Larkin *et al.*, 2007). Functional domains were visualized in Pfam, a database of protein families (Punta *et al.*, 2012).

Phylogenetic analysis For the recent duplications (in-paralogs), we analyzed the protein tree in order to distinguish the new copy from the old. We used ModelGenerator (Keane *et al.*, 2006) to choose the best substitution model for each particular cluster. Then we reconstructed the phylogeny with PhyML v3.0 (Guindon *et al.*, 2010) and calculated the bootstrap values of each branch 100 times. Phylogenetic trees were made

with PHYLIP (Felsenstein, 2005) and drawn with FigTree (`http://tree.bio.ed.ac.uk/software/figtree/`).

Immune classification The 35 genes annotated in Flybase with a function in hemocyte differentiation and proliferation were classified as "hemopoiesis". For the 109 genes from the genome-wide expression study, we followed the immunological categories in (Sackton *et al.*, 2007; Waterhouse *et al.*, 2007): recognition, signaling and effector. We included an extra functional category, namely serine-proteases, which is analogous to the "modulator category" in (Waterhouse et al. 2007). "Recognition" refers to putative pattern-recognition receptors and proteins involved in binding; "signaling proteins" are those that have been characterized in immune signal transduction pathways, namely Toll, Jak/Stat, IMD and JNK; and "effectors" are antimicrobial peptides, phenoloxidases and intermediates in the melanin production.

Positive selection We used PAML 4.4 (Yang, 2007), a package of programs for analyzing sequences using maximum likelihood. This program is based on the phylogenetic comparison of synonymous (dS) and non-synonymous (dN) substitution rates, expressed in the ratio: $\omega = dN/dS$. We applied a maximum likelihood test in two sets of models that allow ω to vary per position: one nearly-neutral (M1a) model where ω is between (0,1) against a model of positive selection where ω is between (0,2). Models M7 and M8 use the same concept but for a continuous beta distribution. We calculated two times the difference in likelihoods between the corresponding nested models (i.e., M1 versus M2, and M7 versus M8), obtained the p-value from a χ^2 distribution with two degrees of freedom (Yang, 2007) and corrected for multiple testing using FDR implementing a bootstrap method. This analysis was applied only to branches that had one orthologous gene copy in each species. For instance, in the big family of Tep, we applied the test independently to *TepI, TepII, TepIII*, etc., rather than to the whole orthologous group. Orthologous groups with multiple copies in one species (i.e., paralogs) were left out of the analysis.

RT-qPCR To compare gene expression between parasitized and control larvae of *D. melanogaster, D. simulans* and *D. sechellia*, RT-qPCR was performed on second-instar larvae parasitized by *A. tabida* (we used the inbred "TMS" line derived from the *A. tabida* strain collected in Sospel) and non-parasitized control larvae (collected in parallel). For each biological replicate, total RNA was extracted from pools of five larvae that were collected at 5 and 50 hours after parasitoid attack. These time points were chosen based on the expression profiles of our 3 target genes (see below) in an earlier micro-array experiment (Wertheim *et al.*, 2005). RNA was extracted and purified using a combination of Trizol (Invitrogen, Carlsbad, CA, USA) and RNeasy (Qiagen, Hilden Germany), according to the manufacturers' instructions. Tissue homogenization and cell lysis were performed using a pestle in 1ml Trizol, and RNA purification on the RNeasy columns included genomic DNA digestion with DNAseI (Qiagen, Hilden Germany). cDNA was synthesized from 10μl RNA, using RevertAid RT (Fermentas). Primers were designed on exon-exon boundaries whenever possible, using the Perlprimer software (Marshall, 2004). A common primer set for all species could be designed for the two endogenous reference genes (Actin5 and Forkhead68A), and two target genes (*IM1* and *PPO3*).

The high divergence of *TepI* necessitated a specific primer set for each species (See Table S3.1 for the primers). Primers were checked for linear amplification efficiencies and optimized. The cDNA template for one of the reference genes (Actin5) had to be diluted (5 hours: 50x; 50hours: 100x) to avoid formation of secondary structures. The qPCRs were performed in total volumes of $25\mu l$ per reaction in an Applied Biosystems 7300 Real Time PCR System, using Absolute QPCR SYBR ®Green ROX mix (Abgene, Hamburg, Germany). Data were analysed, using the algorithm implemented in the statistical package qpcR (version 1.3-6) (Ritz & Spiess, 2008). The median of three technical replicates was obtained for each of five biological samples. Quantification was based on the window-of-linearity method that incorporates individual PCR efficiencies for each sample. The expression of the target genes per biological replicate was standardized to the geometric mean of the two reference genes (Vandesompele *et al.*, 2002). Statistical differences were estimated for the fold-changes between parasitized and control larvae using the permutation method for error estimation. All scripts were run using Python 2.7.3 and R 2.15.1, and are available upon request.

2.3 Results

2.3.1 Phenotypic characterization: Only species of the melanogaster subgroup show encapsulation ability and produce lamellocytes

For the phenotypic characterization we used 11 *Drosophila* species, of which the genomes are publicly available. These species come from different geographical ranges, some being cosmopolitan such as *D. melanogaster* and *D. simulans*, some with large geographical ranges such as *D. annanassae* (Asia and Pacific), *D. yakuba* (Africa), *D. virilis* (Holartic), *D. pseudoobscura*, *D. persimilis* and *D. willistoni* (America), and some species with (very) limited distributions such as *D. erecta* (west Africa), *D. mojavensis* (Mojave desert) and *D. sechellia* (Seychelles Islands) (Powell, 1997; Singh *et al.*, 2009). Species of *Drosophila* are known to act as host for a variety of larval and pupal parasitoids, with members of the genera *Asobara* (Hymenoptera: Braconidae) and *Leptopilina* (Hymenoptera: Figitidae) being the most common threat across the world (Carton *et al.*, 1986; Fleury *et al.*, 2009). We used *Asobara tabida* to test the encapsulation ability of the *Drosophila* species, as this species has an evasive virulence mechanism (some strains, including ours, produce "sticky eggs") that does not require specificity in the host defenses (Eslin & Prévost, 2000). The *A. tabida* distribution is holartic, and it has been found as natural parasitoid of some species of the melanogaster and obscura groups in Europe and America (Eslin & Prévost, 2000; Kraaijeveld & van Alphen, 1993).

The proportion of larvae that successfully encapsulated eggs of the parasitoid wasp *A. tabida* varied significantly among *Drosophila* species (glm, F = 53.37, DF = 51,8 , $P < 2.2e^{-16}$) (Figure 2.1). To ensure that the lack of resistance in some *Drosophila* species was not due to a lack of co-evolutionary history with the holartic *A. tabida* (e.g., a complete lack of species interaction could result in failure to recognize or respond to the immune challenge), we also tested the encapsulation ability against an African *Asobara* species, *A. citri* and screened the literature for additional information. *Drosophila*

species unable to encapsulate *A. tabida* were also unable to encapsulate *A. citri*, or two parasitoid species from the genus *Leptopilina* (Table 2.1). Of all *Drosophila* species tested, only species of the melanogaster subgroup, except *D. sechellia* inside this group, showed any encapsulation ability against *A. tabida* (Figures 2.1).

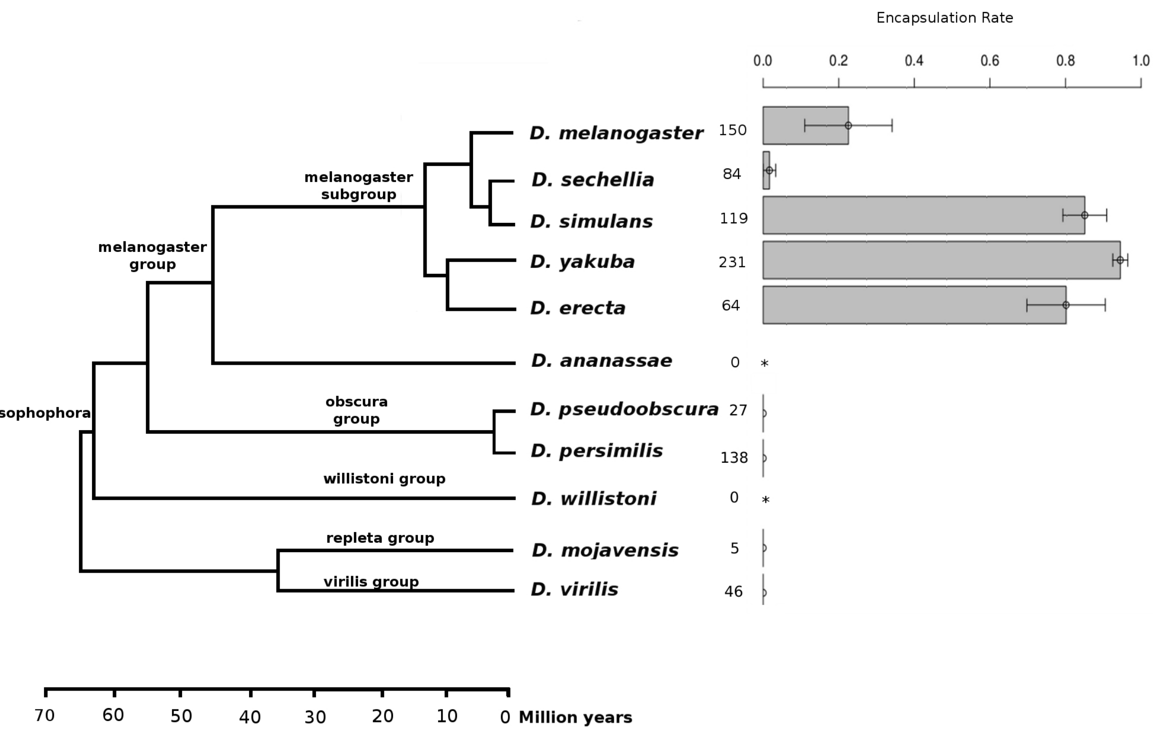


Figure 2.1: **Encapsulation rate of *Drosophila* species against *A. tabida*.** Mean and standard error of the Encapsulation Rate (ER), defined as $ER = c/(c + w)$, where c is the number of adult flies carrying a capsule and w the number of emerged wasps. For each species, we also provide the numbers of parasitized larvae. In some species, *A. tabida* did not develop (asterisks) and other species showed very high mortality rates after parasitization (see Figure S2.7 for more detailed estimations of rates of parasitism, mortality and resistance). The phylogeny is adapted from (Singh *et al.*, 2009).

To further characterize the differences in encapsulation of parasitoids, we investigated two traits that are important for the encapsulation process, the melanization ability and production of lamellocytes. During the dissections of a subset of larvae for each species, we noticed that species unable to encapsulate did not show any signs of melanization around the parasitoid eggs. We verified that all species were able to melanize, independent of the encapsulation process, by pricking the larvae with a fine needle (Table 2.1, Figure S2.1). All species did melanize the site of injury, which indicates that the lack of resistance in some species was not due to a general lack of melanization ability.

Table 2.1: **Phenotypic characterization of cellular immune response.** The cellular immune response of 11 *Drosophila* species against parasitoids, based on experimental assays and published literature. The first two columns list the encapsulation ability against various tested parasitoids species of two distant genera, *Asobara* and *Leptopilina*. The third and four columns refer to evidence of lamellocyte and crystal cells production, respectively. The last columns refers to the ability to initiate a melanization response after injury.

***Drosophila* species**	**Melanotic encapsulation against**		**Lamellocytes**	**Crystal cells**[1]	**Melanization**[1]
	Asobara[1]	***Leptopolina*** [2]			
melanogaster	*tabida*	*boulardi*	yes[1, 3]	yes	yes
simulans	*tabida*	*boulardi*	yes[1, 3]	yes	yes
sechellia	none	none	yes[1, 3]	yes	yes
yakuba	*tabida* *citri*	*boulardi* *hetermotoma*	yes[1, 3]	yes	yes
erecta	*tabida* *citri*	*boulardi* *hetermotoma*	yes[1, 3]	yes	yes
ananassae	none	none	no [1]	yes	yes
persimilis	none	none	no [4]	yes	yes
pseudoobscura	none	none	no [4]	yes	yes
willistoni	none	none	no [*1]	yes	yes
mojavensis	none	none	no [1]	yes	yes
virilis	none	none	no [1]	yes	yes

[*]unusual hemocytes were observed (see Figure S2.2); [1]current study; [2]Schlenke et al 2007; [3]Eslin and Prevost 1998; [4]Havard et al 2009

We confirmed the ability to produce lamellocytes in *D. sechellia*, *D. melanogaster*, *D. simulans*, *D. yakuba* and *D. erecta* from the melanogaster subgroup, and tested *D. ananassae*, *D. willistoni*, *D. mojavensis* and *D. virilis* outside this group. For the obscura group we relied on the detailed characterization in Havard *et al.* (2012). Lamellocytes were produced only by species in the melanogaster subgroup, whereas species outside this group do either not differentiate lamellocytes at all or only a large type of hemocytes with an unusual morphology, that is, not as flat or big as lamellocytes (see Figure S2.2 for details). Lamellocyte production in *Drosophila* therefore appears to be necessary but not sufficient for encapsulation ability, as evidenced by the lack of encapsulation in *D. sechellia*, which produced lamellocytes.

Two additional species (*D. eugracilis* and *D. suzukii*) in the melanogaster group and outside the melanogaster subgroup have been reported to encapsulate wasp eggs and produce lamellocytes (Schlenke *et al.*, 2007; Kacsoh & Schlenke, 2012). More distantly related *Drosophila* species have also been reported to encapsulate parasitoid eggs, yet by means of another type of hemocytes ("pseudopodocytes" in the obscura group) (Havard *et al.*, 2012) or without specifying the involved hemocyte types (Streams, 1968). Two *Drosophila* species in our assay (*D. willistoni* and *D. ananassae*, Figure 2.1) seemed to resist parasitoid development through other mechanisms than encapsulation. We confirmed through dissections that both species were parasitized by *A. tabida*, but no

A. tabida eggs developed in this species, while *A. citri* could develop but never induced melanotic capsules. This suggests either incompatibility of these two species with *A. tabida* or they evolved a different defense mechanism against (some) parasitoids. The combined information across all studied *Drosophila* species indicates that the ability to defend against parasitoids has been gained and lost repeatedly in the *Drosophila* phylogeny, possibly by means of gaining and losing different immunity components, including different types of hemocytes. Due to the uncertainty in the homology of the encapsulation mechanism for more distant species, we focus on the mechanism found in *D. melanogaster* and close relatives. Our current knowledge indicates that a sub-lineage inside the melanogaster group shows: 1) encapsulation of several parasitoid species mediated by the differentiation of lamellocytes, and 2) loss of resistance in *D. sechellia*.

2.3.2 Comparative Genomics

To associate the striking dichotomy that we found across the 11 *Drosophila* species in both lamellocyte differentiation and encapsulation ability, with changes and variation in their genomes, we applied comparative genomic approaches on a list of "candidate genes". We explored the genomic variation of genes in hemopoiesis pathways on 35 protein coding genes with GeneOntology annotation of "hemocyte differentiation", "hemocyte proliferation" or "regulation of hemocyte differentiation" in Flybase (version FB201204) (Zettervall *et al.*, 2004; Williams, 2007; Fauverque & Williams, 2011; McQuilton *et al.*, 2012) or identified as inducers of lamellocyte differentiation through lineage tracing studies (Avet-Rochex *et al.*, 2010; Stofanko *et al.*, 2010). Because the genetic mechanisms that induce and regulate the proliferation and differentiation of hemocytes upon parasitization have not been fully elucidated, we also analysed 109 genes that were previously found to be differentially expressed after parasitoid attack in *D. melanogaster* (Wertheim *et al.*, 2005; Schlenke *et al.*, 2007; Kraaijeveld & Wertheim, 2009). Lamellocyte differentiation is strongly induced by parasitoid attack, and therefore the genome-wide transcriptional response after parasitoid attack can help to identify genes involved in this process. Genes were classified in 5 immunological categories, partially following Sackton *et al.* (2007); Waterhouse *et al.* (2007); Kraaijeveld & Wertheim (2009): i) "hemopoiesis", containing the 35 genes annotated in Flybase with a function in hemocyte differentiation, regulation of differentiation and proliferation ii) "recognition", containing putative pattern recognition receptors; iii) "signaling", containing genes characterized in immune signal transduction pathways (Toll, Jak/Stat, IMD and JNK); iv) "effectors", coding for antimicrobial peptides, phenoloxidases and mediators in the melanin production; v) "proteases", containing serine-type endopeptidases with mostly unknown immune function, but sometimes referred to as modulators. The full list of analysed genes and their classification is included in Table **??**. A subset of 71 of the total 144 genes have also been reported as part of the humoral response against microparasites or of a more general stress response, and were analyzed in a previous comparative genomics study in the same *Drosophila* species (Sackton *et al.*, 2007). The 71 overlapping genes comprise most genes of the hemopoiesis class (24 out of 35) and the recognition class (12 out of 15) and all the 17 genes in the signaling category. In the protease class only 1 gene overlapped (out of 45), and in the effector class 17 (out

of 32) overlapped. This partial overlap signifies both a shared actuation and regulatory control of humoral and cellular immune responses against macro- and microparasites, as well as substantial differences down-stream in the reaction cascades.

2.3.3 Orthologs

Of the 144 candidate genes, 96 genes fell into the single-copy-ortholog (SCO) category, which is representative for the proportion of SCO in the *D. melanogaster* genome ($\approx$ 50%) (*Drosophila* 12 Genomes Consortium, 2007). Paralogs (PAR) and lineage restricted (LR) genes were found for 22 and 26 proteins, respectively (Table S2.3).

The candidate genes in the five immunity categories (recognition, signaling, effectors, proteases and hemopoiesis) were not uniformly distributed over the three homology classes, SCO, PAR and LR (χ^2=45.5, df=8, P-value=3.517e^{-6}) (Figure 2.2). A schematic view of the position of the genes in the hemopoiesis and immune pathways is presented in Figures 2B-C. All genes but one in the hemopoiesis class (*Hemese*, a cellular receptor), and most genes in the signaling class (16 out of 17) were SCO. Effector proteins had the largest proportion of PAR (14 out of 32) and proteases the largest proportion of LR (17 out of 45). The previous comparative study by Sackton *et al.* (2007) already showed that several of these genes are highly conserved. This is likely caused by strong constraints acting on developmental pathways in general (Artieri *et al.*, 2009; Rebeiz *et al.*, 2011), where changes in gene regulation suffice to create interspecies variation. Our data are consistent with the hypothesis that signaling genes are highly conserved in long-term evolutionary scales, as these genes most likely evolve under strong constraints (Sackton *et al.*, 2007; Waterhouse *et al.*, 2007), and effector genes and proteases diversify mainly through gene duplication (Sackton *et al.*, 2007; Waterhouse *et al.*, 2007; *Drosophila* 12 Genomes Consortium, 2007).

2.3.4 Sequence divergence

Genomic variation can be quantified by the coding substitutions that have accumulated in a gene. We applied tests for signatures of positive selection using the models of codon substitution implemented in PAML (Yang, 2007) to a subset of 124 genes (i.e., excluding conserved paralogs and alignments with multiple copies of one gene in one species). The majority of the hemopoiesis genes were highly conserved, except for five genes (*Ser, Dpp, ush, cher* and *sgg*) involved in hemocyte differentiation. Of the 92 candidate genes that are induced upon parasitization (excluding the conserved paralogs), 23 showed signs of positive selection (Table 2). Fourteen of the genes under positive selection are proteases, and three of these proteases *(CG4259, CG18477* and *CG6639*), are expressed primarily in hemocytes (Irving *et al.*, 2005). Using electronic prediction (in pfam: Punta *et al.* (2012)), we found that in 4 of the 14 proteases, the sequence variation led to changes in the functional domain among species (Table S2.5).

Of the seven putative recognition proteins under positive selection, five are also involved in the humoral response, whereas *αPS4* is exclusive to the cellular response (Irving *et al.*, 2005). A second recognition protein exclusive to the cellular response, *Lectin-24A* (Keebaugh & Schlenke, 2012), was found to be significant in the PAML analysis, but not after FDR correction. Recognition proteins that were significant for

Table 2.2: **Candidate genes showing positive selection.** Sequence divergence in genes in the hemopoiesis pathway or in 92 genes that were over-expressed after parasitization. The genes showing positive selection, based on models of codon substitution, were allocated to the five immune categories as described in the main text. The first four proteases (indicated by an asterisk) show among-species differences in functional domains (Figure S2.5).

Recognition	Signaling	Effector	Proteases	Hemocytes
Hml, Corin, αPS4, TepI, TepII, TepIV	*nec*	*yellow-f Cyp309a1*	*CG11313**, *CG30414**, *CG4259*,* *CG9673**, *CG12951, CG17278, CG18477, CG17572, CG31780, CG3916, CG30090, CG6639, CG9676, Jon65Aiii*	*Ser, Dpp, ush, cher, sgg*

positive selection share a common pattern: they were expressed later in the response against wasp attack, suggesting that they act down-stream in the reaction cascade. This contrasts with recognition proteins that show high conservation, both in terms of ortholog numbers and protein-coding substitutions, such as PGRPs and GNBPs (peptidoglycan recognition proteins and gram-negative binding proteins, respectively), which are expressed early during the response and can be thus considered to be up-stream in the cascade (Figure S2.4).

2.3.5 Lineage specific gains and losses

Among all the 26 lineage restricted genes, only 5 have a homolog outside the melanogaster group (Table S2.3), and eleven LR genes appear in the closed interval between the melanogaster group and subgroup, i.e. the interval that contains species able to encapsulate by means of lamellocytes (Figure 2.3). Genes can be restricted to a certain lineage due to duplications in a specific branch, to *de novo* appearance, or because they have diverged from their orthologs beyond recognition (Tautz & Tomislav, 2011). For 4 of the LR (*yellow-f, PPO3, αPS4* and *TepI*), we established that they are recent duplications (Figure 2.4). For the remaining LR, additional outgroups would be necessary to detect the timing of the duplication event. Nonetheless most genes appear to be part of large gene families, suggesting a combination of duplication and rapid accumulation of coding mutations.

Three of the LR, *TepI, PPO3* and *CG11313*, showed large-scale differences in the sequence of *D. sechellia*, the only representative of the melanogaster subgroup unable to encapsulate. These patterns might be associated with the loss of the encapsulation trait, for example, through relaxed stabilizing selection. A detailed examination of *TepI* revealed a major deletion of 4 exons in *D. sechellia*, which are all present in the remaining species (Figure S2.3, note that the predicted gene product did not fully correspond to our sequenced transcript). The *PPO3* gene showed a disproportionately long phylogenetic distance to the other species (Figure 2.4C). Pairwise estimations of substitution rates suggest a neutral substitution rate in *D. sechellia*, while this gene seems to be under stabilizing selection in the other species (Table S2.2). A closer look into the alignment

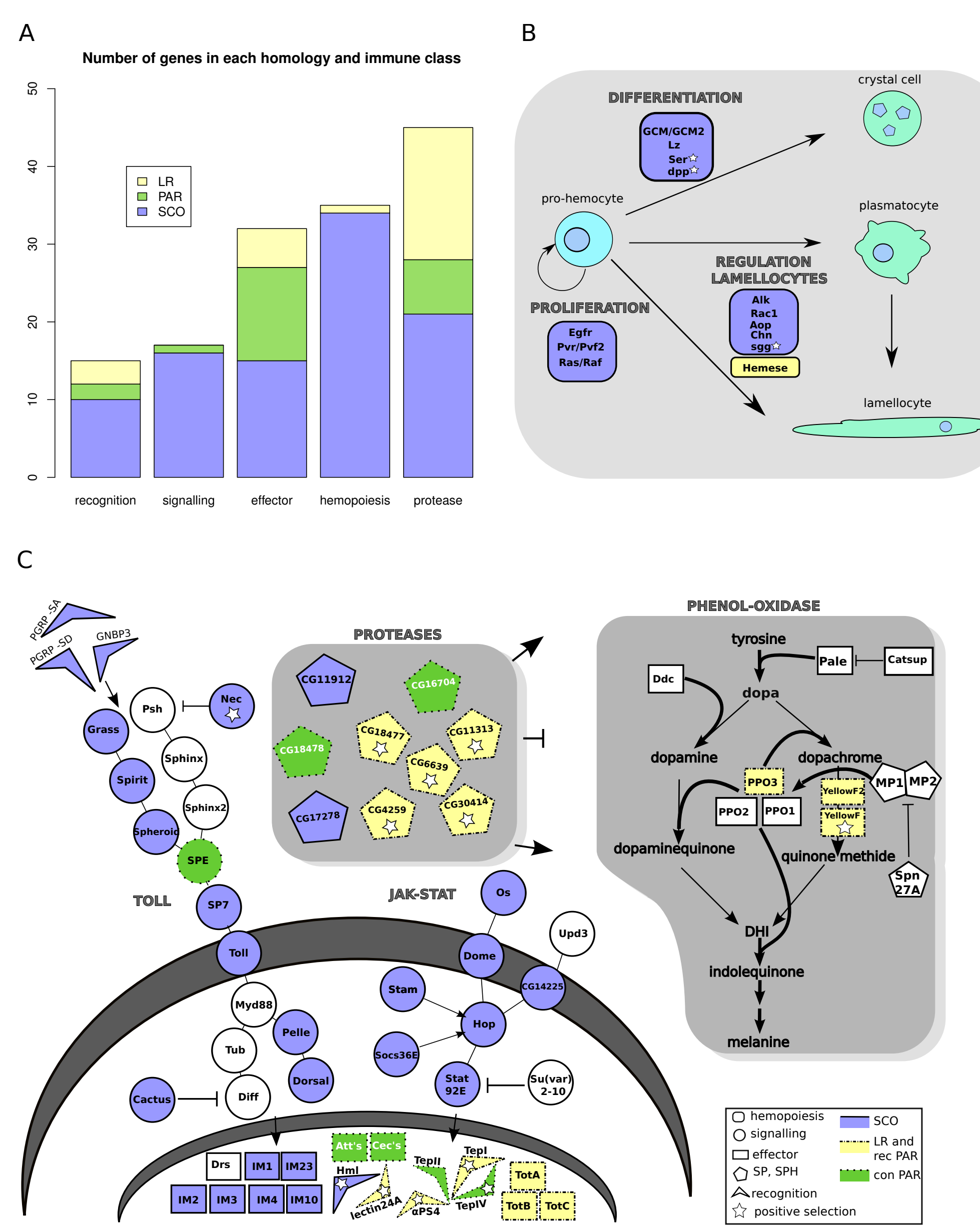


Figure 2.2: **Distribution of proteins in homology and immune categories.** A) LR: lineage-restricted, PAR: paralogs, SCO: single-copy-orthologs; B) Schematic representation of pathways controlling hemopoiesis. Adapted from (Meister, 2004; Zettervall *et al.*, 2004; Williams, 2007); C) Schematic representation of immune pathways expressed under parasitoid attack. Non-filled shapes correspond to proteins known to be in the pathway, but that were not found to be differentially expressed after parasitoid attack in microarray studies. These genes can still be involved in the encapsulation response. Chemical compounds are shown in plain text. Adapted from Schlenke *et al.* (2007) and Tang (2009).

of the *PPO3* protein shows that of the 3 domains predicted through Pfam, the first 2 are lost in *D. sechellia* (Figure S2.6). Of the three pro-phenoloxidase (PPO) coding-

genes in the *D. melanogaster* genome (Tang, 2009), the products of *PPO1* and *PPO2* are primarily expressed in crystal cells, while the expression of *PPO3* is restricted to lamellocytes (Irving et al. 2005). All 11 species produced crystal cells and possessed the genes *PPO1* and *PPO2*, but only the species that produced lamellocytes possessed the gene *PPO3*. Finally, *CG11313* showed a lack of clip domain in *D. sechellia*, which is present in the rest of the species of the melanogaster subgroup (Figure S2.5). Clip is a regulatory domain that controls the proteinase action during activation and regulation of protease cascades (Piao *et al.*, 2005). Although the specific immune function has not yet been described for this gene, its high rate of amino acid substitutions suggests directional selection. Possibly, the loss of the clip domain in *D. sechellia* is accompanied by a new function rather than loss of function.

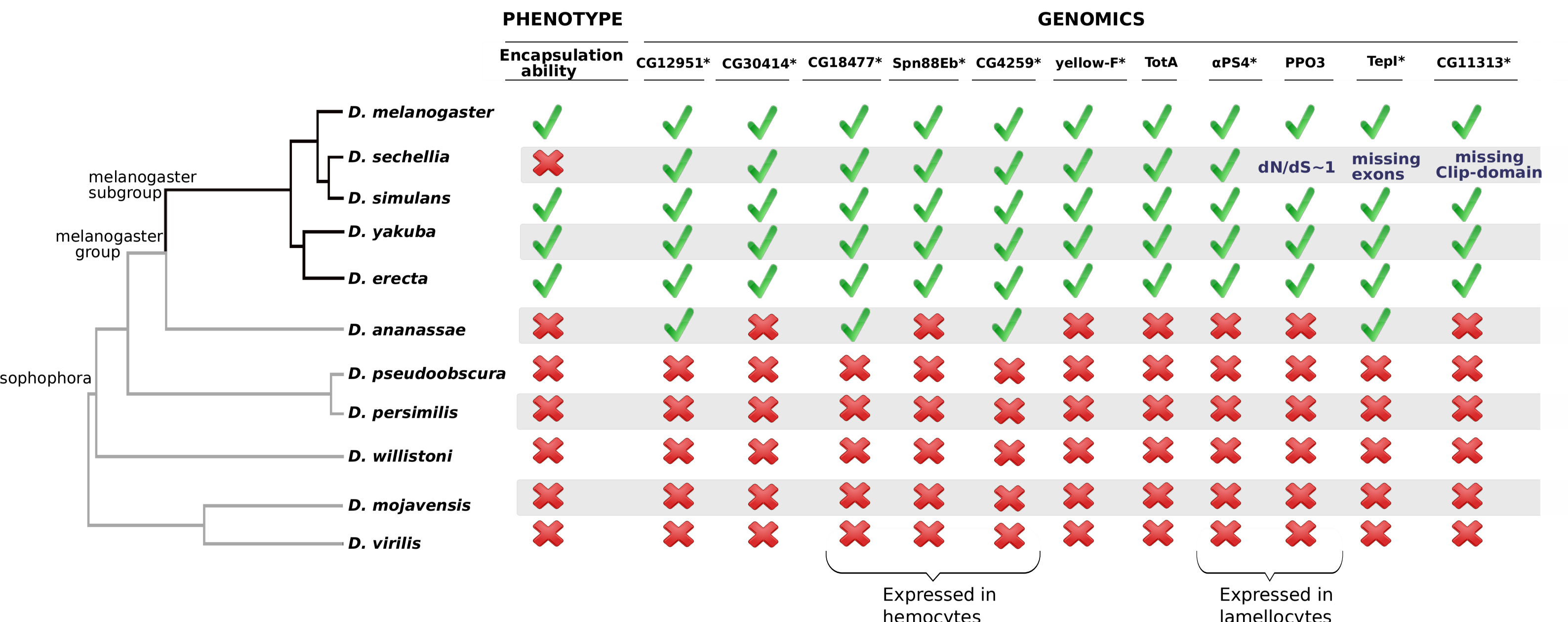


Figure 2.3: **Correlations for phenotypes and Lineage-Restricted genes.** Correlation of phenotypic characterization and the pattern of presence-absence and major genomic changes of 11 newly aqcuired genes between the closed interval of melanogaster group and subgroup. Asterisk indicates genes under positive selection.

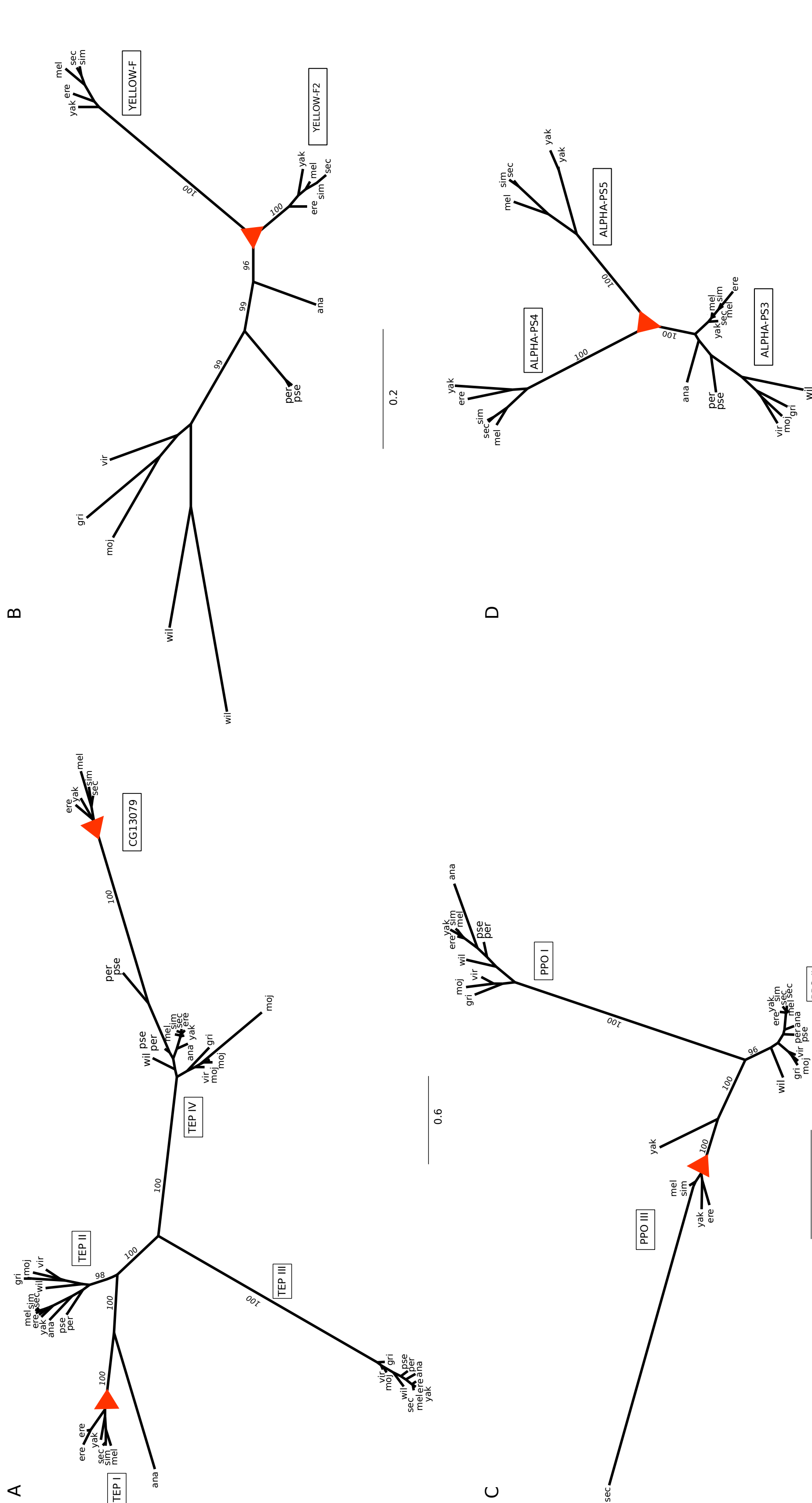


Figure 2.4: **Phylogeny of orthologous groups with recent duplications.** A) *Tep*, B) *yellow* , C) *PPO* and D) αPS. Triangles depict expansions in the melanogaster group and subgroup. Bootstrap values are shown for major subgroups.

2.3.6 Comparative expression of *TepI, PPO3* and *IM1*

In order to i) test for species differences in the (level of) activation of the well-established signal transduction pathways in the immune response against parasitoids (Toll, Jak/Stat and Prophenoloxidase) and ii) gain insight into the relation between the substantial genomic changes in *D. sechellia* and this activation, we performed RT-qPCR assays. We compared the fold changes in expression of *TepI, PPO3* and *IM1* for larvae at two time points (5 and 50 hours) after parasization, among the three sister species, *D. melanogaster, D. simulans* and *D. sechellia*. *TepI* and *PPO3* are the diverged targets of the Jak/Stat pathway and Phenoloxidase cascade, respectively, while *IM1* is a conserved target of the Toll pathway (Figure 2.5). Apart from its role as indicator for the activation of the Toll pathway, *IM1* could also be considered a more general indicator for immune activation, as it is induced in response to a variety of immune challenges (Kraaijeveld & Wertheim, 2009).

Five hours after parasitization, *IM1* was induced in larvae of all three species, indicating that all species activated the Toll pathway and responded to the immune challenge. *TepI* was strongly induced five hours after parasitization in *D. simulans*, and in *D. melanogaster* at very low levels at 5 hours and strongly at 50 hours; indicating that *D. melanogaster* and *D. simulans* species activated the Jak/Stat pathway, but *D. simulans* did so faster. In *D. sechellia*, *TepI* was expressed only at 50 hours, but at similar levels in control and parasitized groups. *PPO3* was not differentially expressed five hours after parasitization in any *Drosophila* species, but was up-regulated at 50 hours after attack in *D. melanogaster* and *D. simulans*. Interestingly, no expression of *PPO3* was found in *D. sechellia*, which is consistent with a loss of function for *PPO3*.

2.4 Discussion

From the species we tested, those outside the melanogaster subgroup were unable to encapsulate eggs of the parasitoids *A. tabida* or *A. citri*, and also did not produce lamellocytes, a specialized type of blood cell important for the encapsulation process. Importantly, the production of lamellocytes and the presence versus absence of encapsulation ability among the 11 surveyed *Drosophila* species is not specific to the *Asobara* parasitoids, but most likely representative for parasitoid wasps in general, as evidenced by very similar patterns among *Drosophila* exposed to the distantly related *Leptopilina* parasitoids (Schlenke *et al.*, 2007). Lamellocytes were previously found to be lacking in some *Drosophila* species that did not mount immune responses against parasitoid wasps (Havard *et al.*, 2009), which was considered a loss of the trait. Conversely, our study combined with data on other species (Schlenke *et al.*, 2007) indicates that lamellocyte-mediated encapsulation is not a common trait, shared among all *Drosophila* species, but appears to be restricted to only a subset of species. Older references reported encapsulation ability outside the melanogaster group, in *D. algonquin* from the obscura group (Nappi, 1970), and in a distantly related species of the subgenus Dorsilopha, *D. busckii* (Streams, 1968), but it appears that the mechanisms are not likely to be the same. In some of the species of the obscura group that lack lamellocytes, including the aforementioned *D. algonquin*, the encapsulation process is mediated by a different type of hemocyte, the pseudopodocytes (Havard *et al.*, 2012). Although hemocytes

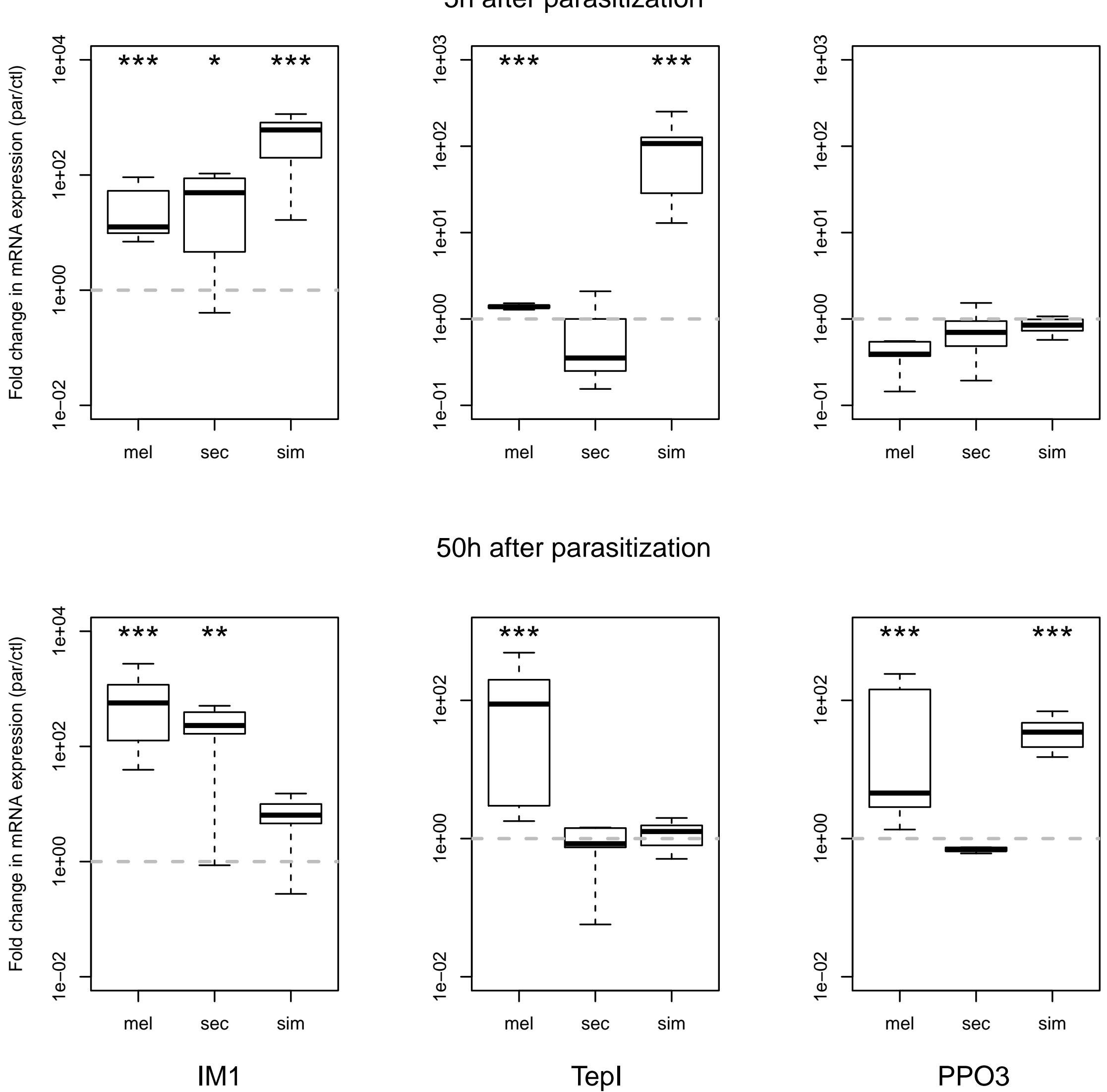


Figure 2.5: **Fold changes in expression of *IM1*, *TepI* and *PPO3* after parasitization.** The ratio between parasitized and control expression levels is calculated and normalized by two reference genes. Boxplots depict the distribution of the replicates and the error estimated through permutation. The dotted grey line describes the value for which the ratio is one (i.e no over-expression). Significance level: $0.05^{*}, 0.005^{**}, 0.001^{***}$

have traditionally been identified through morphology, the use of molecular markers is helping to resolve some of the controversies from the morphological classification of the different hemocyte types. We found that some of the commonly used markers for lamellocytes are genes restricted to the clade able to produce lamellocytes (*αPS4* and *Hemese* (Kurucz *et al.*, 2007; Havard *et al.*, 2012)). This could indicate that blood cells involved in encapsulation in more distantly related species are of a different type as was also found in the obscura group and might explain why no labeling is observed in species of this group when using the available antibodies built against *D. melanogaster* hemocytes (Havard *et al.*, 2012).

Outside *Drosophila*, encapsulation has also been reported in the orders Lepidoptera and Orthoptera (among others) (Strand & Pech, 1995). While less is known about encapsulation in Orthoptera, the encapsulation process in Lepidoptera is one of the functions of granulocytes and plasmatocytes, which do not seem to be the equivalent of lamellocytes in *Drosophila* (Ribeiro & Brehélin, 2006). There is also much variation for mechanisms underlying encapsulation in Dipteran species outside the *Drosophila* genus. In mosquitoes, encapsulation occurs by a sheath of melanin in the absence of a multicellular layer, which is referred to as humoral encapsulation (Vey, 1993). The clear division of function between phagocytic and adhesive cells has not been found in mosquitoes (Castillo *et al.*, 2006). In house flies (Musca), nematodes are also encapsulated by a sheath of melanin, which is then covered by a syncytial mass of host hemocytes, probably of oenocytoid origin (Nappi & Stoffolano, 1971). This variety in the blood cell types among insects reflects the plastic nature of the hemolymphatic tissue, and makes it difficult to establish the homology of the mechanism. To fully understand whether lamellocyte-mediated encapsulation represents an acquired novel trait or whether it has been lost multiple times during evolution requires the investigation of additional species, additional strains for both host and parasitoid species, as well as rigorous phylogenetic comparisons of the type of blood cells, encapsulation process and genes involved.

In this study, we focused on the evolutionary genomics underlying the striking phenotypic variation in *Drosophila*, and investigated the gain/loss and diversification of genes that underlie lamellocyte differentation and melanotic encapsulation. Using a comparative genomics approach we show that the presence of lamellocytes and encapsulation ability is associated with the evolution of various novel genes as well as rapid divergence in (new) protein-coding genes (Figure 2.2). We followed up on genes associated with hemopoieses and genome-wide expression studies after parasitoid attack to identify genes putatively involved in the melanotic encapsulation response. Although we do not claim complete inclusion of all relevant genes for encapsulation ability, and we are likely to miss non-coding regions or genes with small effects (i.e., genes that were not significant in the expression study or whose phenotypic effects are not yet identified or imperceptible), we obtained a more comprehensive list of candidate genes that reflects the process of differentiation and proliferation of blood cells upon parasitization, as well as other aspects of the encapsulation defenses. Our comparative analyses revealed that, except for *Hemese*, all hemopoiesis-associated genes are highly conserved and present in all species independently of their resistance. Only five of the 35 hemopoietic genes showed signs of positive selection, and these five are associated with the process of hemocyte differentation. This relatively low proportion is not too surprising, considering that genes involved in hemocyte proliferation and differentiation are also implicated in a variety of other biological and developmental processes, and their evolution is therefore likely to be highly constrained. In contrast, of the genes differentially expressed after parasitoid attack, 25 were novel genes, of which only five have homologs outside the melanogaster group, and 23 genes were significant for positive selection, mostly proteases and recognition genes.

In an attempt to identify candidate genes underlying the evolution of parasitoid resistance, we specifically focused on the novel genes. Although it would be tempting to hypothesize that the acquisition of the only LR gene in the hemopoiesis pathway,

Hemese, is responsible for the origin of lamellocytes, this may be premature. *Hemese* is expressed in all hemocytes, while inhibition of its expression by RNAi enhances both the proliferation of hemocytes and the production of lamellocytes after parasitoid attacks (Kurucz *et al.*, 2003). It therefore appears that *Hemese* functions as a negative regulator of lamellocyte differentiation, fine-tuning the activation and recruitment of hemocytes, rather than in initiating lamellocytes differentiation (Kurucz *et al.*, 2003). The other genes known for lamellocyte differentiation are common to all 11 species, indicating that existing (hemopoiesis) genes have been co-opted for the acquisition/evolution of a new type of hemocytes. Of the remaining 25 novel genes, 13 were significant for positive selection (Supplementary Table 4), and of these, seven (*TepI, αPS4, Lectin-24A, CG4259, CG18477, Spn88Eb7* and *PPO3*) are mainly or exclusively expressed in hemocytes or lamellocytes (Irving *et al.*, 2005). Four of the novel genes were derived from recent duplications, and most others also appear to be (new) members of large gene families. The combined patterns suggests neo-functionalization of duplicated genes, where they evolved new functions associated with lamellocyte differentiation and melanotic encapsulation. The signature of positive selection in the duplicated genes may reflect the neo-functionalization process itself, where the sequences evolve to optimize their new function, while it could also reflect the strong selection pressures that may occur in host-parasite co-evolution. Although detailed functional studies of these genes are required to confirm their precise role in the cellular immune response (currently under research), we hypothesize that they may be instrumental in the evolution of parasitoid resistance in the *Drosophila* lineage.

Of the novel genes, three show considerable changes exclusive to *D. sechellia*, which secondarily lost resistance. Three genes (*TepI, PPO3* and *CG11313*) show a loss of putative functional domains in *D. sechellia* (Figures S2.3, S2.5 and S2.6). Our expression study indicated that *TepI* was expressed, but not significantly induced after parasitization in *D. sechellia*, thus it is not clear the degree to which *TepI* retained some functionality in this species. *PPO3* seemed to accumulate coding mutations at a neutral rate in *D. sechellia*, whereas the gene is under strongly purifying selection in the other *Drosophila* species. These changes suggest a release of the selection pressure for this gene, and the complete lack of expression of this gene in *D. sechellia* strongly supports that its function is lost. Especially the three genes that show a loss of a functional domain for protein interactions in *D. sechellia* could provide strong candidates for genes involved in the secondary loss of the encapsulation ability in this species, although these molecular signatures could also reflect a relaxation from balancing selection. Fast changes and loss of genes in *D. sechellia* has been shown to occur during its resource specialization on "noni" fruit (McBride, 2007). An interesting question is whether the lack of resistance against parasitoid wasps is also a consequence of the specialization to this resource that is toxic to other *Drosophila* species. Our preliminary results indicate that this fruit indeed is toxic to parasitoids too, which would imply that *D. sechellia* may have lost its immunological resistance to parasitoids, because it is living in an enemy-free niche.

Previous genomic studies argued that divergence in genes involved in antagonistic host-parasite interactions should happen more often in: 1) immune pathways that are targeted and suppressed by parasites (which is apparently the case for IMD and RNAi) and 2) receptors that are in direct contact with the pathogens (Obbard *et al.*, 2009b).

Parasitoid counter-defense strategies include the injection of immuno-suppressive virulence genes coming from DNA viruses (Bitra *et al.*, 2012) and the production of RhoGAP toxins by the parasitoids that induce changes in morphology and adhesion properties of host hemocytes (Colinet *et al.*, 2007). Unfortunately, the immune suppressive effects of parasitoids remains much less understood than the immune response of the host, and even for the latter, the molecular mechanisms for parasitoid recognition are not known. The rapid evolution of certain immune genes within the recognition class in our analyses suggests that the position of genes in the reaction cascade is also important for their evolutionary dynamics. Of the 15 recognition genes in our candidate list, six genes were under positive selection. All these genes are expressed at later stages during the immune response, suggesting that they act down-stream in the reaction cascade, e.g. by directing the cellular response towards the foreign body. In contrast, four recognition genes with high conservation in terms of both number of orthologs and amino-acid sequence (e.g, PGRPs) are upregulated immediately after the immune challenge (Figure 2.2, Figure S2.4), suggesting they act up-stream, triggering the reaction cascade. Unfortunately, for the remaining 5 recognition genes no expression profile was available for early time points. The divergent evolutionary patterns for the up-stream and down-stream recognition genes could be the consequence of different constraints. The effects of genes that act up-stream is amplified along the cascade, and changes in their protein-coding sequence can have profound consequences on the triggered response (Sackton *et al.*, 2010). The high conservation both in ortholog number and coding sequences could thus be the consequence of selection acting to preserve a mechanism that evolved even before the diversification of insects. Other receptor genes that act down-stream in the immune response (*Teps, lectin-24A, αPS4*), would be less constrained by this amplification effect, having thus more potential to change.

Our study on the cellular immune response complements the insights that previous genomic studies on the humoral and RNAi immune responses have established in *Drosophila* (Sackton *et al.*, 2007; Obbard *et al.*, 2009b). Consistent with these studies, we find that most of the protein-coding genes involved in the immune response show high conservation, both in terms of number of orthologs and coding substitutions. Similarly, we find that effector genes diversify mainly through gene duplication. Different to previous studies, we combined a comprehensive list of candidate genes associated with hemopoieses and the response to parasitoid attack. We found that an important number of the up-regulated genes are fast evolving genes or novel genes while most of the hemopoietic genes are highly conserved. Our study also highlights the importance of proteases in the evolution of the cellular immune response. Proteases were not only the largest class of proteins (45), but also the one containing most of the duplicated genes and genes under positive selection (17 and 14, respectively). At present proteases appear to be fundamental mediators in regulatory processes (Jang *et al.*, 2008). Our finding of both high rates of duplication and protein-coding substitution indicates that once a new protease copy arises, it can diversify to generate new outcomes of exisitng pathways. Such rapid change suggests that proteases are “easily” recruited in existing pathways, and in the case of the cellular immune response, this rapid change may play a pivotal role in coordinating differentiation and movement of cells on which the cellular response relies.

An important question that remains to be explored is under what circumstances the

ability to encapsulate evolved in a certain group and why it was lost in some species. The molecular mechanisms for the emergence of novel traits and, more dramatically, the loss of traits that were thought to be essential is currently a hot topic (Rebeiz *et al.*, 2011; Star *et al.*, 2011; Johnson & Tsutsui, 2011). These studies have profited enormously from genomics approaches, because only through this whole genome approach, genes are studied in the genomic context where they evolved.

In conclusion, through a combination of phenotypic and genomic characterizations we provide an important step towards understanding the evolution of the cellular resistance against parasitoids in *Drosophila* species. We highlight specific protein-coding genes that are likely to be important in the acquisition and subsequent loss of this trait, bridging the gap between phenotype and genotype. Understanding the detailed processes underlying the evolution of the encapsulation ability in *Drosophila* may also give insights into the evolution of immune traits in general. *Drosophila* has been long recognized as an excellent model organism for revealing the molecular mechanisms of innate immunity and hemopoiesis also in vertebrates (Williams, 2007). Interestingly, the immune response of vertebrates relies largely on a variety of differentiated blood cells. We showed that a combination of co-option and neofunctionalization is likely to have contributed to the acquiring of the new immunity component in the cellular immune response, and that particular gene families (serine-type proteases, Teps and lectins) could be of special interest for the processes of hemocyte differentiation, proliferation and activation. It would be of great interest to study the role of these gene families in the evolution of the large versatility in blood cells in vertebrates and invertebrates.

2.5 Acknowledgements

We would like to thank Wen-Juan Ma for providing the wasps; Anna Rensink, Eveline Verhulst and Ammerins de Haan for assistance with the qPCR; and Daniel van der Post, Kirsten Jalvingh and Sylvia Gerritsma as well as four anonymous reviewers for commenting on an earlier version of the manuscript. This work was supported by a VIDI grant to BW from the Netherlands Organisation for Scientific Research (NWO) [grant 864.08.008].

Table 2.3: **Lineage restricted genes associated with the cellular immune response** Lineage restricted (LR) genes in the hemopoiesis pathway ("hemo") or among the genes that are over-expressed after parasitization in microarray studies (Wertheim *et al.*, 2005; Schlenke *et al.*, 2007). The genes are allocated to 5 immune categories, as indicated in the main text. The presence or absence of orthologs is indicated for the 22 genes (out of the 26 LR genes in total) that had a copy in at least 3 species. Only five genes have a copy outside the melanogaster group. Asterisks denote genes that are also under positive selection.

Drosophila	Effector					Protease													Recognition			Hemo
	TotA	TotB	TotC	PPO3	yellow-f⋆	Spn88Eb ⋆	CG11313 ⋆	CG6639 ⋆	CG30414 ⋆	CG4259 ⋆	CG12951 ⋆	CG18477 ⋆	Jon65Aii⋆	Jon25Bii	CG30090 ⋆	CG18563	CG4793	CG3117	lectin-24A	αPS4⋆	TepI ⋆	Hemese
melanogaster	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
simulans	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
sechellia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓
yakuba	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓
erecta	✓	✗	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✗	✓	✓	✗
ananassae	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✓	✗
persimilis	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗
pseudoobscura	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗
willistoni	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗
mojavensis	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
virilis	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗

2.6 Supplementary Material Chapter 2

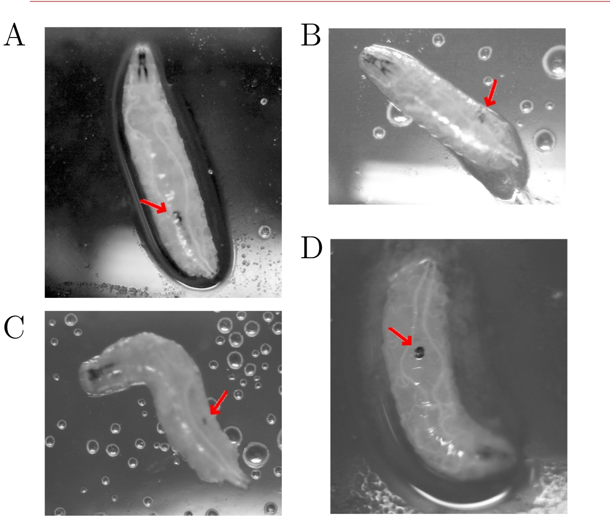


Figure S2.1: **Examples of melanization in larvae.** A: *D. annanassae* after 2h, B: *D. pseudoobscura* after 4h, C: *D. simulans* after 2h , D: *D. virilis* after 2h

Table S2.1: **Insect strains information**

Species	**abbreviation**	**Reference**	**Location**	**Date of collection**
D. ananassae	ana	14024 – 0371.13	Hawaii (USA)	1945
D. erecta	ere	14021 – 0224.01	Unknown	Unknown
D. melanogaster	mel	14021 – 0231.36	Unknown	Unknown
D. mojavensis	moj	15081 – 1352.22	California (USA)	2002
D. persimilis	per	14011 – 0111.49	California (USA)	1997
D. pseudoobscura	pse	1401 – 0121.94	Colorado (USA)	1996
D. sechellia	sec	14021 – 0248.25	Cousin Island (Seychelles)	1980
D. virilis	vir	15010 – 1051.87	Unknown	Unknown
D. willistoni	will	14030 – 0811.24	Guadaloupe Island (France)	Unknown
D. yakuba	yak	14021 – 0261.01	Liberia	1983
A. tabida			Sospel (France)	1994
A. citri			Ivory Coast	1995

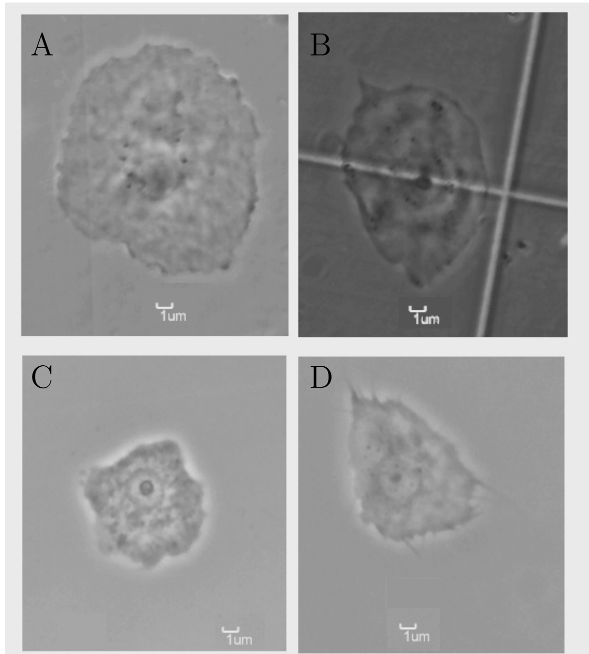


Figure S2.2: **Example of lamellocytes in parasitized third-instar larvae.** A: *D. melanogaster*, B: *D. simulans*, C,D: *D. willistoni.* The morphology of the cells in *D. willistoni* share some features with lamellocytes, such as a flat shape and the presence of pseudopodia, however their size is not as big as a typical lamellocyte and a "halo" of light suggests that they are not completely flat

Table S2.2: **Pair-wise estimation of dN and dS for *PPO3***

sp1	sp2	S	N	t	kappa	omega	dN +- SE	dS +- SE
sim	mel	298.5	1681.5	0.1380	1.6559	0.0489	0.0117 +- 0.0027	0.2393 +- 0.0351
ere	mel	320.2	1659.8	0.4208	1.6559	0.0928	0.0543 +- 0.0059	0.5857 +- 0.0767
ere	sim	310.6	1669.4	0.4058	1.6559	0.0979	0.0553 +- 0.0059	0.5650 +- 0.0763
yak	mel	309.0	1671.0	0.4685	1.6559	0.0923	0.0616 +- 0.0063	0.6675 +- 0.0931
yak	sim	299.1	1680.9	0.4380	1.6559	0.0952	0.0599 +- 0.0062	0.6295 +- 0.0864
yak	ere	320.3	1659.7	0.3629	1.6559	0.1827	0.0702 +- 0.0067	0.3841 +- 0.0499
sec	mel	331.2	1648.8	1.2615	1.6559	0.4032	0.3371 +- 0.0172	0.8360 +- 0.0941
sec	sim	321.7	1658.3	1.1933	1.6559	0.4505	0.3320 +- 0.0170	0.7369 +- 0.0802
sec	ere	342.4	1637.6	1.6761	1.6559	0.2710	0.3813 +- 0.0189	1.4070 +- 0.2733
sec	yak	331.6	1648.4	1.7598	1.6559	0.2202	0.3682 +- 0.0183	1.6724 +- 0.4014

Rates estimated between two species (sp1 and sp2) by the method of Yang and Nielsen (2000) Yang & Nielsen (2000) implemented in the program yn00 of PAML: number of synonymous (S) and non-synonymous (N) sites in a sequence, time (t), numbers of synonymous (dS) and non-synonymous (dN) substitutions per site, omega (dN/dS) and *kappa* (transition/transversion ratio).

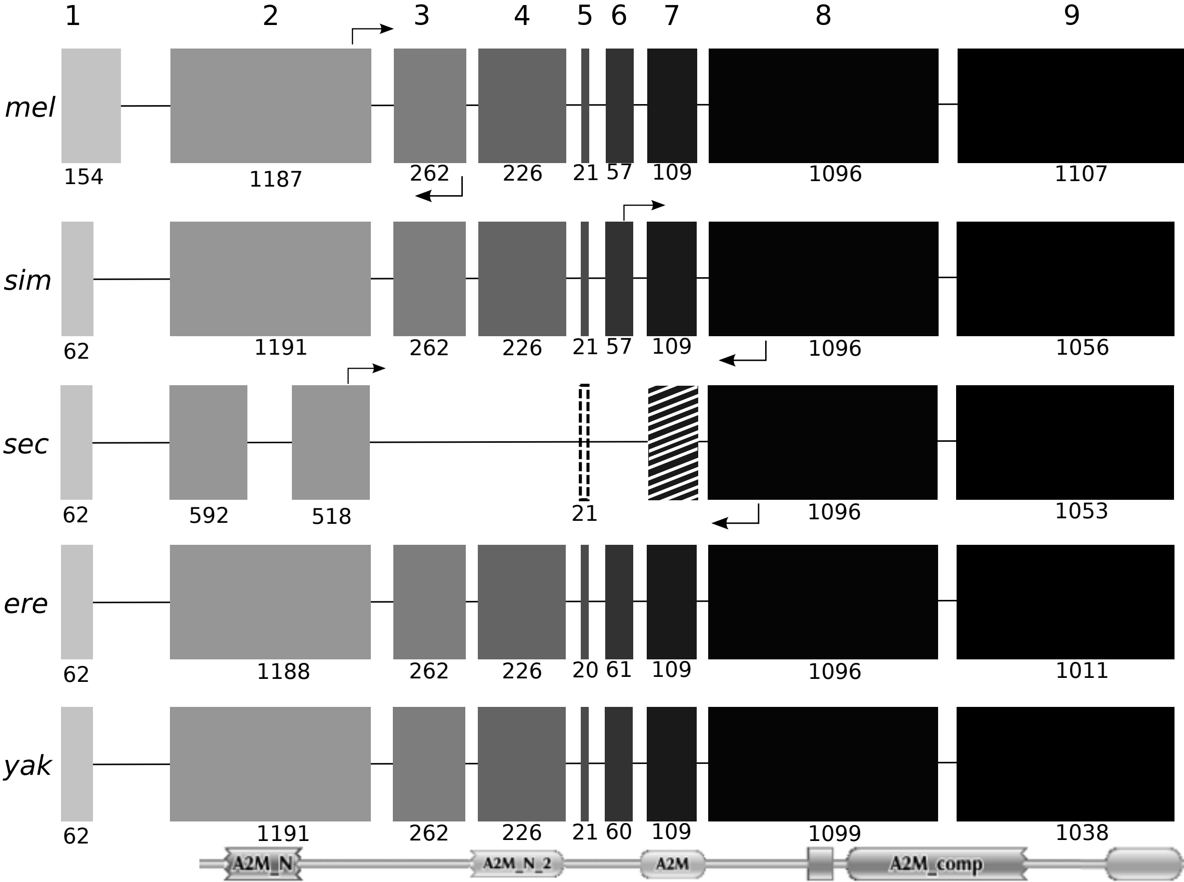


Figure S2.3: **Scheme of TepI exon-intron alignment** Exons are represented by squares and their respective approximate functional domains from Pfam is shown at the bottom (A2M:α2Macroglobulin family, N: N terminal region, comp: complement component, recep: receptor, square: thiol-ester bond-forming region). Arrows depict the position of the primers for the species, for which expression was measured (*D. melanogaster, D. simulans, D.sechellia*). We sequenced the qPCR product of *D. sechellia* because the amplicon size was considerably larger than predicted, and found two differences with respect to the prediction: predicted exon 5 is not found in the mRNA sequence (non-filled square), and a predicted intron in position 7 was found in the sequenced mRNA (striped square)

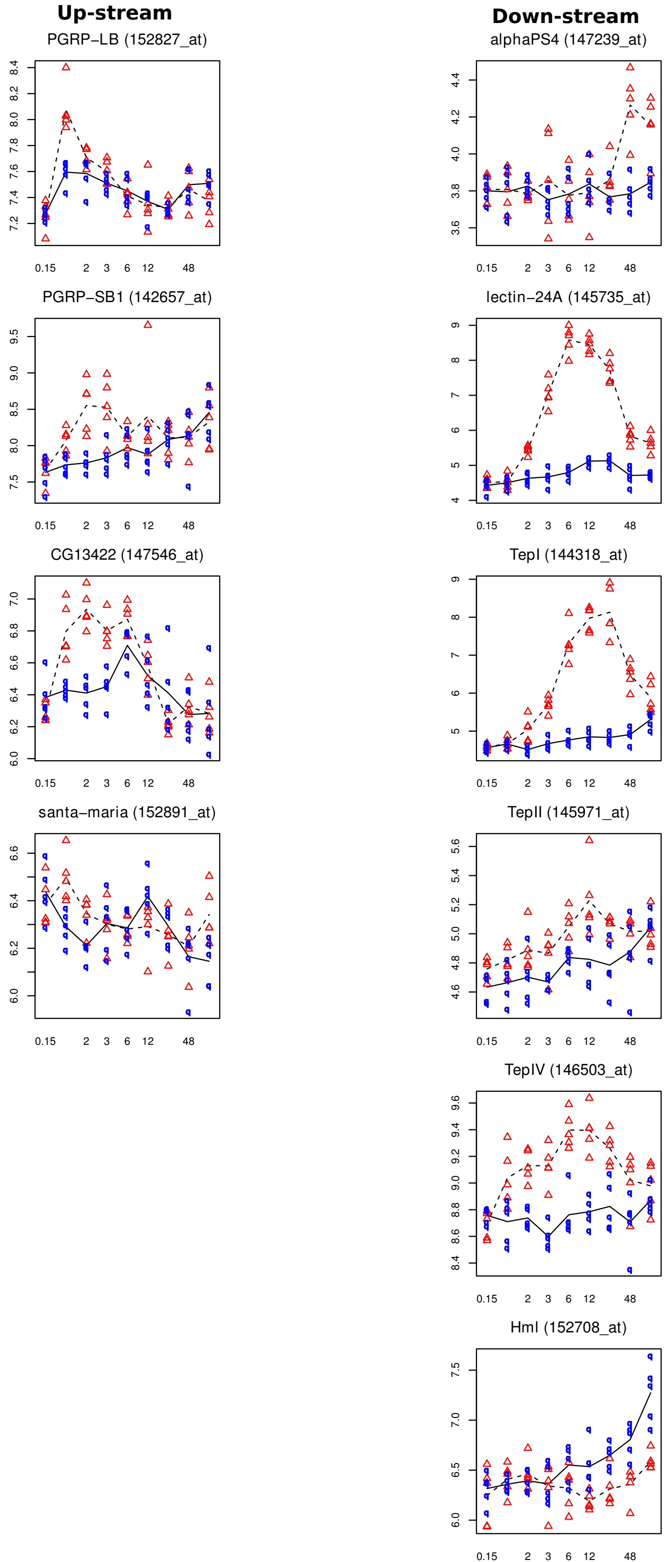


Figure S2.4: **Two types of receptors** Receptors that are up-stream the signalling pathways express earlier (i.e expression peak is in the first 3 hours after parastization), whereas activated receptors down-stream the signalling pathways express later (i.e expression peak after 5 hours). Although our list of candidate genes included data from both studies, Wertheim *et al.* (2005) and Schlenke *et al.* (2007), the expression pattern is based only on Wertheim *et al.* (2005), because in Schlenke *et al.* (2007) there are no measurements earlier than 5 hours. Reproduced with permission from Wertheim *et al.* (2005).

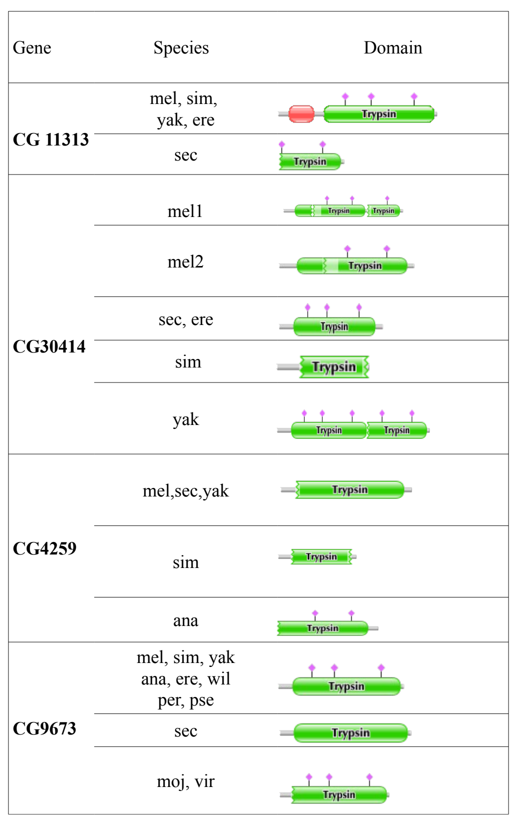


Figure S2.5: **Serine proteases under positive selection with differences in functional domains** Green: trypsin, red: clip domain. Each pink stick correspond to one residue of the catalytic triad: histidine, cystein and serine

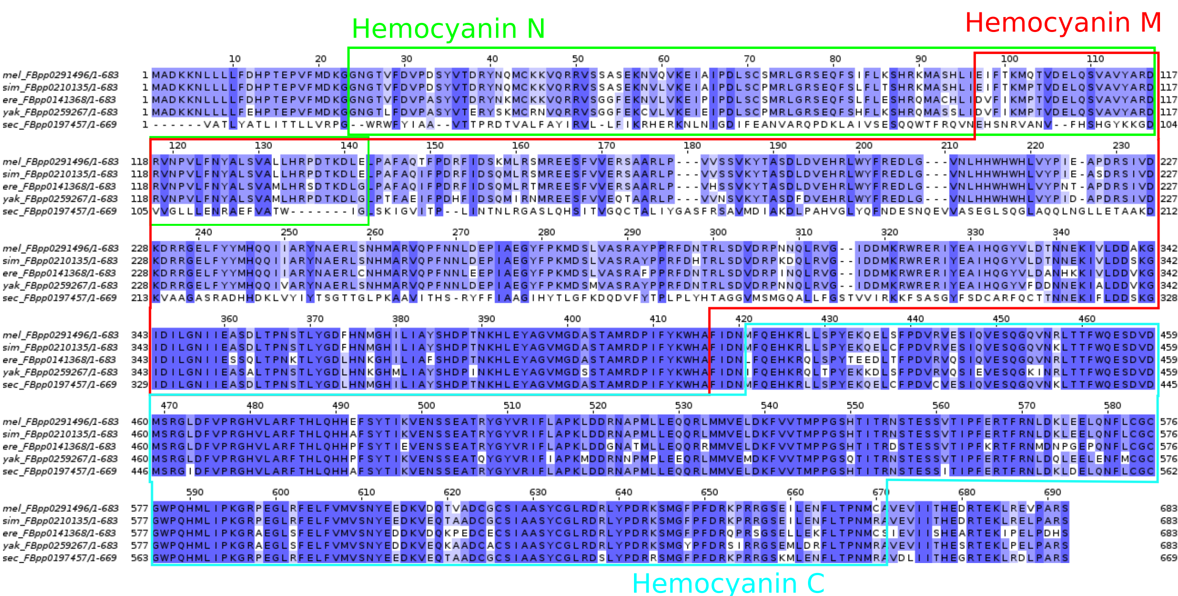


Figure S2.6: **Multiple alignment of PPO3** The functional domains predicted by PFAM are highlighted

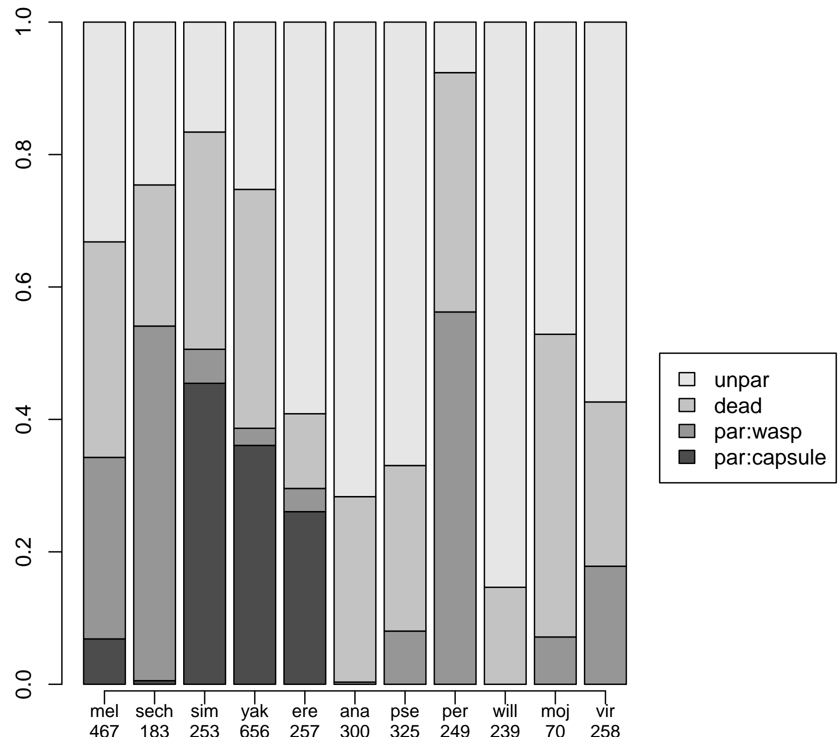


Figure S2.7: **Summary of the parasitization experiment** The total of individuals tested are pooled for all replicates and stages (dissected larvae and adults). On the x-axis are the species with the total number of individuals tested. On the y-axi the outcome for each individual as proportions of: parasitized individuals that formed a capsule (par:capsule), parasitized individuals that led to wasp emergence (par:wasp), individuals that died (dead) estimated from the difference between number of pupae and emerging adults (flies or wasps), and unparasitized individuals (unpar) where no signs of capsules were found.

Name	CG	Fbgn	Fbpp	Immune	mel	sim	sec	yak	ere	ana	per	pse	wil	moj	vir	gri	Homology	Sackton	par_to	M7vsM8	FDR
AttA	CG10146	FBgn0012042	FBpp0086567	effector	2	2	1	2	2	1	3	2	1	1	2	1	par	par	AttB		
AttB	CG18372	FBgn0041581	FBpp0086568	effector	2	2	1	1	2	1	3	3	1	1	2	1	par	par	AttA		
AttD	CG7629	FBgn0038530	FBpp0082908	effector	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
proPO59	CG42640	FBgn0261363	FBpp0291496	effector	1	1	1	2	1	0	0	0	0	0	0	0	rec_par	par	PPO1_2	1.00	0.81
cact	CG5848	FBgn0000250	FBpp0080402	signalling	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.49	0.81
CecC	CG1373	FBgn0000279	FBpp0084980	effector	2	2	2	1	2	1	0	4	2	1	4	4	par	par	CecA2		
CG11313	CG11313	FBgn0039798	FBpp0289924	protease	1	1	1	1	3	0	0	0	0	0	0	0	lr			0.00	0.00
CG11912	CG11912	FBgn0031248	FBpp0077758	protease	1	1	1	1	1	1	1	1	1	2	1	1	sco			0.66	0.81
CG12951	CG12951	FBgn0037677	FBpp0081536	protease	1	1	1	1	1	1	0	0	0	0	0	0	lr			0.02	0.07
et	CG14225	FBgn0031055	FBpp0074526	signalling	1	1	1	1	1	1	1	1	0	1	1	1	sco	lr		0.25	0.51
CG15065	CG15065	FBgn0040734	FBpp0085848	effector	1	1	1	2	1	1	2	1	2	1	1	3	par			1.00	0.81
CG16704	CG16704	FBgn0031558	FBpp0077239	protease	1	1	1	1	1	2	2	2	1	3	1	4	par			1.00	0.81
CG16712	CG16712	FBgn0031561	FBpp0077235	protease	1	1	1	1	1	1	1	1	0	1	1	1	sco			0.97	0.81
CG16713	CG16713	FBgn0031560	FBpp0077236	protease	1	1	1	1	1	1	1	1	1	2	1	4	par			1.00	0.81
CG17278	CG17278	FBgn0046763	FBpp0083449	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.01	0.04
CG17475	CG17475	FBgn0038481	FBpp0082858	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
CG17477	CG17477	FBgn0038479	FBpp0082859	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.99	0.81
CG17572	CG17572	FBgn0032753	FBpp0080715	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.00	0.00
CG18477	CG18477	FBgn0028864	FBpp0080315	protease	2	1	1	1	1	2	0	0	0	0	0	0	lr		CG31780	0.00	0.00
CG18478	CG18478	FBgn0028517	FBpp0080316	protease	2	0	0	1	0	0	0	0	0	0	0	0	lr		CG31827		
CG18563	CG18563	FBgn0032639	FBpp0290124	protease	1	1	0	1	0	0	0	0	0	0	0	0	lr			0.10	0.28
CG2736	CG2736	FBgn0035090	FBpp0072310	recognition	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.20	0.45
CG30090	CG30090	FBgn0050090	FBpp0086417	protease	1	1	1	1	1	0	1	1	0	0	0	0	lr			0.00	0.00
CG30371	CG30371	FBgn0050371	FBpp0087874	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
CG30414	CG30414	FBgn0050414	FBpp0071941	protease	1	1	1	1	1	0	0	0	0	0	0	0	lr			0.00	0.00
CG3117	CG3117	FBgn0031471	FBpp0289585	protease	1	1	1	1	0	0	0	0	0	0	0	0	lr			0.65	0.81
CG31266	CG31266	FBgn0051266	FBpp0289337	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
CG31269	CG31269	FBgn0051269	FBpp0112313	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
CG31780	CG31780	FBgn0051780	FBpp0080317	protease	2	1	1	1	1	2	0	0	0	0	0	0	lr		CG18477	0.00	0.00
CG31827	CG31827	FBgn0051827	FBpp0080318	protease	2	0	0	1	0	0	0	0	0	0	0	0	lr		CG18478		
CG32374	CG32374	FBgn0052374	FBpp0076539	protease	2	1	1	2	1	2	1	1	1	0	1	1	par		CG32376		
CG32376	CG32376	FBgn0052376	FBpp0076538	protease	2	1	1	2	1	2	1	1	1	0	1	1	par		CG32374		
CG32483	CG32483	FBgn0052483	FBpp0072504	protease	2	2	1	2	2	1	2	2	2	3	2	2	par		CG3344		
CG33127	CG33127	FBgn0053127	FBpp0077712	protease	1	0	0	0	0	0	0	1	0	0	0	0	lr			0.12	0.31
CG3344	CG3344	FBgn0035154	FBpp0072505	protease	2	2	1	2	2	1	2	2	2	3	2	2	par		CG32483		
CG3505	CG3505	FBgn0038250	FBpp0082425	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.49	0.81
CG3916	CG3916	FBgn0038003	FBpp0082032	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.01	0.04
CG4053	CG4053	FBgn0038482	FBpp0292474	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
CG4259	CG4259	FBgn0031389	FBpp0077526	protease	1	1	1	1	1	1	0	0	0	0	0	0	lr			0.00	0.00
CG4653	CG4653	FBgn0030776	FBpp0074095	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.97	0.81
CG4793	CG4793	FBgn0028514	FBpp0289839	protease	1	1	1	1	1	0	1	1	0	0	0	0	lr			0.32	0.62
CG5246	CG5246	FBgn0038484	FBpp0082846	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
CG6041	CG6041	FBgn0029826	FBpp0070826	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
CG6639	CG6639	FBgn0032638	FBpp0080557	protease	1	1	1	1	0	0	1	2	0	0	0	0	lr			0.01	0.04
Spn88Eb	CG6687	FBgn0038299	FBpp0082594	protease	1	1	1	1	1	0	0	0	0	0	0	0	lr			0.00	0.00
CG9240	CG9240	FBgn0030669	FBpp0073876	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.07	0.21
CG9673	CG9673	FBgn0030775	FBpp0074136	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.02	0.07
CG9733	CG9733	FBgn0039759	FBpp0085000	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
Corin	CG2105	FBgn0033192	FBpp0087983	recognition	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.00	0.00
Cyp12a4	CG6042	FBgn0038681	FBpp0083145	effector	3	3	3	3	3	4	5	6	5	3	3	4	par		Cyp12a5_CG13604		
Cyp12a5	CG11821	FBgn0038680	FBpp0083144	effector	3	3	3	3	3	4	5	6	5	3	3	4	par		Cyp12a4_CG13604		
Cyp12e1	CG14680	FBgn0037817	FBpp0290498	effector	1	1	1	1	1	1	1	1	2	1	1	1	sco			0.66	0.81
Cyp309a1	CG9964	FBgn0031432	FBpp0077429	effector	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.00	0.00
Cyp4e3	CG4105	FBgn0015035	FBpp0079515	effector	3	3	3	3	2	3	4	4	2	1	2	1	par		Cyp4e2_Cyp4e1		
Cyp6a17	CG10241	FBgn0015714	FBpp0086581	effector	2	2	2	2	1	3	2	2	3	2	1	1	par		Cyp6a23		
Cyp9c1	CG3616	FBgn0015040	FBpp0072228	effector	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
Cyp9f2	CG11466	FBgn0038037	FBpp0082070	effector	1	1	1	1	1	1	1	1	1	1	2	1	sco			0.83	0.81
Dhpr	CG4665	FBgn0035964	FBpp0076258	effector	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
dome	CG14226	FBgn0043903	FBpp0074525	signalling	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
dl	CG6667	FBgn0260632	FBpp0080558	signalling	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.21	0.46
emb	CG13387	FBgn0020497	FBpp0079278	signalling	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
Fmo-2	CG3174	FBgn0033079	FBpp0085481	effector	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
grass	CG5896	FBgn0039494	FBpp0084481	signalling	1	1	1	1	1	1	1	1	1	1	1	2	sco	sco		0.69	0.81
GNBP3	CG5008	FBgn0040321	FBpp0076237	recognition	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
CG13422	CG13422	FBgn0034511	FBpp0085532	recognition	1	1	1	1	1	1	1	1	1	1	1	1	sco	par		1.00	0.81
Hml	CG7002	FBgn0029167	FBpp0075495	recognition	1	1	1	1	1	1	2	2	4	1	1	1	sco	sco		0.00	0.00
hop	CG1594	FBgn0004864	FBpp0073313	signalling	1	1	1	1	1	1	1	1	2	1	1	1	sco	par		0.97	0.81
IM1	CG18108	FBgn0034329	FBpp0085843	effector	2	2	2	0	2	1	1	1	2	1	1	0	par	par	IM2		
IM2	CG18106	FBgn0025583	FBpp0085845	effector	2	2	2	0	2	1	1	1	2	1	1	0	par	par	IM1	1.00	0.81
IM3	CG16844	FBgn0040736	FBpp0085846	effector	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
IM4	CG15231	FBgn0040653	FBpp0071487	effector	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
IM10	CG18279	FBgn0033835	FBpp0086780	effector	2	1	1	1	1	1	1	1	1	1	2	1	sco	par		1.00	0.81
IM23	CG15066	FBgn0034328	FBpp0085908	effector	1	0	1	1	1	1	1	1	1	1	1	1	sco	sco		0.33	0.62
Jon25Bii	CG8869	FBgn0031654	FBpp0078691	protease	1	1	0	1	1	1	1	1	1	0	0	0	lr			0.23	0.48
Jon65Aii	CG6580	FBgn0035666	FBpp0076750	protease	1	1	1	1	1	1	0	1	1	0	0	0	lr			0.78	0.81
Jon65Aiii	CG6483	FBgn0035665	FBpp0076749	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.00	0.00
lectin-24A	CG3410	FBgn0040104	FBpp0077228	recognition	1	1	1	0	0	0	0	0	0	0	0	0	lr			0.04	0.14
Mtk	CG8175	FBgn0014865	FBpp0086518	effector	1	1	1	1	1	1	1	1	1	0	1	1	sco	sco		1.00	0.81
mthl2	CG17795	FBgn0035623	FBpp0076796	effector	4	4	7	5	6	1	1	1	4	1	1	2	par		Mthl1_3_4		

Name	CG	Fbgn	Fbpp	Immune	mel	sim	sec	yak	ere	ana	per	pse	wil	moj	vir	gri	Homology	Sackton	par_to	M7vsM8	FDR
nec	CG1857	FBgn0002930	FBpp0088123	signalling	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.01	0.04
pll	CG5974	FBgn0010441	FBpp0084549	signalling	1	2	1	1	1	1	1	1	1	1	1	1	sco	sco		0.97	0.81
Pxn	CG12002	FBgn0011828	FBpp0072827	effector	1	1	2	1	1	1	1	1	1	1	2	1	sco			1.00	0.81
PGRP-LB	CG14704	FBgn0037906	FBpp0081872	recognition	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
PGRP-SA	CG11709	FBgn0030310	FBpp0073358	recognition	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.84	0.81
PGRP-SB1	CG9681	FBgn0043578	FBpp0075107	recognition	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
PGRP-SD	CG7496	FBgn0035806	FBpp0076519	recognition	1	1	2	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
Rel	CG11992	FBgn0014018	FBpp0088375	signalling	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.19	0.45
santa-maria	CG12789	FBgn0025697	FBpp0079088	recognition	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
Sp7	CG3066	FBgn0037515	FBpp0081237	signalling	1	1	2	1	1	1	1	1	1	1	1	1	sco	par		0.08	0.23
SPE	CG16705	FBgn0039102	FBpp0083832	signalling	1	2	2	2	2	1	1	1	1	1	1	1	par	sco			
spheroide	CG9675	FBgn0030774	FBpp0074137	signalling	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.15	0.37
spirit	CG2056	FBgn0030051	FBpp0071174	signalling	1	1	1	1	1	2	1	1	1	2	1	1	sco	par		1.00	0.81
Stam	CG6521	FBgn0027363	FBpp0079732	signalling	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.64	0.81
Stat92E	CG4257	FBgn0016917	FBpp0088487	signalling	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
Tep1	CG18096	FBgn0041183	FBpp0080369	recognition	1	1	1	1	1	1	0	0	0	0	0	0	rec_par	lr	TepII_III_IV_V	0.00	0.00
Tep2	CG7052	FBgn0041182	FBpp0079133	recognition	5	4	5	5	6	4	4	4	3	6	3	5	par	par	TepI_III_IV_V	0.00	0.00
Tep4	CG10363	FBgn0041180	FBpp0080795	recognition	5	4	5	5	6	4	4	4	3	6	3	5	par	par	TepI_II_III_V	0.00	0.00
Tequila	CG4821	FBgn0023479	FBpp0076252	protease	1	2	1	1	1	1	2	2	1	2	2	1	par			0.37	0.69
Tl	CG5490	FBgn0003717	FBpp0084431	signalling	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
TotA	CG31509	FBgn0028396	FBpp0083378	effector	1	2	2	1	2	0	0	0	0	0	0	0	lr	lr		0.96	0.81
TotB	CG5609	FBgn0038838	FBpp0083380	effector	1	1	1	2	0	0	0	0	0	0	0	0	lr	lr		0.07	0.21
TotC	CG31508	FBgn0044812	FBpp0083379	effector	1	1	1	0	0	0	0	0	0	0	0	0	lr	lr		1.00	0.81
epsilonTry	CG18681	FBgn0010425	FBpp0087256	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.22	0.47
lambdaTry	CG12350	FBgn0043470	FBpp0087222	protease	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
yellow-c	CG4182	FBgn0041713	FBpp0080260	effector	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.20	0.45
yellow-f	CG18550	FBgn0041710	FBpp0082191	effector	1	1	1	1	1	0	0	0	0	0	0	0	rec_par	par	Yellow-f2	0.00	0.00
yellow-f2	CG8063	FBgn0038105	FBpp0082190	effector	2	2	2	2	2	1	1	1	2	1	1	1	par	par	Yellow-f		
yellow-g	CG5717	FBgn0041709	FBpp0072780	effector	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
alphaPS4	CG16827	FBgn0034005	FBpp0086503	recognition	1	1	1	1	1	0	0	0	0	0	0	0	rec_par		AlphaPS3_5	0.00	0.00
Adgf-A	CG5992	FBgn0036752	FBpp0089360	hemocyte	1	1	1	2	1	1	1	1	1	1	1	1	sco			0.44	0.79
dpp	CG9885	FBgn0000490	FBpp0077451	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.00	0.00
brm	CG5942	FBgn0000212	FBpp0075278	hemocyte	1	1	1	1	1	1	1	1	1	1	1	2	sco	sco			
phl	CG2845	FBgn0003079	FBpp0070385	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.32	0.62
srp	CG3992	FBgn0003507	FBpp0082669	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
zfh1	CG1322	FBgn0004606	FBpp0085063	hemocyte	1	1	2	1	1	1	1	1	1	1	1	1	sco				
Su(H)	CG3497	FBgn0004837	FBpp0080261	hemocyte	1	1	1	1	1	1	1	2	1	1	1	1	sco	sco		0.07	0.21
Ser	CG6127	FBgn0004197	FBpp0084498	hemocyte	1	1	1	1	1	1	1	2	1	1	1	1	sco	sco		0.00	0.00
lwr	CG3018	FBgn0010602	FBpp0077741	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
gcm2	CG3858	FBgn0019809	FBpp0079435	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
wg	CG4889	FBgn0004009	FBpp0079060	hemocyte	1	0	1	1	1	1	1	1	1	1	1	1	sco			0.75	0.81
Nc	CG8091	FBgn0026404	FBpp0076076	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.65	0.81
N	CG3936	FBgn0004647	FBpp0070483	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.65	0.81
gcm	CG12245	FBgn0014179	FBpp0079451	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.92	0.81
Pvf2	CG13780	FBgn0031888	FBpp0079051	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.12	0.31
Dsor1	CG15793	FBgn0010269	FBpp0071248	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.12	0.31
lz	CG1689	FBgn0002576	FBpp0071255	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
ytr	CG18426	FBgn0021895	FBpp0072143	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
Rac1	CG2248	FBgn0010333	FBpp0072614	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
l(3)mbn	CG12755	FBgn0002440	FBpp0076770	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.15	0.37
mxc	CG12124	FBgn0260789	FBpp0071267	hemocyte	1	0	1	1	1	1	1	1	1	0	1	1	sco			0.32	0.62
Pvr	CG8222	FBgn0032006	FBpp0079244	hemocyte	1	1	1	1	1	1	1	1	1	1	1	1	sco			0.07	0.21
kn	CG10197	FBgn0001319	FBpp0086595	hemocyte_reg	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.60	0.81
Ras85D	CG9375	FBgn0003205	FBpp0081600	hemocyte_reg	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.05	0.17
Egfr	CG10079	FBgn0003731	FBpp0071570	hemocyte_reg	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.81	0.81
ush	CG2762	FBgn0003963	FBpp0077723	hemocyte_reg	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.00	0.00
RpS6	CG10944	FBgn0261592	FBpp0071087	hemocyte_reg	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		0.21	0.46
cher	CG3937	FBgn0014141	FBpp0088480	hemocyte_reg	1	1	1	1	1	1	1	1	1	2	2	1	sco	sco		0.00	0.00
dom	CG9696	FBgn0020306	FBpp0071528	hemocyte_reg	1	2	1	1	1	1	2	1	1	1	1	1	sco	sco			
edl	CG15085	FBgn0023214	FBpp0085834	hemocyte_reg	1	1	1	1	1	1	1	1	1	1	1	1	sco	sco		1.00	0.81
pnt	CG17077	FBgn0003118	FBpp0088656	hemocyte_reg	1	1	1	1	1	1	1	1	0	1	1	1	sco	sco		0.38	0.69
sgg	CG2621	FBgn0003371	FBpp0070449	hemocyte_reg	1	0	1	1	1	1	1	1	1	1	1	1	sco			0.00	0.00
Cyt-b5	CG2140	FBgn0264294	FBpp0087979	hemocyte_reg	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
chn	CG11798	FBgn0015371	FBpp0110305	hemocyte_reg	1	1	1	1	1	1	1	1	1	1	1	1	sco			1.00	0.81
He	CG31770	FBgn0028430	FBpp0080189	hemocyte_reg	1	1	1	1	0	0	0	0	0	0	0	0	lr	lr		0.69	0.81

Table S2.3: **Lineage restricted genes**

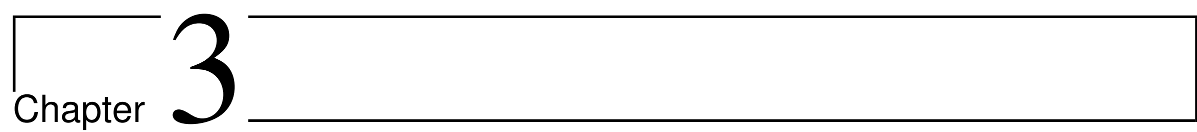
	Effector					Protease														Recognition			
	TotA	TotB	TotC	PPO3	yellow-f⋆	Spn88Eb ⋆	CG11313 ⋆	CG6639 ⋆	CG30414 ⋆	CG4259 ⋆	CG12951 ⋆	CG18477 ⋆	Jon65Aii⋆	Jon25Bii	CG30090 ⋆	CG18563	CG4793	CG3117	CG18478	lectin-24A	αPS4⋆	He	TepI ⋆
mel	1	1	1	1	1	1	1	1	2	1	1	2	1	1	1	1	1	1	2	1	1	1	1
sim	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
sec	2	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1
yak	1	2	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
ere	2	0	0	1	1	1	3	0	1	1	1	1	1	1	1	0	1	0	0	0	1	0	1
ana	0	0	0	0	0	0	0	0	0	1	1	2	1	1	1	0	0	0	0	0	0	0	1
per	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
pse	0	0	0	0	0	0	0	2	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0
wil	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
moj	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
vir	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pathway	JakStat			PPO		non-identified																	JakStat

⋆ Genes under positive selection

For all lineage restricted genes, the numbers of orthologs in each species is given according to the OrthoMCL algorithm Li *et al.* (2003). From the 26 LR found, 23 are listed in this table and the remaining 3 are duplications inside *D. melanogaster* (Table **??**).

Table S2.4: **List of primers for RT-qPCR**

Amplicon name	Direction	Sequence 5′ → 3′	Amplicon size (bp)
PPO3	F	AAGAGGATAAGGTTGACCAGAC	251
	R	GCTTCTCTTATTCAGGAGCGA	251
IM1	F	TCCACTGTCGCCCGATCC	92
	R	CTTGGGTTGAAACTTCCTACTTGC	92
FH68A	F	GAGCAGAAGAGCCCCTACCT	84
	R	AATGAAACCCTGACGTGGAC	84
ACT5C	F	CACACCAAATCTTACAAAATGTGT	83
	R	AATCCGGCCTTGCACATG	83
TEP1 *D. sechellia*	F	ACTAAGAAGTGGCTGTTATACC	190
	R	AGTTGGTGGTAATAAAGAACGG	190
TEP1 *D. simulans*	F	AAGATGAGTTAGCAAAGAACTCAG	282
	R	AAGTTGGTGGTAATAAAGAACGG	282
TEPI *D. melanogaster*	F	AGTCCCATAAAGGCCGACTGA	101
	R	CACCTGCATCAAAGCCATATTG	101



Genetic variation of the immune receptor *Tep1* among natural populations of *Drosophila melanogaster*

Laura Salazar Jaramillo*, Sylvia Gerritsma*, Jelmer Beerda, Carmen van de Beld, Ammerins de Haan and Bregje Wertheim

*Joint first authors

Abstract

Immune receptors are at the interface of host-pathogen interactions and are therefore likely to diverge fast due to co-evolution. In an earlier study, we found divergence in some receptors that are up-regulated after parasitoid attack in several *Drosophila* species. The level of expression of one of these receptors, *Tep1*, correlated with the level of immunological resistance of the species. To test whether the large observed divergence among species in these receptors is also associated with high levels of polymorphisms among populations, we sequenced DNA fragments of 5 immune receptors. For this, we used eight field lines of *D. melanogaster* that differ in resistance against the parasitoid *A. tabida*. Only one of the five genes, *Tep1*, showed considerable sequence variation. This was reflected in rare polymorphisms scattered throughout the gene and high heterozygosity, particularly in four SNP sites. Expression of this gene was measured in four of the field lines after parasitoid attack, and it was up-regulated in all lines, with subtle differences in the timing and level of expression among lines. A RNAi knock-down of *Tep1* gene, suggested an effect of *Tep1* on the composition of the differentiated hemocytes and the encapsulation ability, although the precise function of *Tep1* remains to be determined. The observed large genetic variation in *Tep1* may reflect the co-evolutionary dynamics with parasites, and resembles the hyper-variability that is observed in the Major Histocompatibility Complex.

3.1 Introduction

Immune genes are among the most rapidly changing genes in the genomes (Christophides *et al.*, 2002; Nielsen *et al.*, 2005; Sackton *et al.*, 2007; Obbard *et al.*, 2009b). Both the high parasite

diversity and the dynamic co-evolution between parasites and hosts imposes unrelenting and diversifying selection pressures on the hosts immune system. High diversity has been found in immune genes (Christophides *et al.*, 2002; Nielsen *et al.*, 2005; Sackton *et al.*, 2007; Obbard *et al.*, 2009b; McTaggart *et al.*, 2012), particularly in immune receptors that bind directly to parasites (Hughes & Nei, 1989; Little & Cobbe, 2005; Jiggins & Kim, 2006; Dishaw *et al.*, 2010). This interaction between receptors and parasites is a crucial step to either trigger or guide the immune response to an efficient clearance of the parasite. Studying immune receptors is essential to the understanding of mechanisms for non-self recognition, the activation of immune signalling pathways, and the overall effectiveness of the immune response.

One of the best documented examples of diversity in an immune receptor is the Major Histocompatibility Complex (MHC), which harbours an unprecedented level of diversity in terms of both nucleotide substitutions among species and polymorphisms among populations (Maruyama & Nei, 1981; Parham *et al.*, 1989; Hughes & Yeager, 1998; Klein *et al.*, 2007). The MHC encodes cell-surface glycoproteins that bind antigens derived from pathogens and present them to T-lymphocytes to trigger the immune response against parasites (Penn & Ilmonen, 2005). The variation in MHC genes is generally concentrated in the amino acid residues that are part of the peptide-binding region. Evidence shows that this is the outcome of balancing selection rather than a consequence of higher mutation rates (Hughes & Yeager, 1998). To explain the great diversity present in MHC two selective forces have been suggested : 1) selection favouring heterozygosity and 2) frequency-dependent selection (Hughes & Nei, 1989; Borghans *et al.*, 2004). The precise mechanism by which diversity in MHC confers an advantage remains elusive, regardless of the efforts to establish the implications of decrease in diversity on fitness in laboratory and wild populations (Sommer, 2005).

Genes in the innate immune system also show high levels of polymorphism and signs of adaptive evolution across species, closely resembling MHC (Watson *et al.*, 2005; Dong *et al.*, 2006; Jiggins & Kim, 2006; Dishaw *et al.*, 2010). The innate immune system consists of humoral factors that are released to attack pathogens, and cellular components, such as specialized blood cells that can phagocytize or encapsulate foreign bodies. Striking similarities have been found in the molecular genetics of vertebrate and invertebrate's innate immune systems, despite the large phylogenetic distance. Studies carried out mostly in *D. melanogaster* and humans have found homology in immune pathways, as well as in the mechanisms to regulate cell proliferation and migration (e.g JAK/Stat, Notch, NF-κB) (Williams, 2007; Fauverque & Williams, 2011). In a previous comparative genomics study on the evolution of the cellular immune response in *Drosophila*, we showed that genes involved in hemopoiesis are highly conserved across *Drosophila* species in contrast to the large genetic diversity and paralogy found in other subsets of immune genes, especially those that are likely to interact directly with the pathogens or parasites (Salazar-Jaramillo *et al.*, 2014).

Different types of parasites can attack *Drosophila*, from micro-parasites (fungal and bacterial) to macro-parasites (parasitoid wasps, nematodes and mites) (Kraaijeveld & Wertheim, 2009). The immune response against micro- and macro- parasites differ to some degree because micro-parasites can be neutralized by effector molecules and phagocytized, while macro-parasites are too big to be phagocytized and need to be sequestered by a multicellular layer of specialized cells. Macroparasites such as parasitoid wasps infect *Drosophila* by injecting an egg, which develops at the cost of the host. For the fly larva to survive, the parasitoid egg needs to be encapsulated and melanized by means of specialized hemocytes (blood cells) that proliferate and differentiate upon infection. In *D. melanogaster* three types of blood cells have been described: 1) plasmatocytes which perform phagocytosis of bacteria and other small pathogens and are also recruited in the cellular capsules around parasitoid eggs, 2) lamellocytes, which are large, adhesive and flat cells and important for the formation of the cellular

layer around the parasitoid egg; and 3) crystal cells, which store the precursors of the melanin that is deposited on invading pathogens (Pech & Strand, 1996; Williams, 2007; Fauverque & Williams, 2011)

While great progress has been made in revealing the mechanisms for recognition of micro-parasites by pattern-recognition peptides (e.g PGRPs, GNBPs) (Ferrandon *et al.*, 2007), the recognition of macro-parasites is less well understood. Several hemocyte receptors showing high diversity have been suggested to play an important role in the encapsulation of macro-parasites, such as *Lectin-24A* (Keebaugh & Schlenke, 2012) and $\alpha PS4$ (Salazar-Jaramillo *et al.*, 2014), although their function and binding mechanisms are not well understood.

A different type of receptors in invertebrates is the Thioesther-containing protein (Tep), which have been described in *Drosophila* (Jiggins & Kim, 2006; Aoun *et al.*, 2011), mosquitoes (Blandin & Levashina, 2004; Obbard *et al.*, 2009a) and *Daphnia* (Little & Cobbe, 2005). These Teps contain a central hypervariable region, which shows homology to domains in vertebrates that are important for binding to the parasite (i.e., the bait domain of the $\alpha 2$ macroglobulin and the anaphylatoxin domain of C3 (Lagueux *et al.*, 2000)). Teps are secreted extracellularly and bind to pathogens (Stroschein-Stevenson *et al.*, 2006), acting as an opsonin to enhance phagocytosis by plasmatocytes.

In *Drosophila* there are six Tep genes encoded in the genome (Aoun *et al.*, 2011), from which *Tep1* and *Tep2* have been found to be under positive selection (Jiggins & Kim, 2006; Sackton *et al.*, 2007; Salazar-Jaramillo *et al.*, 2014). What makes it a particularly interesting receptor for macro-parasite recognition is our previous finding that *Tep1* arose as a recent duplication in the *melanogaster* group (Salazar-Jaramillo *et al.*, 2014). The monophyletic clade that has the *Tep1* gene contains a number of *Drosophila* species capable of lamellocyte-mediated encapsulation, while species outside the clade do not produce lamellocytes and are mostly not able to survive after parasitoid infection (Salazar-Jaramillo *et al.*, 2014). Within this clade the level of immunological resistance against parasitoid infection varies largely among species, including one species, *D. sechellia*, where the ability to encapsulate has secondarily been lost. Interestingly, a deletion of four exons in *Tep1* was found in *D. sechellia*, while the exon structure was conserved among the other species of the clade. Moreover, large differences in the level and speed of expression of *Tep1* were found when comparing sister species with different levels of resistance (Salazar-Jaramillo *et al.*, 2014).

Tep1 is expressed in *Drosophila* in hemocytes (lamellocytes and plasmatocytes (Irving *et al.*, 2005)), lymph glands (the hemopoeitic organ in *Drosophila* larvae) and larval fat body (the equivalent of the vertebrate liver) (Aoun *et al.*, 2011), which are the main tissues involved in larval immunity. It has been found that *Tep1* is up-regulated after infection with bacteria and parasitoid wasps, suggesting that it plays a role in the immune response against both micro-parasites and macro-parasites (Lagueux *et al.*, 2000; DeGregorio *et al.*, 2002; Wertheim *et al.*, 2005; Schlenke *et al.*, 2007). Whereas Teps function as opsonins to bacteria, we here hypothesize that *Tep1* may have an important function in the immune response against parasitoids, for example, in guiding the cells that form the capsule towards the parasitoid egg.

In this study we characterized the genetic variation in *Tep1* and compared it to other receptors that are up-regulated after parasitoid attack (two *PGRPs, Lectin-24A* and $\alpha PS4$). Whereas we previously compared species that showed large variation in the immune response to parasitoids, we now exploit the large variation that exists among natural populations of *D. melanogaster* (Kraaijeveld & Godfray, 1999; Gerritsma *et al.*, 2013). We used eight field lines of *D. melanogaster* collected in Europe with well-characterized variation in resistance against one of its natural parasitoids, the wasp *Asobara tabida* (Gerritsma *et al.*, 2013). In Europe, this variation has been associated to local abundance and virulence of the parasitoids and the presence of alternative hosts (Kraaijeveld & Godfray, 1999). We investigated whether the

phenotypic variation could be associated to the genetic variation in immune cellular receptors among the field lines. We expected to find large sequence diversity among field lines and to find variation in the level and speed of expression of *Tep1*, based on a previous inter-species comparison. We tested this by 1) identifying single-nucleotide-polymorphism, heterozygosity and haplotypes of a sequence fragment for *PGRP-SB1, PGRP-LB, Tep1, αPS4* and *Lectin-24A* and 2) measuring the expression of *Tep1* in a subset of the natural populations for different time points in parasitized and control groups. Additionally, we tested whether *Tep1* is essential for the immune response against parasitoids through RNA interference of mRNA.

3.2 Materials and Methods

Insect lines The *D. melanogaster* lines were collected from natural populations in Europe in the summer of 2009. Lines showed substantial genetic differentiation, as indicated by an average pair-wise F_{ST} value of 0.124 ± 0.015 (S. Gerritsma, unpublished data). More details on these lines and measurements of resistance against *Asobara tabida* can be found in Gerritsma *et al.* (2013) (Table 3.1). In short, they differ in their ability to encapsulate *A. tabida* eggs, measured as the percentage of parasitized larvae that had fully melanized a parasitoid egg, 96 hours post-parasitization (scored by dissections of the parasitized larvae). All flies were kept in mass cultures at 20 ℃ under a dark: light regime of 12:12 in bottles containing 30 mL standard medium (26 g inactivated yeast, 54 g sugar, 13 ml nipagin 8.5 mM solution, dissolved in 1 L).

For the RNAi experiment we used the GAL4/UAS system. The UAS-Tep and the w1118 control lines were obtained from the Vienna Drosophila RNAi Center (Tep, transformant ID: 30873, w1118: Transformant ID 60100) (Dietzl *et al.*, 2007). The ubiquitous DaGAL4 driver line was kindly provided by Pascale Dijkers (University Medical Center Groningen). These lines were maintained in vials in uncrowded conditions and at room temperature.

Two strains of the parasitoid wasp *A. tabida* were used in the experiments. The line used in the expression experiment, “TMS”, is an inbred line of moderate virulence. It was established as an isofemale line in 2010 from a cross between two lines, one originally from Sospel (France) and one from Pisa (Italy). This line has been been maintained on *D. melanogaster* at 20 ℃ under a dark:light regime of 12:12. The line used in the RNAi experiment (HK) is a low virulence line originally collected in Hertenkamp, Leiden (the Netherlands). This line has been maintained on *D. subobscura* at 20 ℃ under a dark:light regime of 12:12.

DNA amplification and sequencing DNA was extracted from 6 females per field line (for 8 field lines, Table 3.1) using a high salt DNA extraction protocol without chloroform, based on Aljanabi & Martinez (1997). Briefly, tissue was homogenized in 400 μl homogenizing buffer (0.4 M NaCl, 10 mM Tris-HCl pH 8.0, 2 mM EDTA) using tip-melted filter tips. After homogenization, 40 μl of 20 % SDS and 8.5 μl of 10 *mg/ml* proteinase K (200 μM final concentration) were added and mixed well. The samples were incubated for 1h at 55 ℃, after which 190 μl of 6 M NaCl (35 g NaCl saturated in 100 ml dH2O) was added to each sample. Samples were vortexed for 30 s at full speed and then centrifuged for 30 min at 10000 g RT. The supernatant was transferred to new tubes and an equal volume of ice-cold isopropanol was added to each sample, vortexed and incubated for 1h at -20 ℃. Samples were centrifuged for 30 min at 15000 g RT. The supernatant was removed and the pellet washed 3x with 70 % ethanol, dried and suspended in 3.8 μl sterile dH2O.

Pairs of primers were designed to amplify variable regions of *Tep1, GNBPlike, PGRP-SB1,PGRP-LB, Lectin-24A* and *αPS4*. The gene primers for *GNBPlike* did not lead to any

product and were therefore excluded from further analysis. The amplicons were each of approximately 500 bp along the regions of interest of the genes (for *Tep1* three fragments were sequenced). The primers were designed using the program PerlPrimer v1.1.21 (Marshall, 2004) (Table S1). For the PCR reactions, a 1:10 dilution of the extracted DNA was used as template. After amplification with a standard PCR (3 minutes on 94 ℃ 35 cycles of 94 ℃ for 25 s, melting temperature for 45 s and 72 ℃ for 45 s, 72 ℃ for 7 min), products were purified from excess primers, dNTPs and polymerases by adding the following reaction mix: 0.08 μl ExoI (sExonuclease I, 20 U/μl), 0.12 μl FAP (FastAP thermosensitive alkaline phosphasate, 1 U/μl) and 3.8 μl sterile dH2O to 5μl of the PCR product. This was then heated to 37 ℃ for 30 min to activate the enzymes after which the mix was heated to 80 ℃ for 15 min to deactivate the reaction. The product was sequenced by the commercial company GATC Biotech, Germany.

Sequence products were processed with CLC Genomics Workbench using the function "second peak calling" (25 %) to identify heterozygotes, and aligned with ClustalW (Larkin *et al.*, 2007). Sequence errors and low quality bases were manually removed from the sequences. A consensus sequence was created by combining the forward and reverse sequences and by concatenating multiple fragments of one gene (only for *Tep1*). The gene sequences were aligned to the reference transcript (obtained from Flybase) using GMAP (Wu & Watanabe, 2005). Gene Sequences were manipulated and analysed with customized scripts in Python and R (R Development Core Team, 2008). Population genetic parameters were estimated for all individuals pooled using Egglib (Mita & Siol, 2012) and the R package Adegenet (Jombart, 2008). Genetic diversity estimators, F_{ST} and G_{ST} (Nei, 1973), were calculated using GenAlEx 6.5 (Peakall & Smouse, 2012). Deviation from neutrality of the SNPs was tested using an F_{ST} outlier analysis implemented in LOSITAN (Antao, 2008; Beaumont & Nichols, 1996). Default parameter settings were used for the analysis: 50,000 simulations, confidence interval of 0.95, false discovery rate of 0.1, 8 populations, subsample size of 12 loci (SNPs), simulated F_{ST} of 0.09554 and an attempted F_{ST} of 0.095. Loci were considered as candidates for positive selection above a probability level of 0.95.

To test copy number variation in *Tep1*, we performed a qPCR on genomic DNA of 23 samples on 5 different populations (6 Bay, 5 Kal, 1 Arl, 5 Sta, 6 Got; see table 1 for the abbreviations), taking as endogenous control the gene $\alpha Tub84B$. All samples were diluted to the final concentration of 20 $ng/\mu l$. Two technical replicates were used for all lines. The data were analysed similarly as for mRNA (see below subsection "Gene expression").

Gene expression Two highly resistant field lines (collected in Gotheron, France and Kaltern am See, Italy) and two low resistant lines (collected in Saint Andrews, UK and Bayreuth, Germany) (Table 3.1) were used to test differential expression of *Tep1* at four time points after parasitization using quantitative real-time PCR (RT-qPCR). Larvae that were not exposed to the parasitoid were used as controls and were sampled at the same time points.

Fifty first-instar *D. melanogaster* larvae of each field line were transferred to a petri dish with an agar base and a thin layer of yeast. When larvae reached second-instar stage, individuals from the treatment group were exposed to one parasitoid female. At least 40 larvae were collected, for which oviposition was observed for at least 10 seconds. The wasps were replaced after successful parasitization of eight larvae, and the total period of oviposition per line per time point was approximately 30 minutes to minimize variation in larval development. Parasitized larvae were transferred to a new petri dish to allow development for a fixed period of hours (h) (3 h, 6 h, 24 h, 48 h) when sampling took place. Each sample per time point per line consisted of five pooled larvae. Samples were snap-frozen in liquid nitrogen and stored

at -80 ℃ until RNA extraction. For each time point three biological replicates were collected per line. In the control group larvae were treated similarly as the parasitized group except no wasp was introduced.

RNA was extracted in 1 ml of Trizol (Invitrogen, Carlsbad, CA, USA) using a pestle for tissue disruption. Purification was performed with QiaGen RNeasy plus mini kit (Qiagen, Hilden Germany) according to manufacturer's protocol. To ensure that no genomic DNA was present in the samples, genomic DNA digestion was performed with DNAse (QiaGen) on the columns. cDNA was synthesized from 10 μl of purified RNA using Revert Aid RT (Thermo scientific). The qPCRs were performed in total volumes of 25 μl per reaction in an Applied Biosystems 7300 Real Time PCR System, using Absolute QPCR SYBR Green ROX mix (Abgene, Hamburg, Germany). Three technical replicates were used for each sample.

Gene expression of *Tep1* was analysed using the R package qpcR (1.3-7.1) (Ritz & Spiess, 2008). Forkhead domain 68A (fh68A) and α-Tubulin at 84B (αtub84B) were used as endogenous reference genes. The median of three technical replicates was obtained for each of three biological samples. Quantification was based on the window-of-linearity method that incorporates individual PCR efficiencies for each sample to calculate the initial concentration (N0) for genes. The N0 of the target gene per biological replicate was standardized to the geometric mean of the N0 of the two reference genes (Vandesompele *et al.*, 2002). Statistical differences were estimated with an ANOVA on the standardized initial concentration of *Tep1*.

Knock-down: GAL4/UAS RNAi To induce gene silencing, DaGAL4 males were crossed with UAS-*Tep1* females, which produced progeny with the genotype DaGAL4/UAS-*Tep1*. To control for genetic background effects, males from the same DaGAL4 driver line were crossed to females from the w1118 line (progeny with genotype GAL4/+) and males from the construct line UAS-*Tep1* with females w1118 (progeny with genotype UAS/+). Crosses were performed at 25 ℃.

To assess the phenotypic effect of the *Tep1* knock-down, the encapsulation ability was measured and circulating hemocytes counted. The experiment was performed twice (15 and 45 parasitized larvae in experiment 1 and 2, respectively), using a modified protocol from (Gerritsma *et al.*, 2013). Briefly, eggs were collected within four hours of oviposition and kept at 25 ℃. Two days after oviposition, at the second-instar stage, a parasitoid wasp was introduced and the oviposition behaviour was observed. Only larvae that were parasitized were collected in the treatment group and non-exposed larvae were collected in the control group. Larvae were transferred to 20 ℃ because the parasitoid does not develop well at higher temperatures. At 72 hours after parasitization, the hemolymph of three pooled larvae was collected (approximately 0.2 μl for the pooled set) and diluted in 8 μl Ringer's solution with 0.5 % Giemsa solution (KaryoMAX® Giemsa Stain Stock Solution) to stain the blood cells. Cells were counted using a Neubauer Improved hemocytometer slide (0.1 mm depth) under a phase-contrast microscope (10x40 magnification). After bleeding, the parasitized larvae were immediately dissected to assess the presence of a wasp egg and to score the percentage of melanization around the egg and development to the second-instar stage. To verify the extent of gene silencing, we performed RT-qPCR of larvae 50 hours after parasitization (three biological replicates, and five pooled larvae per biological replicate), using the same protocol and analysis as described above (see subsection "Gene expression"). The expression of the gene Forkhead domain 68A (fh68A) was used as endogenous control.

3.3 Results

3.3.1 Sequence variation

Fragments of five immune receptors were sequenced for eight field lines that varied in resistance against *A. tabida* (Tables 3.1 and 3.2). The expression of these receptors is induced by parasitoid infection (Wertheim *et al.*, 2005). One of these receptors, *PGRP-LB1*, had no polymorphic sites for the 48 individuals that we tested. Three of these receptors had 1-4 polymorphic sites, including only one non-synonymous substitution. In contrast, *Tep1* showed 29 polymorphic sites, of which the majority were exonic (n=23), and of this, 17 non-synonymous SNPs (Figure 3.2). No fixation was found in any of the polymorphic sites, all contained a mixture of homozygous and heterozygous individuals for each site (Figure 3.1).

Standard population genetic parameters were determined for *Tep1* for the pool of individuals of all field lines (Table 3.3). Although there was an indication for a skew towards rare genetic variants (negative value for Tajima's D), this was not significant. An F_{ST} outlier analysis, implemented in LOSITAN, confirmed that there was no deviation from neutrality (Figure S3.2). Of the 13 haplotypes found (Table 3.3), none was common or a high frequency haplotype, nor was any haplotype exclusive to a field line.

Estimation of the genetic differentiation of populations based on the SNPs in *Tep1* using pairwise comparisons of G_{ST} and F_{ST} showed substructuring among some populations (Table 3.1). There was no clear association between the combined minor allele frequencies across all 29 polymorphic positions and either the level of resistance or the geographic origin of the lines (Figure 3.2).

An excess of observed heterozygosity with respect to expected values was found for four sites (Figure 3.2 and Figure S3.1). Two of these sites lead to a non-synonymous change. A possible scenario for an excess of heterozygosity is the presence of multiple gene copies for *Tep1*. We tested this hypothesis through DNA qPCR of a fragment of *Tep1* and found no significant deviation from a 1:1 ratio with respect to the endogenous control (t = -1.6633, df = 24.737, P-value = 0.1089).

Table 3.1: **Information of field lines: collection site, resistance levels and *Tep1* sequence diversity.** Asterisks indicate significance (P<0.05) after 999 permutations.

Line	Location	% Resistance	F_{ST} (Wright)							
			BAY	STA	GRO	BRE	INN	KAL	ARL	GOT
Bayreuth (BAY)	Germany	3.7	-	0.065	0.048	0.031	0.071	0.039	0.038	0.035
St Andrews (STA)	Scotland	3.3	0.036	-	0.066	0.020	0.055	0.026	**0.093***	0.032
Groningen (GRO)	Netherlands	15.6	0.015	**0.041***	-	0.043	**0.062***	0.022	0.049	0.049
Bremen (BRE)	Germany	25.9	-0.006	-0.006	0.011	-	0.049	0.014	0.043	0.011
Innsbruck (INN)	Austria	27.9	0.038	0.029	**0.032***	0.016	-	0.035	**0.072***	0.058
Kaltern (KAL)	Italy	44.4	0.005	0.001	-0.008	-0.018	0.005	-	0.049	0.022
Arles (ARL)	France	45.5	0.000	**0.064***	0.015	0.005	**0.038***	0.015	-	0.039
Gotheron (GOT)	France	46.4	0.000	0.007	0.018	-0.022	0.026	-0.009	0.003	-
			BAY	STA	GRO	BRE	INN	KAL	ARL	GOT
			Pairwise G_{ST} (Nei)							

Table 3.2: **Polymorphic sites of five immune receptors.** Sequence fragments were obtained for 48 individuals (six individuals for eight populations). Details on amplicon size are provided in table S3.1.

	Polymorphic sites			Heterozygous sites
	Exons		Introns	
	Syn	NonSyn		
Tep1	6	17	6	29
Lectin24-A	1	0	0	1
αPS4	3	0	0	2
PGRP-LB1	0	0	0	0
PGRP-SB1	2	1	0	3

Table 3.3: **Summary of population genetic parameters for *Tep1* across 48 individuals from eight populations**

π	θW	Tajima D	Haplotype	
			Number	Diversity
0.0045	0.006	-0.7 n.s	13	0.67

π: average pairwise difference, θW: Watterson estimate of $4N\mu$, Tajima's D: estimators of deviation from neutrality. n.s: non-significant P-value, estimated with the python package *eggcoal* from 100 simulations with parameters obtained from the data.

3.3.2 Gene expression

The previous finding that resistance levels among different species correlated significantly with differences in the expression of *Tep1* (Salazar-Jaramillo *et al.*, 2014), motivated us to test whether a similar pattern could be found among four field lines with very different levels of resistance (Sta, Bay, Kal, Got) against *A. tabida* (Table 3.1). We performed RT-qPCR on groups of parasitized larvae and the corresponding non-parasitized controls along a time course. *Tep1* was induced after parasitization in all populations (ANOVA; F=84.6; df=1,61; P-value=0.011). The level of expression of both control and parasitized individuals changed over time (ANOVA; F=4.02; df=3,61; P-value=$4e^{-13}$), but this induction was only marginally different among the populations (ANOVA; F=2.19; df=3,61; P-value=0.0982) (Figure 3.3, Table S6). Thus, the level, or speed of expression may be correlated with the resistance level of the populations, but this correlation is either weak or the differences between populations are too subtle to be detected with our samples sizes.

3.3.3 Knock-down

To assess the importance of *Tep1* in the immune response against parasitoid wasps, we attempted to do a knock-down of *Tep1* using the GAL4/UAS system. The GAL4/UAS system is a technique that allows knock-down of a gene by expressing a construct containing a complementary sequence to the target gene. This induces the formation of a double stranded structure with mRNA of the target gene, which is degraded by the anti-RNAi machinery of

the organism. The expression of the construct is under the control of an exogenous transcription factor that is situated on another construct. By crossing flies with the two types of construct, the knock-down of target genes can be induced (Duffy, 2002).

One of our experiments gave significantly reduced expression (although weak) in the crosses containing the construct (UAS-*Tep1*) compared to the cross without the *Tep1* construct (GAL4/+) (ANOVA; F=5.63; P-value=0.02; Table S3.2; Figure 3.4, panel 1). For both the GAL4/UAS-*Tep1* cross and the +/UAS-*Tep1* cross, a knock-down of *Tep1* was observed in the control (unparasitized) treatment and in the parasitized treatment, while in the GAL4/+ cross, parasitization seemed to induce *Tep1* expression as is normally observed. The knock-down in the +/UAS-*Tep1* cross indicates that there is some leaky expression of the construct (i.e., expression even without the exogenous GAL4 transcription factor).

Interestingly, the knock-down cross (GAL4/UAS-*Tep1*) showed a decrease in the encapsulation rate (glm; F=17.10; df=2,25; P-value=0.0; Table S3.3; Figure 3.5, panel 1) and significantly increased the lamellocytes in the parasitized group (ANOVA; F=6.85; df=1,22; P-value=0.016; Table S3.4) and plasmatocytes (ANOVA; F=19.52; df=2,22; P-value=0.0; Table S3.4) in both the control and parasitized groups (Figure 3.6, panel 1). This suggests an effect of *Tep1* on the production of circulating hemocytes or on the involvement of *Tep1* in recruiting circulating hemocytes into the cellular capsule. Unfortunately, upon repeating this experiment with larger sample sizes, the knock-down of *Tep1* was less successful and variation between replicates was too large to show a clear pattern (Figures 3.4, 3.5 and 3.6, panel 2). Based on these patterns in the two knock-down experiments combined, we cannot yet draw any firm conclusions on the functionality of *Tep1* after parasitoid attack.

3.4 Discussion

Our aim was to study the genetic variation of immune receptors involved in the cellular response in field lines of *D. melanogaster* to parasitoid wasps. Many studies on genetics of *Drosophila* immunity are carried out using pathogens that are not known to infect *Drosophila* in nature (Keebaugh & Schlenke, 2012), which makes it difficult to interpret the genetic variation in relation to the ecological context where organisms live. Here we used 8 field lines with a clear difference in resistance against one of its natural parasites, the parasitoid wasp *A. tabida*. The variation in resistance among these field lines has been shown to have a strong genetic component (Gerritsma *et al.*, 2013).

We first characterized the genetic variation of five immune receptors that had previously been found to change expression after parasitoid attack. Only *Tep1* showed considerable polymorphism levels. The low level of sequence variation in the other receptors was surprising given that polymorphisms in *PGRP*s and *GNBP*s (involved in the recognition of microbial cell wall) have been associated with resistance to bacterial load (Sackton *et al.*, 2010) and adaptive evolution has been reported for *Lectin-24A* (Keebaugh & Schlenke, 2012). The high diversity found in *Tep1* confirms previous work on Tep genes in *Drosophila* (Jiggins & Kim, 2006) and in other arthropods (Little & Cobbe, 2005; Obbard *et al.*, 2009a). In contrast to other studies, we did not find fixation of a particular SNP in any of the populations nor a dominant or exclusive haplotype to a population. Instead, we found high levels of heterozygosity, which

may be in part the result of using outbred field lines instead of isofemale lines. The power of our analysis is, however, limited by the small sample size (six individuals in each population), thus the inference of population parameters should be treated with caution.

The excess of heterozygotes at four sites within *Tep1*, of which two lead to non-synonymous changes, is intriguing. We tested for multiple gene copies of *Tep1* in our lines, and found none in 23 tested samples. Although only speculative at this stage, it is possible that the high heterozygosity at specific sites constitutes a signature of the interaction with parasites. As in the case of MHC, it remains unclear what type of selection would be responsible for such high variation and how is it maintained. Theoretical models inspired on MHC suggested that heterozygote advantage alone is insufficient to account for the observed high degree of polymorphism in MHC (Borghans *et al.*, 2004), whereas negative frequency dependent selection can produce a large variety of rare alleles and high heterozygosity levels. It remains to be determined if the same would apply to the high variability and excess of heterozygotes in *Tep1*.

We also examined the variation in expression of *Tep1* in four of the eight field lines. We confirmed that this gene is consistently up-regulated after parasitization in all field lines, and shows subtle differences in the expression profile among lines. While the increase in expression of *Tep1* after parasitization would be consistent with an induction upon parasitoid attack, this increase could also reflect the proliferation of hemocytes (where *Tep1* is expressed), a process that typically occurs after parasitization. Currently, we cannot disentangle the two effects, but further expression studies on a broader repertoire of genes, including hemocyte-specific genes and other Tep genes, should help to tease apart the role of its up-regulation.

Finally, we tried to establish whether *Tep1* is essential for the immune response using UAS/GAL4 lines. Unfortunately our experiments did not lead to conclusive results due to the weak or absent knock-down of the gene in part of the experiment. Failure to obtain a strong knock-down could at least be partly explained by the temperature, at which the experiments were carried out. The UAS/GAL4 system is optimized for temperatures above 25 ℃ (Duffy, 2002), but after parasitization the flies needed to be maintained at 20 ℃ due to the high sensitivity of the wasps to higher temperature. Additionally, the highly-inbred laboratory strains suffered from very low resistance, which required the use of an avirulent wasp strain. This led to low parasitization rates, and high variability among parasitized larvae. Our first small-scale experiment showed weak but significant knock-down. Interestingly, there was a considerable increase in lamellocytes for the parasitized group in the knock-down cross. Although this may at first seem counter-intuitive, this result could be interpreted as the failure of lamellocytes to adhere to the parasitoid egg, thus increasing the number of circulating lamellocytes. A similar knock-down study on *Tep1* had previously found that this gene was not strictly required for responding against bacterial and fungal infections (Aoun *et al.*, 2011), and it was proposed that its function can be compensated by other Tep proteins, acting thus redundantly. The effect of RNAi against *Tep1* for responding to parasitoid wasps was not tested in that study. Upon repeating our experiment with larger sample sizes, we failed in getting consistent knock-down of *Tep1*. Thus, our study gives some tentative indications that *Tep1* may be required for responding to parasitoids, but confirmation will require additional optimization of these RNAi experiments.

Studying the molecular mechanisms of recognition of foreign bodies is fundamental to the understanding of immunity and its great diversity (Litman *et al.*, 2010). The study of MHC has enormously contributed to this field. Most animals, however, lack antibody-based immunity and yet possess an efficient immune system able to recognize and eliminate foreign bodies. In invertebrates, highly variable gene families have been described recently (e.g., *Dscam* in insects and *FREPs* in molluscs), which may be involved in pathogen-specific immune response, but

their exact role and mechanism of recognition remains unknown (Dong *et al.*, 2006; Bowden *et al.*, 2007). The study of immune molecules with large genetic variation in invertebrates may help in understanding the evolution of hypervariable immune molecules, including MHC. Here, we focused on *Tep1*, one hypervariable molecule, which has recurrently been associated with the immune response. We found consistent up-regulation after parasitoid attack of *Tep1* among natural populations, and subtle differences in expression between natural populations. We found that this gene harboured levels of polymorphism much larger than other immune receptors, and high heterozygosity levels concentrated at particular sites. Although the exact role of *Tep1*, particularly in the immune response against parasitoid wasps, remains to be elucidated, this could reflect the history of co-evolutionary interactions with parasites.

3.5 Acknowledgements

We would like to thank Nikkoletta Sidiropoulou for help during gene sequencing and its analysis, Pascale Dijkers for providing the GAL4 line, Jean-Christophe Billeter and his lab for sharing the facilities to carry out the knock-down experiment and for general discussion and assistance during the experiment; Anna Rensink, Corneel Vermeulen and Louis van de Zande for assistance and discussion about the molecular techniques, and Leo Beukeboom and the Evolutionary Genetics group in Groningen for discussion and comments on the manuscript.

3.6 Supplementary Material Chapter 3

Table S3.1: **Pairs of primers used for DNA sequencing, gene expression (mRNA) and copy-number-variation (DNA)**

	Forward	Reverse	Amplicon	Temp(°C)
sequencing				
αPS4	CTTTGCTGGTAATGGATCGGTG	TTCCGAGCAGATCTCTTCGT	480	52
Lectin24-A	GCAGACGTTTCCAACATATTGG	CGCCCAGGAAGTATGATTTCG	411	52
PGRP-SB1	ATCAAGAACATTCAGTCGGATCAC	TGATCTCGTTGTACAGAGCATCAC	330	55
PGRP-LB	CGTCATCATCCATCACTCGT	AAATGAAGAGAAGGAAGAAGG	347	51
GNBPlike	CGGAGATGTTCTGTACTATTGG	AATCCCTTCGGTGAGTTGAC	–	52
Tep1 A	TGACTAATGTGCTAGTGGTGG	GGGAAGCATTATGTGCTCGG	532	61
Tep1 B	AAGACTGTGGTTAGAAACTTGGAC	TCAAAGTCCCTCTCGATATCGAC	569	61
Tep1 C	GAACAATCTCGCCACGTACTC	AGGTCAACGTATCGGGAACAG	547	52
qPCR				
FD68A-mRNA	GCTAGTCCACGTCAGGGT	GTCTGGAACAGATCCTGT	86	55
Tep1-mRNA	AGTCCCATAAAGGCCGACTGA	CACCTGCATCAAAGCCATATTG	101	55
αTub84B-mRNA	GTTTGTCAAGCCTCATAGCC	TGGATAGAGATACATTCACGCA	120	55
Tep1-DNA	CATTCAAGAAACCCGTAAAGG	CTAACCACAGTCTTAAATTCAG	283	55
αTub84B-DNA	TGATACTTCGACGCATAACTG	GGATAGAGATACATTCACGCTG	243	55

Table S3.2: **ANOVA (two-way) of normalized expression of *Tep1*.** Differences in the log-transformed ratio of *Tep1* and the endogenous control were tested among three crosses and two treatments (parasitized and control) for two RNAi experiments.

A) experiment1					
	df	Sum Sq	Mean Sq	F value	Pr(>F)
cross	2	2.39	1.20	5.63	0.0230
treatment	1	0.00	0.00	0.00	0.9606
cross:treatment	2	0.28	0.14	0.65	0.5411
Residuals	10	2.13	0.21		
B) experiment2					
cross	2	1.60	0.80	0.32	0.7340
treatment	1	0.13	0.13	0.05	0.8245
cross:treatment	2	0.92	0.46	0.18	0.8350
Residuals	12	30.18	2.52		

Table S3.3: **GLM of encapsulation for the RNAi experiment.** Differences in the encapsulation of the parasitoid egg were tested among three crosses for two RNAi experiments

A) experiment1						
	df	Deviance	Resid. Df	Resid. Dev	F	Pr(>F)
NULL			27	19.42		
cross	2	9.59	25	9.83	17.10	0.0000
B) experiment2						
NULL			76	11.53		
cross	2	0.32	74	11.21	1.26	0.2906

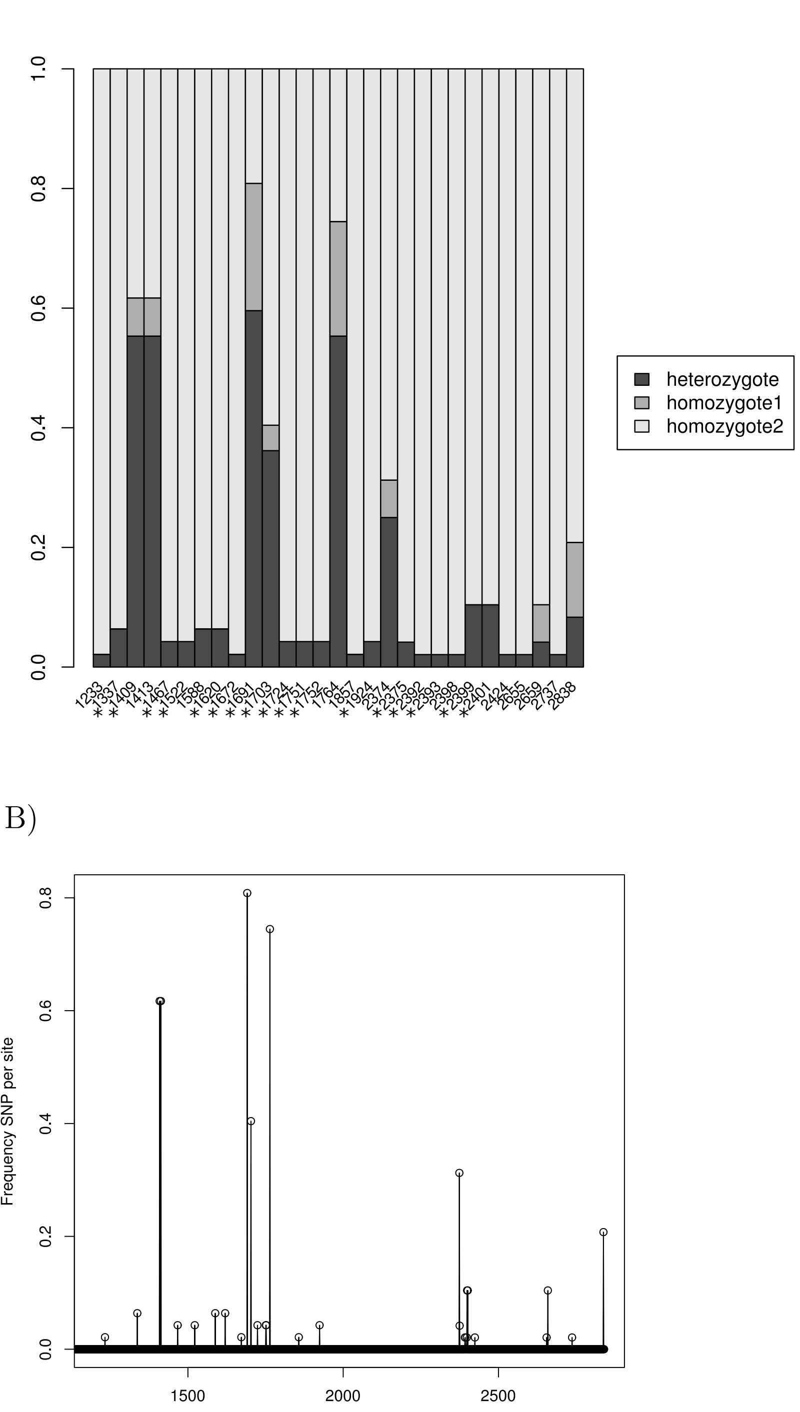


Figure 3.1: **Distribution of *Tep1* SNPs and genotypes.** A) Distribution of genotype proportions (heterozygous and homozygous) for each SNP position and B) Proportion of the 48 individuals containing the minor allele for each SNP position (indicated on the x-axis) and a schematic representation of the exonic structure of the gene (in dark-gray the sequenced fragment).

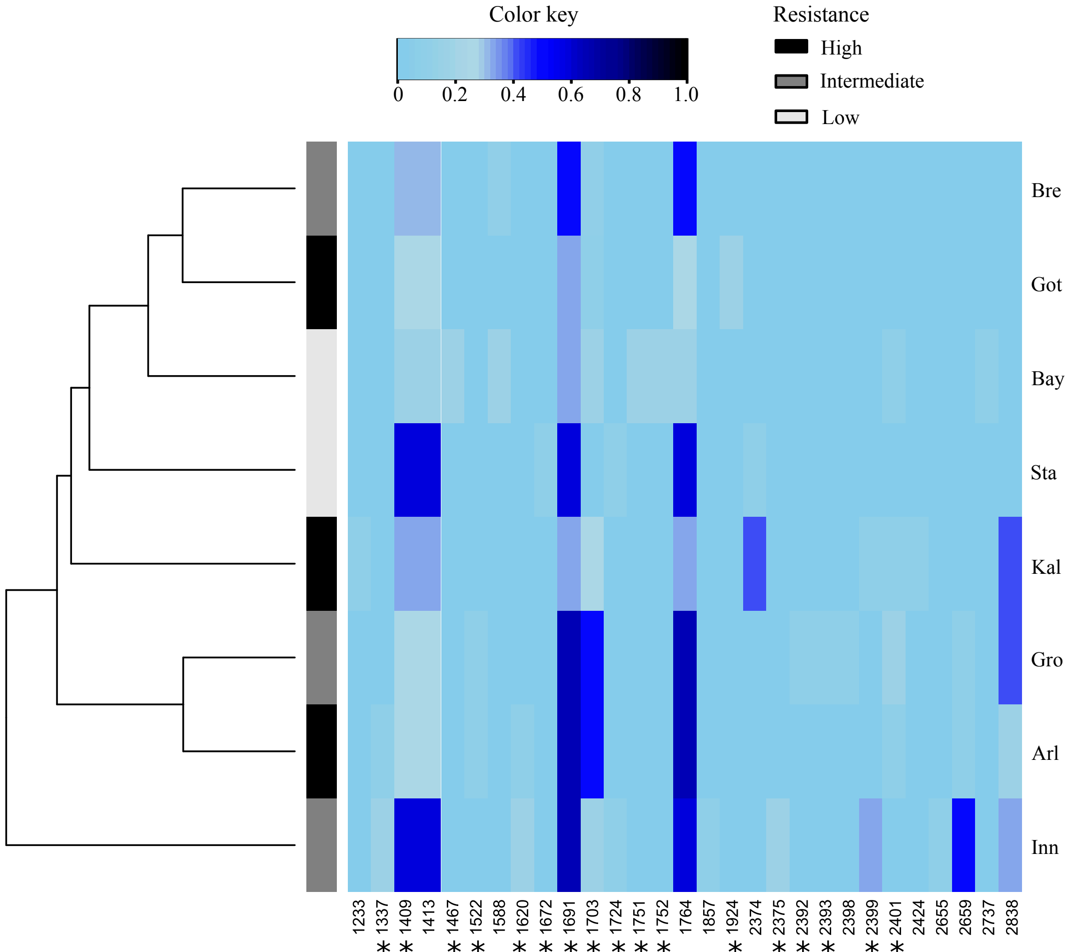


Figure 3.2: **Heatmap of the *Tep1* allele frequencies** The figure shows the clustering of eight field lines, according to the minor allele frequencies in each of the 29 polymorphic positions. The level of parasitoid resistance for each field line is provided by colour coding on the left of the heatmap.

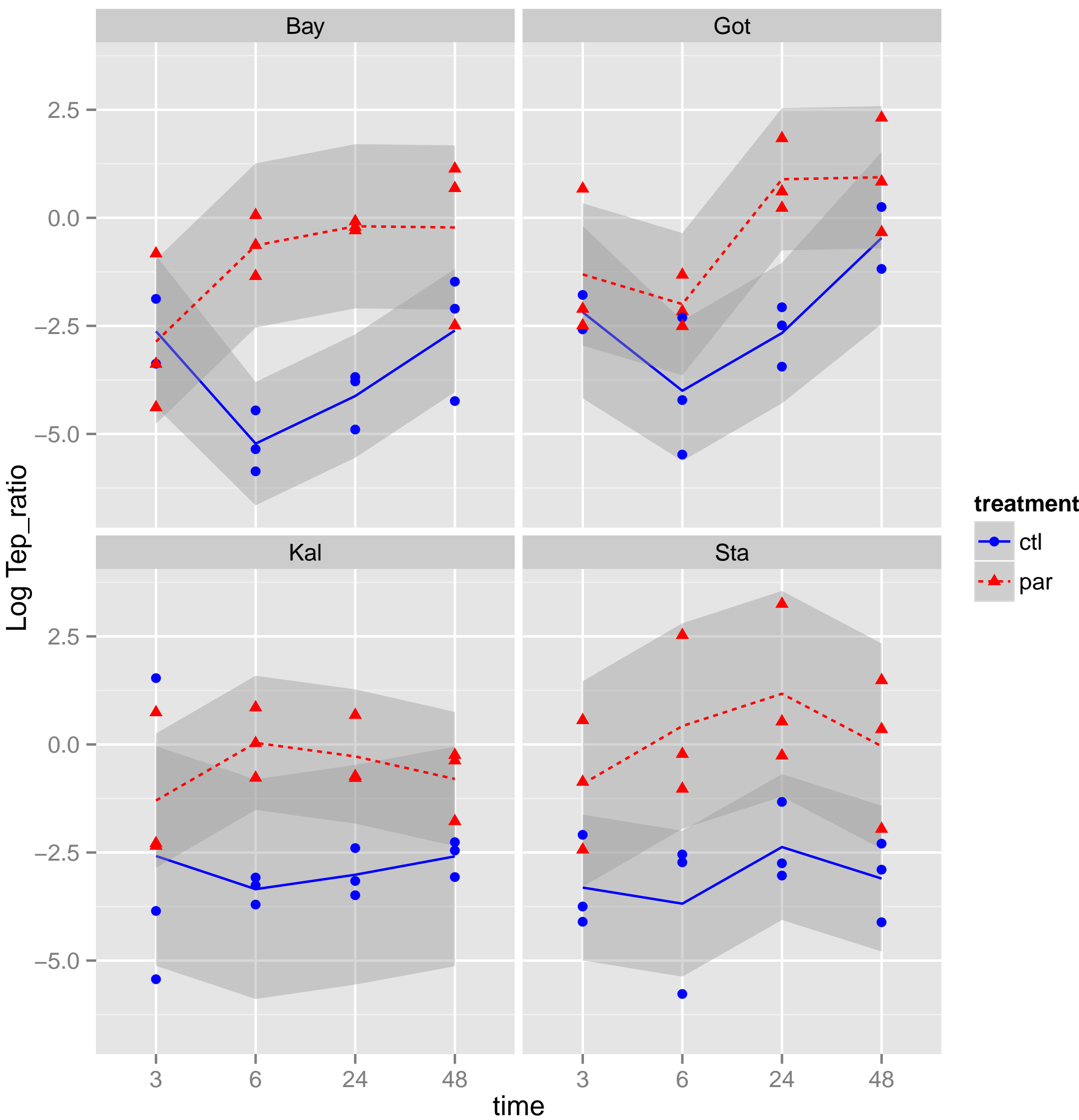


Figure 3.3: **Time series of *Tep1* expression in four field lines** Log-transformed normalized expression of *Tep1* for four time points in parasitized and control groups of four populations that differ in resistance: High ("Got" and "Kal") and low ("Sta" and "Bay")

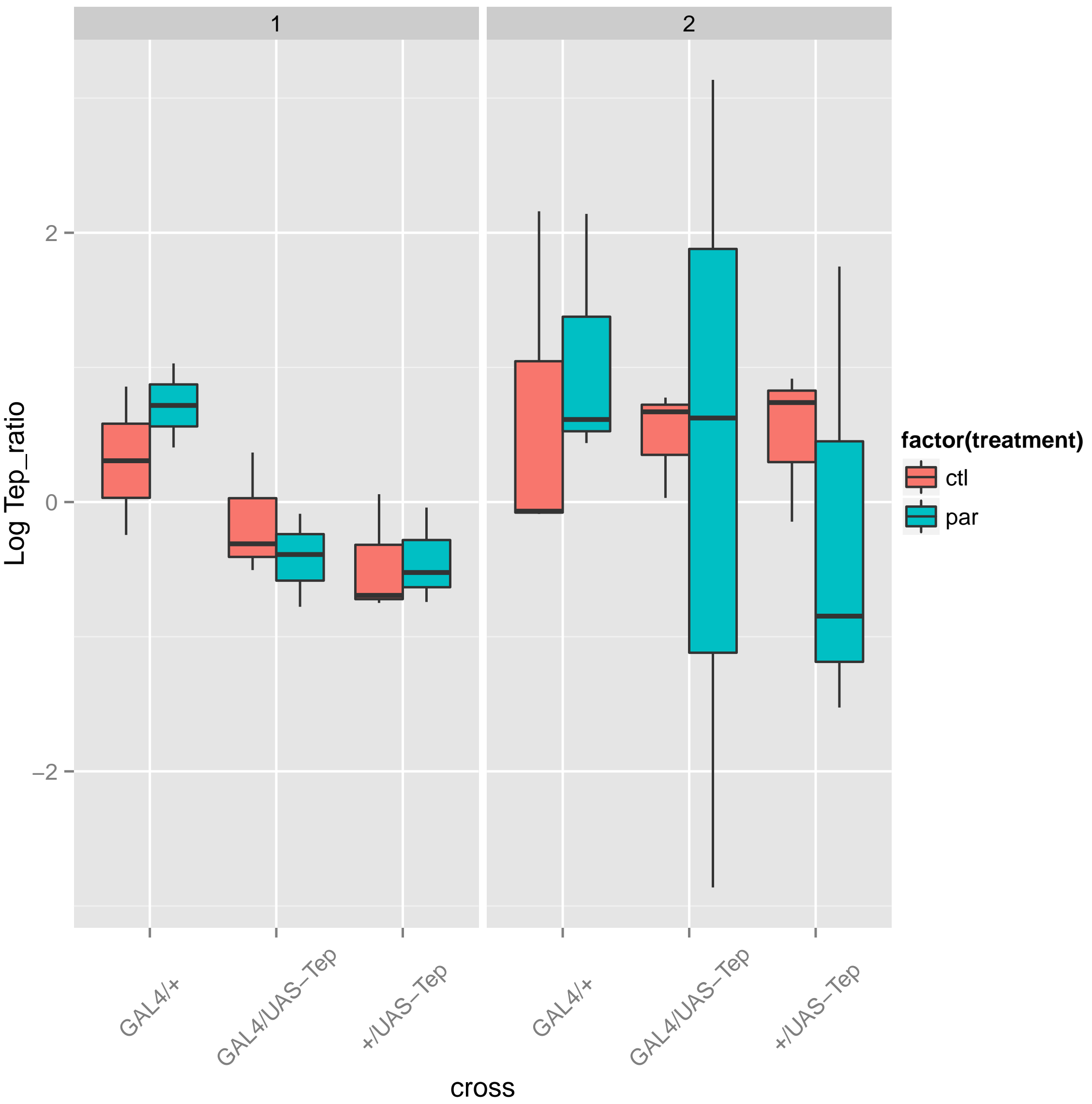


Figure 3.4: **Expression of *Tep1* in RNAi experiment** Log-transformed normalized expression of *Tep1* in experiment 1 and 2 (respectively, panels 1 and 2). The x-axis contains the crosses (the knock-down cross in the middle of two control crosses)

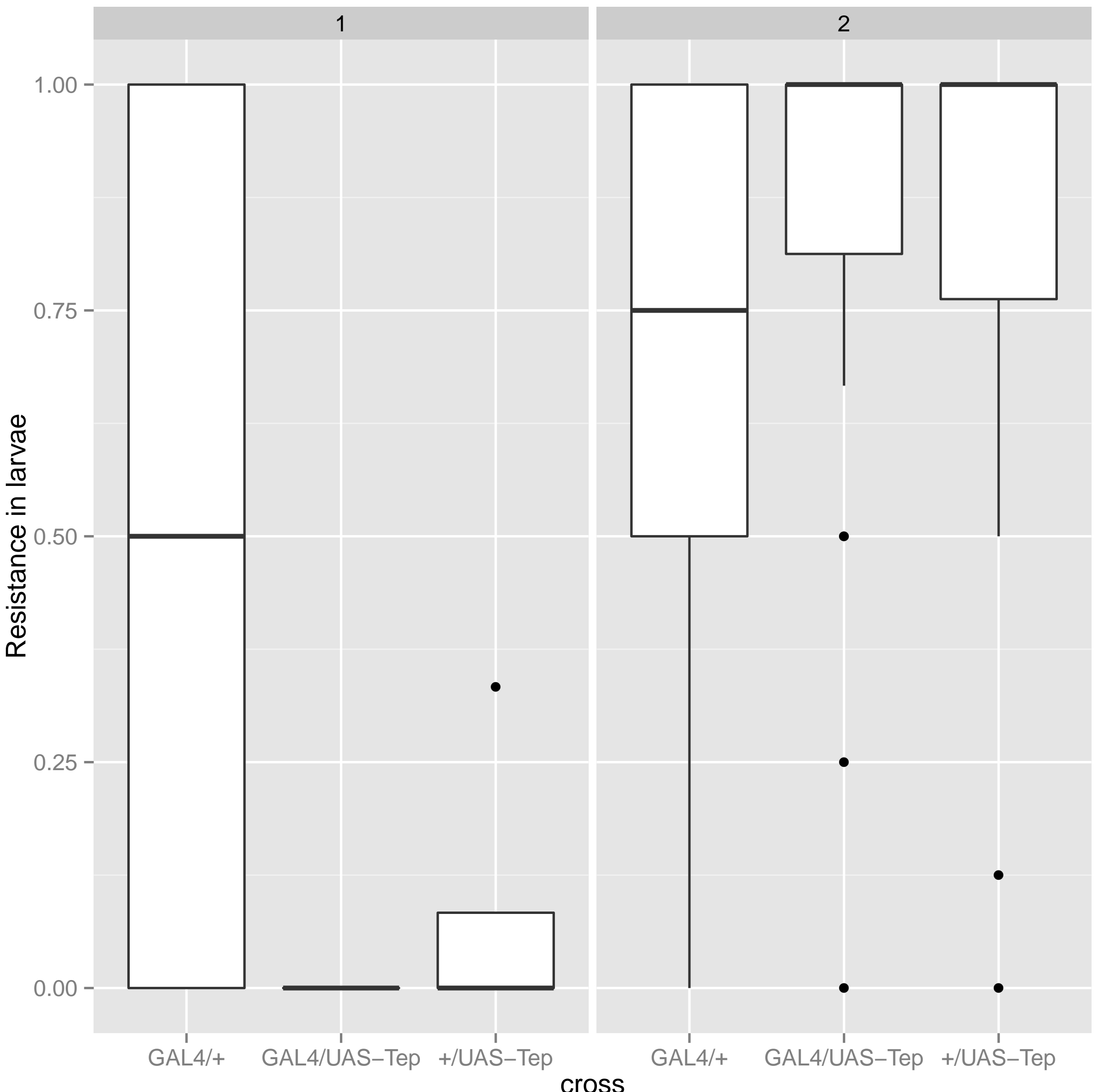


Figure 3.5: **Encapsulation ability in RNAi experiment.** Resistance estimated from dissections of larvae in experiment 1 and 2 (respectively, panels 1 and 2). The x-axis contains the crosses (the knock-down cross in the middle of two control crosses)

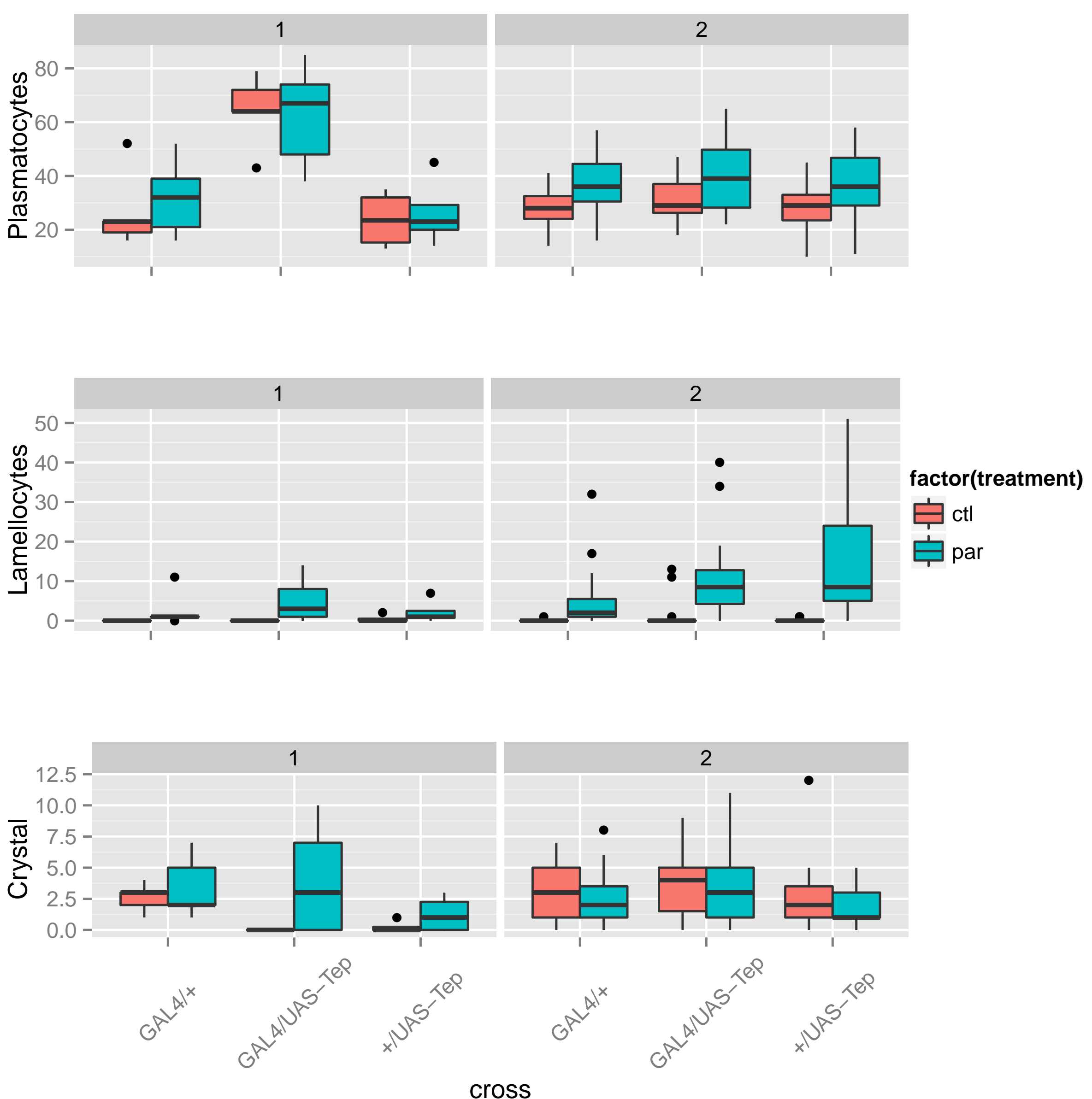


Figure 3.6: **Hemocyte counts in RNAi experiment.** Counts for each of three types of hemocytes in experiment 1 and 2 (respectively, panels 1 and 2) in parasitized and control larvae. The x-axis contains the crosses (the knock-down cross in the middle of two control crosses)

Table S3.4: **Statistics on the counts of the three types of hemocytes for the RNAi experiment.** Two-way ANOVA on square-root transformed hemocyte counts tested among three crosses and two treatments (parasitized and control) for two RNAi experiments.

	df	Sum Sq	Mean Sq	F value	Pr(>F)
Plasmatocytes					
A) experiment1					
treatment	1	26.04	26.04	0.12	0.7316
cross	2	8418.75	4209.38	19.52	0.0000
treatment:cross	2	1.19	0.59	0.57	0.5655
Residuals	22	4745.10	215.69		
B) experiment2					
treatment	1	2780.89	2780.89	28.30	0.0000
cross	2	281.56	140.78	1.43	0.2420
treatment:cross	2	56.55	28.27	0.29	0.7504
Residuals	150	14741.22	98.27		
Lamellocytes					
A) experiment1					
treatment	1	78.89	78.89	6.85	0.0157
cross	2	9.43	4.72	0.41	0.6689
treatment:cross	2	14.43	7.22	0.63	0.5437
Residuals	22	253.35	11.52		
B) experiment2					
treatment	1	3666.03	3666.03	57.23	0.0000
cross	2	741.82	370.91	5.79	0.0038
treatment:cross	2	637.12	318.56	4.97	0.0081
Residuals	151	9672.26	64.05		
Crystal Cells					
A) experiment1					
treatment	1	28.00	28.00	5.31	0.0310
cross	2	22.50	11.25	2.14	0.1420
treatment:cross	2	15.60	7.80	1.48	0.2493
Residuals	22	115.90	5.27		
B) experiment2					
treatment	1	8.31	8.31	1.69	0.1952
cross	2	52.61	26.30	5.36	0.0056
treatment:cross	2	9.02	4.51	0.92	0.4011
Residuals	151	740.71	4.91		

Table S3.5: **ANOVA (three-way) of normalized expression of *Tep1* for field lines.** Differences in the log-transformed ratio of *Tep1* and the endogenous control were tested among 4 time points, 4 field lines (Got, Bay, Kal, Sta) and two treatments (parasitized and control).

Factor	df	SumSq	MeanSq	F-value	Pr(>F)
time	3	23.10	7.70	4.024	0.0112*
treat	1	161.77	161.77	84.557	3.99e-13 ***
pop	3	12.58	4.19	2.191	0.0982 .
time:treat	3	21.13	7.04	3.681	0.0167*
time:pop	9	22.79	2.53	1.323	0.2439
treat:pop	3	5.69	1.90	0.992	0.4027
time:treatment:pop	9	8.08	0.90	0.469	0.8897
Residuals	61	116.71	1.91		

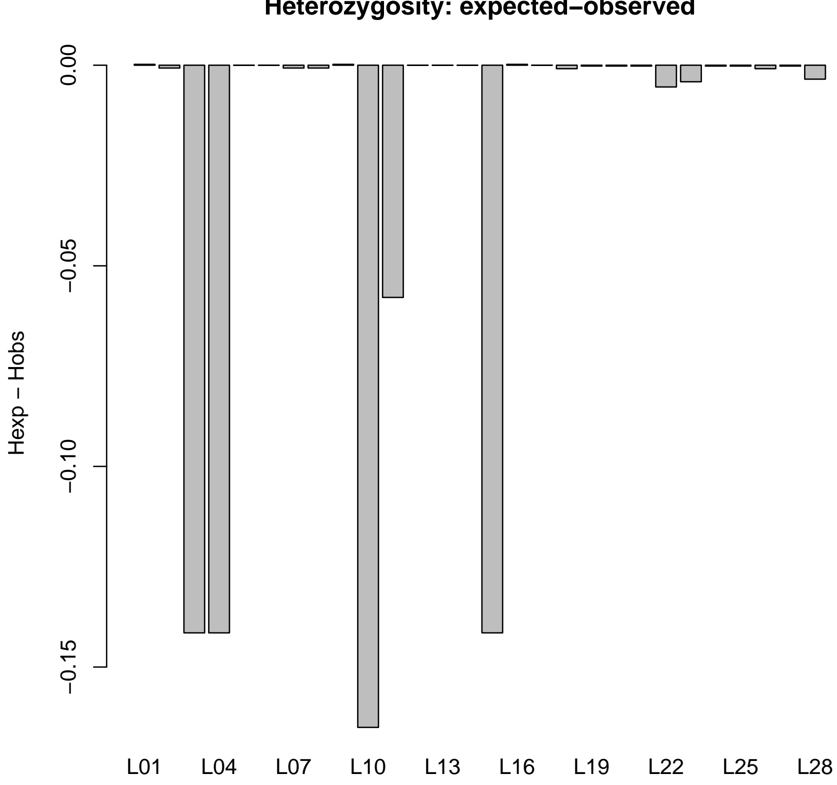


Figure S3.1: **Excess of heterozygosity in *Tep1*.** For each SNP position, the expected heterozygosity was calculated and compared to the observed heterozygosity

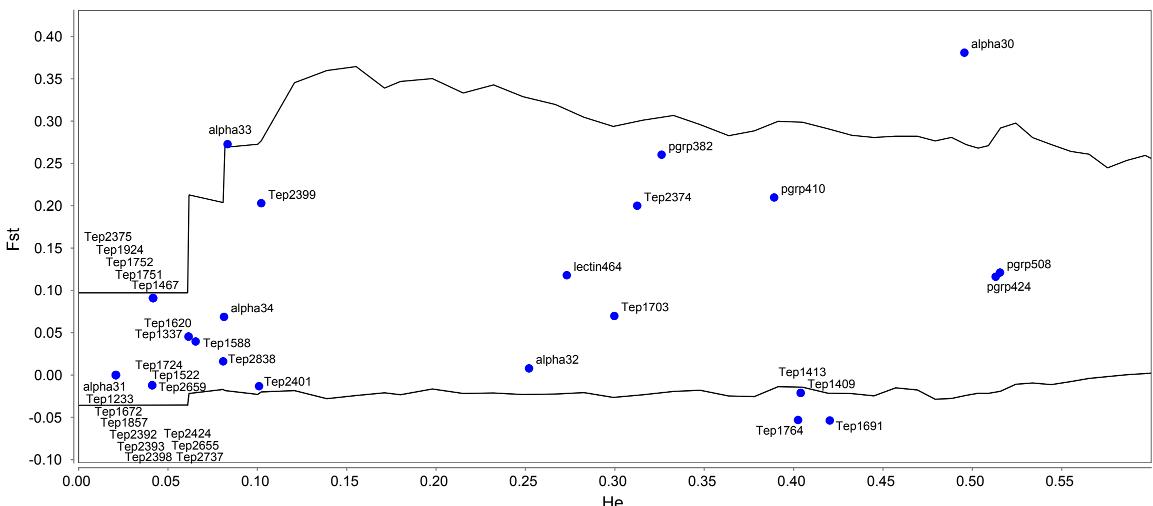


Figure S3.2: **F$_{ST}$ outlier analysis of deviation from neutrality in SNPs of all genes**. SNPs for four receptors (one gene did not show any polymorphisms) are labeled with the receptor name and the SNP site. All SNPs located between the 95% confidence interval lines do not show evidence for any type of selection. Although one SNP from *αPS4* (“alpha30”) is located above the top line, the probability to be under positive selection is below 0.995

Inter- and intra-species variation in the genome-wide gene expression of *Drosophila* in response to parasitoid wasp attack

Laura Salazar Jaramillo, Kirsten Jalvingh, Ammerins de Haan, Ken Kraaijeveld, Henk Buermans and Bregje Wertheim

Abstract

Parasitoid resistance in *Drosophila* differs considerably, both among species and within species. Among species, resistance evolved in a sublineage of *Drosophila* and was lost in at least one species within this lineage. Within *D. melanogaster*, resistance differs among geographic populations and it can be experimentally selected to higher levels. While the mechanisms of resistance are fairly well documented in *D. melanogaster*, much less is known for closely related species, which share similarities but also differences in their immune response. Here, we studied the inter- and intra-species variation in gene expression after parasitoid attack in *Drosophila*. We used RNAseq after parasitization of four closely related *Drosophila* species of the melanogaster subgroup and lines of *D. melanogaster* experimentally selected for increased resistance. We found a set of genes that is consistently up-regulated after parasitoid attack in the species and lines tested. These genes are largely lineage-restricted to the melanogaster subgroup, supporting the hypothesis that the ability to immunologically resist against parasitoid attack relies largely on new genes. Some of these genes show no up-regulation or expression in the species unable to raise an immune response against parasitoids. At the intra-species level, selection for increased resistance showed a constitutive increase in expression of a lamellocyte-specific gene in selected lines. The remainder of the differentially expressed genes were down-regulated in selection lines compared to non-selected lines in the samples that were not parasitized (constitutive difference in expression). The lack of constitutive induction of immune or hemopoiesis genes in selected lines suggests that modulation (indirect regulation) of the response may be more important than pre-activation of immune pathways. Among artificially selected lines, we found differential exon usage, indicating that selection acted on the expression of transcript isoforms.

4.1 Introduction

Comparative biology has contributed enormously to the understanding of evolution (Sanford *et al.*, 2002). In the last decade the comparative approach has extended to genomes, enabling us to study how phenotypic diversity is encoded in the genome. Although the relationship between genotype and phenotype is highly complex, the current Next Generation Sequencing (NGS) technology has made a great contribution in unraveling the genotype-phenotype map. It allows for the inspection of whole genomes and transcriptomes in a relatively unbiased way, and it enables investigations beyond model species. We can now extend the comparison of traits that have been long studied in one model species to closely related species, and to lines experimentally selected for changes in traits, in order to better understand the evolution of that trait (Ekblom & Galindo, 2011).

One trait that shows remarkably fast evolution, and dramatic changes among species, is the resistance to parasites. The hosts' ability to defend against parasites has to continuously evolve and re-adjust to the co-evolving parasites (Lambrechts *et al.*, 2006). These hosts' defense mechanisms often consist of specific immune responses that effectuate the clearance of the parasite. When species migrate or colonize into different habitats or niches, they may encounter new or different parasites. When they start co-evolving with these local parasite communities, they may change the investment in immunity, or even acquire novel immunity traits in the arms' race with the parasites (Schmid-Hempel, 2005).

A type of parasite that imposes a very strong selection on its hosts is parasitoids. Parasitoids are insects that lay eggs on or in other arthropods (usually insects), and kill this host during the development of the parasitoid. Parasitoids are among the most abundant and species-rich arthropod groups, and exert a strong selection pressure on their hosts (Godfray, 1994). Insect species are often natural hosts to different parasitoid species and resistance against parasitoids has been described in several insect lineages (Pech & Strand, 1996). However, hosts show large variation in their ability to resist parasitoids, both among closely related species and among natural populations (Eslin & Prévost, 1998; Kraaijeveld & Godfray, 1999; Suma *et al.*, 2012; van Nouhuys *et al.*, 2012).

Resistance to parasitoid in terms of an immune response has been found in various *Drosophila* species, but not in all species and with pronounced differences among species (Havard *et al.*, 2009; Eslin *et al.*, 2009). We showed that a particular type of immune response, lamellocyte-mediated encapsulation, is restricted to a monophyletic group inside the melanogaster group, and has been secondarily lost in one species within this group, *D. sechellia* (Salazar-Jaramillo *et al.*, 2014). Also within a single species, *D. melanogaster*, resistance differs among populations from different geographic locations (Kraaijeveld & Godfray, 1999; Gerritsma *et al.*, 2013), and can be experimentally selected, resulting in a powerful increase in resistance within a few generations (Kraaijeveld & Godfray, 1997; Wertheim *et al.*, 2011; Jalvingh *et al.*, 2014).

The mechanism of this immunological resistance against parasitoids (for simplicity we refer here to it simply as resistance) has been well documented in *D. melanogaster*. It involves the recognition of the parasitoid egg by the host, a subsequent increase and differentiation of hemocytes (blood cells in invertebrates) that surround the egg, forming a multicellular capsule, which is melanized. The differentiation of hemocytes is a critical step in the process, and leads to three main types of cells: 1) lamellocytes, which are responsible for forming the capsule around the egg, and are usually not present in unparasitized larvae 2) plasmatocytes, which phagocytize small pathogens and can differentiate further into lamellocytes (Stofanko *et al.*, 2010) and 3) crystal cells, which contain the melanin that is deposited in the capsule (Lemaitre & Hoffman, 2007).

Variation in resistance among and within species of *Drosophila* is associated with variation in the numbers of hemocytes they possess and can induce after parasitization (Eslin & Prévost, 1998). Lamellocytes are only found in the group of species that is able to resist parasitoid attack, including *D. sechellia* (although at very low levels) (Salazar-Jaramillo *et al.*, 2014; Eslin & Prévost, 1998). Species with a high level of resistance typically maintain higher constitutive hemocyte loads, and/or can more strongly induce the production of hemocytes after parasitization (Eslin & Prévost, 1998). In contrast, in natural populations of *D. melanogaster*, both the constitutive hemocyte load and induced production of hemocyte types varies greatly, but this does not necessarily correlate with the level of resistance (Gerritsma *et al.*, 2013). In lines selected for higher resistance, however, the constitutive level of hemocytes did show a significant increase (Kraaijeveld *et al.*, 2001; Wertheim *et al.*, 2011; Jalvingh *et al.*, 2014).

Great progress has been made in understanding the genetic basis of parasitoid resistance in *D. melanogaster*. Microarray studies revealed approximately 150 genes that significantly changed in expression in *D. melanogaster* across different time points after attack by *Asobara tabida* (Wertheim *et al.*, 2005) and *Leptopilina* species (Schlenke *et al.*, 2007). As a complement to this induced response, a subsequent microarray study identified nearly 900 genes that changed in their constitutive expression after selecting for increased parasitoid resistance (Wertheim *et al.*, 2011). Interestingly, a very limited overlap (28 genes) was found between the differential expression of the evolved and induced response, implying that the evolved response to parasitoid attack did not consist of a simple pre-activation of the inducible response, but involved a different set of genes (Wertheim *et al.*, 2011). In this latter study, gene expression in the selected lines was not tested after parasitization leaving open the question whether artificial selection for increased resistance also affected gene expression during the inducible response. Whole-genome sequencing of experimentally selected lines for parasitoid resistance identified genome regions that showed a significant signature of selection (Jalvingh *et al.*, 2014). These selected regions were associated with 98 candidate genes, of which 5 genes were also found in the ∼150 that changed expression in response to parasitoid attack and 29 genes overlapped with the ∼900 genes that changed expression after selection for increased parasitoid resistance (Jalvingh *et al.*, 2014).

While information on the immune response to parasitoid attack has been accumulating for *D. melanogaster*, little is still known for closely related species, which show similar mechanisms for resistance against parasitoids. In a recent comparative genomics study on candidate genes for parasitoid response and hemocyte differentiation from *D. melanogaster* we revealed that hemopoiesis-associated genes are highly conserved and present across the *Drosophila* genus, independently of level of resistance. In contrast, 11 of the ∼150 genes that are differentially expressed during the response to parasitoids are novel genes, specific to the *Drosophila* sublineage capable of lamellocyte-mediated encapsulation (Salazar-Jaramillo *et al.*, 2014). These novel genes are predominantly expressed in hemocytes and several of these sequences show signatures of positive selection. Currently, we do not know the extent of the similarities and differences among closely related *Drosophila* species in the activation and expression levels of the induced genes. This includes the expression pattern in *D. sechellia*, the species that has secondarily lost the resistance trait.

In this study, we characterize the genome-wide gene expression during the immune response in *Drosophila* species and lines that vary in their ability to resist against parasitoid wasps. To study the inter- and intra-specific variation in their immune response, we used RNAseq of different *Drosophila* species and experimentally selected lines of *D. melanogaster*, for which the genomes are fully sequenced. In previous studies, we phenotypically characterized the immune responses against parasitoids for these lines and species (Jalvingh *et al.*, 2014; Salazar-Jaramillo *et al.*, 2014). Comparison between experimentally selected lines and non-selected lines allows

us to understand short-term evolutionary processes, which acts most likely through selection on standing variation. Comparison among species allows us to study long-term evolution, such as the acquisition and loss of a trait, which may involve major genome changes (e.g, gene duplication, loss and fast divergence of genes). Using transcriptome sequencing we aim to understand 1) the genes involved in the induced response in four closely related species, 2) the similarities and differences in the immune response compared to *D. melanogaster* and 3) the effect of short-term selection processes in the induced response against parasitoids. To address these questions we performed an RNAseq experiment to compare *D. melanogaster*, *D. simulans*, *D. sechellia* and *D. yakuba*. For *D. melanogaster* we used two selected lines for increased resistance to parasitoids and two non-selected (control) lines. We analysed gene expression in fly larvae after parasitoid attack and their respective controls at two different time points, five and fifty hours after parasitization, which reflect the start and end of the immune response (Figure 4.1).

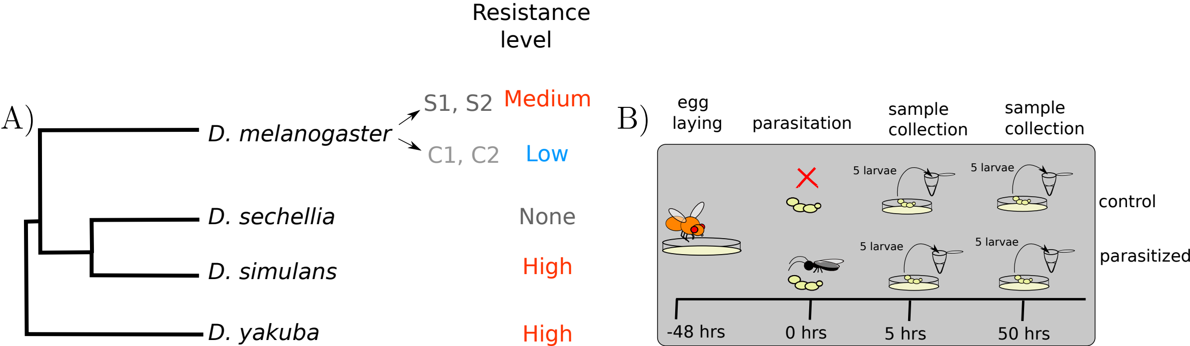


Figure 4.1: **RNAseq experimental design** A: species and lines used for RNAseq and their level of resistance against parasitoids and B: schematic representation of sample collection for RNAseq.

4.2 Materials and Methods

Insect strains For the comparison within a species, two lines of *D. melanogaster* selected for increased resistance and two control lines were used. These lines and the artificial selection procedure are fully described in Jalvingh *et al.* (2014). Briefly, the common base population for these lines was originally collected in Leiden, The Netherlands and had a low level of resistance. For each selection line the second-instar larvae were exposed to a moderately virulent strain of the parasitoid *A. tabida* for 24 hours. After pupation each individual was manually checked under a stereo-microscope and only those pupae that contained a visible capsule (sign of parasitization and resistance) and survived to adulthood were taken to the next generation. Selection was executed for five generations. Alongside each selection line, a matched control line (“non-selected line”) was cultured in parallel. Lines differed significantly in their resistance, with selected lines showing an average resistance of 50 % and non-selected lines 20 % at the end of the experimental selection procedure. The sampling for the RNAseq experiment was performed 31 generations after the discontinuation of the experimental selection procedure, while the selection lines were still significantly more resistant than the control lines (Jalvingh et al, in preparation).

For the comparison across species, genome project strains were used for *D. simulans*, *D. yakuba* and *D. sechellia* from the Drosophila Stock Center (San Diego University) (*Drosophila* 12 Genomes Consortium, 2007). The immune response of these strains against parasitoids is described in Salazar-Jaramillo *et al.* (2014).

For the parasitizations, the *A. tabida* parasitoid strain, "TMS", was used. This inbred line was established as an isofemale line in 2010 from a cross between two lines, one originally from Sospel (France) and one from Pisa (Italy). It has a moderate level of virulence.

Flies were reared at 20 ℃ under a dark:light regime of 12:12 and 50 % relative humidity in quarter-pint bottles containing 30 mL standard medium (26 g dried yeast, 54 g sugar, 17 g agar and 13 mL nipagine solution per litre). The parasitoid *A. tabida* has been maintained on *D. melanogaster* at 20 ℃ under a dark:light regime of 12:12.

Parasitization Fifty second-instar *Drosophila* larvae of each species were exposed to one *A. tabida* female. Larvae for which parasitoid oviposition was observed for at least 10 seconds were transferred to a new petri dish to allow development for a fixed period of time (5 hours or 50 hours) after which sampling took place. A control group was treated in the same way, except no wasp was introduced. Three biological replicates, each consisting of five larvae, were taken per species or line, treatment and time point. This resulted in a total of 84 samples for RNAseq analysis.

RNA extraction Larvae were collected, immediately snap frozen in liquid nitrogen and stored at -80 ℃. The RNA was extracted for pools of five larvae, using ZR Tissue & Insect RNA MicroPrep^{TM} kit (Zymo Research), according to the manufacturers' instructions. The RNA was eluted in RNase free water. Quality control was performed with Nanodrop of (260/280 and 260/230 close or above 2) and Bioanalyzer. Samples having a Bioanalyzer concentration of preferably $2\mu g$, but at least $1\mu g$ in $15\mu l$ and showing strong, distinct peaks corresponding to 18S and 28S rRNA were accepted for sequencing

Sequencing Sequencing was performed in June-August 2012 in the Leiden Genome Technology Center. Three pooled libraries were constructed, containing one biological replicate for each species, line, treatment (parasitized or control) and time point (5h or 50h after parasitization). Each sample in the libraries was individually barcoded (barcodes were randomly assigned). Strand specific RNAseq libraries were generated using the method described by (Parkhomchuk *et al.*, 2009) with minor modifications. In short, mRNA was isolated from 500 ng total RNA using oligo-dT Dynabeads (LifeTech 61002) and fragmented to 150-200 nt in first strand buffer for 3 minutes at 94 ℃. Random hexamer primed first strand was generated in presence of dATP, dGTP, dCTP and cTTP. Second strand was generated using dUTP instead of dTTP to tag the second strand. Subsequent steps to generate the sequencing libraries were performed with the NebNext kit for Illumina sequencing with minor modifications, i.e., after indexed adapter ligation to the dsDNA fragments, the library was treated with USER enzyme (NEB M5505L) in order to digest the second strand derived fragments. After amplification of the libraries, samples with unique sample indexes were pooled and paired-end 2x100 bp sequenced on 1 single HiSeq2000-v3 lane. Each pooled library was sequenced on 2 lanes, and a 7th lane was used to re-run some of the failed samples.

Analysis Quality control of the raw reads was performed with the Fastx toolkit (`http://hannonlab.cshl.edu/fastx_toolkit/`). Reads were trimmed based on a quality score of Phred Score smaller than 20 using FastQC 0.013 (`http://www.bioinformatics.babraham.`

ac.uk/projects/fastqc/). Filtered reads were mapped using Gsnap (parameters: -B 3 -t 6 -A sam) (Li *et al.*, 2009) to the respective reference genome downloaded from FlyBase (dmel5.51, dsimV2, dsec1.3 and dyak1.3). The alignments were run in the Millipede Cluster of Groningen University. Sam and bam files were manipulated with Samtools (Wu & Watanabe, 2005) and duplicated reads were removed with Picard Tools 1.79 (http://picard.sourceforge.net/). Counts were calculated with HTSeq-counts (Anders *et al.*, 2014) based on uniquely mapped and unambiguous reads only. Differential expression of the counts was analysed using the Bioconductor package EdgeR (Robinson *et al.*, 2010) and differential exon usage was analysed using the Bioconductor package DEXSeq (Anders *et al.*, 2012). Annotation files were modified to match the annotation file of *D. melanogaster*, using the orthology file from Flybase (*gene_orthologs_fb_*2013_06.*tsv*).

4.3 Results

The RNAseq experiment consisted of 84 samples from four species, (*D. melanogaster*, *D. simulans*, *D. sechellia* and *D. yakuba*). For *D. melanogaster* four lines were used, two selected for increased parasitoid resistance and two non-selected, and one line each for the other three species. Larvae were compared between parasitized and non-parasitized treatments at two time points, 5 and 50 hours (5h and 50h) after parasitization. Three biological replicates were obtained for each line, treatment and time point. These lines and species differed largely in the level of resistance to parasitoids (Table 4.1).

Table 4.1: **Summary of samples**. Four species of *Drosophila* were used in the RNAseq experiment. Species differ in the levels of resistance against parasitoid attack. For *D. melanogaster* four lines were used, two selected for increased resistance against parasitoids and two non-selected.

Species	**Line**	**Abbreviation**	**Resistance (%)**
D. melanogaster	non-selected (C1)	C1	20
	non-selected (C2)	C2	20
	selected (S1)	S1	55
	selected (S2)	S2	50
D. simulans		sim	80
D. sechellia		sec	0
D. yakuba		yak	90

We obtained good quality reads for 83 out of 84 samples (average phred score 34); for one *D. yakuba* sample, 50h after parasitization, sequencing failed. The number of aligned and mapped reads varied considerably among samples as shown in Figure S4.1. The values for the counts were higher for *D. melanogaster* than for the other species, which may reflect differences in annotation files used to guide the counts from the alignments. While for *D. melanogaster* an average of 50% of aligned reads could be mapped, for the remaining species this percentage was between 10% and 20%, with *D. simulans* showing the lowest values.

We implemented a set of statistical models to analyse the counts obtained for each gene, summarized in Table S4.1. For all models we corrected for batch effect, as we found covariance among samples within pooled libraries. We first tested the effect of parasitization for each species and time point independently, by defining a model with parasitization as explanatory

variable (Model PAR). The number of differentially expressed genes (DEG) was (at 5h and 50h, respectively): *D. melanogaster*(71;51), *D. simulans* (5;0), *D. sechellia* (9;4) and *D. yakuba* (2;12). It is not surprising that the DEG number was larger in *D. melanogaster*, as there were data for four lines (24 samples), while for the remaining species, there were only data for one line (6 samples). When mapped to their homologs in *D. melanogaster*, these sets of genes overlapped largely among the species. Only *D. yakuba* showed a majority (9 out of 14) of species-specific genes, whereas this fraction was smaller in *D. sechellia* (5 out of 13) and *D. simulans* (1 out of 5) (Table S4.3).

4.3.1 Inter-species comparison

In order to statistically compare species without introducing a bias due to differences in the number of lines, we used only one line,"C1", of *D. melanogaster*. Genes of all four species were annotated to their homologue in *D. melanogaster*. Statistical analysis was applied to 9592 genes that were present in all species. The significance was considered at $FDR < 0.05$. We tested the effect of parasitization (Model PAR) and the interaction of parasitization and species (PAR_SP). We did not include a model with only species as explanatory variable as variation among species was very large, resulting in DEG not necessarily related to the response against parasitoids. We also increased the values for the prior degrees of freedom of the tagwise normalization method, given the large biological variation found among species. This had an important effect on the number of DEG (Figure S4.2), and we chose a value of 20, which represents a slightly larger value than the residual degrees of freedom.

At 5 hours after parasitization model PAR and PAR_SP led to 17 and 32 significantly DEG, respectively, of which 15 genes overlapped between the two sets. A heatmap of the DEG (Figure 4.2) shows that control samples of *D. melanogaster* and *D. simulans* clustered together and parasitized samples of *D. melanogaster* and *D. simulans* clustered together. For the other two species, however, clustering was by species rather than parasitization (control and parasitized samples of *D. sechellia*, and control and parasitized samples of *D. yakuba*). This reflects a lack of, or lower, response to parasitoid attack in *D. sechellia* for several genes (e.g., *Tep1, IM23, CG15067, CG18557*), as well as substantial differences in the gene expression and response to parasitization in *D. yakuba* (e.g., *CG9989, Spn88Eb, CG4715, edin, IM4*). At 50 hours after parasitization model PAR and PAR_SP led to 20 and 51 DEG, respectively, of which 14 genes overlapped between the two sets. A heatmap of DEG at this time point (Figure 4.3) shows that for *D. simulans* and *D. yakuba* the clustering was by species.

In general, *D. melanogaster* and *D. simulans* showed a more similar response while that of *D. sechellia* and *D. yakuba* differed more. *D. sechellia* showed down-regulation (or no expression) in some of the immune genes, and in general very low levels of expression in all sets. *D. yakuba* showed low levels of expression compared to *D. melanogaster* and *D. simulans*, particularly at the 50h time point. During observations of the response in *D. yakuba* we found that this species is slighter faster in the encapsulation than the other species. Thus the different expression profile in *D. yakuba* may be related to a faster and therefore already completed immune response. Moreover, *D. yakuba* is phylogenetically more distant from *D. melanogaster* and *D. simulans*, which is reflected in both the different expression profile and the larger number of DEG that were species-specific.

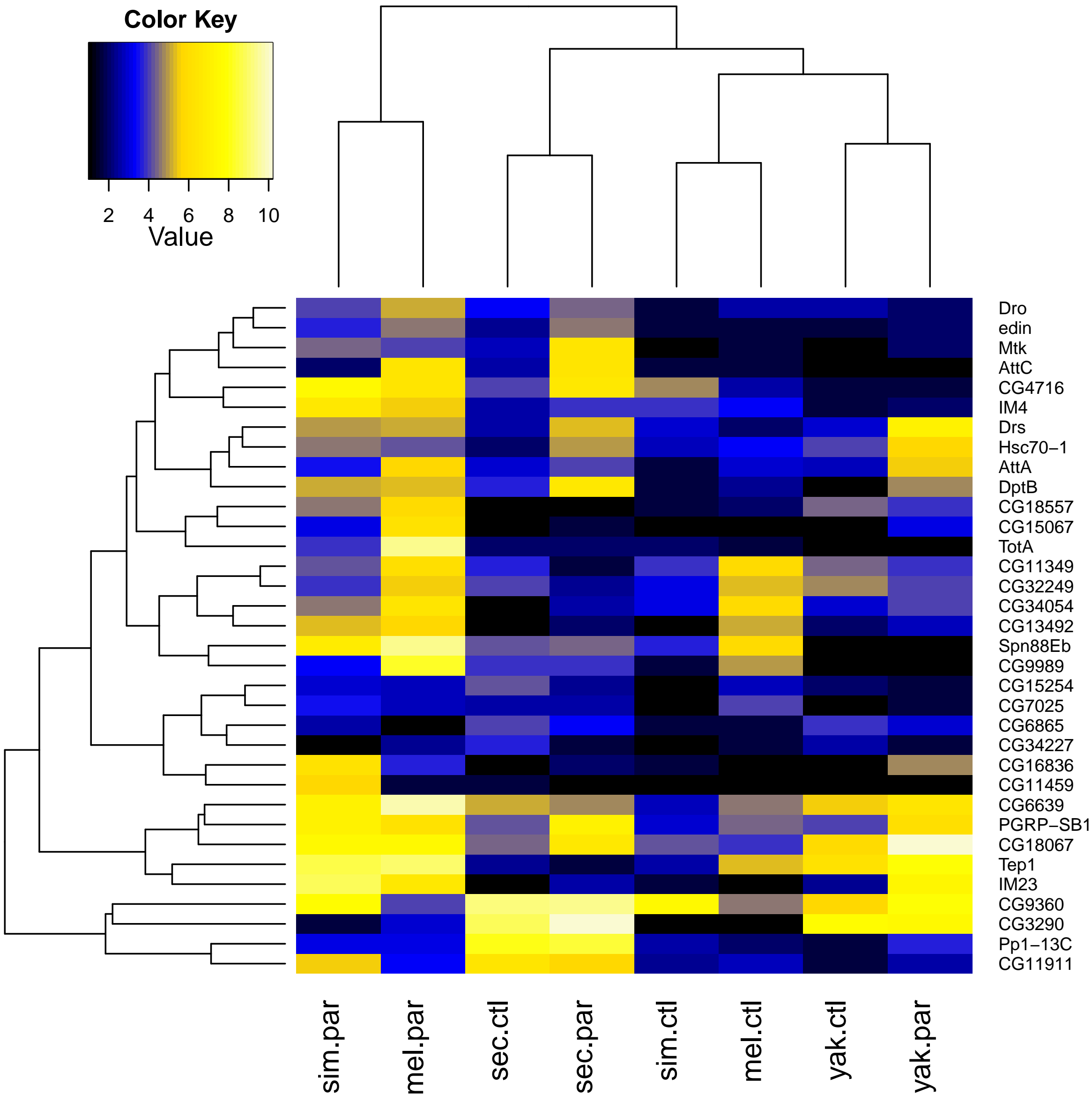


Figure 4.2: **Differentially expressed genes in all species 5h after parasitization** For each gene (rows), the median of the Log2(counts per million) of three replicates was obtained for each species and treatment (parasitized vs control) (columns).

In order to test the effect of analysing the response to parasitoids attack across species with very variable expression and response, we repeated the analysis but added one species at the time to the analysis starting from *D. melanogaster* (Figure S4.3). At the 5h time point, the addition of the *D. simulans* data to *D. melanogaster* data increased the number of DEG. This implies that the similarity in response of the two species increased the power of the statistical analysis, resulting in higher sensitivity to detect changes in gene expression after parasitoid attack. Adding the other two species, however, decreased the number of DEG, implying that the differences in the transcriptional regulation of the response to parasitoid attack among species were obscuring the similarities of *D. melanogaster* and *D. simulans*. At the 50h time point, the analysis of *D. melanogaster* only resulted in the largest set of DEG, and the intersection of DEG decreased with the addition of the other species.

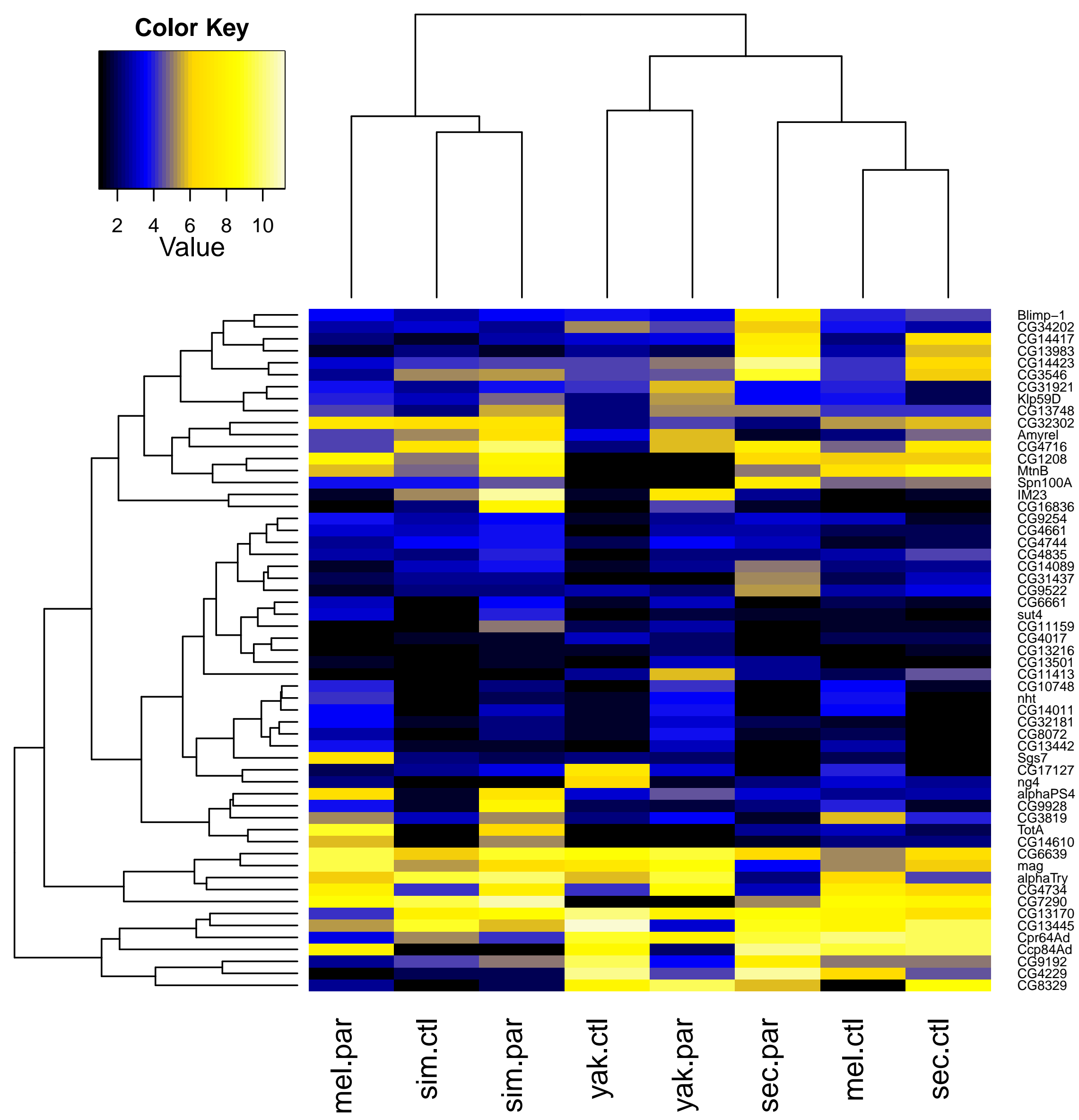


Figure 4.3: **Differentially expressed genes in all species 50h after parasitization** For each gene (rows), the median of the Log2(counts per million) of three replicates was obtained for each species and treatment (parasitized or control) (columns).

4.3.2 Intra-species comparison

The comparison across lines of *D. melanogaster* resulted in 71 and 51 DEG, at 5h and 50h respectively, in response to parasitization (model PAR in table S4.1). The four lines of *D. melanogaster* initiated the response in a similar way against the parasitoid attack (parasitized samples clustered in one set) (Figure 4.4). At 50h (Figure 4.5), the parasitized samples of the two selection lines clustered separately from the other (parasitized and control) samples. This seems to reflect a higher level of expression in several of the genes that are induced after parasitoid attack.

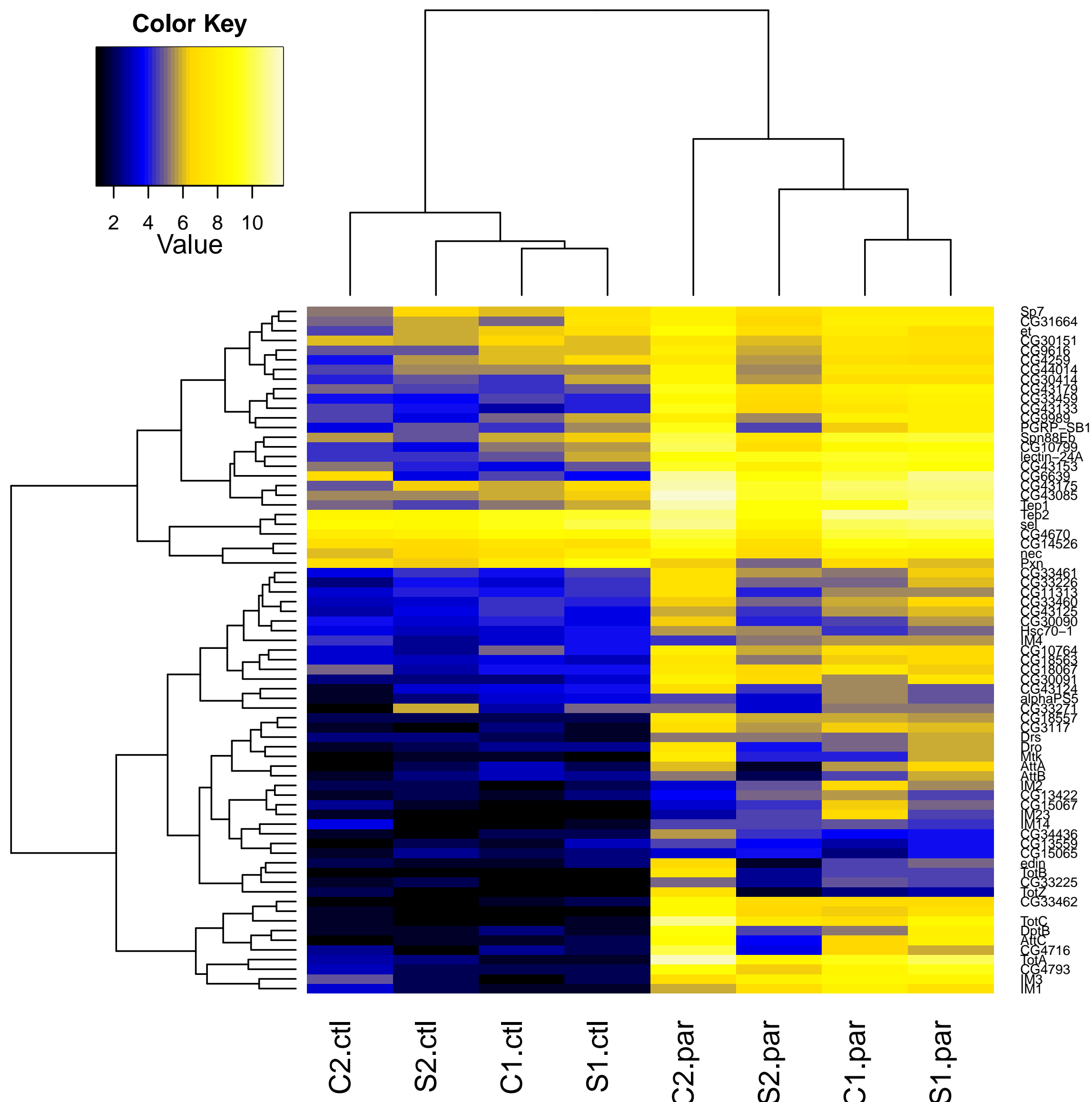


Figure 4.4: **Differentially expressed genes in *D. melanogaster* 5h after parasitization** For each gene (rows), the median of the Log2(counts per million) of three replicates was obtained for each line and treatment (parasitized vs control) (columns).

To better understand the effect of selection, we tested two models. One that considers constitutive differences in expression (before parasitization) (Model SEL), and another that considers the interaction between selection and parasitization (Model PAR_SEL). These two models were tested for each time point independently. At 5h, Model SEL and Model PAR_SEL did not show any DEG, whereas they did at 50h (Model SEL: 27 and Model PAR_SEL: 1). Interestingly, the DEG in Model SEL are mostly down-regulated in selected lines with respect to non-selected lines, except for *proPO59* (*PPO3*), which was up-regulated for the selected lines even before parasitization (Figure 4.6a). The only gene that was significant for the interaction (Model PAR_SEL) was *CG43666*. It only showed up-regulation in *D. melanogaster*

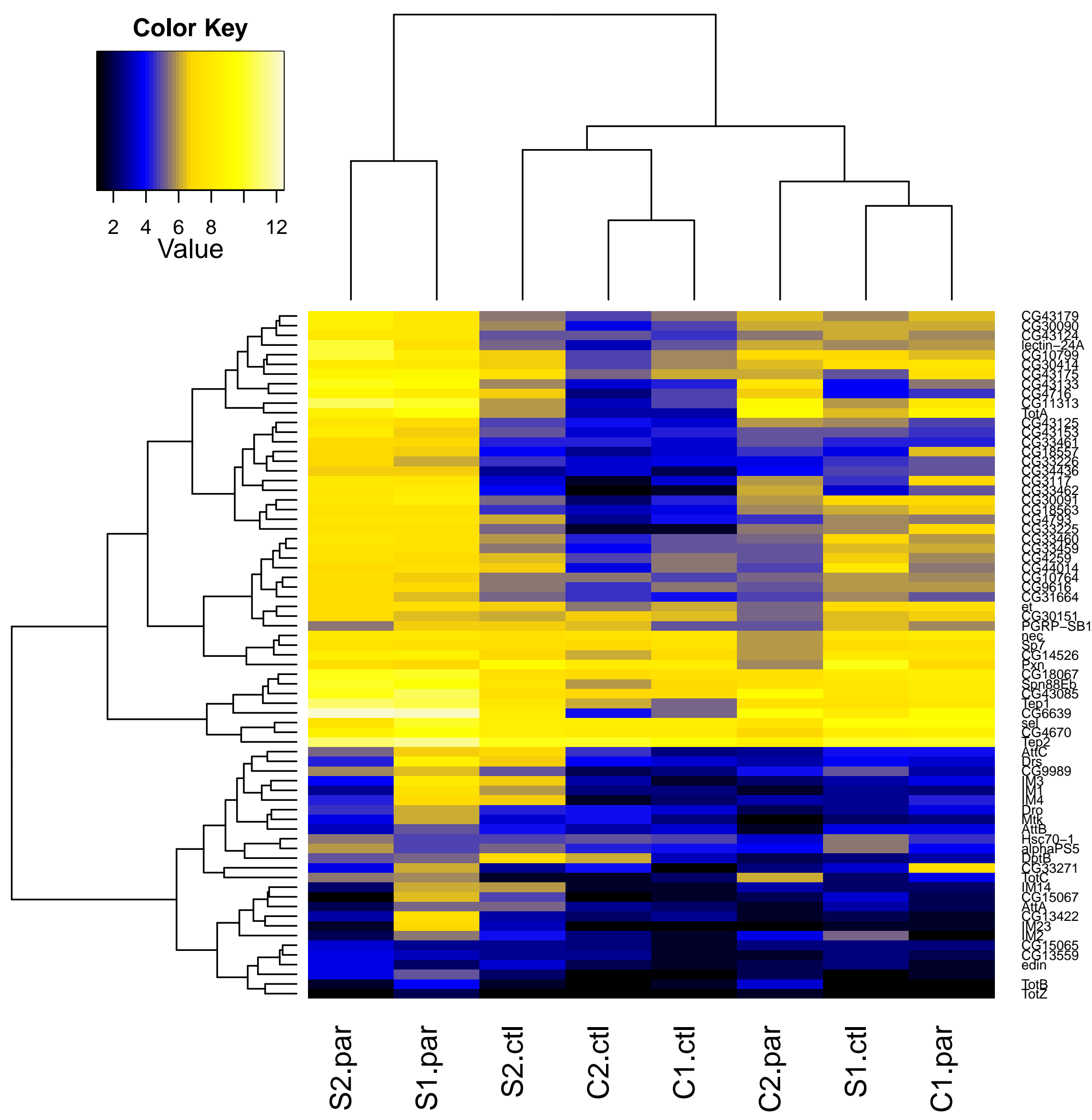


Figure 4.5: **Differentially expressed genes in *D. melanogaster* 50h after parasitization** For each gene (rows), the median of the Log2(counts per million) of three replicates was obtained for each line and treatment (parasitized vs control) (columns).

line “S2” (Figure 4.6b).

We also measured the differential exon usage, which reflects the expression of transcript isoforms (Anders *et al.*, 2012). Exon usage was compared between parasitized versus non-parasitized (Model EXON_PAR), selected versus non-selected (Model EXON_SEL) and the interaction between these two conditions (Model EXON_PAR_SEL) with both time points pooled. Model EXON_PAR led to four genes with significant differential exon usage: *CG15065,CG43133,Tep2* and *fok* (Figure 4.7). Model EXON_PAR_SEL led to 6 significant genes, of which 3 genes overlapped with Model EXON_PAR: *px, jp, alt, CG15065, CG43133,*

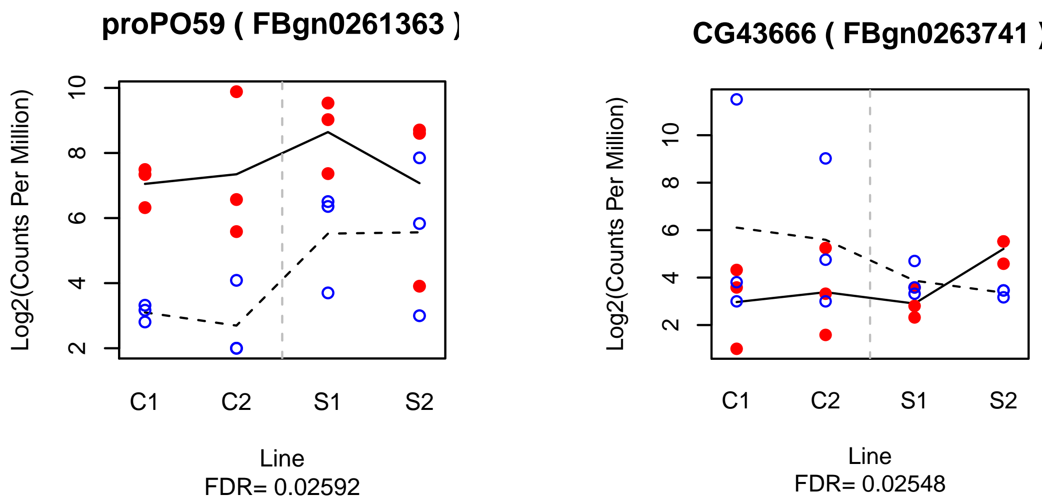


Figure 4.6: **Effect of selection in *D. melanogaster***. *PPO3 (proPO59)* was constitutively up-regulated in selected lines. *CG43666* was the only gene with a significant interaction between parasitization and selection. Red: parasitized values, blue: control values. Lines indicate the mean of controls (continous) and parasitized (dashed)

fok (Figure 4.7). Model EXON_SEL did not lead to any significant genes. These results showed that 6 genes have splicing variants that are expressed differentially in response to parasitoid attack in selected lines compared to the non-selected lines.

4.4 Discussion

We used RNAseq to compare the intra- and inter-species differences in genome-wide gene expression after parasitoid attack in four *Drosophila* species, *D. simulans*, *D. sechellia*, *D. yakuba* and *D. melanogaster*. For the latter species, we compared two lines selected for higher resistance to parasitoids with two non-selected lines. These species and lines differ greatly in their ability to immunologically resist parasitoid wasps, and we found that these differences were, to a large extent reflected in the expression profiles. Some genes showed a similar response in all four species. This was particularly the case 5 hours after parasitization. Another set of genes responded similarly among species able to mount the immune response, while they were not induced in the immunity-deficient *D. sechellia*. A different set of genes showed a species-specific response. This was particularly the case 50 hours after parasitization, when the response was more divergent across both species and lines.

We obtained in total 191 DEG from all statistical models and time points (Table S4.3). The majority of these genes had functional annotations within seven broad general categories, including various significantly over-represented annotations in this list of DEG (Table S4.2). These genes were clustered based on the mean fold changes in expression of parasitized versus control samples in Figure 4.8. This figure shows four general clusters of genes (G1-G4).

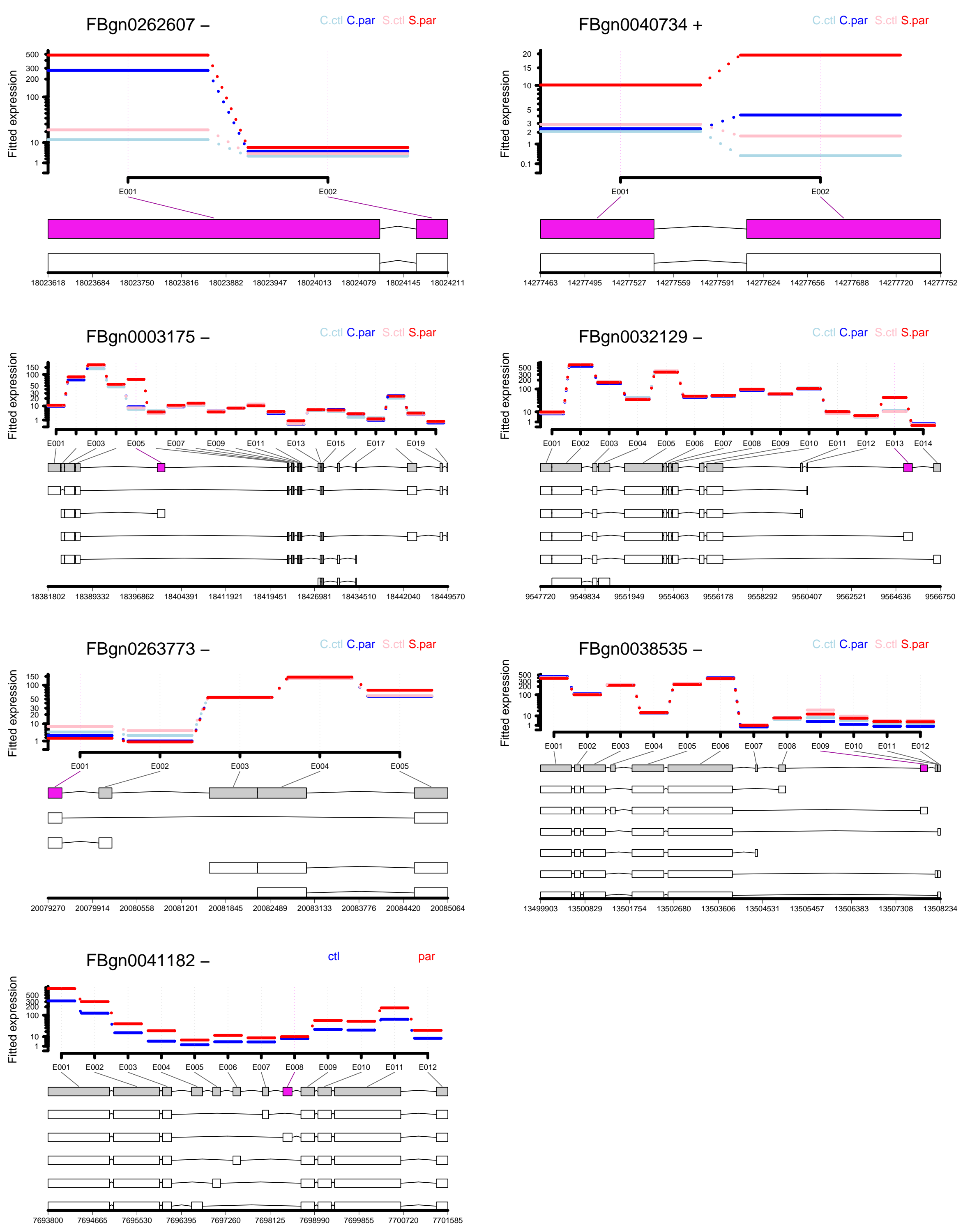


Figure 4.7: **Differential exon usage.** Differential expression of transcript isoforms was tested within *D. melanogaster*. Exons, for which significant differences are found, are coloured. Genes FBgn0262607 (*CG42133*), FBgn0040734 (CG15065), FBgn0263773 (*fok*), FBgn0003175 (*px*), FBgn0032129 (*jp*), FBgn0038535 (*alt*) showed significant differences in exon usage for treatment (parasitization, control) and lines (selected and non-selected lines). Gene FBgn0041182 (*Tep2*) was siginificant only when considering treatment

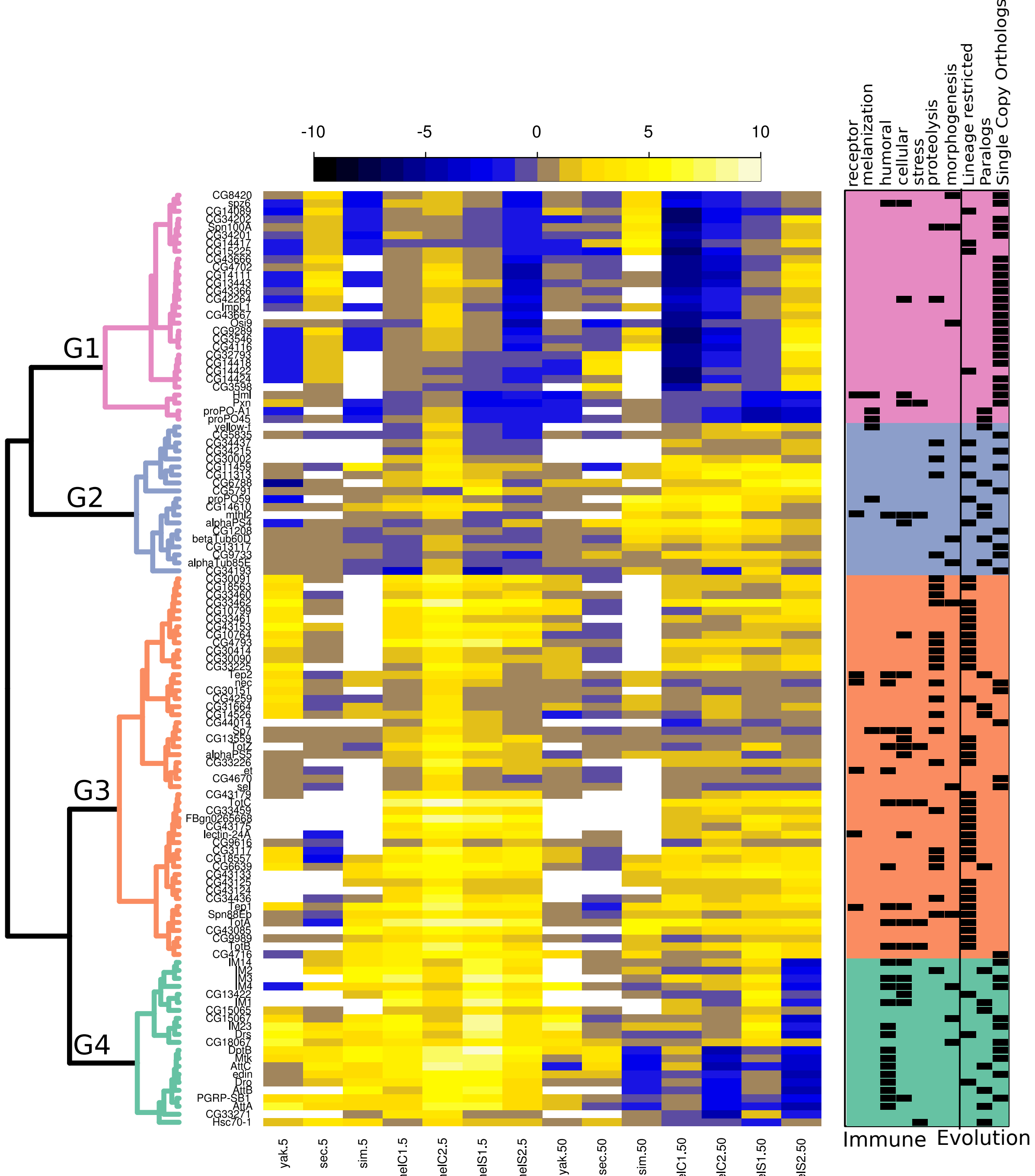


Figure 4.8: **Heatmap of fold changes in all DEG** Log2 of the average fold changes in mapped reads (counts per million) between parasitized and control (unparasitized) samples for all species, lines and time points. White cells in the heatmap indicate unknown values. The annotation matrix at the right is composed of two functional categories: immune and evolution (black black donote the categories that apply to the gene).

Cluster G1 is composed of genes with highly heterogeneous response, suggesting that these genes were differentially expressed in only a subset of the species and lines. Cluster

G2 is composed of genes that show a slight up-regulation at 5h time point and stronger up-regulation at 50h. The set of genes in cluster G2, therefore, seem to be especially important for the finalization of the response. Clusters G3 and G4 are composed of genes that show up-regulation in most species at 5h. Cluster G4, however, contains a set of genes that show down-regulation at 50h, while in cluster G3 most genes remain up-regulated at the later time point. Interestingly, *D. sechellia* showed very low levels of expression in most genes, particularly in cluster G3.

When focusing on the evolutionary relationship of genes, it seems evident that the set of genes in cluster G1 consists of poorly annotated genes, which are mostly single copy orthologs, that is, genes present in all species of *Drosophila* with one copy. In contrast, the set of genes that are up-regulated and immune annotated genes (especially genes of cluster G3) are largely lineage restricted genes. The early-up-regulated genes (cluster G4) are more often single copy orthologs and paralogs. The genes in cluster G4 have mostly annotations in humoral immunity, while the G3 cluster contains many genes with humoral and cellular immunity, stress and proteolysis annotations.

One question that arises is why so many up-regulated genes are lineage restricted? These are genes that have homologs only in a subset of *Drosophila* species, and are usually the product of gene duplication and fast divergence, presumably due to the acquisition of new functions. It was previously found in *Drosophila* that immunity relies largely on "new" genes (Singh *et al.*, 2009). More generally, it has been shown that in the *Drosophila* lineage the proportion of duplicated genes correlates with environmental heterogeneity (Makino & Kawata, 2012). Here we hypothesize that environmental heterogeneity would mean exposure to a greater array of new parasites and hence novel selection pressures on the immune system. This may have facilitated adaptation to new niches, where duplicated copies can rapidly acquire new functions, and thus diversify fast.

The expression profile of *D. sechellia* was generally very different to the other species. *D. sechellia* is a species of *Drosophila* restricted to the Seychelles islands in the Indean Ocean, where it specialized on the fruit *Morinda citriflora* (commonly known as noni). This fruit is toxic to *D. sechellia*'s sister species, *D. melanogaster* and *D. simulans*, and may also be to parasitoid wasps (Salazar-Jaramillo *et al.*, 2014; Legal *et al.*, 1994). Previously, we described large genomic changes in *Tep1* and *PPO3*, as well as low or no expression of these two genes after parasitoid attack in *D. sechellia*, and associated it to loss of function (Salazar-Jaramillo *et al.*, 2014). Here we confirm that *Tep1* is not up-regulated in *D. sechellia*, while it is highly up-regulated in all other species and lines. We also confirm the lack of expression of *PPO3* in *D. sechellia*, while it was expressed in all other species, although it did not show up-regulation in *D. yakuba.* This gene is exclusively expressed in lamellocytes (Irving *et al.*, 2005) and has been found to spontaneously activate melanization in the absence of proteolytic enzymes (Nam *et al.*, 2008), contrary to the other two pro-phenoloxidases expressed in the crystal cells. Lamellocytes are usually not present in unparasitized larvae and they differentiate upon parasitoid attack, suggesting that the regulation of *PPO3* is tightly associated to the differentiation of lamellocytes. Intriguingly, selected *D. melanogaster* lines showed constitutive higher expression (before parasitization) in *PPO3*, which could reflect the higher level of constitutive lamellocyte load found in selected lines (Jalvingh et al. submitted).

When comparing selected to non-selected lines of *D. melanogaster*, a considerable fraction of the differentially expressed genes are down-regulated in the control (unparasitized) treatment in the selected lines. Some of these genes are annotated with diverse developmental processes (here collectively referred to as "morphogenesis"). While some of them may have an immune function, others may reflect differences, for example, in developmental speed. Differences in developmental speed could also have led to the less homogeneous expression

pattern found at the later time point (50h) compared to the earlier time point (5h). At the 50h timepoint, the parasitized samples of the two selection lines clustered separately from the other samples. This pattern suggested a higher level of expression in approximately half of the genes that were induced after parasitization, although statistical analysis did not indicate any significant differences in the response to parasitoid attack between the selected and control lines. Some differences between selected and non-selected lines were not found in the level of expression but in the difference in transcript exon usage after parasitization. This was the case for seven genes, *CG15065, CG43133, Tep2, Fok, px, jp* and *alt*. For five of these genes splicing variants had been reported, but not in *CG15065*. Two of these genes (*Tep2* and *CG15065*) have immune annotation, while three (*px, jp* and *alt*) are associated with developmental processes, and the annotations for *fok* and *CG43133* are unknown.

The comparison between selected and non-selected lines confirms the previous finding that selection for higher resistance does not act on a pre-activation of the immune response (Wertheim *et al.*, 2011). Selection may target mainly developmental and cellular pathways, which have an impact on the coordination and speed of the response. In this study we were also able to report transcript variants that express differentially in selected lines compared to non-selected lines. Most of these genes were not significantly up- or down-regulated, meaning that selection may have acted on exon usage.

The technique used here allows us to only focus on mRNA, but it is likely that an important part of the selection process involves post-transcriptional regulation. Additionally we cannot identify tissue-specific responses as we used whole larvae, while important differences may be only perceptible at the level of tissues (Chintapalli *et al.*, 2007). Some technical aspects of the RNAseq procedure can have important implications in the analysis for this study. For example, certain steps in the process of mRNA-seq library preparation can dramatically impact the amount of reads obtained and consequently the gene coverage. As these aspects are external to the biological variation and can therefore affect the estimation of expression, we used a conservative approach, which discarded reads that were duplicated, non-uniquely mapped or ambiguous. This reduced extensively the amount of mapped reads used in the differential expression analysis. Additionally, the quality of the reference genomes and genome annotations for the species differed, and this may influence the amount of mapped reads.

Of the 191 genes found differentially expressed over all statistical designs, 15 were not differentially expressed in response to parasitoid attack in *D. melanogaster*, but in another species. Of these 15, the expression of one gene is species-specific to *D. simulans*, while the other 14 are either *D. sechellia* or *D. yakuba* species-specific. Of the remaining 176 genes that were differentially expressed, 51 overlapped with previous microarray studies on the constitutive or induced response of *D. melanogaster* against *A.tabida* (Wertheim *et al.*, 2005, 2011). Additionally, 8 genes overlapped with genomic regions with a significant signature of selection in the selected lines used in this study, as identified through DNA sequencing.

By comparing the inter- and intra-species variation in gene expression after parasitoid attack our study provides new insights into the evolution of the immune response against parasitoids. We showed that during the early stages after parasitization, a general and mostly humoral response occurs that is consistent across species, independent of their level of resistance. Additionally, we found genes that are up-regulated in the species able to resist against parasitoid attack while no up-regulation was found in the one species that is deficient for the response. Especially this set of genes may be of interest to understand the evolution of the response. At the intra-species level, we found constitutive expression of the lamellocyte-specific gene, *PPO3* in the selected lines. The remaining differentially expressed genes in selected lines compared to non-selected lines were down-regulated in the absence of parasitization and these were mostly involved in developmental pathways. The lack of constitutive induction of im-

mune or hemopoiesis genes may suggest that modulation of the response (indirect regulation via e.g., developmental pathways) is more important than pre-activation of immune pathways. We found differential exon usage, suggesting that selection may act on the expression of transcript isoforms. In conclusion, our results suggest that the immune response against parasitoids involves the expression of a core set of genes, which are largely lineage restricted to the melanogaster subgroup. This supports the hypothesis that the ability to encapsulate has evolved in a group of species within *Drosophila* (Salazar-Jaramillo *et al.*, 2014).

4.5 Acknowledgements

We would like to thank Bob Dröge and Ger Strikwerda for support and assistance with the Millipede Cluster of Groningen University. The Molecular Genetics group in Groningen University for facilitating the Bioanalyzer to measure the quality of the RNA. We-Juan Ma for providing the wasps, Daniel van der Post for help during sample collection and for comments on the manuscript. Irina Pulyakina, Paris Veltsos and Mike Ritchie for useful discussion on RNAseq and Leo Beukeboom for commenting on the manuscript.

4.6 Supplementary Material Chapter 4

Table S4.1: **Statistical models used in the analysis**.

Model	Description	Data	Package
PAR	DEG: effect of parasitization	- intra species (each sp) - inter-species (all spp, one line per spp)	edgeR
PAR_SP	DEG: effect of interaction between parasitization and species	inter species (all spp)	edgeR
SEL	DEG: effect of selection	*D. melanogaster*	edgeR
PAR_SEL	DEG: effect of interaction parasitization and selection	*D. melanogaster*	edgeR
EXON_PAR	DEU: effect of parasitization	*D. melanogaster*	DexSeq
EXON_SEL	DEU: effecto of to selection	*D. melanogaster*	DexSeq
EXON_PAR_SEL	DEU: effect of interaction between parasitization and selection	*D. melanogaster*	DexSeq

DEG: differentially expressed genes, DEU: differential exon usage

Table S4.2: **Functional categories**. Seven general categories were compiled from Gene Ontology subcategories. Asterisks indicate significant overrepresentation of that category according to the analysis in DAVID (Huang *et al.*, 2009)

Functional category	Annotation
Humoral	defense response bacterium*
	defense response to fungus*
Cellular	phagocytosis
	cell adhesion
	regulation hemocyte proliferation
	hemolymph coagulation
	transmembrane transport
Stress	response to heat*
	response to water deprivation*
	response to UV*
	protein folding
	response to oxidative stress
Receptor	plasma membrane*
	glycoprotein*
	transmembrane*
Melanization	dopamine metabolism*
	melanin biosynthesis*
	pigmentation*
Proteolysis	regulation proteolysis*
	regulation endopeptidases*
Morphogenesis	chaeta development
	imaginal disc-derived
	multicellular organism reproduction
	axonogenesis
	mitotic spindle assembly
	centrosome duplication
	lateral inhibition
	microtubule binding*

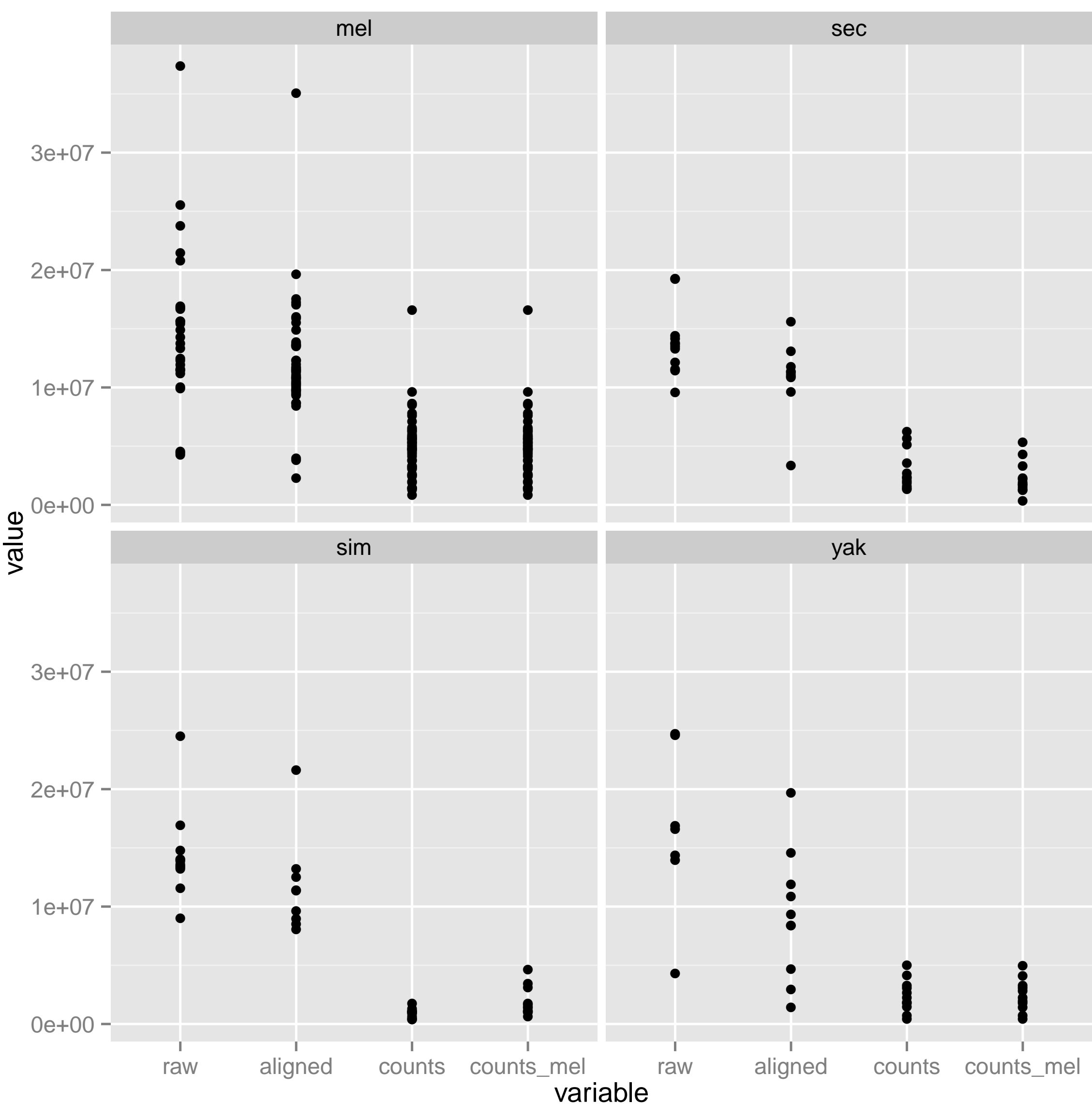


Figure S4.1: **Overview of reads and counts** Number of raw reads, aligned reads, reads mapped to the species' reference genome and reads mapped to *D. melanogaster* reference genome

Table S4.3: **Differentially expressed genes**

FB	Symbol(mel)	5h	50h	Model	OtherStudies
FBgn0011828	Pxn	mel	mel	PAR	Microarray induced
FBgn0033367	proPO45		mel	PAR	
FBgn0085231	CG34202		mel	SEL	
FBgn0014865	Mtk	sec,mel,spp	sec,mel	PAR+PAR_SP	Microarray induced
FBgn0002930	nec		mel	PAR	Microarray induced
FBgn0259149	CG42264		mel	SEL	
FBgn0052185	edin	mel,spp		PAR+PAR_SP	
FBgn0034331	CG15067	mel,spp		PAR+PAR_SP	
FBgn0029645	CG14422		mel	SEL	
10FBgn0261363	proPO59		mel	PAR+SEL	Microarray induced
FBgn0085230	CG34201		mel	SEL	
FBgn0031055	et	mel		PAR	Microarray induced
FBgn0029167	Hml		mel	PAR	Microarray induced+selection
FBgn0262531	CG43085	mel, sim	mel	PAR	
FBgn0038299	Spn88Eb	mel,spp	mel	PAR	Microarray induced
FBgn0030880	CG6788		intra mel	PAR	
FBgn0050091	CG30091	mel	mel	PAR	
FBgn0039759	CG9733		intra mel	PAR	Microarray induced
FBgn0036861	CG14089		intra mel	SEL	
20FBgn0052793	CG32793		intra mel	SEL	
FBgn0261362	proPO-A1		mel	PAR	
FBgn0029716	CG3546		intra mel	SEL	
FBgn0039795	Spn100A		intra mel	SEL	
FBgn0085466	CG34437		intra mel	SEL	
FBgn0053459	CG33459	mel	mel	PAR	
FBgn0053461	CG33461	mel	mel	PAR	
FBgn0028396	TotA	mel,spp	mel,spp	PAR	Microarray induced
FBgn0035623	mthl2		mel	PAR	Microarray induced
FBgn0034407	DptB	mel,sec,spp		PAR+PAR_SP	
30FBgn0031389	CG4259	mel		PAR	Microarray induced+selection
FBgn0041183	Tep1	mel,spp	mel	PAR+PAR_SP	Microarray induced
FBgn0050090	CG30090	mel	mel	PAR	Microarray induced+selection
FBgn0034551	CG15225		intra mel	SEL	DNAseq(0.05)
FBgn0037416	Osi9		intra mel	SEL	
FBgn0262607	CG43133	mel	mel	PAR	
FBgn0040653	IM4	mel		PAR	Microarray induced
FBgn0037664	CG8420		mel	SEL	
FBgn0001256	ImpL1		mel	SEL	
FBgn0262588	CG43125	mel	mel	PAR	
40FBgn0262808	CG43179	mel	mel	PAR	
FBgn0025644	CG14424		mel	SEL	
FBgn0034870	CG13559	mel		PAR	
FBgn0041182	Tep2	mel		PAR	Microarray induced
FBgn0043578	PGRP-SB1	mel, sec, sim, spp		PAR+PAR_SP	Microarray induced
FBgn0040734	CG15065	mel		PAR	Microarray induced+selection
FBgn0034329	IM1	mel		PAR	Microarray induced
FBgn0085244	CG34215		mel	PAR	
FBgn0041710	yellow-f		mel	PAR	Microarray induced
FBgn0262587	CG43124	mel	mel	PAR	
50FBgn0262794	CG43175	mel	mel	PAR	
FBgn0003888	betaTub60D		mel	PAR	DNAseq(0.05)
FBgn0050151	CG30151	mel		PAR	
FBgn0001216	Hsc70-1	mel,sim,sec,spp		PAR+PAR_SP	DNAseq(0.05)
FBgn0041581	AttB	mel		PAR	Microarray induced
FBgn0025645	CG3598		mel	SEL	
FBgn0069056	CG33226	mel	mel	PAR	
FBgn0040104	lectin-24A	mel	mel	PAR	Microarray induced+selection
FBgn0034328	IM23	mel,spp	spp	PAR+PAR_SP	Microarray induced
FBgn0044812	TotC	mel	mel	PAR	Microarray induced
60FBgn0263260	sel	mel		PAR	
FBgn0031470	CG18557	mel,spp	mel	PAR	
FBgn0034005	alphaPS4		mel,spp	PAR+PAR_SP	Microarray induced
FBgn0260474	CG30002		mel	PAR	
FBgn0053460	CG33460	mel		PAR	

FB	Symbol(mel)	5h	50h	Model	OtherStudies
FBgn0040736	IM3	mel		PAR	Microarray induced
FBgn0025583	IM2	mel		PAR	Microarray induced
FBgn0010388	Dro	mel,spp		PAR+PAR_SP	
FBgn0003132	Pp1-13C	spp		PAR	
FBgn0054054	CG34054	spp		PAR	DNAseq
70FBgn0053462	CG33462	mel	mel,yak	PAR	
FBgn0029649	CG4116		mel	SEL	DNAseq
FBgn0085222	CG34193		mel	PAR+SEL	DNAseq
FBgn0053225	CG33225	mel	mel	PAR	
FBgn0027578	CG14526	mel		PAR	
FBgn0037386	CG1208		mel,spp	PAR	Microarray selection
FBgn0037992	CG4702		mel	SEL	
FBgn0263109	CG43366		mel	SEL	
FBgn0264776	CG44014	mel		PAR	
FBgn0265668		mel	mel	PAR	
80FBgn0040354	CG14418		mel	SEL	
FBgn0032639	CG18563	mel	mel	PAR	Microarray induced
FBgn0263742	CG43667		mel	SEL	
FBgn0031471	CG3117	mel	mel	PAR	Microarray induced
FBgn0035056	spz6		mel	SEL	
FBgn0040582	CG5791		mel	PAR	
FBgn0051664	CG31664	mel	mel	PAR	Microarray selection
FBgn0263741	CG43666		mel	SEL+PAR_SEL	
FBgn0012042	AttA	mel, yak		PAR+PAR_SP	Microarray induced
FBgn0039593	CG9989	mel,spp		PAR	Microarray selection
90FBgn0040353	CG14417		mel	SEL	
FBgn0085465	CG34436	mel	mel	PAR	
FBgn0032638	CG6639	mel,spp	mel,spp	PAR+PAR_SP	Microarray induced
FBgn0034548	CG13443	sec	mel	SEL	
FBgn0034512	CG18067	mel	mel, yak, spp	PAR+PAR_SP	DNAseq(0.05)
FBgn0037477	CG14610		mel	PAR	
FBgn0033821	CG10799	mel	mel	PAR	Microarray selection
FBgn0038214	CG9616	mel		PAR	
FBgn0067905	IM14	mel		PAR	
FBgn0032058	CG9289		mel	SEL	
100FBgn0050414	CG30414	mel	mel	PAR	Microarray induced
FBgn0028514	CG4793	mel	mel	PAR	Microarray induced+DNAseq
FBgn0037396	CG11459		mel	PAR	
FBgn0003886	alphaTub85E		mel	PAR	
FBgn0044809	TotZ	mel		PAR	
FBgn0037515	Sp7	mel		PAR	Microarray induced
FBgn0053271	CG33271	mel		PAR	
FBgn0036350	CG14111		mel	SEL	Microarray selection
FBgn0039798	CG11313	mel	mel	PAR	Microarray induced
FBgn0038838	TotB	mel		PAR	Microarray induced
110FBgn0033814	CG4670	mel		PAR	
FBgn0034221	CG10764	mel		PAR	DNAseq
FBgn0034511	CG13422	mel		PAR	Microarray induced
FBgn0034880	alphaPS5	mel		PAR	
FBgn0262683	CG43153	mel		PAR	
FBgn0032140	CG13117			PAR	
FBgn0038682	CG5835		mel	PAR	Microarray selection, DNAseq
FBgn0034662	CG13492	spp		PAR_SP	
FBgn0036817	CG6865	spp		PAR_SP	
FBgn0085256	CG34227	spp		PAR_SP	
120FBgn0031930	CG7025	spp		PAR	
FBgn0032472	CG9928		spp	PAR	Microarray selection
FBgn0033826	CG4734		spp	PAR_SP	
FBgn0035607	CG4835		spp	PAR_SP	Microarray selection, DNAseq
FBgn0034684	CG13501		spp	PAR	
FBgn0036532	CG13445		spp	PAR_SP	
FBgn0020506	Amyrel		spp	PAR_SP	
FBgn0052302	CG32302		spp	PAR	Microarray selection
FBgn0032143	CG4017		spp	PAR + PAR_SP	
FBgn0036070	CG8072		spp	PAR	

FB	Symbol(mel)	5h	50h	Model	OtherStudies
130FBgn0003863	alphaTry		spp	PAR_SP	
FBgn0036403	CG6661		spp	PAR_SP	
FBgn0030429	CG4661	spp	PAR		
FBgn0031722	CG14011		spp	PAR_SP	
FBgn0035513	Cpr64Ad		spp	PAR	
FBgn0036639	CG4229		spp	PAR	
FBgn0028560	sut4	spp	spp	PAR	
FBgn0035193	CG9192		spp	PAR	
FBgn0035022	sp	CG11413		PAR_SP	
FBgn0036996	mag		spp	PAR_SP	Microarray selection
140FBgn0052181	CG32181		spp	PAR	
FBgn0033703	CG13170		spp	PAR_SP	
FBgn0028513	CG32181		spp	PAR	
FBgn0032472	CG9928		spp	PAR+PAR_SP	
FBgn0033834	CG4744		spp	PAR	
FBgn0034827	Klp59D		spp	PAR	
FBgn0041103	nht		spp	PAR	
FBgn0051921	CG31921		spp	PAR	
FBgn0030332	CG9360	spp		PAR	
FBgn0041579	AttC	sec,spp		PAR	
150FBgn0040735	CG16836	spp	spp	PAR	
FBgn0037396	CG11459	spp		PAR	
FBgn0033820	CG4716	spp	spp	PAR	
FBgn0034711	CG3290	spp		PAR+PAR_SP	
FBgn0010381	Drs	sim,sec,spp	yak	PAR+PAR_SP	
FBgn0184771		sim		PAR	
FBgn0169233	Lsp1gama	sec		PAR	
FBgn0175569	Lcp1	sec		PAR	
FBgn0178907	Hsc70	sec		PAR	
FBgn0180858	Hsc70	sec		PAR	
160FBgn0180001	Hsp22	sec		PAR	
FBgn0236594		yak	yak	PAR	
FBgn0230583			yak	PAR	
FBgn0234953	CG2233		yak	PAR	
FBgn0232022	CG11413		yak	PAR	
FBgn0242665	CG11029		yak	PAR	
FBgn0239561	CG12971		yak	PAR	
FBgn0234115	CG43182		yak	PAR	
FBgn0237724	CG8543		yak	PAR	
169FBgn0238875	CG8541		yak	PAR	
FBgn0028949	CG15254	spp		PAR_SP	
FBgn0031249	CG11911	spp		PAR_SP	
FBgn0035550	CG11349	spp		PAR_SP	
FBgn0052249	CG32249	spp		PAR_SP	
FBgn0003377	Sgs7		spp	PAR_SP	
FBgn0031792	CG13983		spp	PAR_SP	
FBgn0002869	MtnB		spp	PAR_SP	Microarray selection
FBgn0036833	CG3819		spp	PAR_SP	
FBgn0004780	Ccp84Ad		spp	PAR_SP	
FBgn0010296	ng4		spp	PAR_SP	
FBgn0029646	CG14423		spp	PAR_SP	
FBgn0030587	CG9522		spp	PAR_SP	
FBgn0032299	CG17127		spp	PAR_SP	
FBgn0033355	CG13748		spp	PAR_SP	
FBgn0033591	CG13216		spp	PAR_SP	
FBgn0034539	CG11159		spp	PAR_SP	
FBgn0034546	CG13442		spp	PAR_SP	
FBgn0035625	Blimp-1		spp	PAR_SP	DNAseq
FBgn0036022	CG8329		spp	PAR_SP	
FBgn0036327	CG10748		spp	PAR_SP	
FBgn0036949	CG7290		spp	PAR_SP	
FBgn0051437	CG31437		spp	PAR_SP	

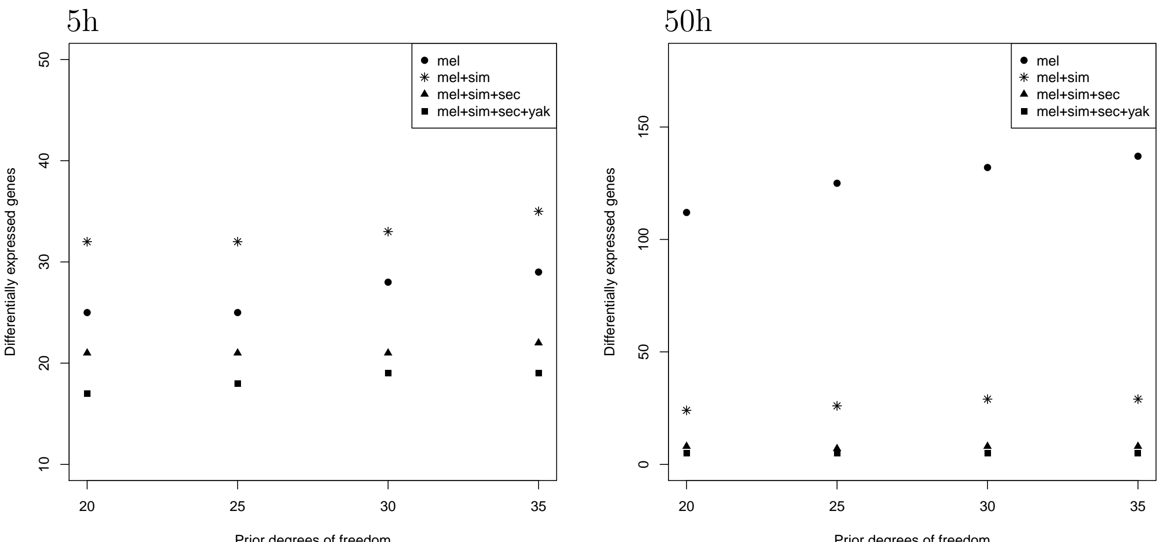


Figure S4.2: **Relationship between differentially expressed genes and prior degrees of freedom**

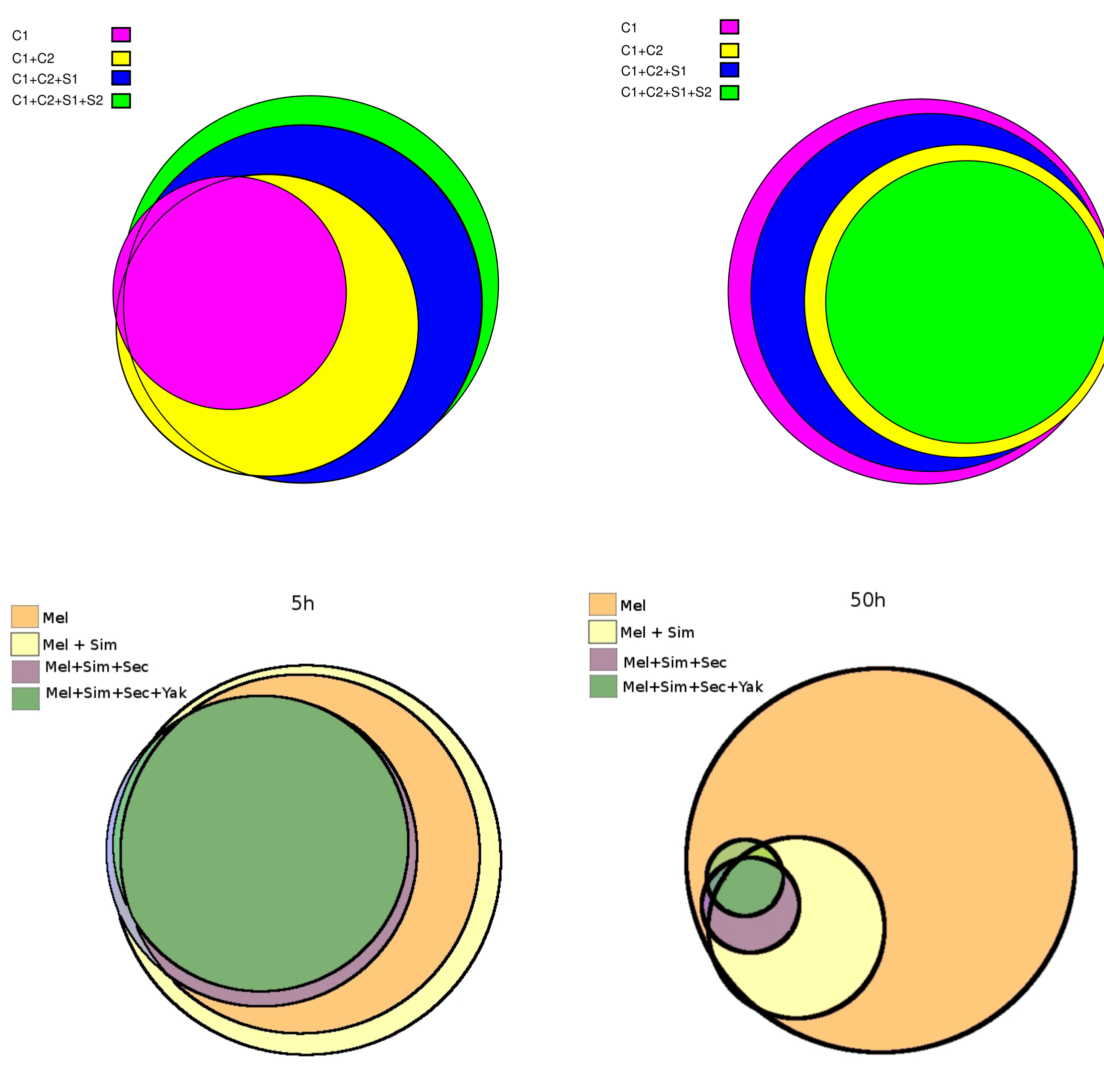
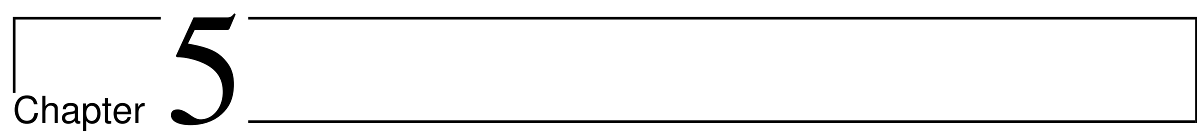


Figure S4.3: **Venn diagrams of DEG when adding one species at the time**



From genomes to natural history: does *Drosophila sechellia* escape parasitoid attack by feeding on a "toxic" resource?

Laura Salazar Jaramillo and Bregje Wertheim

Abstract

Drosophila sechellia is a species restricted to the Seychelles islands, where it specialized on the fruit *Morinda citriflora* (noni). This fruit is known to be toxic to *D. sechellia*'s sister species, *D. melanogaster* and *D. simulans*. *D. sechellia* has secondarily lost the ability to mount an immunological response against wasp attack, while its sister species can resist. We hypothesized that *D. sechellia*'s specialization on noni fruit may have protected it from parasitoid wasps, which could lead to relaxation in the selection pressure to maintain parasitoid resistance. In order to test this hypothesis, we performed a laboratory experiment using a commercial extract of noni fruit, and a field study of *Drosophila* species and parasitic wasps present in noni in the natural reserve of Cousin island in the Seychelles. The laboratory experiments indicated that larvae in substrate with noni extract were more likely to escape parasitism than in standard substrates. The field study showed that among *Drosophila* species, only *D. sechellia* developed on noni fruit throughout different stages of maturity. At later stages of decomposition, a slightly larger diversity of insects was attracted to noni, including at least one more species of *Drosophila* and parasitic wasps from the *Leptopilina* genus. Together these results suggest that the adaptation of *D. sechellia* to the ripe noni provided it of an environment with reduced interspecific competition and reduced risk of parasitoid attack.

5.1 Introduction

Ecological dynamics of host shifts can drastically change the selective pressure that animals experience from their environment. Evolving the ability to overcome plant toxins, for example, may yield access to new food sources, alter the competitive dynamics among species and/or provide an enemy-free space (Feder, 1995; Matzkin, 2012). Successful host shifts are sometimes followed by a rapid diversification process, and can therefore also promote biodiversity. For

all these reasons, host shifts are considered to be of major importance in the ecology and evolution of organisms (Nyman, 2010).

Changes in life history traits associated with host shifts include the recognition of specific stimulants (Caillaud & Via, 2000), mechanisms to cope with toxic chemical compounds (Matzkin, 2012) or shifts in the life cycle to coincide with the timing of host availability (Dambroski & Feder, 2007). When correlated traits include reproductive isolation, these changes could eventually lead to ecological speciation (Matsubayashi *et al.*, 2009). The study of the genomic changes associated with host shifts and their correlated traits may help us in revealing the mechanisms of adaptation and its relationship to ecological divergence.

The sequencing of *Drosophila* specialist species, particularly *D. sechellia*, made a great contribution to the study of molecular and genomic consequences of host shifts (Whiteman & Pierce, 2008). *D. sechellia* is a species of *Drosophila* restricted to the Seychelles islands in the Indian Ocean, where it specialized on the fruit *Morinda citriflora* (commonly known as noni) (Louis & David, 1986; Gerlach, 2009). This fruit is known to be toxic to *D. sechellia*'s sister species, *D. melanogaster* and *D. simulans*, which have also been reported from the Seychelles.

In recent studies, we found that *D. sechellia* has lost the ability to resist parasitoid wasp attack, while the four other species of the clade where *D. sechellia* belongs to can resist the parasitoids through innate immune defenses (Salazar-Jaramillo *et al.*, 2014). Using comparative genomics, we also found large DNA sequence changes in some immunity genes of this species that would be consistent with loss of function for these genes (Salazar-Jaramillo *et al.*, 2014). Furthermore, *D. sechellia* showed remarkable differences in the genome-wide expression profile after parasitization compared to its sister species. This was particularly the case for those immune genes previously described with large sequence changes (**chapter 4**). Thus, we hypothesized that by specializing its feeding and breeding on noni fruit, *D. sechellia* may be protected from parasitization, which consequently may have caused a relaxation of the selection pressure to maintain resistance to parasitoids.

In this chapter, we test the hypothesis that by specializing on noni fruit as oviposition substrate, *D. sechellia* may be experiencing a parasitoid-free habitat. We used 1) a laboratory experiment to test the effect of the exposure to a commercial extract of noni fruit on *Drosophila* species survival and parasitization and 2) a field survey of *Drosophila* species and parasitic wasps present in the natural reserve of Cousin island in the Seychelles.

5.2 Materials and Methods

Laboratory experiment We tested the effect of the exposure to noni on the survival of *D. melanogaster* and *D. simulans* (*D. sechellia* did not lay enough eggs for an appropriate comparison), and on the parasitization by the parasitoid *Asobara citri*. The strains of insects are the same as used in **chapter 2** (Table S2.1). Briefly, the fly species are the genome project strains from the Drosophila Stock Center (San Diego University) (Drosophila 12 Genomics Consortium 2007). The parasitoid strain of *A. citri* was collected in Ivory Coast, West Africa, in 1995, and has been maintained on *D. melanogaster* at 25℃.

For the survival and parasitization assays, adult flies were placed on petri dishes of 70mm diameter filled with standard medium (26 g dried yeast, 54 g sugar, 17 g agar and 13 mL nipagine solution per litre) and a layer of either noni extract ("Hawaiian Health Ohana" brand) or yeast. They were incubated overnight for egg-laying, after which the adult flies were removed from the petri dishes. Eggs were transferred to new petri dishes at standardized density (25 eggs/petridish), and kept at 25℃. After two days, when the larvae were in the second-instar stage, a female of *A. citri* was introduced to half of the petri dishes for 24

hours, while the rest of cultures were used as unparasitized control (n = 4 petri dishes per fly species (*D. melanogaster*, *D. simulans*), per medium type (noni, yeast) and per parasitization treatment (parasitoid, control); 2 control petri dishes (no parasitoids) for *D. melanogaster* on noni are missing due to insufficient egg laying).

After pupation, the fly pupae were transferred to vials to complete development away from the potential toxic compounds. This would reflect the natural behaviour of *Drosophila* fruit flies, which often pupate outside the oviposition substrate (Sameoto & Miller, 1968). The number of adult flies and wasps that emerged from the petri-dishes was scored. The adult flies were collected and squashed between 2 glass slides under a stereo-microscope to assess whether they had been parasitized and had encapsulated the parasitoid egg, following similar protocols as described in Salazar-Jaramillo *et al.* (2014).

Generalized Linear Models (glm) were implemented in R 3.1.0 (R Development Core Team, 2008) to analyse: 1) the host survival, as the number of adult flies of *D. simulans* and *D. melanogaster* (response variable) that developed in each of two media, yeast and noni extract, and in two conditions, parasitized and non-parasitized (controls); 2) the wasp development (response variable) as the emergence of wasps in two types of hosts, *D. simulans* and *D. melanogaster* and two media, yeast and noni extract. We used quasibinomial distribution of the errors and logit-link function.

Field study We collected flies and parasitic wasps on Cousin island in the Seychelles, using two approaches: 1) a choice assay: baits of noni, papaya and banana were used to score egg laying and 2) a field collection of fruit flies and wasps on noni. The field study took place during March-April 2014 throughout the island. Cousin Island is a natural reserve characterized by the presence of indigenous and endemic forest (mixed pisonia, noni and ochrosia), whereas invading plant species are manually removed. For the choice assay, 21 baits were placed containing a substrate of North Carolina instant medium and a layer of either banana, papaya or noni. Each set of three fruit types were placed in 7 different locations across the island and left during 24 hours to allow flies to lay eggs. Banana trees are not present in Cousin and papaya trees are removed as they are considered invasive.

For the field inventory, we collected noni fruit that had fallen of the plant, at different stages of maturity. A total of 20 noni fruit were collected. The fruits were placed in plastic containers to capture larvae that would leave to the fruit surface for pupation. The containers and fruits were left open at the site of collection to enable further oviposition by insects, for a period varying from 1-5 days depending on the stage of maturity of the noni fruit. Thereafter, the containers were closed with a piece of gauze to ensure that all insects that emerged from the fruit could be retrieved. The containers were brought to the field station and checked regularly for any emerging insects. Emerged adult insects from both approaches were collected and preserved in 70% alcohol for taxonomic identification.

5.3 Results

5.3.1 Laboratory experiment

D. melanogaster and *D. simulans* showed significant differences in survival, but this was not related to the exposure to the commercial noni extract. The survival of *D. simulans* was significantly higher than the survival of *D. melanogaster*, both in the control samples (where no parasitoid was introduced) (glm; F=15.65; df=1,12; P-value=0.00225) and in the parasitization samples (glm; F=7.4; df=1,14; P-value=0.0175). The mortality was significantly

higher in the parasitization samples compared to the control samples (glm; F=5.5; df=1,26; P-value=0.0286), but again, this was not affected by medium type (noni versus yeast medium) (Table 5.1a). Thus, *D. melanogaster* had a lower survival than *D. simulans* on both medium types, and both species suffered a higher mortality when exposed to the parasitoid *A. citri*. The fly survival was, however, not significantly affected by the exposure to noni.

The effect of the medium was significant on the parasitoid wasps (Table 5.1b, Figure 5.1). The proportion of wasps that emerged from the petri dishes was significantly lower on noni medium than on yeast medium (glm; F=12.73; df=1,13; P-value=0.00343). This reflected primarily a larger proportion of hosts that were left unparasitized on noni than on yeast (glm; F=14.33; df=1,13; P-value=0.00227). In addition, the parasitoids showed a trend towards having a lower parasitization rate on *D. simulans* than *D. melanogaster* (glm; F=4.48; df=1,14; P-value=0.054).

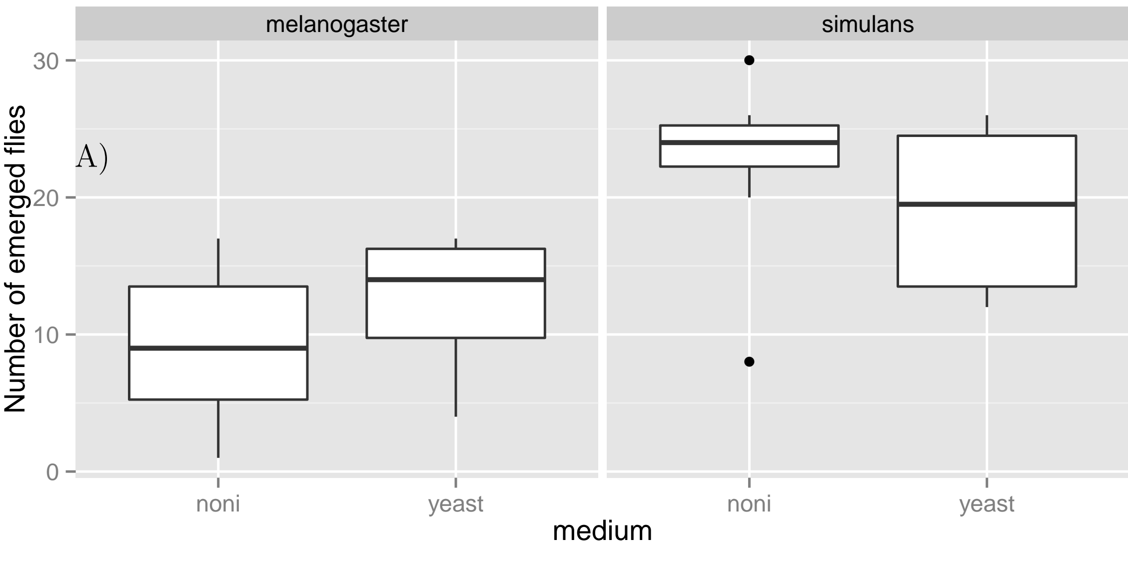
Another striking observation was a high mortality on noni medium among the parasitoids that were used for the parasitization treatment. When wasps were removed after 24 hours of parasitization, over half of the individuals parasitizing on noni extract were found dead, whereas none of the individuals on yeast had died. This will have contributed at least partially to the higher number of fly larvae that escaped parasitism. In three of the eight petri-dishes with noni substrate and parasitoid treatment, none of the (surviving) larvae had been parasitized, while in all eight petri-dishes on yeast, at least some larvae were parasitized. These results indicate that prolonged exposure to noni is likely to be lethal for the *Asobara* wasp, but not for the *Drosophila* fruit flies. Although these patterns are completely in agreement with our predictions of a parasitoid-free environment in noni fruit, two considerations need to be taken into account. We used a commercial extract of the noni fruit and we could only test the response of *D. melanogaster* and *D. simulans* against a parasitoid wasp not known to occur in the Seychelles.

Table 5.1: **GLM on the effect of different media (noni vs yeast) on insect development** The development to adulthood was compared for a) the hosts *D. melanogaster* and *D. simulans* and b) the parasitoid wasp, using as explanatory variables the host species and the medium. The model had a quasibinomial distribution of the error and a logit link function

	a) Host						b) Wasp					
	df	Deviance	Resid. df	Resid. Dev	F	Pr(>F)	df	Deviance	Resid. df	Resid. Dev	F	Pr(>F)
			13	88.29					15	169.85		
species	1	49.63	12	38.66	15.65	**0.0023**	1	24.20	14	145.65	4.52	0.0533
medium	1	1.15	11	37.51	0.36	0.5596	1	68.18	13	77.47	12.73	**0.0034**

5.3.2 Field study

All our choice assays needed to be carried out in the field, as it is not allowed to extract live material from the Seychelles. This limited enormously our sample sizes, as many samples were lost due to uncontrolled conditions (e.g., rain and animal invasion). Of the 21 baits that were placed in the field, only 9 were not damaged or invaded by other animals, and these included only one bait of papaya, three of banana and five of noni. From these baits, we scored the total number of individual flies that developed to adulthood and the number of species that could be distinguished from morphological features and male genitalia (taxonomical identification needs to be approved by expert advice). Both the number of individuals and the diversity of



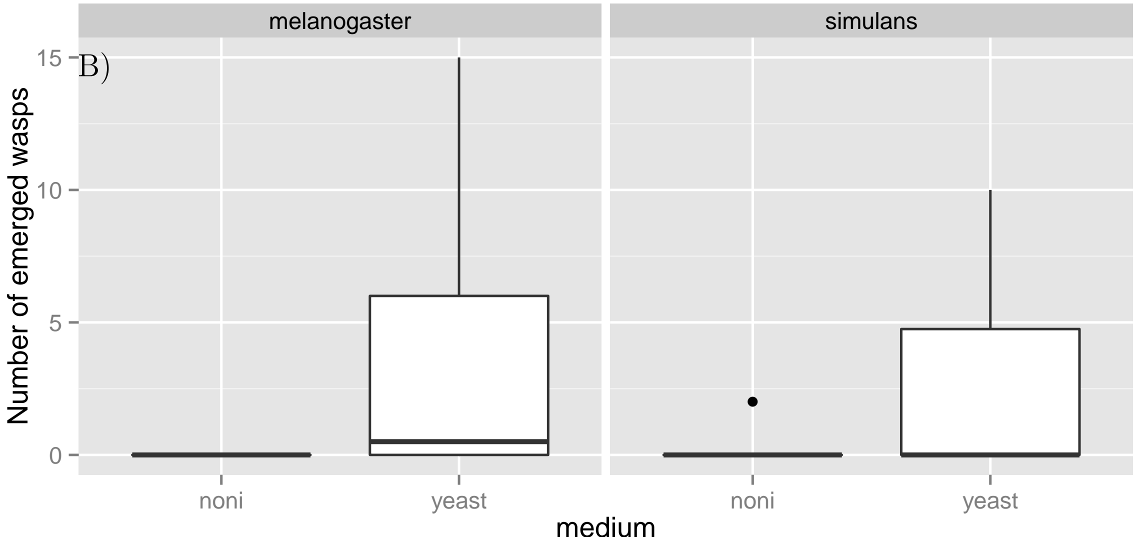


Figure 5.1: **Effect of noni extract on the survival of flies and wasps.** The development to adulthood was compared for A) the host species *D. melanogaster* and *D. simulans* and B) the parasitoid wasp *A. citri.* The significance of the explanatory variables host *species* and *medium* were tested in a GLM with quasibinomial distribution of the error and logit link function.

species were lower in noni compared to other fruits, despite noni being predominant on the island (Figure 5.2).

For the inventory of insects that naturally occur on noni fruit, 20 noni were collected at various stages of maturation (Table 5.2). Our characterization of fruit fly and wasp community that developed from the noni revealed that the stage of maturity was decisive for the number

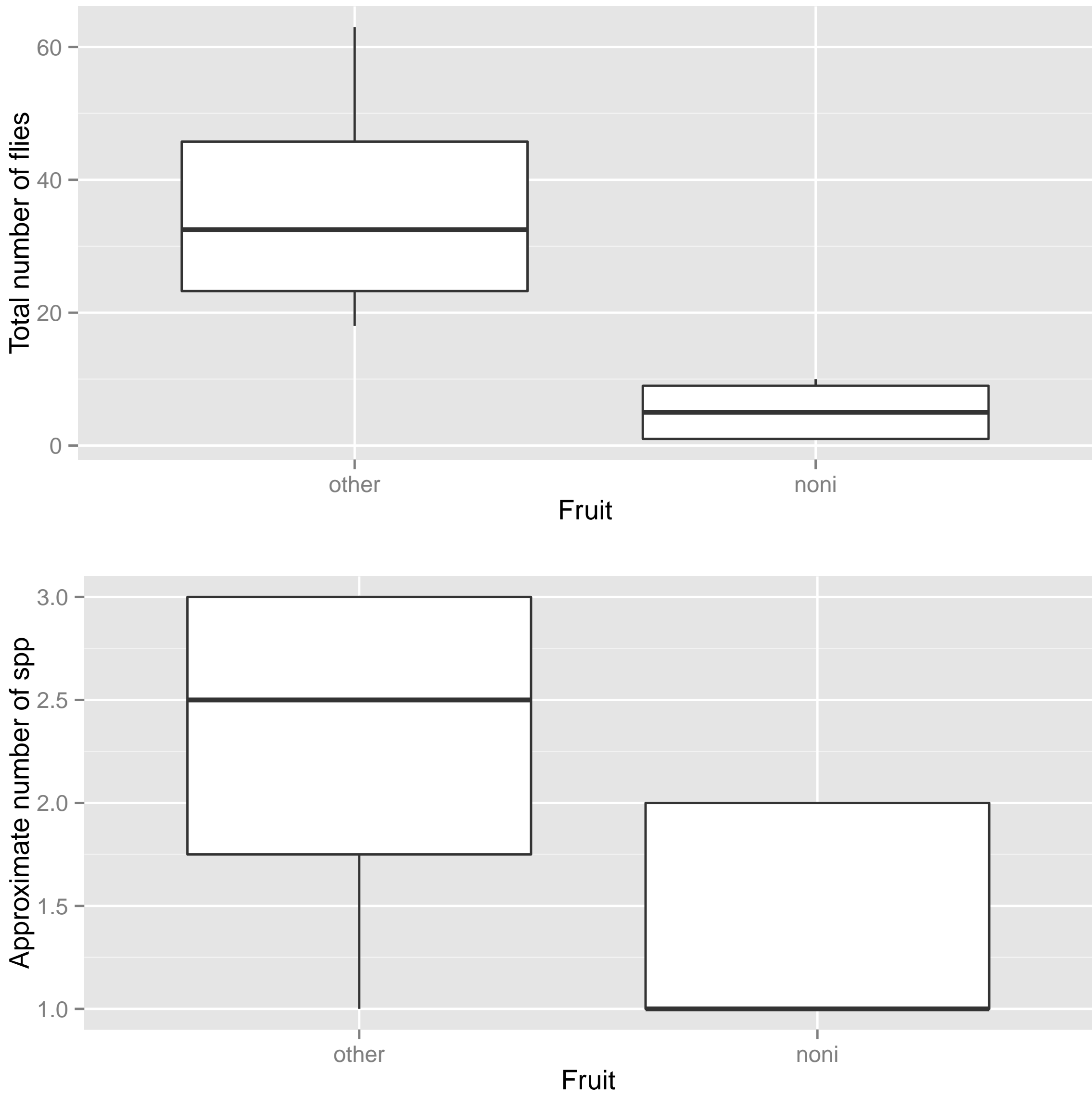


Figure 5.2: **Abundance and diversity of fruit flies in a choice assay.** Three different baits (banana, papaya and noni) were placed in trios on Cousin island (Seychelles). The *Drosophila* flies that emerged from these baits were counted and identified.

and type of species that emerged from it. Ripe noni of less than one week yielded only *D. sechellia* with large numbers of individuals. In later stages, after approximately 2 weeks, other species besides *D. sechellia* were found to develop from noni fruit (Table 5.2). From this late stage two noni fruit also yielded wasps of the genus *Leptopilina*

Behavioural observations during the collections confirmed these patterns. On ripe fruit, dozens of fruit flies would typically be seen, and collection of some of these adults indicated that these belonged predominantly to *D. sechellia*. At later stages of maturation of the noni fruit, adults of other fruit fly species were observed on the noni. At this later stage, we also found parasitoid wasps foraging on the noni fruit. Seven weeks after the collection of over-ripe noni, development of *Leptopilina* wasps was observed.

Table 5.2: **Inventory of insects collected in noni.** This table describes the stage of noni at the moment of closing the containers (noni stage), number of containers at a certain stage (replicates), number of *D. sechellia* males, number of *Drosophila* individuals that are not *D. sechellia* and number of wasps

noni stage	replicates	*D. sechellia*	other *Drosophila*	wasps
green	1	0	0	0
ripe	11	63	0	0
over-ripe	5	46	6	9
mouldy	3	7	0	0

An important mortality factor for the *Drosophila* larvae on noni were moulds. After heavy rain, the surface of noni fruit became covered with filamentous fungus (Figure 5.3) and very few adults emerged from such fruit.

Figure 5.3: **Stages of maturity of noni.** When the fruit falls to the ground it is first occupied by *D. sechellia*. During decomposition, other *Drosophila* species and parasitoid wasps are attracted to the fruit. After heavy rains, it can be overtaken by moulds

5.4 Discussion

The consequences of specialization of *D. sechellia* on noni (*Morinda citriflora*) have been extensively studied in the context of life history traits, as it provided a good model system to gain insights into the evolution of specialization and tolerance to a “toxic” resource (Jones, 1998; McBride, 2007; Matsuo, 2007). A study on the biochemical basis of the toxicity of noni revealed that *D. sechellia* was five to six times more resistant than *D. melanogaster* to one of the toxic compounds, octanoic acid, present in high concentration in the ripe stage of the fruit, but less abundant in the rotten and green stages (Legal *et al.*, 1994).

We hypothesized that one of the consequences of *D. sechellia* specialization on noni was that it would escape from parasitization by parasitoid wasps. Some results of our laboratory experiment supported this hypothesis. Noni substrate significantly decreased the number of parasitoids that developed, which was mostly because a significantly higher number of the larvae on noni was not parasitized. This may have been caused by the high mortality among parasitoid wasps during parasitization. Although compelling, these data were not conclusive, as we did not have access to a strain of parasitoid wasps that naturally occurs on the Seychelles.

Our field choice assay in the Seychelles showed that the diversity of insects developing on noni was indeed lower than on banana, despite the latter not being present on the island. In our field study of the insect community that naturally occurs on noni, we also found an ecological succession of species following the maturity and decomposition of the noni on the ground. Among *Drosophila* species, only *D. sechellia* was found to develop on noni throughout all different stages, and it was more abundant during the earlier stages of maturity. At

later stages, during decomposition, a slightly larger diversity of insects was attracted to noni, including at least one more species of *Drosophila* and a species of parasitoid wasp. Thus, it seems that the specialization of *D. sechellia* to noni is specific to an early stage of maturity of the fruit, when it is toxic to other insects. Besides reducing the inter-species competition, this also reduced the risk of parasitoid attack. Further studies are needed to complete the picture of parasitoid wasps and *Drosophila* species that are present on noni during different ripening stages of the fruit. Another interesting question concerns the level of host resistance and parasitoid virulence in this natural community, as it may be affected by the differential effects of the noni compounds on both taxa.

The idea that ecological specialization on new resources reduces the strength of competition between species has been fundamental to understand diversity (Poisot *et al.*, 2011). The consequences of this process on inter-related life history traits have also been extensively studied. Specialization can lead to trait loss, such as, for instance, the loss of eyes and pigmentation in cave animals (Poulson & White, 1969) and the loss of metabolic pathways in close associations with hosts or endosymbionts (Visser *et al.*, 2010; Payne & Loomis, 2006). Specialization has a great impact on an organism's life history, and we are starting to identify the signatures that this process leaves on the genome. In this chapter, we tested a hypotheses on the ecology of a species, inspired on a previous comparative genomics study. Conversely, using the ecological information of species can be crucial to the interpretation of genomic data.

5.5 Acknowledgements

We would like to thank Jesse Jorna and Ralph Bennen for observations and discussion on the laboratory experiment. Jan Komdeur and Sjouke Kingma for information and contacts in Cousin. Nature Seychelles and Seychelles Bureau of Standards for facilitating the collection of the material. Alex Underwood for help during the collection in Cousin. The staff and volunteers on Cousin island for the logistics during the field study. This work was supported by a KNAW Ecology Fund grant to LSJ.

[...] This thinker observed that all the books, no matter how diverse they might be, are made up of the same elements: the space, the period, the comma, the twenty-two letters of the alphabet. He also alleged a fact which travelers have confirmed: In the vast Library there are no two identical books. From these two incontrovertible premises he deduced that the Library is total and that its shelves register all the possible combinations of the twenty-odd orthographical symbols (a number which, though extremely vast, is not infinite): Everything: the minutely detailed history of the future, the archangels' autobiographies, the faithful catalogues of the Library, thousands and thousands of false catalogues, the demonstration of the fallacy of those catalogues, the demonstration of the fallacy of the true catalogue, the Gnostic gospel of Basilides, the commentary on that gospel, the commentary on the commentary on that gospel, the true story of your death, the translation of every book in all languages, the interpolations of every book in all books [...]
Jorge Luis Borges, The library of Babel

Chapter 6

General Discussion

6.1 Overview

Genomes contain key information about the natural history of organisms. Changes in selection pressures from the environment or from interactions with other organisms leave molecular signatures in the genome. Learning "how to read" this information requires addressing fundamental issues in evolutionary biology such as common ancestry, adaptive versus neutral evolution and stabilizing versus directional selection (Kondrashov, 1999). Interpreting the history of species from its genome involves mapping the phenotype to the genotype, studying how the phenotype affects ecological interactions and *vice versa*, which selection pressures this interaction generates on the phenotype, and how the resulting evolutionary processes are "archived" in the genome.

In this thesis I focus on one important aspect of the natural history of *Drosophila*, the interaction with parasitoid wasps. Parasitoids are widespread parasites that exert a strong selection pressure on their hosts, as they develop by consuming them, often before the host reaches the reproductive stage (Godfray, 1994). A successful immune response to such lethal infections seems an important trait for the fitness of the host. The ability to survive a parasitoid attack is, however, not ubiquitous, and shows large variation among insect species. To understand the mechanisms and the evolution of the ability to survive parasitoid attack, I studied the large variation in the immune response against parasitoids in *Drosophila*.

I approached this question using comparative genomics, molecular tools, phenotypic assays and observations in the field. In this chapter I will integrate the different findings of the preceding chapters to highlight which important insights we have gained on the evolution and the mechanisms of the ability to encapsulate parasitoid eggs in *Drosophila*. I place these findings in the context of the current state of the research field and suggest future research that could further progress this field.

6.2 The encapsulation ability in *Drosophila* species

At the start of this project it was well established that within the *Drosophila* genus there is large variation in encapsulation ability, including the complete absence in some species of the *obscura* group (Eslin & Prévost, 1998; Havard *et al.*, 2009). This immunodeficiency was associated with the lack of an important cellular immunity component, the lamellocytes, and was considered a trait loss (Eslin & Doury, 2006). Since some species in the *obscura* group are natural hosts of parasitoid wasps, the ability to encapsulate cannot only depend on the

natural exposure to parasitoids.

Among populations of *D. melanogaster* the variation in the encapsulation ability is also large. This variation has been associated to local differences in the abundance and virulence of the parasitoids and the presence of alternative hosts (Kraaijeveld & Godfray, 1999). Moreover, increase in encapsulation ability can be experimentally selected for, and shows a powerful response in only a few generations (Kraaijeveld & Godfray, 1997; Wertheim *et al.*, 2011; Jalvingh *et al.*, 2014). There is, thus, a crucial genetic component underlying the ability to mount a successful immune response against the parasitoid, and an important ecological component creating different selection pressures, which contributes to the variation in the response.

In **chapter 2** we showed that, contrary to what was previously assumed, the encapsulation and the production of lamellocytes is not a trait that is shared among most *Drosophila* species. Instead, encapsulation by lamellocytes is restricted to a sublineage of *Drosophila*, and inside this sublineage encapsulation is absent (and lamellocytes present) in one species, namely *D. sechellia*. This may imply that lamellocytes and encapsulation evolved in a shared ancestor after the split of the melanogaster group. Although in our phenotypic survey we found that only species of the melanogaster subgroup were able to encapsulate wasp eggs, two species (*D. eugracilis* and *D. suzukii*) in the melanogaster group and outside the melanogaster subgroup have been reported to also encapsulate wasp eggs and produce lamellocytes (Schlenke *et al.*, 2007; Kacsoh & Schlenke, 2012). In some of the species of the *obscura* group the encapsulation process is not mediated by lamellocytes but a different type of hemocyte, the pseudopodocytes (Havard *et al.*, 2012), suggesting that blood cells involved in encapsulation may have evolved multiple times. Other distantly related *Drosophila* species (e.g., *D. busckii*) have also been reported to encapsulate parasitoids (Streams, 1968), but the mechanism and blood cells involved are unknown. These combined findings strongly suggest that encapsulation by differentiated blood cells may have evolved multiple times within the genus *Drosophila*.

In the phenotypic assay in **chapter 2** two species of *Drosophila*, *D. ananassae* and *D. willistoni* did not show encapsulation ability, but also did not sustain development of *A. tabida*, while *A. citri* could develop. Besides immune responses, hosts can show other forms of resistance. For instance, behavioural responses can affect their susceptibility to be attacked (Carton & Sokolowski, 1992; Lefévre *et al.*, 2011) or the interaction with their microbiota can enhance their ability to respond (Hurst & Hutchence, 2010). More studies are needed to establish why certain parasitoid species do not develop in some host species.

6.3 Evolving encapsulation ability: a trait involving new genes

To associate the evolution of lamellocytes and encapsulation ability to changes in the genome, I used a comparative genomics approach on 11 sequenced *Drosophila* species (**chapter 2**, Salazar-Jaramillo et al., 2014). I focused on two sets of genes that were derived from studies in *D. melanogaster*: a set of 35 genes associated with hemocyte differentiation and proliferation (Fauverque & Williams, 2011; Zettervall *et al.*, 2004; Williams, 2007; McQuilton *et al.*, 2012; Stofanko *et al.*, 2010; Avet-Rochex *et al.*, 2010) and a set of 109 candidate genes coming from microarray studies (Wertheim *et al.*, 2005; Schlenke *et al.*, 2007). We showed that hemopoiesis-associated genes are highly conserved and present in all species independently of their resistance. This suggests that all the species possess the same set of (known) hemopoeisis genes, but in the above-mentioned sublineage some of these genes have been co-opted to differentiate a new type of hemocyte, the lamellocyte. In contrast, the set of 109 differentially

expressed genes during the response to parasitoids in *D. melanogaster* contained 11 novel genes, specific to the *Drosophila* sublineage capable of lamellocyte-mediated encapsulation. These novel genes arose via duplications and are predominantly expressed in hemocytes. The majority of these newly duplicated genes showed signatures of positive selection, suggesting that they may have been recruited for new functions in the new hemocyte type and/or the encapsulation response (Salazar-Jaramillo *et al.*, 2014).

Based on these findings, we proposed that the mechanism of parasitoid encapsulation found in *D. melanogaster* evolved in a subgroup of species and involved duplicated genes. This is also in agreement with the idea that species of *Drosophila* rely largely on "new genes" for their immune response, based on a review of comparative genomics studies in *Drosophila* (Singh *et al.*, 2009). Further support for this hypothesis was provided in **chapter 4** with an RNAseq experiment on the transcriptional response after parasitoid attack in *D. melanogaster*, *D. simulans*, *D. sechellia* and *D. yakuba*. A set of genes was up-regulated across all four species, largely composed of genes annotated with immune function. A subset of these genes showed up-regulation in all species, independent of their resistance. This subset is, thus, likely to reflect a general defensive and stress response that is not specific to parasitoid attack, but more general after an immune challenge. Interestingly, another subset of genes showed up-regulation in all species except *D. sechellia*, for which they showed down-regulation or no expression, consistent with its lack of immune response against parasitoids. Especially the set of genes up-regulated in all species except *D. sechellia* may be of particular interest to understand specific genes involved in the immune response against parasitoid attack. Genes differentially expressed after parasitoid attack across species showed a large proportion of lineage restricted and paralogous genes. This set also overlapped with previous expression studies after parasitoid attack in *D. melanogaster*. This RNAseq study highlighted that the conclusion drawn from *D. melanogaster*, that a large proportion of up-regulated genes during the immune response are lineage restricted, extends to closely related species. Nonetheless, the transcriptional response also showed several distinct differences among species, suggesting that "fine-tuning" of the immune response may have followed different evolutionary trajectories in the various species.

A following question that arises is how these new genes are incorporated in an existing gene interaction network, and how they differentiate to acquire new functions? Here I will describe the documented cases for two genes, *PPO3* and *yellow-f*, where expression changes have occurred between two gene copies. These two genes were found as new copies in the group able to encapsulate in **chapter 2** and to be significantly up-regulated after parasitization in **chapter 4**.

The first case is *PPO3*, one of the three pro-phenoloxidases present in the genome of the melanogaster subgroup species. Contrary to *PPO1* and *PPO2*, *PPO3* is spontaneously activated inducing melanization in absence of proteolytic enzymes (Nam *et al.*, 2008). Tight regulation of the phenoloxidase cascade is necessary, given the toxic nature of products and by-products of the melanization pathway (Cerenius *et al.*, 2008). Expression of *PPO3* is restricted to the lamellocytes, while the other two PPOs are expressed in crystal cells (Irving *et al.*, 2005). Lamellocytes are usually only differentiated following parasitization. Thus, the spontaneous activation of this new copy may have contributed to a local and more rapid response upon parasitization, while its regulation became subject to cell differentiation.

The second case is the iso-enzymes *yellow-f* and *yellow-f2*, which act on intermediates of the melanine production. Both proteins can perform the same reaction but *yellow-f* is involved in melanization reactions during larval and early pupal stages, while *yellow-f2* plays a predominant role in melanization reactions during later pupal and adult stages (Han *et al.*, 2002). Melanization is involved in various processes in *Drosophila*, ranging from immune

responses against bacteria and against parasitoids, as well as cuticle pigmentation (Lemaitre & Hoffman, 2007). The difference in expression pattern could mean that, while the function is retained in both copies, they specialized to different developmental stages. The subsequent divergence of one of the copies specialized in the larval stage, *yellow-f*, possibly allowed for the incorporation of the melanization reaction during the immune response after parasitoid challenge.

While we found that gene duplicates were crucial in the evolution of the immune response against parasitoid attack, duplicated genes were not found equally distributed among gene categories (**chapter 2**). Genes coding for effector proteins (i.e., antimicrobial peptides, phenoloxidases and intermediates in the melanin production) and for proteases (i.e., enzymes that activate other proteins through cleavage) show a large proportion of duplicated genes (Salazar-Jaramillo *et al.*, 2014; Sackton *et al.*, 2007; Waterhouse *et al.*, 2007; *Drosophila* 12 Genomes Consortium, 2007). In contrast, most genes belonging to signalling pathways showed single copy orthologs (Salazar-Jaramillo *et al.*, 2014; Sackton *et al.*, 2007; Waterhouse *et al.*, 2007). Similarly, all genes involved in hemocyte proliferation and differentiation, except one, *Hemese*, were found as single copy orthologs (Salazar-Jaramillo *et al.*, 2014). *Hemese* is a cellular receptor expressed in all hemocytes. Its role appears to be the negative regulation of lamellocyte differentiation, thus preventing an over-production of lamellocytes (Kurucz *et al.*, 2003). Hemocyte associated genes are likely to be involved in different developmental pathways. Developmental genes have been found to be more essential than non-developmental genes and more likely to be sensitive to dosage (Makino *et al.*, 2009). Moreover, it has been found that younger genes are less likely to be essential than older genes (Chen *et al.*, 2012). This together with our own findings, suggests that duplicated copies are more likely to be retained for genes that act at the "periphery" (e.g., that act as modulators) of long established networks composed by essential genes.

An interesting question is what ecological conditions allow the maintenance of duplicated copies and how they evolve new functions. In the *Drosophila* lineage the proportion of duplicated genes has been shown to significantly correlate with habitat heterogeneity (Makino & Kawata, 2012). Particularly, genes restricted to the melanogaster subgroup (i.e., new genes that mostly diverged fast) are significantly enriched for proteins involved in behaviour, response to biotic stimulus, symbiosis and parasite response (Heger & Ponting, 2007). It is, thus, possible that species exposed to more heterogeneous environments (e.g., generalists) have retained duplicates at particularly high rates that could have entailed adaptations to new conditions. Habitat heterogeneity also implies exposure to new pathogens, and thus new niches related to the interaction with parasites, where new gene copies can rapidly acquire new functions. By specializing on a resource (e.g., the case of *D. sechellia*), species may experience relaxation of the selection pressure on some of the newly acquired genes.

6.4 Fast evolving genes

Acquisition of new gene functions may be reflected in the high rate of substitution found in the duplicated genes. The category containing the largest proportion of duplicated genes, as well as the largest proportion of genes with signatures of positive selection in the genomic analysis in **chapter 2** was serine-proteases. Serine-proteases belong to a large and highly diverse family of proteins, which play important physiological roles by activating specific proteins through proteolytic cleavage. These proteins are usually synthesized as pro-enzymes (zymogens) and need to be activated through hydrolysis to reveal their active site. The active site is characterized by a catalytic triad consisting of histidine, aspartate and serine amino acid residues,

which allows them to specifically form a peptide bond and activate other zymogens, thus generating a reaction cascade for a rapid response (Rawlings & Barrett, 1994; Muhlia-Almazán *et al.*, 2008; Page & Cera, 2008). The genome of *Drosophila* comprises approximately 200 serine-proteases and serine-protease homologues (that may have lost their catalytic function (Rossa *et al.*, 2003)). Duplication followed by rapid divergence is likely to be responsible for the expansion of this gene family and its wide spectrum of substrate specificity (Rossa *et al.*, 2003). Serine-proteases homologues, despite the lack of catalytic domains, have also been found to be important in the regulation of the immune response (Kambris *et al.*, 2006; Patterson *et al.*, 2013). Immune responses against micro-parasites (bacterial and fungal) and macro-parasites (parasitoid) are at least partially coordinated through the same immunity pathways (e.g. Toll, JakStat). However, it has been suggested that specific downstream protease cascades may separate the immune response against micro-parasites and macro-parasites (Shah *et al.*, 2008).

A different category with a large proportion of genes under positive selection was immune receptors. Immune receptors are crucial molecules at the interface of the host-pathogen interaction. Due to their physical interaction with parasites, receptors are particularly likely to be under co-evolution with parasites. Yet, not all immune receptors appear to change under similar evolutionary rates. While some immune receptor genes rapidly evolve, as expected under a co-evolutionary arms' race, other receptors are highly conserved (Jiggins & Kim, 2006; Little & Cobbe, 2005). Among the fast changing receptors in *Drosophila* are Teps (thioester binding protein) (Jiggins & Kim, 2006; Sackton *et al.*, 2007; Salazar-Jaramillo *et al.*, 2014) and some Lectins (Salazar-Jaramillo *et al.*, 2014), whereas some carbohydrate binding proteins, such as GNBPs and PGRPs (Gram negative binding proteins and peptidoglycan recognition protein, respectively) are highly conserved (**chapter 2**, Jiggins & Kim (2006); Sackton *et al.* (2007)). To explain this dichotomy, it has been suggested that protein-protein interactions have more potential for co-evolution than carbohydrate-protein interactions (Little & Cobbe, 2005). This hypothesis would explain the difference in evolutionary rates found between Tep genes (protein-protein interaction) and GNBPs/PGRPs (protein-carbohydrate), but it would pose a challenge for the fast evolutionary rates found in some Lectins (protein-carbohydrate). Alternatively, it has also been proposed that those genes that activate the immune response have a large effect due to the amplification of the signal and are therefore more constrained than genes acting later in the response (Sackton *et al.*, 2010). This scenario could imply that two types of activation exist: 1) an early and more general activation of immune pathways, which strongly amplifies the signal and is triggered by highly conserved receptors (e.g GNBPs and PGRPs) and 2) a later and more specific response with a local effect, such as phagocytosis of micro-parasites or encapsulation of macro-parasites triggered by the more divergent receptors (Teps and Lectins).

Fast divergence of genes has also been correlated to tissue-specific expression. In multicellular organisms, genes expressed in restricted cell types evolve significantly more rapidly than ubiquitously expressed genes (Duret & Mouchiroud, 2000; Larracuente *et al.*, 2008). Consistent with this, we found that some of the candidate genes in **chapter 2**, which are mainly or exclusively expressed in hemocytes (*TepI, αPS4, Lectin-24A, CG4259, CG18477, Spn88Eb7*), showed signatures of positive selection. Regulation of gene expression requires the activity of trans-acting factors and cis-acting motif sequences, which modulate nearby genes. Cis-regulatory regions have shown extensive variation within natural populations and has been suggested to constitute an important part of the genetic basis for adaptation (Fay & Wittkopp, 2008; Wray, 2007). In this thesis we focused on coding regions. An important follow-up study, however, is to quantify the substitution rates of non-coding regions and associate it with the expression data of **chapter 4**.

Of 35 genes associated with hemopoiesis, 5 genes, *Ser, Dpp, ush, cher* and *sgg*, had signatures of positive selection. These 5 genes were annotated as regulators of hemocyte differentiation, and two of them, *cher* and *sgg* (Zettervall *et al.*, 2004), specifically of lamellocyte differentiation. This suggests that the evolution of lamellocytes may have driven fast divergence in genes involved in its regulation.

6.5 Encapsulation ability among lines of *D. melanogaster*

The RNAseq experiment in **chapter 4** confirmed the previous interpretation that selection for increased resistance does not act on a pre-activation of the immune response (i.e immune genes were not constitutively up-regulated) (Wertheim *et al.*, 2011). This is perhaps not surprising, for raising the expression of immune proteins or to constitutively activate immune pathways can either result in the damage of the host tissues or carry high energetic costs (Kraaijeveld & Wertheim, 2009).

Selection may, instead, exploit alternative mechanisms to effectuate a faster and more efficient response. The increase in hemocyte load among lines selected for high resistance, for example, has been recurrently reported (Kraaijeveld & Godfray, 1997; Wertheim *et al.*, 2011; Jalvingh *et al.*, 2014). We found that the lamellocyte-specific gene *PPO3* was constitutively up-regulated in selected lines compared to non-selected lines, which is consistent with an increased hemocyte load prior to infection. Another mechanism that appears to be exploited by selection is the differential expression of isoforms or splice variants (without necessarily up-regulating the gene expression) as shown in **chapter 4**. The detection of splice variants is greatly facilitated through the current RNAseq technology. We found that some transcript variants expressed predominantly in selected lines compared to non-selected lines during parasitization. Alternative splicing amplifies the protein diversity encoded in the genome (Hallegger *et al.*, 2010; Nilsen & Graveley, 2010), and may, thus, constitute an important source of variation upon which selection can act.

Experimental selection is a very useful method to test the evolutionary potential of a phenotype, but may not necessary emulate what happens in nature. For instance, the increase in total hemocyte load found among selected lines (Kraaijeveld *et al.*, 2001) was not found among the field lines collected in Europe (Gerritsma *et al.*, 2013). Within populations of *D. melanogaster* a minimum hemocyte load was needed for a successful response, and above this threshold, the ratio lamellocytes to crystal cells correlated to resistance, independently of the total load of hemocytes. Natural populations are exposed to a combination of pathogens and environmental conditions, which may create different selection pressures and constraints on the levels of hemocytes. Thus the variation in the levels of hemocytes of the field lines can be the product of the local conditions that natural populations are adapted to, including the co-existence with parasites (Gerritsma *et al.*, 2013).

In order to identify genes that were possibly engaged in host–parasite co-evolution, we focused on immune receptors in **chapter 3**. Immune receptors are likely to mediate the direct physical contact between the parasite and the host, and some may therefore show signatures of selection (Little *et al.*, 2004; Jiggins & Kim, 2006; Ellis *et al.*, 2012). Two types of co-evolutionary dynamics are typically distinguished: 1) an “arms-race” where hosts and parasites continually accumulate adaptive mutations and 2) “Red Queen” dynamics, where host- and parasite- genotype frequencies fluctuate over time, due to frequency dependent selection (Woolhouse *et al.*, 2002). Molecular signatures of selection are usually interpreted as the outcome of one of these co-evolutionary dynamics. “Arms-race” dynamics are expected

to lead to recurrent fixation of alleles and transient polymorphism (i.e., positive selection), whereas "Red Queen" is expected to lead to cycles of allele frequency changes (i.e., balancing selection) (Tellier *et al.*, 2014).

We studied the genetic diversity in 5 immune receptor genes among field lines of *D. melanogaster*, collected in Europe and with different levels of resistance against *A. tabida*. We sequenced fragments of *PGRP-SB1,PGRP-LB, Lectin24A,* $\alpha PS4$ and *Tep1* in 48 individuals in total. We found little sequence variation among the field lines in any of these receptors, except for *Tep1*. This is in contrast to the pattern for substantial sequence divergence among species, where three of these immune receptor, *Tep1*, *Lectin-24A* and $\alpha PS4$ showed signatures of positive selection across species (**chapter 2**). Additionally, *Tep1* was identified as a duplicated gene under positive selection with large-scale genomic changes in *D. sechellia* (**chapter 2**). This gene showed up-regulation after parasitization in species able to resist, while no up-regulations was found in *D. sechellia*, consistent with loss-of-function (**chapter 2** and **chapter 4**). In the field lines, however, we found up-regulation of the expression of the gene in all lines and only subtle differences in expression among the field lines.

Although we found large genetic diversity in *Tep1*, population genetic analysis on *Tep1* showed that no haplotype was exclusive to any population, and that the distribution of SNPs did not deviate from neutrality. Interestingly, we found very high heterozygosity, particularly at four positions in the gene. We tested and rejected the hypothesis that high heterozygosity was the outcome of multiple gene copies. The high level of heterozygosity in specific loci could be a direct consequence of the interaction with parasites, although the details of this molecular interaction remain unknown.

The two general models addressing co-evolutionary dynamics described above may in fact be considered as two extreme cases of a large range of possible interactions (Agrawal & Lively, 2002). Selective processes do not only depend on the direct co-evolutionary interaction (i.e., host-parasite), but may be influenced by other ecological factors (e.g., fluctuating environments or competition), which generates a mosaic of conditions that introduces an extra level of variation to the selection pressures (Thompson, 2005; Laine, 2009). Thus, to interpret the pattern found in *Tep1*, a deeper understanding is needed of its function, as well as the ecological conditions the field lines are adapted to. It may also be useful to contrast the genetic diversity in *Tep1* with the genetic diversity of other members of the gene family (e.g., *Tep2*) as well as non-immune genes (e.g., mitochondrial genes). This would improve our understanding of the extent to which the molecular patterns in these Tep genes correspond to neutral processes (e.g., population structuring), Red Queen dynamics or whether they reflect particular adaptations to the parasite or ecological communities.

6.6 Contrasting long- and short-term evolution

On a longer time scale (i.e., across species) the evolution of the response against parasitoids involved the duplication of genes and differentiation of new cell types. During this process, it is likely that duplication of genes and co-option of existing genes jointly resulted in new functions. For instance, the existing hemopoiesis pathways were modulated, presumably, by new genes, to produce a new cell type. Meanwhile, some duplicated genes shifted expression to this new cell type, such as *PPO3*. On a shorter time scale (i.e., among lines) the standing variation is exploited. Five generations of experimental selection for increased resistance resulted in signatures of selection across multiple regions across the whole genome, but did not result in fixation of alleles for a particular selection line or locus (Jalvingh *et al.*, 2014), nor in the pre-activation of the immune response (Wertheim *et al.* (2011), **chapter 4**). Experimental

selection, however, did result in differences in expression of transcript isoforms between selected and non-selected lines after parasitization (**chapter 4**). This suggests that transcript isoforms are a source of variation, that can be exploited by short-term selection.

Modulation of hemocytes seems to be a fundamental component in the process of encapsulation. Particularly, the coordination of hemocyte differentiation is an important determinant of the successful encapsulation response among field lines of *D. melanogaster* (Gerritsma *et al.*, 2013). Experimental selection resulted in an increase of the constitutive level of hemocytes, particularly lamellocytes (Jalvingh et al, submitted). Among species of the melanogaster group, a strong correlation between encapsulation ability and hemocyte load has been found (Eslin & Prévost, 1998). It seems, thus, that once the machinery to mount the encapsulation response was acquired, fine-tuning of this response became the target of selection by changing the numbers of hemocytes.

On a sequence level, genes with high divergence among species may be expected to show high polymorphism among populations, according to the neutral theory of molecular evolution (Kimura, 1983). Deviations from this expectation are considered to be a signature of selection (McDonald & Kreitman, 1991). We tested this correlation for five genes, and found that only one gene, *Tep1*, showed both high divergence and polymorphism among field lines. None of these polymorphisms showed fixation of alleles in a particular line, but we did find a high level of heterozygosity. Such pattern could be the outcome of the interaction with parasites (e.g. negative frequency dependent selection), similar to what has been suggested for MHC. *Tep1* showed a very different exon structure in *D. sechellia* compared to the remaining species of the melanogaster subgroup, and no up-regulation after parasitoid attack. These molecular signatures across species and lines found in *Tep1* may exemplify the complexity of interactions with parasites, and thus be important to infer co-evolutionary processes.

6.7 Loss of resistance in *D. sechellia*

The implications of sequencing the genome of the island specialist *D. sechellia* were profound, as it provided a unique opportunity to study the consequences of ecological specialization on the genome. Island ecosystems have been for a long time acknowledged as special habitats where evolution can be studied when species are physically isolated. The species *D. sechellia* is restricted to the Seychelles islands in the Indian Ocean, where it specialized on the fruit *Morinda citriflora* (commonly known as noni). This fruit has been shown to be toxic in the ripe stage to *D. sechellia*'s sister species, *D. melanogaster* and *D. simulans* (Legal *et al.*, 1994). Genomic changes associated to its specialization in the noni have been described for smell and taste receptors (McBride, 2007).

We found large sequence changes in some immune genes present in the genome of this species (**chapter 2**), and a remarkable difference in the genome-wide expression profile after parasitization (**chapter 4**). The lack of resistance against parasitic wasps and the changes in genes that are presumably important during this immune response, led us to hypothesize that *D. sechellia*'s specialization on noni fruit may have effectively protected it from parasitoid wasps. This could cause the relaxation in the selection pressure to maintain parasitoid resistance (**chapter 2**). In order to test this hypothesis, we performed a field study of *Drosophila* species and parasitic wasps present in noni in the natural reserve of Cousin island in the Seychelles during March-April of 2014 (**chapter 5**). We found that the diversity of insects developing in banana baits was larger than that found in noni, despite the fact that banana is not present in the island. Among *Drosophila* species, only *D. sechellia* was found to develop on noni fruit throughout different stages of maturity, and it was more abundant during

the earlier ripening stages, just after falling to the ground. At later stages of decomposition, we found a slightly larger diversity of insects attracted to noni, including at least one more species of *Drosophila* (to be identified) and parasitic wasps from the *Leptopilina* genus. Currently, it is not clear which *Drosophila* species in noni is host to this wasp. Finally, laboratory experiments indicated that larvae in substrate with noni extract were more likely to escape parasitism than in standard substrates (**chapter 4**). Thus, it seems plausible that the adaptation of *D. sechellia* to tolerate the toxins of the ripe noni provided it of an environment with reduced interspecific competition and reduced risk of parasitoid attack.

There are different mechanisms by which a trait can be lost (Lahti *et al.*, 2009). For instance, the loss of sexually selected male traits seems to be taxonomically widespread, as a consequence of a significant cost in terms of predators (Wiens, 2001). A trait can also be lost when its function is provided by an ecological partner (e.g., metabolic pathways provided by hosts or endosymbionts) (Ellers *et al.*, 2012) or as consequence of pleiotropy (e.g., trade-offs between fecundity and resistance to starvation and desiccation in *D. melanogaster* (Hoffmann *et al.*, 2007)). Finally, neutral factors can also contribute to the decay of a trait. This happens after recurrent mutation and genetic drift on a trait under relaxed selection (Lahti *et al.*, 2009). We hypothesized in **chapter 5** that *D. sechellia* specialization on a toxic stage of the noni fruit provided it with an environment of relaxed selection to maintain the immune response against parasitoid attack. This does not imply that some of the factors mentioned above did not also contribute to the loss of the trait, and perhaps enhanced the speed of loss.

6.8 Future work

We proposed that the ability to encapsulate parasitoids is a trait that evolved multiple times, even inside the genus *Drosophila*. This trait has been extensively described in other insects too. However, for some non-*Drosophila* species (e.g., Lepidopteran), the blood cells involved in the encapsulation response do not appear to be homologs of the *Drosophila* hemocyte types (Lavine & Strand, 2002; Ribeiro & Brehélin, 2006). Therefore, to confirm or reject the hypothesis that this trait has evolved independently in different clades, a good characterization of the physiological response, the types of blood cells involved and the differentially expressed genes in multiple clades is required. Some studies have explored the correlation between retained duplications and functions related to environmental response or to habitat heterogeneity in *Drosophila* (Makino & Kawata, 2012). This implies that habitat heterogeneity can provide conditions were new copies acquire new functions. Studies beyond *Drosophila* can give important insights into the genomic patterns associated with retention of duplicated genes and the acquisition of new functions. This may be particularly important during habitat colonization, for example, by invasive species.

This study, together with others, found a pattern of "conserved versus diversified" immune receptors (Jiggins & Kim, 2006; Little & Cobbe, 2005). An interesting follow up study would be to identify the types of immune receptors that are under rapid change, and why some immune receptors have more potential to evolve than others. A suitable taxonomic group for this comparative approach is the Diptera. Besides the extensively studied *Drosophila*, Diptera contains genera (*Anopheles, Aedes* and *Glossina*) with several species fully sequenced, which provides good statistical power for the estimation of the rates of evolution. This could confirm whether the pattern of high conservation of GNBPs/PGRPs and high divergence of Lectins and Teps is general or a particularity of *Drosophila*.

A possible explanation for the difference in evolutionary rates in some immune genes was that some receptors have a disproportionate effect on the amplification of the immune

signal, which would constrain their evolution (**chapter 2**). To test this, it is possible to formulate a network simulation model of pathway evolution. This could help to understand the evolutionary constraints imposed on certain genes by the topology of the network.

6.9 Conclusions

By studying the response of *Drosophila* against parasitoid wasps, we characterized the genomic and physiological components that enabled some species to mount a successful immune response to a lethal infection, and how this response was lost in a particular ecological setting. We quantified the variation in expression among species and among populations to infer differences in the way selection may act on long and short time scales. We found that the evolution of the encapsulation of parasitoid eggs in a subgroup of species of *Drosophila* involved the duplication of genes and differentiation of new blood cell types. Inside *D. melanogaster* selection acts primarily on the modulation of this response by changing the numbers of circulating hemocyte and selection of transcript isoforms. We proposed that the ecological specialization of *D. sechellia* partially protected it from the attack of parasitoid wasps, resulting in the loss of the ability to encapsulate. In conclusion, through a combination of comparative genomics, phenotypic assays, molecular genetics experiments and observations in the field, we gained important insights into the feedback between an ecological defence mechanism encoded in the genome and ecological interactions, and the important impact of this interaction on the genome.

Bibliography

Drosophila 12 Genomes Consortium 2007. Evolution of genes and genomes in the *Drosophila* phylogeny. *Nature* 450:203–218.

Agrawal, A. & Lively, C. M. 2002. Infection genetics: gene-for-gene versus matching- alleles models and all points in between. *Evolutionary Ecology Research* 4:79–90.

Alföldi, J. & Lindblad-Toh, K. 2013. Comparative genomics as a tool to understand evolution and disease. *Genome Research* 23:1063–1068.

Aljanabi, S. M. & Martinez, I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research* 25:4692–4693.

Anders, S., Pyl, P. T. & Huber, W. 2014. Htseq — a python framework to work with high-throughput sequencing data. *bioRxiv* doi:doi:10.1101/002824.

Anders, S., Reyes, A. & Huber, W. 2012. Detecting differential usage of exons from RNA-seq data. *Genome Research* .

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437:1149–1152.

Antao, T. 2008. Lositan: a workbench to detect molecular adaptation based on a FST-outlier method. *BMC Bioinformatics* 9:323.

Aoun, R. B., Hetru, C., Troxler, L., Doucet, D., Ferrandon, D. & Matt, N. 2011. Analysis of Thioester-containing proteins during innate immune response of *Drosophila melanogaster*. *Journal of Innate Immunity* 3:52–64.

Artieri, C. G., Haerty, W. & Singh, R. S. 2009. Ontogeny and phylogeny: molecular signatures of selection, constraint, and temporal pleiotropy in the development of drosophila. *BMC Biology* 7:42.

Avet-Rochex, A., Boyer, K., Polesello, C., Gobert, V., Osman, D., Roch, F., Augé, B., Zanet, J., Haenlin, M. & Waltzer, L. 2010. An *in vivo* RNA interference screen identifies gene networks controlling *Drosophila melanogaster* blood cell homeostasis. *BMC Developmental Biology* 10:65.

Barabási, A.-L. & Oltvai, Z. N. 2004. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* 5:101–113.

Barrett, R. D. H. & Hoekstra, H. E. 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics* 12:767–780.

Beaumont, M. & Nichols, R. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B* 263:1619–1626.

Bitra, K., Suderman, R. & Strand, M. 2012. Polydnavirus Ank proteins bind NF-*kappa*B homodimers and inhibit processing of relish. *PloS Pathogens* 8:e1002722.

Blandin, S. & Levashina, E. A. 2004. Thioester-containing proteins and insect immunity. *Molecular Immunology* 40:903–908.

Blumberg, D. & Ferkovich, S. 1994. Development and encapsulation of the endoparasitoid, *Microplitis croceipes* (Hym.: Braconidae) in six candidate host species (lep.). *Entomophaga* 39:293–302.

Borghans, J. A. M., Beltman, J. B. & de Boers, R. J. 2004. MHC polymorphism under host-pathogen coevolution. *Immunogenetics* 55:732–739.

Bowden, L., Dheilly, N., Raftos, D. & Nair, S. 2007. New immune systems: pathogen-specific host defence, life history strategies and hypervariable immune-response genes of invertebrates. *Invertebrate Survival Journal* 4:127–136.

Brown, T. 2002. *Genomes*. Wiley-Liss, Oxford.

Caillaud, C. & Via, S. 2000. Specialized feeding behaviour influences both ecological specialization and assortative mating in sympatric host races of pea aphids. *The American Naturalist* 156:606–621.

Carroll, L. S. & Potts, W. K. 2006. Functional genomics requires ecology. *Advances in the Study of Behaviour* 36:137–215.

Carton, Y., Boulétreau, M., van Alphen, J. & van Lenteren, J. 1986. The *Drosophila* parasitic wasps. In: *The genetics and biology of* Drosophila (M. Ashburner, H. Carson & J. N. Thompson, eds.), vol. 3, pp. 347–394. Academic Press (London).

Carton, Y. & Sokolowski, M. 1992. Interactions between searching strategies of *Drosophila* parasitoids and the polymorphic behavior of their host. *Journal of Insect Behavior* 5:161–175.

Castillo, J. C., Robertson, A. E. & Strand, M. R. 2006. Characterization of hemocytes from the mosquitoes *Anopheles gambiae* and *Aedes aegypti*. *Insect Biochemistry and Molecular Biology* 36:891–903.

Cerenius, L., Lee, B. L. & Söderhäll, K. 2008. The proPO-system: pros and cons for its role in invertebrate immunity. *Trends in Immunology* 29:263–271.

Chen, W.-H., Trachana, K., Lercher, M. J. & Bork, P. 2012. Younger genes are less likely to be essential than older genes, and duplicates are less likely to be essential than singletons of the same age. *Molecular Biology and Evolution* 29:1703–1706.

Chintapalli, V. R., Wang, J. & Dow, J. A. T. 2007. Using flyatlas to identify better *Drosophila melanogaster* models of human disease. *Nature Genetics* 39:715 – 720.

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.

Colinet, D., Schmitz, A., Depoix, D., Crochard, D. & Poirié, M. 2007. Convergent use of RhoGAP toxins by eukaryotic parasites and bacterial pathogens. *PLoS Pathogens* 3:e203.

Cordero, O. X. & Hogeweg, P. 2007. Large changes in regulome size herald the main prokaryotic lineages. *Trends in Genetics* 23:488–493.

Dambroski, H. R. & Feder, J. L. 2007. Host plant and latitude-related diapause variation in *Rhagoletis pomonella*: a test for multifaceted life history adaptation on different stages of diapause development. *Journal of Evolutionary Biology* 20:2101–2112.

DeGregorio, E., Spellman, P. T., Tzou, P., Rubin, G. M. & Lemaitre, B. 2002. The Toll and IMD pathways are the major regulators of the immune response in *Drosophila*. *The EMBO Journal* 21:2568–2579.

Dietzl, G., Chen, D., Schnorrer, F., Su, K., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S. *et al.* 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* 448:151–156.

Dishaw, L. J., Ota, T., Mueller, M. G., Cannon, J. P., Haire, R. N., Gwatney, N. R., Litman, R. T. & Litman, G. W. 2010. The basis for haplotype complexity in VCBPs, an immune-type receptor in amphioxus. *Immunogenetics* 62:623–631.

Dong, Y., Taylor, H. E. & Dimopoulos, G. 2006. AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLos Biology* doi:10.1371/journal.pbio.0040229.

Duffy, J. B. 2002. Gal4 system in *Drosophila*: A fly geneticist's swiss army knife. *Genesis* 34:1–15.

Duret, L. & Mouchiroud, D. 2000. Determinants of substitution rates in mammalian genes: Expression pattern affects selection intensity but not mutation rate. *Molecular Biology and Evolution* 17:68–74.

Eddy, S. R. 2012. The C-value paradox, junk DNA and ENCODE. *Current Biology* 22:R898–R899.

Ehrlich, P. & Raven, P. 1964. Butterflies and plants: a study in coevolution. *Evolution* 18:586–608.

Ekblom, R. & Galindo, J. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107:1–15.

Ellers, J., Kiers, E. T., Currie, C. R., McDonald, B. R. & Visser, B. 2012. Ecological interactions drive evolutionary loss of traits. *Ecology Letters* 15:1071–1082.

Ellis, J. S., Turner, M. L. & Knight, E. M. 2012. Patterns of selection and polymorphism of innate immunity genes in bumblebees (Hymenoptera: Apidae). *Genetica* 140:205–217.

Eslin, P. & Doury, G. 2006. The fly *Drosophila subobscura*: A natural case of innate immunity deficiency. *Developmental and Comparative Immunology* 30:977–983.

Eslin, P. & Prévost, G. 1998. Hemocyte load and immune resistance to *Asobara tabida* are correlated in species of the *Drosophila melanogaster* subgroup. *Journal of Insect Physiology* 44:807–816.

Eslin, P. & Prévost, G. 2000. Racing against host's immunity defenses: a likely strategy for passive evasion of encapsulation in *Asobara tabida* parasitoids. *Journal of Insect Physiology* 46:1161–1167.

Eslin, P., Prévost, G., Havard, S. & Doury, G. 2009. Immune resistance of *Drosophila* hosts against *Asobara* parasitoids: Cellular aspects. In: *Parasitoids in* Drosophila (G. Prévost, ed.), vol. 70 of *Parasitoids in* Drosophila, pp. 189–215. Advances in Parasitology.

Fauverque, M.-O. & Williams, M. J. 2011. *Drosophila* cellular immunity: a story of migration and adhesion. *Journal of Cell Science* 124:1373–1382.

Fay, J. & Wittkopp, P. 2008. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* 100:191–199.

Feder, J. L. 1995. The effects of parasitoids on sympatric host races of *Rhagoletis pomonella* (Diptera: Tephritidae). *Ecology* 76:801–813.

Feder, M. E. & Mitchell-Olds, T. 2003. Evolutionary and ecological functional genomics. *Nature Reviews Genetics* 4:649–655.

Felsenstein, J. 2005. Phylip (phylogeny inference package) version 3.6.

Ferrandon, D., Imler, J.-L., Hetru, C. & Hoffmann, J. A. 2007. The *Drosophila* systemic immune response: sensing and signalling during bacterial and fungal infections. *Nature Reviews Immunology* 7:862–874.

Flatt, T. & Heyland, A. (eds.) 2011. *Mechanisms of Life History Evolution.* Oxford University Press, 1 edn.

Fleury, F., Gibert, P., Ris, N. & Allemand, R. 2009. Ecology and life history evolution of frugivorous *Drosophila* parasitoids. In: *Parasitoids in* Drosophila (G. Prévost, ed.), vol. 70, pp. 3–44. Advances in Parasitology.

Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C. & Feldman, M. W. 2002. Evolutionary rate in the protein interaction network. *Science* 296:750–752.

Fussmann, G., Loreau, M. & Abrams, P. A. 2010. Eco-evolutionary dynamics of communities and ecosystems. *Nature Reviews Immunology* 10:543–553.

Gerlach, J. (ed.) 2009. *The Diptera of the Seychelles.* Pensoft, Sofia, Bulgaria.

Gerritsma, S., de Haan, A., van de Zande, L. & Wertheim, B. 2013. Natural variation in differentiated hemocytes is related to parasitoid resistance in *Drosophila melanogaster*. *Insect Physiology* 59:148–158.

Gillespie, J. P. & Kanost, M. R. 1997. Biological mediators of insect immunity. *Annual Review in Entomology* 42:611–643.

Godfray, H. C. J. 1994. *Parasitoids: Behavioral and Evolutionary Ecology.* Princeton University Press.

Gould, S. J. & Lewontin, R. C. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist program. *Proceedings of the Royal Society of London B* 205:581–598.

Guindon, S., Dufayard, J., Lefort, V., Anisimova, M., Hordijk, W. & O., G. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* 59:307–321.

Haddrill, P. R., Bachtrog, D. & Andolfatto, P. 2008. Positive and negative selection on noncoding DNA in *Drosophila simulans. Molecular Biology and Evolution* 25:1825–1834.

Hallegger, M., Llorian, M. & Smith, C. W. J. 2010. Alternative splicing: global insights. *The FEBS Journal* 277:856–866.

Han, Q., Fang, J., Ding, H., Johnson, J. K., Christensen, B. M. & Li, J. 2002. Identification of *Drosophila melanogaster yellow-f* and *yellow-f2* proteins as dopachrome-conversion enzymes. *Bioquemistry Journal* 368:333–340.

Hardison, R. C. 2003. Comparative genomics. *PLoS Biology* 1:e58.

Havard, S., Doury, G., Ravallec, M., Brehélin, M., Prévost, G. & Eslin, P. 2012. Structural and functional characterization of pseudopodocyte, a shaggy immune cell produced by two *Drosophila* species of the *obscura* group. *Developmental and Comparative Immunology* 36:323–331.

Havard, S., Eslin, P., Prévost, G. & Doury, G. 2009. Encapsulation ability: Are all *Drosophila* species equally armed? an investigation in the *obscura* group. *Canadian Journal of Zoology* 87:635–641.

Hawkings, B. A. & Sheehan, W. (eds.) 1994. *Parasitoid Community Ecology*. Oxford University Press, 1 edn.

Heger, A. & Ponting, C. P. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. *Genome Research* 17:1837–1849.

Hoffmann, A. A., Hallas, R., Sinclair, C. & Partridge, L. 2007. Rapid loss of stress resistance in *Drosophila melanogaster* under adaptation to laboratory culture. *Evolution* 55:436–438.

Huang, D., BT, S. & RA, L. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* 4:44–57.

Hughes, A. & Nei, M. 1989. Nucleotide substitution at Major Histocompatibility Complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Sciences* 86:958–962.

Hughes, A. L. & Yeager, M. 1998. Natural selection at Major Histocompatibility Complex loci of vertebrates. *Annual Review of Genetics* 32:415–435.

Hurst, G. D. & Hutchence, K. J. 2010. Host defence: Getting by with a little help from our friends. *Current Biology* 20:806–808.

Irving, P., Ubeda, J.-M., Doucet, D., Troxler, L., Lagueux, M., Zachary, D., Hoffmann, J. A., Hetru, C. & Meister, M. 2005. New insights into *Drosophila* larval haemocyte functions through genome-wide analysis. *Cellular Microbiology* 7:335–350.

Isenbarger, T. A., Carr, C. E., Johnson, S. S., Finney, M., Church, G. M., Gilbert, W., Zuber, M. T. & Ruvkun, G. 2008. The most conserved genome segments for life detection on earth and other planets. *Origins of Life and Evolution of Biospheres* 38:517–533.

Jalvingh, K. M., Chang, P. L., Nuzhdin, S. V. & Wertheim, B. 2014. Genomic changes under rapid evolution: Selection for parasitoid resistance. *Proceedings of the Royal Society B* 281. doi:10.1098/rspb.2013.2303.

Jang, I.-H., Nam, H.-J. & Lee, W.-J. 2008. Clip-domain serine proteases in *Drosophila* innate immunity. *BMB Report* 41:102–107.

Jiggins, F. M. & Kim, K.-W. 2006. Contrasting evolutionary patterns in *Drosophila* immune receptors. *Journal of Molecular Evolution* 63:769–780.

Johnson, B. R. & Tsutsui, N. D. 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* 12:164.

Jombart, T. 2008. adegenet: a r package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405.

Jones, C. 1998. The genetic basis of *Drosophila sechellia* resistance to a host plant toxin. *Genetics* 149:1899–1908.

Kacsoh, B. Z. & Schlenke, T. A. 2012. High hemocyte load is associated with increased resistance against parasitoids in *Drosophila suzukii*, a relative of *D. melanogaster*. *PLoS ONE* 7:e34721–e3472.

Kambris, Z., Brun, S., Jang, I.-H., Nam, H.-J., Romeo, Y., Takahashi, K., Lee, W.-J., Ueda, R. & Lemaitre, B. 2006. *Drosophila* immunity: A large-scale *in vivo* RNAi screen identifies five serine proteases required for Toll activation. *Current Biology* 16:808–813.

Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J. & McInerney, J. O. 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that *ad hoc* assumptions for choice of matrix are not justified. *BMC Evolutionary Biology* 6:29.

Keebaugh, E. S. & Schlenke, T. A. 2012. Adaptive evolution of a novel *Drosophila* lectin induced by parasitic wasp attack. *Molecular Biology and Evolution* 29:565–577.

Kimbrell, D. A. & Bruce, B. 2001. The evolution and genetics of innate immunity. *Nature Review Genetics* 2:256–267.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge.

Klein, J. & Nikolaidis, N. 2005. The descent of antibody-based immune system by gradual evolution. *Proceedings of the National Academy of Sciences* 102:169–174.

Klein, J., Sato, A. & Nikolaidis, N. 2007. MHC, TSP, and the origin of species: From immunogenetics to evolutionary genetics. *Annual Review of Genetics* 41:281–304.

Kohn, M., Fang, S. & Wu, C. 2004. Inference of positive and negative selection on the 5' regulatory regions of *Drosophila* genes. *Molecular Biology and Evolution* 21:374–383.

Kondrashov, A. S. 1999. Comparative genomics and evolutionary biology. *Current Opinion in Genetics and Development* 9:624–629.

Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. 2002. Selection in the evolution of gene duplications. *Genome Biology* 3:research0008.

Koonin, E. V. 2005. Orthologs, paralogs and evolutionary genomics. *Annual Reviews in Genetics* 39:309–338.

Kopp, A. & McIntyre, L. M. 2010. Transcriptional network structure has little effect on the rate of regulatory evolution in yeast. *Molecular Biology and Evolution* pp. 1–8. doi: 10.1093/molbev/msq283.

Kraaijeveld, A. R. & van Alphen, J. J. M. 1993. Successful invasion of North America by palearctic *Drosophila* species (Diptera: Drosophilidae): a matter of immunity to local parasitoids? *Netherlandse Journal of Zoology* 43:235–241.

Kraaijeveld, A. R. & Godfray, H. C. J. 1997. Trade-off between parasitoid resistance and larval competitive ability in *Drosophila melanogaster*. *Nature* 389:278–80.

Kraaijeveld, A. R. & Godfray, H. C. J. 1999. Geographic patterns in the evolution of resistance and virulence in *Drosophila* and its parasitoids. *American Naturalist* 153:S61–S74.

Kraaijeveld, A. R. & Godfray, H. C. J. 2009. Evolution of host resistance and parasitoid counter-resistance. In: *Parasitoids in* Drosophila (G. Prévost, ed.), vol. 70 of *Parasitoids of* Drosophila, chap. 10, pp. 257–277. Advances in Parasitology, 1 edn.

Kraaijeveld, A. R., Limentani, E. C. & Godfray, H. C. J. 2001. Basis of the trade-off between parasitoid resistance and larval competitive ability in *Drosophila melanogaster*. *Proceedings of the Royal Society B* 268:259–261.

Kraaijeveld, A. R. & Wertheim, B. 2009. Costs and genomic aspects of *Drosophila* immunity to parasites and pathogens. In: *Insect infection and Immunity* (J. Rolff & S. Reynolds, eds.), chap. 12, pp. 187–205. Oxford.

Krzemien, J., Oyallon, J., Crozatier, M. & Vincent, A. 2010. Hematopoietic progenitors and hemocyte lineages in the *Drosophila* lymph gland. *Developmental Biology* 346:310–319.

Kurucz, E., Váczi, B., Márkus, R., Laurinyecz, B., Vilmos, P., Zsámboki, J., Csorba, K., Gateff, E., Hultmark, D. & Andó, I. 2007. Definition of *Drosophila* hemocyte subsets by cell-type specific antigens. *Acta Biologica Hungarica* 58:95–111.

Kurucz, E., Zettervall, C., Sinka, R., Vilmos, P., Pivarcsi, A., Ekengren, S., Hegedüs, Z., Ando, I. & D., H. 2003. Hemese, a hemocyte-specific transmembrane protein, affects the cellular immune response in *Drosophila*. *Proceedings of the National Academy of Sciences* 100:2622–2627.

Lackie, A., Tackle, G. & Tetley, L. 1985. Haemocytic encapsulation in the locust *Schistocerca gregaria* (Orthoptera) and in the cockroach *Periplaneta americana* (Dictyoptera). *Cell Tissue Research* 240:343–351.

Lagueux, M., Perrodou, E., Levashina, E. A., Capovilla, M. & Hoffmann, J. A. 2000. Constitutive expression of a complement-like protein in Toll and JAK gain-of-function mutants of *Drosophila*. *Proceedings of the National Academy of Sciences* 97:11427–11432.

Lahti, D. C., Johnson, N. A., Ajie, B. C., Otto, S. P., Hendry, A. P., Blumstein, D. T., Coss, R. G., Donohue, K. & Foster, S. A. 2009. Relaxed selection in the wild. *Trends in Ecology and Evolution* 24:487–496.

Laine, A.-L. 2009. Role of coevolution in generating biological diversity: spatially divergent selection trajectories. *Experimental Botany* 60:2957–2970.

Lambrechts, L., Fellous, S. & Koella, J. C. 2006. Coevolutionary interactions between host and parasite genotypes. *Trends in Parasitology* 22:12–16.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., A., M. P., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R. *et al.* 2007. Clustalw and ClustalX version 2. *Bioinformatics* 23:2947–2948.

Larracuente, A., Sackton, T., Greenberg, A., Wong, A., Singh, N., Sturgill, D., Zhang, Y., Oliver, B. & Clark, A. 2008. Evolution of protein-coding genes in *Drosophila*. *Trends in Genetics* 24:114–123.

Lavine, M. & Strand, M. R. 2002. Insect hemocytes and their role in immunity. *Insect Biochemestry and Molecular Biology* 32:1295–1309.

Lefévre, T., de Roode, J. C., Kacsoh, B. Z. & Schlenke, T. A. 2011. Defence strategies against a parasitoid wasp in *Drosophila* fight or flight? *Biology letters* 8:230–233.

Legal, L., Chappe, B. & Jallon, J. 1994. Molecular basis of *Morinda citriflora* toxicity on *Drosophila*. *Journal of Chemical Ecology* 20:1931–1943.

Lemaitre, B. & Hoffman, J. 2007. The host defence of *Drosophila melanogaster*. *Annual Review in Immunology* 25:697–743.

Letsou, A. & Bohmann, D. 2005. Small flies—big discoveries: Nearly a century of *Drosophila* genetics and development. *Developmental Dynamics* 232:526–528.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Subgroup, . G. P. D. P. 2009. The sequence alignment/map format and samtools. *Bioinformatics* 25:2078–2079.

Li, L., Stoeckert, C. & Roos, D. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research* 13:2178.

Litman, G. W., Rast, J. P. & Fugmann, S. D. 2010. The origins of vertebrate adaptive immunity. *Nature Reviews Immunology* 10:543–553.

Little, T. J. & Cobbe, N. 2005. The evolution of immune-related genes from disease carrying mosquitoes: diversity in a peptidoglycan- and a thioester-recognizing protein. *Insect Molecular Biology* 14:599–605.

Little, T. J., Colbourne, J. K. & Crease, T. J. 2004. Molecular evolution of *Daphnia* immunity genes: Polymorphism in a gram-negative binding protein gene and an α-2-macroglobulin gene. *Journal of Molecular Evolution* 59:498–506.

Louis, J. & David, J. 1986. Ecological specialization in the *Drosophila melanogaster* species subgroup: a case-study of *Drosophila sechellia*. *Acta Oecologica* 7:215–230.

Makino, T., Hokamp, K. & McLysaght, A. 2009. The complex relationship of gene duplication and essentiality. *Trends in Genetics* 25:152–155.

Makino, T. & Kawata, M. 2012. Habitat variability correlates with duplicate content of *Drosophila* genomes. *Molecular Biology and Evolution* 29:3169–79.

Marshall, O. J. 2004. Perlprimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* 20:2471–2472.

Maruyama, T. & Nei, M. 1981. Genetic variability maintained by mutation and overdominant selection in finite populations. *Genetics* 256:441–59.

Matsubayashi, K. W., Ohshima, I. & Nosil, P. 2009. Ecological speciation in phytophagous insects. *Entomologia Experimentalis et Applicata* 134:1–27.

Matsuo, T. 2007. Odorant-binding proteins Obp57d and Obp57e affect taste perception and host-plant preference in *Drosophila sechellia. Proceedings of National Academy of Science* 104:4996–5001.

Matzkin, L. 2012. Population transcriptomics of cactus host shifts in *Drosophila mojavensis. Molecular Ecology* 21:2428–2439.

McBride, C. 2007. Rapid evoultion of smell and taste receptor genes during host specialization in *Drosophila sechellia. Proceedings of National Academy of Science* 104:4996–5001.

McDonald, J. H. & Kreitman, M. 1991. Adaptive protein evolution at the Adh locus in *Drosophila. Nature* 351:652–654.

McKean, K., Yourth, C., Lazzaro, B. & Clark, A. 2008. The evolutionary costs of immunological maintenance and deployment. *BMC Evolutionary Biology* 8:76.

McQuilton, P., Pierre, S. E. S., Thurmond, J. & the FlyBase Consortium 2012. Flybase 101 – the basics of navigating flybase. *Nucleic Acid Research* 40:D706–14.

McTaggart, S. J., Obbard, D. J., Conlon, C. & Little, T. J. 2012. Immune genes undergo more adaptive evolution than non-immune system genes in *Daphnia pulex. BMC Evolutionary Biology* 12:63.

Meister, M. 2004. Blood cells of *Drosophila*: cell lineages and role in host defence. *Current Opinion in Immunology* 16:10–15.

Mita, S. D. & Siol, M. 2012. Egglib: processing, analysis and simulation tools for population genetics and genomics. *BMC Genetics* 13:2947–2948.

Mitsui, H., Achterberg, K. v., Nordlander, G. & Kimura, M. T. 2007. Geographical distributions and host associations of larval parasitoids of frugivorous Drosophilidae in japan. *Journal of Natural History* 41:1731–1738.

Muhlia-Almazán, A., Sánchez-Paz, A. & García-Carreno, F. L. 2008. Invertebrate trypsins: a review. *Journal of Comparative Physiology* 178:655–672.

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A., Mobarry, C. M., Reinert, K. H. J., Remington, K. A. *et al.* 2000. A whole-genome assembly of *Drosophila. Science* 287:2196–2204.

Nachman, M. W., Hoekstra, H. E. & D'Agostino, S. L. 2003. The genetic basis of adaptive melanism in pocket mice. *Proceedings if the National Academy of Sciences* 100:5268–5273.

Nam, H.-J., Jang, I.-H., Asano, T. & Lee, W.-J. 2008. Involvement of pro-phenoloxidase 3 in lamellocyte-mediated spontaneous melanization in *Drosophila*. *Molecules and Cells* 26:606–610.

Nappi, A. 1970. Cellular immune reactions of larvae of *Drosophila algonquin*. *Parasitology* 70:189–194.

Nappi, A. J. & Stoffolano, J. G. 1971. *Heterotylenchus autumnalis* hemocytic reactions and capsule formation in the host, *Musca domestica*. *Experimental Parasitology* 29:116–125.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* 70:3321–3323.

Nielsen, R., Bustamante, C., Clark, A. G., Glanowski, S., Sackton, T. B., Hubisz, M. J., Fledel-Alon, A., Tanenbaum, D. M., Civello, D., White, T. J. *et al.* 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biology* 3:e170.

Nilsen, T. W. & Graveley, B. R. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463:457–463.

van Nouhuys, S., Niemikapee, S. & Hanski, I. 2012. Variation in a host–parasitoid interaction across independent populations. *Insects* 3:1236–1256.

Nyman, T. 2010. To speciate, or not to speciate? Resource heterogeneity, the subjectivity of similarity, and the macroevolutionary consequences of niche-width shifts in plant-feeding insects. *Biological Reviews* 85:393–411.

Obbard, D. J., Callister, D. M., Jiggins, F. M., Soares, D. C., Yan, G. & Little, T. J. 2009a. The evolution of TEP1, an exceptionally polymorphic immunity gene in *Anopheles gambiae*. *BMC Evolutionary Biology* 8:274.

Obbard, D. J., Welch, J. J., Kim, K.-W. & Jiggins, F. M. 2009b. Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genetics* 5:e1000698.

Otterloo, E. V., Cornell, R. A., Medeiros, D. M. & Garnett, A. T. 2013. Gene regulatory evolution and the origin of macroevolutionary novelties: Insights from the neural crest. *Genesis* 51:457–470.

Page, M. & Cera, E. 2008. Serine peptidases: Classification, structure and function. *Cellular and Molecular Life Sciences* 65:1220–1236.

Papp, J. 1977. *Phaenocarpa (Asobara) persimilis* sp.n (Hymenoptera, Braconidae, Alysiinae) from Australia. *Opuscula Zoologica Budapest* 13:1–2.

Parham, P., Lawlor, D. A., Lomen, C. E. & Ennis, P. D. 1989. Diversity and diversification of HLA-A,B,C alleles. *Journal of Immunology* 142:3937–3950.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H. & Soldatov, A. 2009. Transcriptome analysis by strand-specific sequencing of complementary dna. *Nucleic Acids Research* 37:e123.

Patterson, R. A., Juarez, M. T., Hermann, A., Sasik, R., Hardiman, G. & McGinnis, W. 2013. Serine proteolytic pathway activation reveals an expanded ensemble of wound response genes in *Drosophila*. *PLoS ONE* 8:e61773.

Payne, S. H. & Loomis, W. F. 2006. Retention and loss of amino acid biosynthetic pathways based on analysis of whole-genome sequences. *Eukaryotic Cell* 5:272–276.

Peakall, R. & Smouse, P. E. 2012. Genalex 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539.

Pech, L. L. & Strand, M. R. 1996. Granullar cells are required for encapsulation of foreign targets by insect haemocytes. *Journal of Cell Science* 109:2053–2060.

Pelletier, F., Garant, D. & Hendry, A. 2009. Eco-evolutionary dynamics. *Transactions of the Royal Society B* 364:1483–1489.

Penn, D. J. & Ilmonen, P. 2005. Major Histocompatibility Complex (MHC). In: *eLS*, vol. 41. doi:10.1038/npg.els.0003986.

Piao, S., Song, Y.-L., Kim, J. H., Park, S. Y., Park, J. W., Lee, B. L., Oh, B.-H. & Ha1, N.-C. 2005. Crystal structure of a clip-domain serine protease and functional roles of the clip domains. *The EMBO Journal* 24:4404–4414.

Plotkin, J. B. & Kudla, G. 2011. Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics* 12:32–42.

Poisot, T., Bever, J. D., Nemri, A., Thrall, P. H. & Hochberg, M. E. 2011. A conceptual framework for the evolution of ecological specialisation. *Ecology Letters* 14:841–851.

Poulson, T. & White, W. 1969. Cave environment. *Science* 165:971–981.

Powell, J. 1997. *Progress and Prospects in Evolutionary Biology : The Drosophila Model.* Oxford Univeristy Press, 1st edn.

Prevost, G., Eslina, P., Dourya, G., Moreaub, S. & Guillota, S. 2005. *Asobara*, braconid parasitoids of *Drosophila* larvae: unusual strategies to avoid encapsulation without VLPs. *Journal of Insect Physiology* 51:171–179.

Punta, M., Coggill, P., Eberhardt, R., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J. *et al.* 2012. The pfam protein families database. *Nucleic Acid Research* 40:D290–D301.

R Development Core Team 2008. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rawlings, N. D. & Barrett, A. J. 1994. Families of serine peptidases. In: *Proteolytic Enzymes: Serine and Cysteine Peptidases* (A. J. Barrett, ed.), vol. 244 of *Methods in Enzymology*, pp. 19 – 61. Academic Press.

Rebeiz, M., Jikomesa, N., Kassnera, V. A. & Carroll, S. B. 2011. Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proceedings of National Academiy of Science* 108:10036–10043.

Reed, D. A., Luhring, K. A., StaVord, C. A., Hansen, A. K., Millar, J. G., Hanks, L. M. & Paine, T. D. 2007. Host defensive response against an egg parasitoid involves cellular encapsulation and melanization. *Biological Control* 41:214–222.

Ribeiro, C. & Brehélin, M. 2006. Insect haemocytes: What type of cell is that? *Journal of Insect Physiology* 52:417–429.

Richards, S., Liu, Y., Bettencourt, B. R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M. J., Chen, R., Meisel, R. P. *et al.* 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene, and cis-element evolution. *Genome Research* 15:1–18.

Ritz, C. & Spiess, A. N. 2008. qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics* 24:1549–1551.

Robinson, M., McCarthy, D. & Smyth, G. 2010. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.

Rossa, J., Jianga, H., Kanostb, M. R. & Wanga, Y. 2003. Serine proteases and their homologs in the drosophila melanogaster genome: an initial analysis of sequence conservation and phylogenetic relationships. *Gene* 304:117–131.

Rubin, G. M. & Lewis, E. B. 2000. A brief history of *Drosophila's* contributions to genome research. *Science* 287:2216–2218.

Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor, G. L., Miklos, Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R. *et al.* 2000. Comparative genomics of the eukaryotes. *Science* 287:2204–2215.

Rubinstein, M. & de Souza, F. S. J. 2013. Evolution of transcriptional enhancers and animal diversity. *Phylosophical Transactions of the Royal Society B* 368. doi:10.1098/rstb.2013.0017.

Sackton, T. B., Lazzaro, B. P. & Clark, A. G. 2010. Genotype and gene expression associations with immune function in *Drosophila*. *PLoS Genetics* 6:e1000797.

Sackton, T. B., Lazzaro, B. P., Evans, J. D., Dan, H. & Clark, A. G. 2007. Dynamic evolution of the innate immune system in *Drosophila*. *Nature Genetics* 39:1461–1468.

Salazar-Jaramillo, L., Paspati, A., van de Zande, L., Vermeulen, C. J., Schwander, T. & Wertheim, B. 2014. Evolution of a cellular immune response in *Drosophila*: a phenotypic and genomic comparative analysis. *Genome Biology and Evolution* 6:273–289.

Sameoto, D. & Miller, R. S. 1968. Selection of pupation site by *Drosophila melanogaster* and *D. simulans*. *Ecology* 49:177–180.

Sanford, G. M., Lutterschmidt, W. I. & Hutchison, V. H. 2002. The comparative method revisited. *BioScience* 52:830–836.

Schadt, E. E., Friend, S. H. & Shaywitz, D. A. 2009. A network view of disease and compound screening. *Nature Reviews Drug Discovery* 8:286–295.

Schlenke, T. A., Morales, J., Govind, S. & Clark, A. G. 2007. Constrasting infection strategies in generalist and specialist wasp parasitoids of *Drosophila melanogaster*. *PLoS Pathogens* 3:e158.

Schmid-Hempel, P. 2005. Evolutionary ecology of insect immune defenses. *Annual Review of Entomology* 50:529–551.

Shah, P. K., Tripathi, L. P., Jensen, L. J., Gahnim, M., Mason, C., Furlong, E. E., Rodrigues, V., White, K. P., Bork, P. & b, R. S. 2008. Enhanced function annotations for *Drosophila* serine proteases: A case study for systematic annotation of multimember gene families. *Gene* 407:199–215.

Singh, N. D., Larracuente, A. M., Sackton, T. B. & Clark, A. G. 2009. Comparative genomics on the *Drosophila* phylogenetic tree. *Annual Review in Ecology and Evolution Systematics* 40:459–480.

Sommer, S. 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology* 2:16.

Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M. & Bergelson, J. 1999. Dynamics of disease resistance polymorphism at the rpm1 locus of *Arabidopsis*. *Nature* 400:667–671.

Star, B., Nederbragt, A. J., Jentoft, S., Grimholt, U., m, M. M., Gregers, T. F., Rounge, T. B., Paulsen, J., Solbakken, M. H., Sharma, A. *et al.* 2011. The genome sequence of atlantic cod reveals a unique immune system. *Nature* 477:207–210.

Stofanko, M., Kwon, S. Y. & Badenhorst, P. 2010. Lineage tracing of lamellocytes demonstrates *Drosophila* macrophage plasticity. *PLoS ONE* 5:e14051.

Strand, M. R. & Pech, L. L. 1995. Immunological basis for compatibility in parasitoid-host relationships. *Annual Review Entomology* 40:31–56.

Streams, F. A. 1968. Defense reactions of *Drosophila* species (*Diptera: Drosophilidae*) to the parasite *Pseudeucoila bochei* (*Hymenoptera: Cynipidae*). *Annals of the Entomological Society of America* 61:158–164.

Stroschein-Stevenson, S. L., Foley, E., O'Farrell, P. H. & Johnson, A. D. 2006. Identification of *Drosophila* gene products required for phagocytosis of *Candida albicans*. *PLoS Biology* doi:10.1371/journal.pbio.0040004.

Suma, P., Mansour, R., Russo, A., Torre, I. L., Bugila, A. A. A. & Franco, J. C. 2012. Encapsulation rates of the parasitoid *Anagyrus* sp. nr. *pseudococci*, by three mealybug species (Hemiptera: Pseudococcidae). *Phytoparasitica* 40:11–16.

Tang, H. 2009. Regulation and function of the melanization reaction in *Drosophila*. *Fly* 3:105–111.

Tautz, D. & Tomislav, D.-L. 2011. The evolutionary origin of orphan genes. *Nature Reviews Genetics* 12:692–702.

Tellier, A., Moreno-Gámez, S. & Stephan, W. 2014. Speed of adaptation and genomic footprints of host–parasite coevolution under arms race and trench warfare dynamics. *Evolution* doi:10.1111/evo.12427.

Thompson, J. N. 2005. *The Geographic Mosaic of Coevolution.* The University Chicago Press, United States of America.

Vandesompele, J., Preter, K. D., Pattyn, F., Poppe, B., Roy, N. V., Paepe, A. D. & Speleman, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* 3:0034.1–0034.11.

Vey, A. 1993. Humoral encapsulation. In: *Insect Immunity* (J. Pathak, ed.), chap. 5, pp. 59–68. Springer.

Visser, B., Lann, C. L., den Blanken, F. J., Harvey, J. A., van Alphen, J. J. M. & Ellers, J. 2010. Loss of lipid synthesis as an evolutionary consequence of a parasitic lifestyle. *Proceedings of National Academy of Science* 107:8677–8682.

Wagner, A. 2007. Rapid detection of positive selection in genes and genomes through variation clusters. *Genetics* 176:2451–2463.

Wang, W., Brunet, F., Nevo, E. & Long, M. 2002. Origin of *sphinx*, a young chimeric RNA gene in *Drosophila melanogaster*. *Proceedings of National Academy of Science* 2:4448–53.

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.

Watson, F. L., Püttman-Holgado, R., Thomas, F., Lamar, D. L., Hughes, M., Kondo, M., Rebel, V. I. & Schmucker, D. 2005. Extensive diversity of Ig-Superfamily proteins in the immune system of insects. *Science* 309:1874–1878.

Wertheim, B., Kraaijeveld, A. R., Hopkins, M., Boer, M. W. & Godfray, H. C. J. 2011. Functional genomics of the evolution of increased resistance to parasitism in *Drosophila*. *Molecular Ecology* 20:932–949.

Wertheim, B., Kraaijeveld, A. R., Schuster, E., Blanc, E., Hopkins, M., Pletcher, S. D., Strand, M. R., Partridge, L. & Godfray, C. J. 2005. Genome-wide gene expression in response to parasitoid attack in *Drosophila*. *Genome Biology* 6:1–20.

Whiteman, N. K. & Pierce, N. E. 2008. Delicious poison: genetics of *Drosophila* host plant preference. *Trends in Ecology and Evolution* 9:473–478.

Wiens, J. J. 2001. Widespread loss of sexually selected traits: how the peacock lost its spot. *Trends in Ecology and Evolution* 16:517–523.

Williams, M. J. 2007. *Drosophila* hemopoiesis and cellular immunity. *The Journal of Immunology* 178:4711–4716.

Wilson, R. C. & Doudna, J. A. 2013. Molecular mechanisms of RNA interference. *Annual Review Biophysics* 42:217–239.

Wittkopp, P. J. & Kalay, G. 2012. *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Proceedings of the National Academy of Sciences* 13:59–69.

Woolhouse, M. E. J., Webster, J. P., Domingo, E., Charlesworth, B. & Levin, B. R. 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* 32:569–577.

Wray, G. A. 2007. The evolutionary significance of cis-regulatory mutations. *Nature* 8:206–216.

Wu, T. D. & Watanabe, C. K. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21:1859–1875.

Yang, Z. 2006. *Computational molecular evolution*. Oxford University Press, Oxford.

Yang, Z. 2007. Paml 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24:1586–1591.

Yang, Z. & Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* 17:32–43.

Zettervall, C.-J., Anderl, I., Williams, M. J., Palmer, R., Kurucz, E., Ando, I. & Hultmark, D. 2004. A directed screen for genes involved in *Drosophila* blood cell activation. *Proceedings of National Academiy of Science* 101:14192–14197.

Zhang, J. 2003. Evolution by gene duplication: an update. *Trends in Ecology and Evolution* 18:292–298.

Zuk, M. & Stoehr, A. M. 2002. Immune defense and host life history. *The American Naturalist* 160:S9–S22.

Summary

Ecological interactions can drastically affect the fitness of organisms in their natural environment. Among the strongest selective forces in nature are host-parasite interactions, due to their inherent antagonistic relationship. Interactions with parasites can affect the ability of organisms to survive and reproduce, having thus a great impact on their fitness and life history traits. Such interactions can leave signatures of selection in the genomes of organisms, which can be used to understand their natural history. The field of evolutionary genomics aims at comparing genomes across different clades, and can be used to study these signatures in order to infer evolutionary processes.

In this thesis, I used evolutionary genomics to study host-parasite interactions, focusing on *Drosophila* fruit flies and their parasitoids as model system. I concentrated on the defence mechanisms used by *Drosophila* species to fight parasitoid wasps that lay their eggs in *Drosophila* larvae. Parasitoids are widespread parasites among insects. They exert a strong selection pressure on their hosts, as they develop by consuming them, often before the host reaches the reproductive stage. The main defence mechanism in many insect species consists of an immune response that kills the parasitoid egg, called (melanotic) encapsulation. Although it favours the survival of the host from the lethal parasitoid attack, this response is not ubiquitous among all *Drosopila* species, and shows large variation within and among species. To address the evolution and genomic basis of this immune response, I studied the molecular and cellular mechanisms of encapsulation and the effects on the host's fitness, using a combination of experiments, comparative genomics and field observations across different *Drosophila* populations and species.

I first characterized the inter-specific differences in the ability to survive the wasp parasitization across eleven fully sequenced *Drosophila* species. I showed that encapsulation and the production of a type of specialized blood cells, lamellocytes, are restricted to a sub-lineage of *Drosophila*, but that encapsulation is absent (and lamellocytes present) in one species of this sub-lineage, *D. sechellia*. I then identified genomic changes associated with the gains and losses of parasitoid resistance through a comparative genomic analysis on genes annotated with hemopoiesis function and on genes differentially expressed during the immune response against parasitoids. The comparative genomic analysis revealed that hemopoiesis-associated genes are highly conserved and present in all species independently of their ability to produce lamellocytes and to resist parasitoids. In contrast, eleven genes that are differentially expressed during the immune response to parasitoids are novel genes, specific to the *Drosophila* sub-lineage capable of lamellocyte-mediated encapsulation, of which five genes also showed signatures of positive selection. Three of these novel genes exhibited large-scale and presumably loss-of-function sequence changes in *D. sechellia*, consistent with the loss of resistance in this species. The combined phenotypic and genomic characterization lead us to propose that the resistance against parasitoids evolved at least once within the *Drosophila* genus, and probably evolved independently multiple times across insects.

Large variation in the ability to encapsulate parasitoid eggs is not only found among *Drosophila* species, but also among populations within one species. To investigate the rela-

tionship between genetic divergence among species and levels of polymorphisms among populations, I zoomed in on specific genes coding for immune receptors. Immune receptors are at the interface of host-pathogen interactions and can therefore be strongly affected by co-evolutionary processes. DNA fragments of five immune receptors were sequenced in eight field lines of *D. melanogaster* that differed in resistance against the parasitoid wasp *Asobara tabida*. Of the five genes only one, *Tep1*, showed considerable sequence variation among populations, in terms of both polymorphism and heterozygosity. Additionally, variation in the expression of Tep1 was measured in four of the field lines after parasitoid attack. *Tep1* was up-regulated in all lines, with subtle differences in the timing and level of expression among lines. A RNAi knock-down of the *Tep1* gene suggested an effect of *Tep1* on the encapsulation, although its precise function remains to be determined.

To compare the inter- and intra-species variation in the immune response after parasitoid attack, a whole genome expression (RNAseq) study was performed in four closely related *Drosophila* species of the melanogaster subgroup and in lines of *D. melanogaster* experimentally selected for increased resistance. Comparison at the inter-species level confirmed that the differentially expressed genes included a large set of genes that are lineage-restricted to the melanogaster subgroup, supporting the hypothesis that the ability to immunologically resist against parasitoid attack relies largely on new genes. Some of these genes showed no up-regulation or expression in *D. sechellia*, consistent with its loss of immunological response against parasitoids. The expression of other genes was similar among the species that could resist, while we also found substantial differences among species in the level and timing of expression. At the intra-species level, no constitutive induction of immune or hemopoiesis genes was found, except for one lamellocyte-specific gene. This gene showed a constitutively increased expression in selected lines for higher resistance. Variation between selected and non-selected lines was primarily found in the differential expression of transcript isoforms after parasitization, indicating that modulation rather than pre-activation of the immune response plays an important role in inducing the defence reaction.

D. sechellia was shown to have lost the ability to encapsulate parasitoid eggs, which was accompanied by genomic changes and the (lack of) expression of several genes during the immune response. To investigate the potential evolutionary process underlying this trait loss, we investigated ecological factors that may have driven this process. *D. sechellia* is a species endemic to the Seychelles islands, where it feeds and breeds on the fruit *Morinda citriflora* (noni). The noni fruit is known to be toxic in its ripe stage to the sister species *D. melanogaster* and *D. simulans*. We investigated the hypothesis that the noni fruit may also be toxic to parasitoid wasps, hence providing an enemy-free space that may eventually have lead to the loss of the encapsulation ability. In laboratory experiments, we showed that fly larvae in substrate with noni extract were indeed more likely to escape parasitism than in standard substrates, due to increased mortality of parasitoid wasp. A field study on Cousin island in the Seychelles showed that among the locally occurring *Drosophila* species, only *D. sechellia* developed on noni fruit throughout the different ripening stages. At later stages of decomposition, a slightly larger diversity of insects was attracted to noni, including at least one more species of *Drosophila* and parasitic wasps of the *Leptopilina* genus. These results suggest that *D. sechellia* specialization on noni provided this species with an environment of reduced competition and risk of parasitoid attack. This, in turn, may have led to relaxed selection pressure for maintaining the immune response against parasitoid attack.

In conclusion, I characterized the genomic and physiological components that enabled some *Drosophila* species to mount a successful immune response to a lethal parasitoid infection, how this response is modulated among populations and lines of one species and how it was lost in a particular ecological setting. More generally, the findings described here contribute to

understand the genomic basis of evolving a response in an ecological interaction. They also highlight the power of combining evolutionary genomics with ecological information to infer the evolution forces that have shaped the natural history of organisms.

Samenvatting

Ecologische interacties kunnen de fitness van organismes in hun natuurlijke omgeving drastisch beïnvloeden. De interacties tussen parasieten en hun gastheren horen bij de sterkste selectiekrachten in de natuur, omdat zulke interacties altijd tegenovergestelde fitness effecten veroorzaken in de parasiet en gastheer: wat gunstig is voor de een is ongunstig voor de ander. Het vermogen van organismes om te overleven en zich voort te planten kan sterk beïnvloeden worden door interacties met parasieten. Dit kan dus een groot effect hebben op die eigenschappen van organismes die met hun levensloop te maken hebben, oftewel hun life-history traits. Zulke interacties kunnen een "voetafdruk" achterlaten in het genoom, wat gebruikt kan worden om de historie en evolutie van organismes, en hun eigenschappen, te achterhalen. Het vakgebied "evolutionaire genomica" is erop gericht het genoom van verschillende clades (groepen soorten met dezelfde afstamming) te vergelijken, en kan daarbij deze "voetafdrukken" van selectiekrachten bestuderen, met als doel de onderliggende evolutionaire processen op te helderen.

In dit proefschrift heb ik evolutionair-genomische technieken gebruikt om de interacties tussen parasieten en hun gastheren te bestuderen. Ik heb hierbij fruitvliegen uit het genus *Drosophila* en de sluipwespen die hen parasiteren als modelsysteem gebruikt. Ik heb mij toegelegd op het verdedigingsmechanisme gebruikt door *Drosophila* fruitvliegen om zich te verweren tegen sluipwespen, die hun eieren in de larven van *Drosophila* vliegen leggen. Sluipwespen zijn wijdverbreide parasieten van insecten. Ze oefenen een sterke selectiekracht uit op hun gastheer, omdat ze tijdens hun ontwikkeling de gastheer opeten, vaak voordat deze gastheer zich heeft kunnen voortplanten. Het voornaamste verdedigingsmechanisme in veel insectensoorten bestaat uit een immuunreactie dat het ei van de parasitaire wesp onschadelijk maakt, genaamd (melanotische) inkapseling. Hoewel het de overlevingskans van de gastheer vergroot bij een anderszins dodelijke aanval van de parasiet, is deze immuunreactie niet alomtegenwoordig in alle *Drosophila* soorten en is er veel variatie in deze eigenschap, zowel binnen als tussen soorten. Om de evolutie en genomische basis van deze immuunreactie te bestuderen, heb ik gekeken naar de moleculaire en cellulaire mechanismes van inkapseling en de effecten hiervan op de fitness van de gastheer. Hiervoor gebruikte ik een combinatie van experimenten, genomische vergelijkingen en observaties in de natuur aan verschillende *Drosophila* populaties en soorten.

Als eerste karakteriseerde ik in 11 *Drosophila* soorten het vermogen om een aanval door een sluipwesp te overleven. Ik liet zien dat het vermogen tot inkapseling en de productie van één type gespecialiseerde bloedcellen, lamellocyten, beperkt is tot één ondergroep binnen het genus *Drosophila*, en dat in één soort binnen deze ondergroep, *D. sechellia*, het vermogen tot inkapseling afwezig is (maar lamellocyten wel aanwezig zijn). Vervolgens identificeerde ik de genomische veranderingen die geassocieerd zijn met het verkrijgen, en het verliezen, van afweer tegen sluipwespen. Dit deed ik door een vergelijkende genomische analyse toe te passen op genen die geannoteerd zijn met een functie in hemopoiesis (bloedcel productie) en op genen die tot expressie komen gedurende de immuunreactie na parasitering door sluipwespen. De vergelijkende genomische analyse toonde aan dat genen die geassocieerd zijn met hemopoiesis

in hoge mate gelijk gebleven zijn en aanwezig zijn in alle soorten, onafhankelijk van hun vermogen om lamellocyten te produceren of om zich te verdedigen tegen de sluipwesp. In tegenstelling hiermee, zijn er onder de genen die verschillend tot expressie komen gedurende de immuunreactie tegen de sluipwesp elf nieuwe genen. Deze genen komen specifiek voor in de ondergroep van *Drosophila* soorten die in staat zijn zich de verdedigen tegen parasitering door middel van inkapseling van het sluipwesp-ei met lamellocyten. In vijf van deze genen waren ook sporen van positieve selectie zichtbaar. Drie van deze nieuwe genen lieten in *D. sechellia* grote veranderingen in de DNA sequentie zien, die vermoedelijk geassocieerd zijn met verlies van de functie van het gen. Dit is consistent met het verlies van het verdedigingsmechanisme tegen sluipwespen in deze soort. De combinatie van de fenotypische en genomische karakterisatie van de soorten leidt ons tot het voorstel dat de mmuunrespons tegen sluipwespen tenminste eenmaal is geëvolueerd binnen het *Drosophila* genus, en waarschijnlijk binnen de insecten meerdere malen onafhankelijk van elkaar is geëvolueerd.

Niet alleen tussen *Drosophila* soorten wordt er veel variatie gevonden in het vermogen eieren van sluipwespen in te kapselen, maar ook tussen populaties binnen één soort. Om te onderzoeken of de genetische variatie tussen soorten en de hoeveelheid polymorfisme tussen populaties van dezelfde soort aan elkaar gerelateerd is, heb ik een aantal specifieke genen die coderen voor immuun-receptoren nader onderzocht. Immuun-receptoren zijn aanwezig op het raakvlak van parasiet-gastheer interacties, en kunnen daarom sterk worden beïnvloed door co-evolutionaire processen. In acht veldlijnen van *D. melanogaster*, die verschilden in resistentie tegen de sluipwesp *Asobara tabida*, hebben we de DNA volgorde bepaald van delen van vijf immuun-receptoren. Alleen in één van deze vijf genen, *Tep1*, vonden we aanzienlijke variatie in de DNA sequentie van het gen tussen populaties, zowel in polymorfismes als in heterozygositeit. Bovendien hebben we in vier van de veldlijnen variatie in de expressie van *Tep1* gemeten na parasitering door een sluipwesp. Expressie van *Tep1* was in alle lijnen hoger na parasitatie, met subtiele verschillen in de timing en hoeveelheid van expressie tussen lijnen. Het uitschakelen van expressie van *Tep1* door middel van RNAi knockdown suggereerde een effect van *Tep1* op inkapselingsvermogen, hoewel de precieze functie van dit gen nog nader bepaald moet worden.

Om de variatie in het verdedigingsmechanisme na parasitatie door een sluipwesp zowel binnen een soort (intraspecifiek), als tussen soorten (interspecifiek) te vergelijken heb ik de gen-expressie van het gehele genoom bepaald door middel van RNAseq. We vergeleken de gen-expressie na parasitering in vier nauw-verwante *Drosophila* soorten, afkomstig uit de melanogaster ondergroep, en *D. melanogaster* lijnen die experimenteel geselecteerd zijn voor hogere resistentie tegen de sluipwesp. Vergelijking van de expressie data tussen soorten bevestigde dat een grote groep van de genen die verschilden in expressie niveau, alleen voorkomt binnen de melanogaster subgroep. Deze data ondersteunt dus de hypothese dat het verdedigingsmechanisme tegen sluipwespen grotendeels afhangt van nieuwe genen. Sommige van deze genen lieten geen verhoging in expressie zien in *D. sechellia*, wat consistent is met het verlies van het vermogen zich te verdedigen tegen sluipwespen van deze soort. De veranderingen in expressie van andere genen kwamen overeen tussen de soorten die resistentie vertonen, terwijl er ook genen gevonden zijn die substantiële verschillen in expressie en het tijdsverloop van expressie lieten zien tussen deze soorten. Binnen de *D. melanogaster* lijnen, ofwel intraspecifiek, gaf de vergelijking geen indicatie voor een constitutieve verhoging in expressie van immuungenen of van genen die betrokken zijn bij de hemopoiesis, behalve voor één gen, die specifiek tot expressie komt in lamellocyten. Dit lamellocyten-specifieke gen liet een consistente verhoging van expressie zien binnen de lijnen die experimenteel geselecteerd waren voor hogere resistentie tegen sluipwespen. Variatie in expressie tussen de selectie lijnen en de niet-geselecteerde lijnen werd voornamelijk gevonden in ongelijke expressie van splice varianten na parasitatie,

wat er op wijst dat het moduleren van de immuunreactie, en niet de pre-activatie van het afweersysteem, een belangrijke rol speelt in het induceren van de afweerreactie binnen deze soort.

D. sechellia heeft het vermogen verloren om sluipwespen-eitjes in te kapselen. Dit lijkt samen te zijn gegaan met substantiële veranderingen in het genoom en met verandering in (of de afwezigheid van) gen-expressie tijdens de immuun-reactie. Om het evolutionaire proces te onderzoeken dat dit verlies van een eigenschap mogelijk teweeg heeft gebracht, zijn we op zoek gegaan naar de ecologische factoren die dit proces mogelijk aangedreven hebben. *D. sechellia* is een endemische soort op de Seychellen eilanden groep, waar het foerageert en zich voortplant op het fruit *Morinda citriflora* (ook wel noni fruit genoemd). Als het noni fruit rijp is, produceert het een gif wat schadelijk is voor de nauw verwante soorten *D. melanogaster* en *D. simulans*. We hebben de hypothese getest dat het noni fruit mogelijk ook toxisch is voor sluipwespen, en dus een zone creëert voor de fruitvlieg, waarin het gevrijwaard is van deze natuurlijke vijand, wat mogelijk geleid heeft tot het verlies van vermogen tot inkapseling. Experimenten opgezet in het laboratorium, lieten zien dat larven op een medium met noni extract een hogere kans hadden om aan parasitatie te ontkomen dan larven in een standaard medium, doordat de mortaliteit onder sluipwespen verhoogd was. Een veld-experiment op het eiland Cousin in de Seychellen, liet zien dat van de lokaal voorkomende *Drosophila* soorten, alleen *D. sechellia* zich ontwikkelde op het noni fruit, op alle stadia van rijping. Bij latere stadia van ontbinding van het fruit, werd er een iets hogere diversiteit aan insecten aangetroffen, waaronder tenminste nog een soort *Drosophila* en sluipwespen van het genus *Leptopilina*. Deze resultaten suggereren dat de specialisatie van *D. sechellia* op het noni fruit een habitat creërde met gereduceerde competitie van andere insecten en een lagere bedreiging door sluipwespen. Dit kan vervolgens geleid hebben tot een verminderde selectiekracht op het behouden van het verdedingsmechanisme tegen sluipwespen.

Samenvattend, heb ik de genomische en fysiologische componenten gekarakteriseerd die het mogelijk maken voor sommige *Drosophila* soorten om een succesvolle afweer reactie te volbrengen tegen fatale aanvallen van sluipwespen, laten zien hoe deze afweer reactie gemoduleerd is tussen populaties en lijnen afkomstig van één soort en laten zien hoe dit mechanisme verloren is gegaan binnen een bepaalde ecologische setting. De bevindingen beschreven in deze thesis dragen ook bij aan het algehele begrip van de genomische basis van een evolutionaire reactie op een ecologische interactie. Bovendien benadrukken ze de kracht die het combineren van evolutionaire genomica met de beschikbare ecologische informatie geeft, om de evolutionaire krachten te achterhalen die de natuurlijke geschiedenis van organismen vorm geeft.

Resumen

Las interacciones ecológicas pueden afectar drásticamente el valor selectivo (*fitness*) de los organismos en su ambiente natural. Dentro de las fuerzas selectivas más severas en la naturaleza, están las interacciones hospedero-parásito, debido a su relación antagónica. Las interacciones con parásitos pueden influir en la habilidad de los organismos para sobrevivir y reproducirse, generando así un fuerte impacto sobre sus rasgos de vida. Estas interacciones dejan huellas en los genomas de los organismos, las cuales podemos utilizar para entender su historia natural. El área de la genómica evolutiva consiste en estudiar estas huellas con el fin de esclarecer procesos evolutivos por medio de comparaciones entre genomas de diferentes clados.

En esta tesis utilicé la genómica evolutiva para estudiar las interacciones hospedero-parásito enfocándome en moscas del género *Drosophila.* El estudio se concentró en el mecanismo utilizado por *Drosophila* para defenderse del ataque de avispas parasitoides, las cuales depositan sus huevos en las larvas de las moscas. Los parasitoides son parásitos con una amplia distribución geográfica, que ejercen una fuerte presión de selección sobre sus hospederos, ya que los consumen al desarrollarse, incluso antes de que el hospedero alcance su madurez reproductiva. En muchas especies de insectos el principal mecanismo de defensa consiste en una respuesta inmune llamada encapsulación melanótica que ocasiona la muerte del huevo del parasitoide. Pese a que favorece la supervivencia del hospedero frente a la infección letal del parasitoide, esta respuesta no es ubicua en especies de *Drosophila*, y presenta gran variabilidad inter e intraespecífica. Con el fin de investigar las bases genómicas y evolutivas de esta respuesta inmune, en esta tesis estudié los mecanismos moleculares y celulares detrás de la encapsulación y los efectos de ésta en el valor selectivo del hospedero, utilizando una combinación de experimentos, genómica comparativa y observaciones de campo en diferentes especies y poblaciones de *Drosophila.*

El primer paso en este estudio fue caracterizar las diferencias en la habilidad para sobrevivir a la infección del parasitoide entre 11 especies de *Drosophila*, cuyos genomas completos se ecuentran secuenciados. Los resultados mostraron que la encapsulación y la producción de un tipo de células hemolinfáticas llamadas lamelocitos, están restringidas a un sub-linaje de *Drosophila*, pero que la encapsulación está ausente (y los lamelocitos están presentes) en una especie de este sub-linaje, *D. sechellia.* El siguiente paso fue identificar los cambios genómicos asociados a la presencia/ausencia de la resistencia contra los parasitoides por medio de un análisis de genómica comparativa de genes con funciones hemopoyéticas o que presentan expresión diferencial durante la respuesta inmune. El análisis genómico comparativo reveló que los genes asociados a la hemopoyesis muestran un alto nivel de conservación y están presentes en todas las especies, independientemente de su habilidad para producir lamelocitos o para resistir el ataque de los parasitoides. Por el contrario, once genes con expresión diferencial durante la respuesta inmune contra los parasitoides son genes de origen reciente, específicos al sub-linaje de *Drosophila* capaz de encapsular por medio de lamelocitos, y de los cuales cinco contienen huellas de selección positiva. Tres de estos nuevos genes exhiben diferencias a gran escala en sus secuencias que implican presumiblemente pérdida de funciones en *D. sechellia*, lo cual es consistente con la inhabilidad de esta especie para resistir la infección del

parasitoide. La caracterización genómica y fenotípica en conjunto nos condujo a proponer que la resistencia contra los parasitoides evolucionó por lo menos una vez dentro del género *Drosophila*, y probablemente múltiples veces en insectos.

La alta variabilidad en la habilidad para encapsular huevos de parasitoides no solo se encuentra entre especies de *Drosophila* sino también entre poblaciones de una misma especie. Con el fin de investigar la relación entre la divergencia genética entre especies y los niveles de polimorfismos entre poblaciones, me enfoqué en genes específicos que codifican para receptores inmunes. Los receptores inmunes están en la interfase entre las interacciones hospedero-parásito y pueden, por lo tanto, estar fuertemente influenciados por procesos co-evolutivos. Para estudiar estos procesos, se secuenciaron fragmentos de cinco receptores en ocho poblaciones de *D. melanogaster* que difieren en su resistencia a la infección del parasitoide *Asobara tabida*. De los cinco genes, solo uno, *Tep1*, presentó una variación considerable en la secuencia de DNA entre las poblaciones tanto en términos de polimorfismos como de heterocigosidad. Adicionalmente, se midió la variación en la expresión de *Tep1* después del ataque del parasitoide en cuatro de las ocho poblaciones de *Drosophila*. La expresión de *Tep1* fue inducida por el parasitoide en todas las líneas, con diferencias sutiles entre lineas en el tiempo y los niveles de expresión. Los resultados de un *knock-down* por medio de RNAi en *Tep1* sugirieron que existe un efecto de este gen en la encapsulación, aunque su función exacta aún está por determinar.

Con el objetivo de comparar la variación inter e intraespecífica en la respuesta inmune tras el ataque del parasitoide, se realizó un estudio de expresión del genoma completo (RNAseq). Para este estudio se utilizaron cuatro especies hermanas de *Drosophila* pertenecientes al subgrupo melanogaster y cepas de *D. melanogaster* seleccionadas experimentalmente para incrementar su resistencia al parasitoide. La comparación entre especies confirmó que un importante conjunto de genes con expresión diferencial está restringido al linaje del subgrupo melanogaster, apoyando así la hipótesis de que la habilidad para resistir inmunológicamente el ataque del parasitoide depende en gran medida de genes de origen reciente. Algunos de estos genes no presentaron inducción en la expresión en *D. sechellia*, lo cual es consistente con su inhabilidad para responder inmunológicamente a los parasitoides. La expresión de algunos genes fue similar entre especies capaces de resistir, aunque también se encontraron diferencias sustanciales en el tiempo y niveles de expresión. Al interior de *D. melanogaster* no se encontró inducción constitutiva en la expresión de genes inmunes o con función en hemopoyesis, excepto en un gen, cuya expresión es específica a los lamelocitos. Este gen incrementó su expresión de manera constitutiva en la cepas seleccionadas para aumentar la resistencia. Las diferencias entre cepas seleccionadas y no seleccionadas se encontraron principalmente en las isoformas transcripcionales expresadas luego de la parasitación, indicando así que la modulación más que la pre-activación del sistema inmune juega un papel importante en la reacción de respuesta.

En la especie *D. sechellia* se encontró que la habilidad para encapsular los huevos de parasitoides estaba ausente, y que esto estaba acompañado de cambios en el genoma y la (ausencia de) expresión de algunos genes durante la respuesta inmune. Con el propósito de investigar las causas evolutivas que potencialmente conllevaron a la pérdida de este rasgo, se investigaron los factores ecológicos que pudieron estar involucrados en este proceso. *D. sechellia* es una especie endémica a las islas Seychelles, donde se alimenta y reproduce en la fruta *Morinda citriflora* (noni). Esta fruta es conocida por su toxicidad en estado de madurez hacia las especies hermanas *D. melanogaster* y *D. simulans*. La hipótesis a investigar fue si el noni podría también ser tóxico para las avispas parasitoides, permitiendo un espacio libre de ataque, en el que eventualmente la capacidad para resistir desapareció. En experimentos llevados a cabo en el laboratorio, se demostró que las larvas de moscas mantenidas en sustrato de noni tuvieron un probabilidad mas alta de escapar la parasitación por parte de las avispas

que en un sustrato estándar. En el estudio de campo llevado a cabo en la isla Cousin en Seychelles se encontró que entre las especies de *Drosophila* presentes en la isla, la única que mostró un desarrollo hasta la adultez a lo largo de diferentes estadios de madurez de la fruta fue *D. sechellia*. En estadios tardíos de descomposición del noni se encontró un leve incremento en la diversidad de insectos atraídos por la fruta, incluyendo por lo menos una especie más de *Drosophila* y avispas del género *Leptopilina*. Estos resultados sugieren que la especialización de *D. sechellia* en noni le confirió un ambiente con una reducción en la competencia y riesgo de parasitación. Esto, por su parte, pudo haber relajado la presión de selección para mantener la respuesta inmune contra el ataque de avispas parasitoides.

En conclusión, en esta tesis se presenta la caracterización de los componentes genómicos y fisiológicos que le permitieron a algunas especies de *Drosophila* montar una respuesta inmune exitosa contra la infección letal de parasitoides, cómo es la modulación de esta respuesta entre poblaciones y líneas de una misma especie y como desapareció en un determinado contexto ecológico. De manera más general, los resultados que aquí se describen contribuyen a entender las bases genómicas que subyacen a la evolución de la respuesta a una interacción ecológica, y resaltan la fortaleza de combinar genómica evolutiva con información ecológica para inferir la fuerzas evolutivas que han moldeado la historia natural de los organismos.

Curriculum vitae

Laura Salazar Jaramillo was born on July 11 of 1980 in Medellín, Colombia. She attended the German School in Medellín until 1998. She started her study in Biology in 1999 in the Universidad de Antioquia (Medellín, Colombia) and graduated in 2004. During her study, she conducted research in the institute “Programa de Estudio y Control de Enfermedades Tropicales (PECET)” under the supervision of Luz Elena Velásquez on host parasite interactions between freswhater snails and trematodes. In 2006 she started a Master in Theoretical Biology and Bioinformatics in Utrecht University under the supervision of Paulien Hogeweg. During her Master she worked on two projects. The first project was on the dynamics of RNA replicators and its implications for theory on the origin of transcription co-supervised by Nobuto Takeuchi. The second project was on evolutionary rates of change of protein coding genes in *Saccharomyces cerevisiae*, co-supervised by Like Fokens and Berend Snel. She obtained her MSc degree in 2008. In 2009 she started a PhD in the group Evolutionary Genetics in Groningen University under the supervision of Bregje Wertheim and co-supervised by Leo Beukeboom and Louis van de Zande on the evolutionary genomics of the immune response to parasitoids in *Drosophila* species, the results of which are presented in this thesis

List of Publications

Salazar-Jaramillo L, Paspati A, van de Zande L, Vermeulen CJ, Schwander T, Wertheim B (2014) Evolution of a cellular immune response in *Drosophila*: a phenotypic and genomic comparative analysis. *Genome Biology and Evolution* 6(2):273-89.

Casas E, Gómez C, Valencia E, **Salazar L**, Velásquez LE (2008) Estudio de foco de paragonimosis en Fuente Clara, Robledo, área peri-urbana de Medellín, Antioquia. *Biomédica* 28(3).

Takeuchi N, **Salazar L**, Poole A, Hogeweg P (2008) The evolution of strand preference in simulated RNA replicators with strand displacement: implications for the origin of transcription. *Biology Direct* 3:33.

Salazar L, Estrada V, Velásquez LE (2006) Effect of the exposure to *Fasciola hepatica* (Trematoda: Digenea) on life history traits of *Lymnaea cousini* and *Lymnaea columella* (Gastropoda: Lymneaeidae). *Experimental Parasitology* 114(2):77-83.

Acknowledgments

Gabriel García Márquez used to say that the reason why he kept writing was "so that my friends will love me more". I agree with this view that an important motivation to pursue something, especially during hard times, is to not disappoint people you care about. So here, I would like not only thank but also "blame" some people for being part of the motivation to accomplish this project and finish this thesis (starting it was the easy part!).

The first person I would like to thank is Bregje. I really enjoyed working with you and I learned a lot from you during these years, from basic *Drosophila* handling to statistics and writing. Part of the fun in the project was to be able to develop my own ideas, and you gave me all the freedom for that. Although also critical, you were always very supportive when I wanted to try new things (including going to collect flies on a tropical island!). About two years ago you went through the most painful situation a person can go through when Tessa passed away. This happened during the time we were submitting our first paper and just after delivering the samples for RNAseq. This book is in a way touched by that tragedy, but at the same time, by your immense courage to continue with your life and your research, which I really admire. The combination of scientific rigor and a deep care for students makes people like you indispensable in universities. I am therefore very happy about your tenure position and I wish you a bright future as professor.

I want to express my gratitude to my second promotor and my co-promotor, Leo and Louis. Leo, you were actually the one who suggested me the position in Groningen. Thank you for that and also for your effort and patience in correcting my thesis (including all the typos!). Louis, you were involved in so many projects and courses, and yet you were always eager to discuss progress and papers, thank you for your enthusiasm. Tanja and Corneel, you introduced an extremely stimulating atmosphere in the group and your input was fundamental during my PhD. Thank you for all the lively scientific discussions. The experience and motivation from other post-docs (Bart, Monika, Rinaldo, Eveline) and from PhD students (Silvia, Sylvia, Kirsten, Maartje, Kim, Saleta, Tosca, Elzemiek, Wen-Wen, Wen-Juan) provided also invaluable support and fun to this research. I would also like to express my gratitude to the technicians (Rogier, Anna, Ammerins, Maria and Jelmer). The panic that would arise every time one of you was gone for a period of time reflected your essential role in the group (it hardly mattered if anyone else was gone). A special word of gratitude to Emma and Corine for their efficient management of scholarships, contracts and visas. Thanks to all the members of the Evolutionary Genetics group, it was a great pleasure to be part of it!

I would also like to thank professors Frank Jiggins, Louise de Vet and Janine Olsen, as well as the members of the defence committee, for the time and interest in reading this thesis.

During my PhD I had the opportunity to be involved in student projects, which I really enjoyed and learned a lot from. I used the EGR course to try some new crazy idea, so thank you guys (Harm, Jan-Eise, Koen, Sander, Ralph and Jesse) for being great "guinea pigs". Jelmer and Carmen, you were not officially my students but we ended up working together and your projects brought a lot of insights to my own. Angeliki, you were there during the most difficult period and we had to learn together to do qPCR, at the end you became the

expert and taught me back, thanks a lot.

Other people outside the group or university contributed to this reasearch by providing very important support. Jean-Christophe and people in his lab were essential for facilitating the space to do the RNAi experiment and for having always a big smile. Pascale Dijkers (UMCG) for information on RNAi and for providing the GAL4 lines. In the Seychelles, all the staff and volunteers in Cousin island, and especially to Alex Underwood for help in the collection of insects. In Leiden, Ken Kraaijveld and Henk Buermans for assistance in the RNAseq sequencing and analysis.

I was always told that the more north you go, the colder people are. After living in The Netherlands, I disagree with that *cliché* as I mainly encountered warmth and generosity. Kirsten, Sylvia, Kim and Maartje, after going together through this PhD journey, I hope I have been upgraded from colleague to friend (as colombian, you all were my friends right from the beginning!). In particular I have to thank my paranymphs and best office mates, Kirsten and Sylvia. I really enjoyed our office, our conversations, discussions and, most of all, our laughs. Thank you girls for your continuous support and for all the gezelligheid! Kirsten, thank you also for becoming my theater-mate to watch dance. Syl, Laars, Luna and Emma thank you for being my family in Groningen and kindly hosting me every time I came to visit.

Besides the “locals”, in Groningen I also had the pleasure to meet other foreigners. Fran, Saleta and Xulia, I really enjoyed our dinners, the conversations in Spanish, and having you as neighbours. Wen-Juan, even though we did not live together you have been part of my Groningen family, as *the housemate of my housmates is my housemate.* Now we are bonded for ever as we are going through *This* together!

Being far from the family is the hardest thing of living abroad. Luckily, there was this sparkling italian girl, who had the great idea to share a house with two southamericans without none of us having ever met. Not only did this work, but Silvia and Ruth became like my sisters (which was perfect since I grew up with my sister who was now at the other side of the world and whom I missed so much). I have the best memories of our life together in Thorbeckelaan, thank you for those beautiful years! Ruth, I know that no matter where we are, I will always be connected to you through an unpredictable link, as you are a mysterious hub in the world social network. Silvia, besides living together and working in the same group, we are also bonded by our somewhat parallel lives. I hope this parallelism does not mean a an irreconcilable physical distance, I miss our chats (its not the same through gmail), you still have to come to Colombia (and first to Scotland)!

Outside the university, I had great fun dancing in the Ballet school of Wanda Kuijpers. Thank you Wanda and your staff and students for wonderful lessons and performances!

I came the first time to The Netherlands to do my Masters thanks to Paulien Hogeweg. Paulien, I am deeply grateful with you for accepting me in your group, for your effort in finding a scholarship to fund my Masters, and, most of all, for all the insights I learned during your lectures and discussions. In the Theoretical and Bioinformatics group in Utrecht I acquired fundamental concepts and tools that were extremely useful during my PhD, and I am therefore very thankful to the BINF professors, staff and students. Particularly, thanks to Nobuto, Like and Berend for co-supervising my MSc projects. I also encountered a great atmosphere and met good friends (Ingrid, Otto, Nerminka, Nobuto, Anton) who made my first experience living abroad extremely smooth and happy. Thank you also to Maite and Inge for hosting me every time I needed to spend the night in Utrecht.

I am indebted to the van der Post family in Heemstede for their warm hospitality, and to the van der Post and Nilsson families in Botswana for beautiful holidays in the bush. Hartelijk bedankt voor alles!

In Scotland, I would like to thank Mike Ritchie and his lab for discussions and the members

of Laland's lab for social entertainment. I am also very grateful to Mark Blaxter and his lab for welcoming me as a member of their group in Edinburgh.

One of the most pleasant feelings during these years was going back to Colombia for holidays and meet my family and friends. There, *at home*, I have always been spoiled by the unconditional love of my grandparents, my aunts, uncles and cousins. Their affection has been a very important engine to accomplish my goals. *A mi familia en Colombia un agradecimiento de corazón por su apoyo incodincional y su inmenso cariño.* Holidays were also a great time (although never long enough) to see my old friends, *A mis amigos en Colombia muchas gracias por el ratico que sacaron para tomarse una cerveza o un tintico conmigo durante las vacaciones que siempre fueron mucho mas cortas de lo que hubiera deseado* (Ana, Juanse Caro y Vicky, gracias por los comentarios de la tesis y del resumen en español!)

If there is someone I would never want to let down is my parents. Their enthusiasm and endless support have been the most important inspiration. *Un agradecimiento infinito para mi papá y mamá por alcahuetas y por ser la mayor inspiración.* Living abroad also meant that I had to learn to make friends by myself because for long time I actually did not need many, I had my sister, my best friend. *Gracias a Sana por ser, además de mi hermanita, mi mejor amiga.*

Last, there is one person this thesis does not make justice to. Daniel, you are the anonymous co-author of all the chapters. Not only you had to correct my English and provide technical support (computational, scripting, latex), but you also had to discuss about genetics (which was not exactly your field), evolution and cope with my ups and downs. Little did I know that crossing the ocean to go to study also meant finding my puzzle piece!