UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Effect of different host genetic backgrounds in the response against oral bacterial infection

Diogo Filipe Serrano dos Santos Roque

Mestrado em Biologia Evolutiva e do Desenvolvimento

Dissertação orientada por:
Dr. Élio Sucena, FCUL, IGC
Dra. Sara Magalhães, FCUL

2021

Agradecimentos

Agradecer a quem nos acompanha e ajuda é sempre algo difícil de fazer, para algumas pessoas mais do que para outras, para mim muito. Como tal, apesar de esta ser a primeira secção do trabalho, foi a última a ser escrita. Creio que esta hesitação se deve ao quão pequeno um agradecimento parece ao lado de tudo aquilo que as pessoas a quem me dirijo nestes parágrafos fizeram por mim. Ainda assim tentarei ser o mais justo e minucioso possível, mas quero que tenham em mente que palavras não podem conter tudo aquilo que sinto.

Em primeiro lugar queria agradecer ao Élio por me ter recebido, nas circunstâncias que o fez, de forma tão calorosa, e por me ter dado todo o apoio que necessitei durante este tempo. Por ter sido meu orientador neste projeto, e também meu professor, está presente não só agora, mas também estará em todas as minhas futuras conquistas, pois aquilo que com ele aprendi acompanhar-me-á ao longo da minha vida profissional. Não consigo imaginar ter adquirido tanto conhecimento e experiência com outra pessoa, pois o Élio deu-me espaço para fazer o meu próprio caminho, errando e aprendendo, sem nunca me deixar perdido e sem direção. Antes de trabalhar no seu laboratório errar era algo que eu desvalorizava e repudiava, mas aprendi cedo que é não só a mais natural e eficaz, como também melhor forma de aprender e construirmos a nossa própria pessoa. Contudo, também acertei bastante neste lugar, e também ao Élio tenho de agradecer por isso, por nunca ter desistido, e por me ter sempre ajudado em todos os aspetos da vida, por muito complicada que fosse a situação. O Élio fez isto tudo durante uma pandemia que arrasou a vida de toda a gente, e fê-lo a todo o custo. Posso dizer que é não só um grande professor e cientista, mas também uma grande pessoa.

Queria também agradecer à Sara por me ter coorientado neste trabalho, e também por ter sido uma das melhores professoras que tive no meu percurso académico. A minha admiração por ela é imensurável, quer como cientista, quer como mulher. A sua força e determinação são avassaladoras e também ficarão comigo para sempre.

Special thanks to David Duneau, that although was not officially my supervisor, always helped and taught me. I particularly wanted to thank him for always pushing me, harder than anyone, to be the best that I can and to always do good science above everything. Even when I was trying my hardest, I did not always rose up to his challenges, but even then he was understanding and supportive, helping me solve whatever problem I was facing.

Deixo aos meus colegas de laboratório um agradecimento especial por tão bons companheiros e por tornarem este sítio tão especial. Obrigado Tânia por estares sempre presente quando eu precisei e por me teres ensinado tanto. Thank you Priscilla for being this walking ray of sunshine that makes everything seem bearable even when the day seems really dark. Um obrigado ainda mais especial à Catarina, ela que foi uma das melhores coisas que o mestrado me trouxe. A minha *momma bear* que independentemente das circunstâncias está sempre lá para me ajudar e guiar, e a quem eu recorro nos momentos mais difíceis. É uma das pessoas mais inteligentes e altruístas que conheço e por quem tenho mais admiração. Foi com ela que partilhei muitas das melhores aventuras que vivi, e será com ela que partilharei muitas

que ainda estão para vir. Sei que a nossa amizade é eterna e que nada, nem mesmo 1585,76 km de distância o podem mudar.

Agradeço a todas as pessoas no IGC que me ajudaram, mesmo não fazendo parte do meu laboratório ou sendo meus colegas de faculdade. Agradecimento especial ao André que foi incansável a ajudar-me, quer no meu trabalho, quer na minha vida por ter sido um bom companheiro e amigo. É em pessoas como ele que reside o futuro da ciência, no seu espírito de entreajuda, e na sua incansável dedicação em todos os aspetos da sua vida.

Quero também agradecer à minha família. Este é o agradecimento mais difícil de fazer, talvez por ser o mais importante. É por causa deles que sou quem eu sou e devo-lhes tudo. Agradeço especialmente ao meu avô Faustino que sempre acreditou em mim e que, embora tenha percorrido muitos passos desta etapa da minha vida ao meu lado, não pode estar presente no final. Obrigado ao meu irmão e melhor amigo por me aturar na vida, e por ver em mim uma pessoa que ninguém vê, nem mesmo eu.

Quero agradecer à Neide, que é como uma irmã para mim. A ligação que partilhamos é algo muito menos arbitrário que sangue, é escolha. Ela que me vê como sou, todos os meus defeitos, e mesmo assim está sempre ao meu lado. O conforto que ela oferece em todos os momentos da vida é como nada que eu tenha sentido antes. Obrigado por estares presente.

Finalmente queria agradecer às instituições que me acolheram, o Instituto Gulbenkian de Ciência e a Faculdade de Ciências da Universidade de Lisboa. À primeira por ser uma instituição de excelência científica, cheia de pessoas que formam uma comunidade científica e de trabalho aberta e inclusiva na qual me senti integrado. À segunda por ter contribuído para a minha formação académica e me ter proporcionado as melhores oportunidades de aprendizagem através de um ensino de excelência.

Sumário

Os organismos vivos interagem naturalmente no seu habitat de formas que se estendem sobre uma vasta panóplia de interações bióticas, e de entre as quais o parasitismo parece ser a mais comum e bem-sucedida. O seu sucesso deve-se maioritariamente a um delicado e dinâmico equilíbrio entre a transmissão do parasita e a sua virulência, propriedades que são usualmente atribuídas exclusivamente ao patogénio. Contudo, estas características podem variar quer natural quer experimentalmente, dependendo de condições ambientais, da genética do hospedeiro e das comunidades microbianas que constituem a microbiota. Naturalmente, para sobreviver a uma infeção parasítica o hospedeiro tem de ser apto para montar uma resposta imune competente e eficaz. Em traços gerais, a resposta imune é constituída por mecanismos de resistência, que visam limitar a infeção bloqueando ou eliminado o patogénio e reduzindo a carga parasitária, e por mecanismos de tolerância à doença, que resultam numa melhoria no vigor ou estado de saúde do hospedeiro sem implicar uma redução da carga parasitária.

A mosca da fruta, *Drosophila melanogaster*, passa grande parte do seu ciclo de vida em matéria orgânica em decomposição contactando intimamente com microrganismos, muitos deles patogénicos. Como tal, este organismo possui um vasto conjunto de respostas imunes que o protegem de bactérias, fungos, parasitas e vírus. Embora não possua sistema imunitário adaptativo, presente em vertebrados, a *Drosophila* tem sido amplamente usada como organismo modelo para estudar imunidade inata, devido a conservação evolutiva destes mecanismos. As defesas imunes da mosca da fruta consistem em: 1) Imunidade comportamental, que tem como propósito evitar contacto com microrganismos patogénicos para prevenir uma possível infeção; 2) Imunidade epitelial, que consiste em barreiras físicas que protegem contra o estabelecimento de microrganismos no interior do corpo; 3) Imunidade celular que depende da ação de células móveis e sésseis (hemócitos) presentes na cavidade corporal; 4) Imunidade humoral que tem por base a produção e secreção de péptidos antimicrobianos para a hemolinfa.

Um dos múltiplos patogénios naturais capazes de infetar *Drosophila* é *Pseudomonas entomophila,* cujo genoma se encontra totalmente sequenciado e muitos dos seus fatores de virulência identificados, tornando este microrganismo num modelo ideal para estudos de imunidade em *D. melanogaster*. *P. entomophila* causa ativação da imunidade sistémica mesmo quando ingerida, e a sua patogenicidade neste cenário deve-se à capacidade que esta bactéria tem de persistir no sistema digestivo da mosca, excretando múltiplas substâncias tóxicas que causam danos no intestino. Muitos dos seus genes associados à sua entomopatogenicidade foram identificados como, por exemplo, genes que codificam toxinas, explicando o efeito altamente nefasto que esta bactéria tem quando ingerida pela mosca da fruta.

Dada a importância do *background* genético do hospedeiro em face de uma infeção a que responde através de mecanismos de resistência e tolerância à doença, este trabalho tem como ambição compreender qual o contributo relativo de cada um destes mecanismos neste processo. Para este fim, infetámos um conjunto de 75 linhas do *Drosophila Genetic Reference Panel* (DGRP) com *P. entomophila*. O DGRP é um painel que consiste de cerca de 200 linhas isogénicas, sequenciadas na sua totalidade, que permite decompor a variabilidade genética presente numa população natural de *Drosophila,* constituindo uma poderosa ferramenta para

análises de associação genética. Usando este sistema previamente estabelecido no laboratório o nosso objetivo é caracterizar dinâmicas de infeção em linhas do DGRP por *P. entomophila.*

Com a recolha de dados relativos à sobrevivência após infeção, caracterizámos diferentes dinâmicas de sobrevivência dependentes no *background* genético do hospedeiro. Através da análise combinada de sobrevivência e dados referentes a carga parasitária recolhidos, inferimos fenótipos relativos a resistência e tolerância à doença neste conjunto de linhas do DGRP. No futuro, a extensão dos protocolos à totalidade do DGRP, juntamente com alguns ajustes que visam melhorar a qualidade dos dados recolhidos, poderá levar à identificação através de análises de associação genética (GWAS) de genes envolvidos em mecanismos quer de resistência quer de tolerância à doença.

Palavras-chave: Imunidade; *Drosophila melanogaster*; *Pseudomonas entomophila*; Resistência; Tolerância à doença.

Abstract

Organisms coexist naturally in their environment, interacting through a vast array of biotic interactions of which parasitism seems to be the most common and successful. This success comes from a balance between parasite transmission and virulence, traits which are often seen as properties of the pathogen. However, these traits can vary both experimentally or spontaneously, depending on environment, host genetics and the microbial communities forming the microbiota. Therefore, to survive infection it is particularly important that hosts are able to mount strong and competent immune responses that comprise resistance and tolerance mechanisms. The first is responsible for limiting infection by blocking or eliminating the pathogen, decreasing the parasitic load, while the second reveals itself in the increase in health status or fitness of the host without decreasing the parasitic load.

The fruit fly, *Drosophila melanogaster* spends its whole life cycle in decaying organic matter in close contact with pathogenic microorganisms. Therefore, it possesses an array of immune mechanisms to protect itself against infecting bacteria, fungi, parasites or viruses. Although lacking the presence of an adaptive immune system, present in vertebrates, *Drosophila* has been widely used as model system for the study of innate immune defences, due to their evolutionary conservation. Fruit fly immune defences consist of: 1) Behavioural immunity, which involves avoiding the pathogen in order to prevent infection or reduce pathogen exposure; 2) Physical or epithelial immunity, which is composed by physical barriers against the establishment of microorganisms in the body; 3) Cellular component, which consists in the action of free-floating and sessile blood cells (haemocytes), that circulate through the body cavity; 4) Humoral immunity, that relies on the production and secretion of antimicrobial peptides (AMPs) to the haemolymph.

One of the several natural *Drosophila* pathogens is *Pseudomonas entomophila* which has its genome fully sequenced and its virulence factors identified, making it a prime bacterial model for infection in *D. melanogaster*. *P. entomophila* is able to activate a systemic immune response in the fruit fly even through ingestion. Under oral infection, pathogenicity is strictly dependent on its ability to persist in the gut and the excretion of toxic substances that disrupt host physiology. A lot of its genes have also been associated with its entomopathogenicity. For example, genes that encode toxins explain why infection with this bacterium is so deleterious to the gut, causing irreversible damage to flies.

Given that the host genetic background influences infection outcome through defence mechanisms of resistance and disease tolerance, this work aimed at shedding light on the contributions of each mechanism upon oral infection with a pathogen. As such, we infected a subset of 75 Drosophila Genetic Reference Panel (DGRP) lines with a natural pathogen of *D. melanogaster*, *P. entomophila*. The DGRP is a panel that consists of around 200 fully sequenced isogenic lines that allows for the breakdown and analysis of the genetic variability of a natural population, constituting a powerful tool for Genome Wide Association Studies (GWAS). Using this lab established system our objective was to characterize infection dynamics of oral infection by *P. entomophila* in DGRP lines.

By gathering survival data upon infection, we were able to characterize the differences in infection outcome caused by differences in host genetic background. We also created several

categories of survival profile based on mortality dynamics over the course of infection. We also measured bacterial loads such as initial inoculum, bacterial loads upon death and set point bacterial loads, in order to characterize how the host influences pathogen dynamics during infection. By combining all the data, we were able to infer phenotypes of resistance and disease tolerance on a set of DGRP lines. In the future, amplification of the dataset to the full DGRP, together with the identified necessary adjustments to improve data quality, may lead to Genome Wide Association Studies (GWAS) and the identification of genes that are involved in the mechanisms of resistance and disease tolerance.

Keywords: Immunity; *Drosophila melanogaster*; *Pseudomonas entomophila*; Resistance; Disease Tolerance.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AMP	Antimicrobial peptide
DGRP	*Drosophila* Genetic Reference Panel
HPI	Hours post infection
OD	Optical density
BLUD	Bacterial load upon death
SPBL	Set point bacterial load
CFU	Colony forming unit

1 Introduction

1.1 Infection, resistance and tolerance

Organisms coexist naturally in their environment and interact, both intra- and interspecifically, in ways that are sometimes necessary for survival, establishing biotic interactions. Among these, a subset comprises symbiotic relations, characterized by long-term interactions between two organisms[1]. Symbiotic interactions can be: 1) mutualistic, a cooperative relationship that is mutually beneficial to both organisms; 2) commensal, when one organism benefits, while the other is generally unaffected; or 3) parasitic, an antagonistic relationship, and seemingly the most common biological interaction in the animal kingdom, in which one organism benefits at the expense of the other (Fig. 1.1)[2,3].

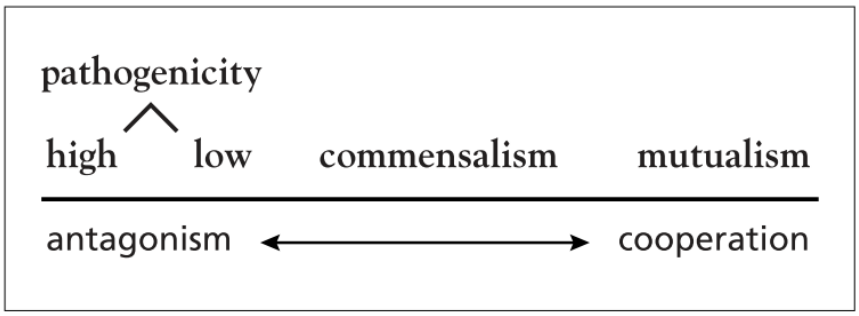


Fig. 1.1 – Continuum of symbiotic relationships. Symbiosis exists in continuum between antagonistic relationships (between host and parasite) and cooperative relationships (between two mutualists). Commensalism is represented as a mixed interaction, in which either of the organisms may benefit. Figure from G. Dimijian, 2000[1].

The success of the parasitic interaction is associated with a higher transmission rate, which is usually the result of increased parasite reproduction, consequently leading to higher virulence. Virulence can be defined as the harm done to the host by the parasite or, from an evolutionary perspective, a quantitative trait that measures the decrease in host fitness due to an infection[4,5]. As such, it is usually seen as a property of pathogens. However, it is known that virulence is not a constant property of the disease-causing pathogen, but that it can vary both experimentally or spontaneously, and it can be enhanced, lost and restored[4]. Work on virulence evolution often uses pathogen-induced mortality rates as a proxy for virulence, since it represents the extent to which the parasite induces damage on its host[6]. Yet, virulence proxies do not represent virulence as it is defined in theoretical literature, making it hard to compare data with virulence theory in order to take conclusions[7]. To go around this issue and understand the concept of virulence, one must always take into account under what ecological circumstances a microorganism is capable of causing disease in a host[4,8,9], considering that pathogenicity itself is highly dependent on the ecological context, the infection route, inoculum size and the genetic background of the host[4]. For example, the fungus *Candida albicans,* that is commonly found in humans, can be both a commensal saprophyte or a pathogen at different times depending on context[10]. Even genetically identical microorganisms (i.e. clones), like the uropathogenic *Escherichia coli*, can either be highly pathogenic or not depending on the

environment[11]. The immunocompetence of the host also plays a role in determining if an infection is pathogenic or not, as several species of microorganisms are only able to cause disease in immunocompromised organisms[12]. These examples reinforce that idea that virulence is not a fixed property of either pathogen or host, but rather a dynamical trait determined by specific host-parasite interactions in a given ecological context and by the genetic background of both pathogen and host.

The host genome is particularly relevant, as it directly impacts resistance and tolerance, the two mechanisms an organism can invoke to recover its health and increase its fitness when challenged with an infection. Resistance refers to a battery of mechanisms that directly attack the pathogen, expelling, containing or killing it, in order to block invasion or eliminate the invading microorganism, decreasing parasitic load[13] (resistance mechanisms will be further explored in the next section). On the other hand, tolerance consists of (still poorly understood) mechanisms that seem to rely on several stress and damage responses to limit the impact caused by the pathogen, decreasing the virulence of an infection, without affecting parasite load[4,14,15]. It is then important to note that, although both mechanisms contribute to promote host health, they differ in their effect on pathogen fitness: while resistance mechanisms have a negative effect on pathogen fitness, tolerance mechanisms have a neutral to positive effect. Therefore, contribution of resistance and tolerance to host's health recovery after infection could potentially be assessed experimentally by measuring a parameter of host health (e.g., survival, tissue damage) and examining its relationship with pathogen burden in the relevant tissues[16].

Ultimately, host defence is the combination and balance of these two mechanisms that are not totally independent. The best example of this dependence is immunopathology, the self-induced damage that happens when the immune response contributes to a decrease in host fitness[14]. In humans, there is evidence that immunopathology is significant for the etiology of various diseases ranked high priority by WHO, including tuberculosis, malaria, Chagas disease and leprosy[17,18]. Experiments using *Plasmodium* infection in rodents allowed the quantification of the amount of damage caused by immunopathology[19]. However, immunopathology is generally considered as having a single causative agent: either the resistance mechanisms that generate the pathology, the organisms responsible for the diseases or the organs that are affected. This mindset fails to capture how and to what extent resistance and tolerance are dependent of one another. In a review article, Janelle S. Ayres and David S. Schneider[14] stressed this issue and tackled the problem by defining classes of immune mechanisms in a spectrum of effects, spanning from those in which resistance and tolerance are absolutely linked to those where they are completely independent. The mechanisms in which there is a complete dependence between tolerance and resistance are usually associated to molecules (effectors or regulators) that take part in both resistance and tolerance pathways. Because of this, it is predicted that there is a trade-off between resistance and tolerance[14]. For example, in the *Drosophila melanogaster*'s immune response, reactive oxygen species (ROS) are a key component when fighting infections, but the oxidative activity of these molecules can also induce self-damage and even death (in extreme cases), resulting in a decrease in tolerance[20,21].

In the context of an infection, when considering only interactions between host and pathogen, one risks missing other important host-microbe interactions. In recent years it has become clear that the microbiota influences every aspect of host physiology, including outcome of infection[22]. Host-microbiota interactions are the object of constant change as some

resident microorganisms can have a role in protecting the host from invaders, while others may be the cause of some disease, depending on the context[23]. It has even been shown that changes in the genetic composition of the microbiome can severely affect disease risk[24,25] and some diseases have also been proven to be associated to changes in host microbiota (for example, bacterial vaginosis and the increased risk of acquiring HIV[22]).

In summary, although pathogens can seriously disable the host using a wide range of virulence factors (e.g., adhesins, invasins, toxins), pathogenesis is the result of an interplay between pathogen, environment, host genetics and the microbial communities forming the microbiota. In this interplay, the mounting of a strong and competent immune response is of utmost importance for survival, but can also have a significant negative impact on host fitness (through immunopathology), which could explain why these responses are often suboptimal regarding resistance mechanisms[26]. In the next section, the mechanistic activation of the immune system of the model organism *D. melanogaster* will be introduced. This is a widely used model organism in biological research, that due to widespread use across multiple research areas, is ideal to address questions related to immunology[27]. In fact, D. melanogaster stands out as a model system for invertebrate (innate) immunity, and has been used in studies in the field for decades[28–32].

1.2 The immune system of *Drosophila melanogaster*

The cellular immune response involves the action of free-floating and sessile blood cells (haemocytes), that circulate through the body cavity. These cells can be separated in three different types, based on their structure and function: plasmatocytes, crystal cells and lamellocytes. Plasmatocytes make up most circulating cells and are responsible for the removal of dead cells and microorganisms; crystal cells are present in small relative numbers among all haemocytes and are nonphagocytic cells involved in the melanization process (wound healing and hypoxic response); finally, lamellocytes are a large, flat and adherent cell type that performs encapsulation and neutralization of any object that is too large to be phagocytosed. This type of cells is exclusive of larval stages and rarely observed in healthy larvae, due to its differentiation being induced, almost solely upon infection by parasitoid wasp eggs (Fig. 1.2, top panel)[27–30].

The humoral immune response, relies on the production and secretion of antimicrobial peptides (AMPs), both systemically in the fat body, the functional equivalent of the mammalian liver and adipose tissue, and locally in the gut, trachea and reproductive tract, for example[27,35,37,38]. Studies on insect's immunity allowed for the identification of 20 immune-inducible AMPs, which have been described and classified in seven groups[27,39,40], based on their activities against different types of microbes, which will be explained in more detail next. The fact that *D. melanogaster* is such a powerful genetic tool facilitated the discovery and dissection of the mechanisms regulating AMPs production[27]. The current model states that the production of each AMP is regulated at the transcriptional level through binding of the nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) transcription factors of the Rel

family (Fig. 1.2, bottom panel)[27,41]. Different AMPs are regulated by a balanced activity of two distinct signalling pathways, Toll and Imd (for immune deficiency). The Toll pathway is mainly triggered by sensing of Lys-type peptidoglycan (characteristic of cell walls of most Gram-positive bacteria) and β-1,3-glucan (structural cell wall components of most fungi), resulting, for example, in the activation of Defensin (which is mainly active against Gram-positive bacteria) and Drosomycin (a potent antifungal AMP). On the other hand, the Imd pathway is activated by sensing of meso-diaminopimelic acid DAP-type peptidoglycan (present in Gram-negative bacteria and some Gram-positive bacteria like *Bacillus* and *Clostridium*), and results in the transcription of AMPs such as Diptericin and Drosocin (effective against Gram-negative bacteria) (Fig. 1.2, bottom panel)[27,41].

Although there is an apparent specificity associated to each pathway, the relationship between them is not fully known, with some immune-induced genes being solely dependent on one pathway (for example Diptericin for Imd, DIM1 for Toll), while others can apparently be induced by both cascades (for example Drosomycin)[27,42]. Finally, besides Toll and Imd, other pathways are also relevant for the immune response, even if not directly activated by recognition of microbial ligands. The c-Jun N-terminal kinase (JNK) and the Janus kinase/signal transducers and activator of transcription (JAK/STAT) pathways, which are responsive to tissue damage and generally associated with stress response in *D. melanogaster*[27], were also shown to be involved in the regulation of the humoral response[43]. Interestingly, the JAK/STAT pathway seems to be particularly important for defence against viruses, as deficient flies are more susceptible to infection with *Drosophila C virus* (DCV), without difference in resistance to bacterial infection[44]. Although the production of AMPs comprises most of the humoral response, the fat body also produces other immune-induced humoral factors (e.g., opsonins, drosophila immune molecules, Turandot proteins, and ROS). Altogether, the systemic antimicrobial response is a complex process that results in dramatic changes in gene expression, leading to the production of AMPs and other immune-induced factors that make up humoral immunity[27,35].

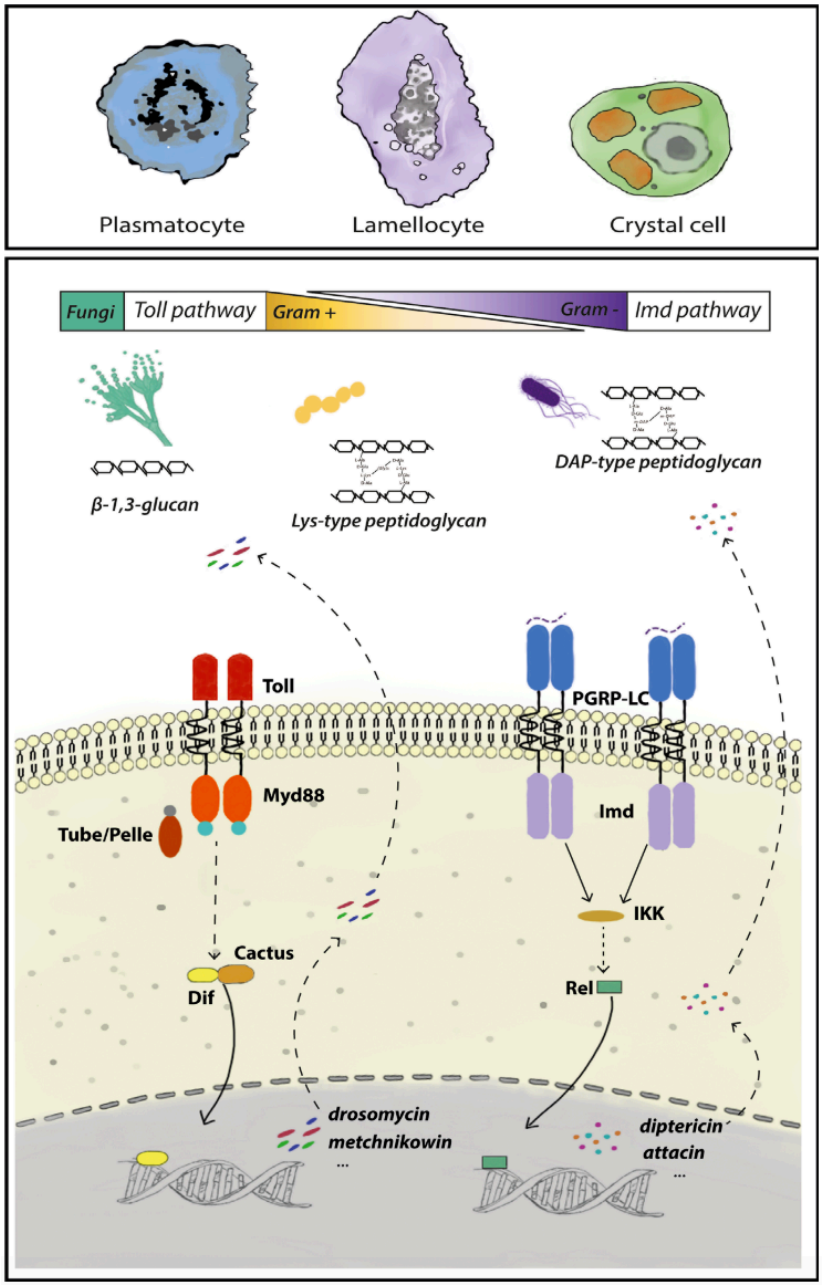


Fig. 1.2 – Cellular and humoral immunity of *D. melanogaster*. The top panel represents the three different types of haemocyte: plasmatocytes, lamellocytes and crystal cells. The bottom panel shows a simplified representation of the two canonical pathways activated by microorganisms that trigger AMP expression: the Toll pathway is triggered by sensing of Lys-type peptidoglycan (characteristic for cell walls in the majority of Gram-positive bacteria) and β-1,3-glucan (structural cell wall components of most fungi); the Imd pathway is mainly activated by sensing of meso-diaminopimelic acid DAP-type peptidoglycan (present in Gram-negative bacteria and some Gram-positive bacteria like *Bacillus* and *Clostridium*). Figure from Nunes et al., 2020[41].

As mentioned previously, the systemic induction of AMP production in the fat body is a rapid and potent process. However, in fact, the first line of humoral defence relies on the barrier epithelia, that besides working as physical barriers also provide chemical protection with the local production of AMPs and ROS. This AMP production has been demonstrated

(using GFP reporter transgenes) in multiple epithelia that contact the external environment, such as the epidermis, reproductive system, and the respiratory and digestive tracts[38,42,45]. This production of AMPs is referred to as local immune response (as opposed to the systemic response) and, surprisingly, is exclusively dependent on the Imd pathway for both Gram-negative and Gram-positive bacteria[38]. Until date, there is no evidence for the induction of local AMP production via the Toll pathway in response to Gram-positive bacteria or fungi[27]. Infections with bacteria also induce the rapid production of ROS in the gut, and the dynamic cycle of synthesis and elimination of ROS seems to be essential for survival. Furthermore, the ROS-dependent immunity is not dependent on the Imd pathway, providing an additional barrier against ingested microorganisms[46]. The immunity of the gut will be revisited in the next section of this work.

1.3 *D. melanogaster* gut morphology and immunity

The digestive tract plays an obvious and central role in nutrient absorption and, like previously mentioned, functions as one of the first lines of defence against invading microorganisms, and provides home to the microbiota. It also performs a key role in neural and endocrine regulation and signalling (modulating the activity of other organs), and maintaining energy homeostasis[37]. The gut of *D. melanogaster* consists of an epithelial layer, surrounded by visceral muscles, nerves and tracheae. Lengthwise it is divided into foregut, midgut and hindgut, and the arrangement of the different types of cells (epithelial, muscle and nervous) is different in each part. Each compartment of the gut is subdivided into even more specialized anatomical regions, each with its own distinct metabolic and digestive functions[47]. For example, the foregut contains an anatomical region, the crop, found only in adults, that functions as a storage organ and that allows them to feed on solid food much less frequently than larvae, ingesting only liquid through their proboscis most of the time. Although some digestion may occur in this region, both digestion and absorption are predominantly accomplished in the midgut[48]. Regarding physical barriers, the foregut and hindgut are lined by an impermeable cuticle, while the midgut it is covered by a peritrophic matrix[37]. Hence, the digestive tract functions as a selective barrier that protects the fly against damaging agents, such as toxins and pathogens (Fig. 1.3).

Upon ingestion, the gut content is first broken down by digestive enzymes before it is absorbed by the intestinal epithelium. *D. melanogaster* possesses a vast enzymatic repertoire, that is thought to be optimized to digest decaying fruits[37], among which are lysozymes, that are capable of hydrolysing peptidoglycan, a major component of bacterial cell walls. Although it was thought that these molecules did not play a role in immunity[30], today it has been shown that lysozymes are also produced in the fat body and tracheal system in response to infection, and have an impact in decreasing the bacterial load[49–51]. The digestion is also affected by physicochemical conditions of the gut (for example the pH)[37]. In *D. melanogaster,* different anatomical regions of the gut have different pH, which helps controlling the growth of microorganisms that are ingested with the food and facilitate the gut local immune response[37,47]. Furthermore, the enzymatic activity of the microbiota also influences

digestion[37,52]. Although fruit fly microbiota shows relatively low bacterial diversity it is known that this microbial community can be beneficial to *D. melanogaster* under poor dietary conditions, because it is often necessary to convert decaying matter into the dietary factors on which the fly depends[52]. Moreover, it is also known that gut microbial communities play an important role in infection susceptibility and outcomes, with experiments showing that some microorganisms can be beneficial upon infection by parasites, while others may be detrimental[22].

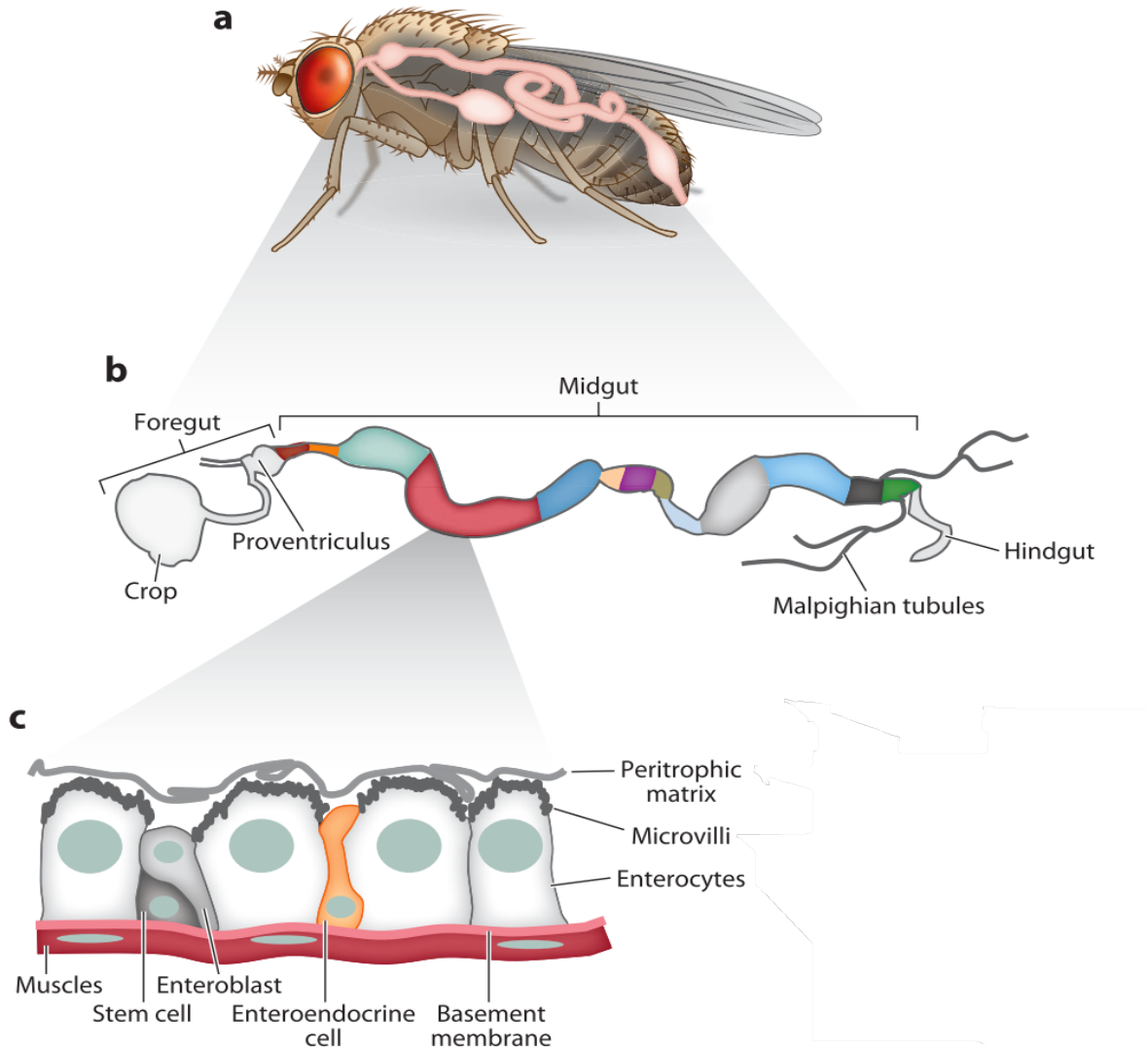


Fig. 1.3 – The *D. melanogaster* digestive tract. (a) Reconstruction of the digestive tract within the body cavity. (b) The digestive tract is divided into three domains: foregut, midgut, and hindgut, each further subdivided into genetically distinct compartments (illustrated by different colours in the case of the midgut). (c) The midgut is composed of an epithelium surrounded by two layers of visceral muscles. The midgut epithelium consists of enterocytes, enteroendocrine cells (EEC) and progenitor cells. Figure adapted from Lemaitre, B & Miguel-Aliaga, I, 2013[37].

As mentioned before, the foregut and hindgut are coated by an impermeable cuticle, while the midgut is protected by a peritrophic matrix. In terms of composition, organization and exact role of these barriers not much is known. Nevertheless, these physical barriers have an obvious role in protecting the epithelium and body cavity from potentially harmful compounds (like toxins) and pathogenic bacteria[53,54]. It is known that the peritrophic matrix is essential to survival since none of the mutant flies devoid of it are viable[53,54]. A mutant fly for the *drosocrystallin* gene (a structural element of the peritrophic matrix) shows higher susceptibility to ingested entomopathogenic bacteria due to reduced matrix thickness and higher permeability[54]. Besides these physical protective barriers, the gut is also armed with potent immune responses. Experiments show that ingestion of Gram-negative bacteria triggers the expression of AMPs in specific anatomical regions along the digestive tract[38]. Stress response programs and increased epithelial renewal are also important mechanisms in dealing with infection (usually associated with disease tolerance), that can be deployed to repair the epithelium and secure the integrity of the gut barrier[55].

Another previously mentioned line of defence is the production of ROS, most notably in the foregut and hindgut regions. Here too, there is a mechanism of negative regulation to reduce the severity of the immune response, and consequently damage caused by autoimmunity[37]. Although there are multiple evidences for the existence of a local humoral response in the gut, there is no indication of the existence of cellular immunity associated with this organ[56]. It becomes obvious that the gut is well-protected against parasitic invasion, with physical and chemical barriers and strong immune responses that result in bacterial clearance. Even the crop produces detoxifying enzymes (like glutathione S-transferase) that function as a first chemical barrier to bacteria before the passage of food to the midgut[37].

To conclude, one needs to note the importance of excretion, not only from a nutritional perspective, but also as a clearance mechanism upon ingestion of pathogenic bacteria. Experiments using food dyes revealed that in *D. melanogaster* a digestion cycle has a duration of less than one hour. However, this duration is affected by nutrient availability and even infection status, with some pathogens increasing the time it takes for food to travel the entire length of the digestive tract, and others reducing it[57]. The implications of these processes on survival are still understudied. It is also important to note that intestinal transit, defecation rate, and the contents of the excreta are subject of complex homeostatic regulation, and that alterations to this homeostatic equilibrium might impact other aspects of intestinal homeostasis, such as stem cell renewal, immunity, and senescence[37]. Since 2005, when it was first identified, *Pseudomonas entomophila* has been largely used to study *D. melanogaster*'s gut immunity, mainly due to its high pathogenicity when ingested (causing massive destruction of the gut epithelium[58]).

1.4 *Pseudomonas entomophila*, a natural pathogen of *D. melanogaster*

P. entomophila was first isolated from a single *D. melanogaster* female collected in Calvaire (Guadeloupe), and since then its genome has been fully sequenced and its virulence factors identified, making it a pristine model bacterium for infection in *D. melanogaster*[59]. This

bacterium has the unique ability to activate a systemic immune response in the fruit fly, even through ingestion[58]. The pathogenicity of this bacterium when ingested is strictly dependent on its ability to persist in the gut and the excretion of toxic substances that disrupt *D. melanogaster* physiology. Several genes in the genome of *P. entomophila* have been associated with its entomopathogenicity. For example, genes that encode for toxins, other insecticidal proteins and proteases are directly responsible for its virulence, and cell surface associated proteins are also direct contributors to pathogenicity by allowing adhesion to the surface of gut epithelial cells and consequently effective gut colonization[59].

Furthermore, in a model of septic injury (i.e., direct pricking of a culture pellet into the body cavity), *P. entomophila* kills flies in less than 24 hours. However, the outcome of oral infection is highly dependent on the bacterial concentration and can range from killing flies within 2-3 days (for OD_{600} between 50 and 100), to being non-lethal (for $OD_{600} = 1$)[58], which suggests that only high dosages of *P. entomophila* are pathogenic for *D. melanogaster* in an oral infection setting[60]. Interestingly, although behavioural avoidance plays an important role in immunity in insects, it has been shown that *D, melanogaster* does not particularly avoid the ingestion of *P. entomophila* when the bacterium is mixed in a sucrose solution[45,58,60], and it is hypothesised that the same happens in nature since this bacterium occurs in high concentrations in decaying fruit[61]. One of the most common observed phenotypes after ingestion of high doses of *P. entomophila* is the cessation of food uptake, resulting in a visible accumulation of bacteria in the fly crop[45,58]. It is still unknown if this food uptake blockage is directly induced by *P. entomophila* through the use of virulence factors or if it is a host response to intestinal damage inherent to the infection process.

Even though the gut of *D. melanogaster* is considered a hostile environment for ingested bacteria, *P. entomophila* has been shown surprisingly to remain in the digestive tract for at least 16 hours post infection (hpi), due to its ability to grow even in pH as low as 3 and its high resistance to lysozymes activity[61]. However, for *P. entomophila* to endure the adverse conditions found in the fly gut, it must survive not only to its adverse physicochemical conditions, but also the local immune defences of this organ, namely the production of AMPs, such as Diptericin, Attacin and Drosocin. Experiments show that infection with *P. entomophila* leads to the expression of AMP-encoding genes in the gut, meaning that this bacterium is specifically recognized by the local immune system[62]. There are also studies showing that preventive localized expression of AMPs in the gut confers a degree of resistance to this bacterium, indicating that *P. entomophila* is sensitive to the activity of AMPs[45]. The ability of this bacterium to resist the innate immune response is closely linked with the damage it does to the gut epithelium, since by damaging epithelial cells the bacterium is able to decrease the gut's local immune response. However, it is generally accepted that the damage *P. entomophila* causes in the gut epithelium is not driven through its direct contact with the epithelial layer[58,62], but rather through a strong activation of pathways responsible for repair and stress response in *D. melanogaster*, which are responsible for reducing the production of AMPs by gut epithelial cells, consequently increasing the survival of the bacteria present in the digestive tract[62]. This reduction of AMPs production results in the inhibition of the repair mechanisms necessary to face the damage caused by infection[60,63]. The cellular damage sustained by the epithelium is caused by two main factors: toxins produced by the bacterium and ROS produced by the host itself (which are actually the main gut damage source during the infection). This is because,

although ROS have been shown to have an important role in surviving oral infections, the production of such molecules is exacerbated during infection by *P. entomophila*[62,64]. Studies have seen epithelial damage from enterocytes displaying abnormal microvilli to cell death and noted that after 6 hpi, the mucus that protects the digestive epithelium was missing, suggesting that there might be passage of bacteria from the gut into the body cavity. However, the peritrophic matrix that lines the midgut is still present after 12 hpi, and after 6 hpi there is no bacteria in the haemolymph of infected flies[58], although bacteria could potentially cross the gut barrier at later stages.

In summary, during the interaction between *P. entomophila* and *D. melanogaster,* not only the immune system fails to eliminate the bacteria, but its activation is also immensely deleterious to the gut, causing irreversible damage to the host.

1.5 Objectives

We may be led to think that illness due to infection is somehow a straightforward process: when infected by a pathogen, an individual reacts using its immune system. This immune response either clears the pathogen and health is restored or the organism dies. However, the relationship between infection and health is extremely complex. The interaction between host resistance and disease tolerance is a dynamic process that, together with pathogen virulence, is responsible for infection outcome[65]. Greater tolerance means that individuals can be infected with larger numbers of pathogens before suffering from severe illness and death[15]. Resistance consists of the host's ability to reduce parasite load. However, the relative contribution of these two mechanisms remains clouded due to the difficulty in studying them individually. Knowing that the host genetic background plays a decisive role in infection outcome, we pretend to shed light on the individual contributions of host's resistance and tolerance upon infection.

In this work, we will infect isogenic lines of the Drosophila Genetic Reference Panel (DGRP) with the natural pathogen of *D. melanogaster*, *P. entomophila*. The DGRP allows breakdown and analysis of the genetic variability of a natural population with relation to a phenotype of choice. It is a panel that consists of around 200 fully sequenced isogenic lines constituting a powerful tool for Genome Wide Association Studies (GWAS)[66]. Using this lab established system, the main goals of this project are:

I. To explore the infection dynamics of oral infection by *P. entomophila* in DGRP lines;

II. To determine relative contributions of resistance and disease tolerance upon oral infection of DGRP lines with *P. entomophila*.

We will follow mortality of infected flies from each line over the course of three days. During this period, multiple live flies will be collected to measure bacterial loads (BL) in specific timepoints. In addition, bacterial load upon death (BLUD) will be determined. This opens the possibility for conducting Genome Wide Association Studies (GWAS) that can present candidate genes that can influence the phenotypes of resistance and disease tolerance.

2 Materials and Methods

2.1 Fly Stocks and Husbandry

For all experiments *Drosophila melanogaster* was reared under constant temperature of 25ºC (±1℃), 60–70% humidity, light-dark cycle of 12:12 hours, and fed with a standard cornmeal-agar medium. At days two or three after eclosion females were isolated in groups of 80, and remained as such until the end of the experiments. The experiments were conducted using only three to four days-old fertilized females. We used a wild-type outbred strain (Mel1) as reference[67]. The population was kept in laboratory cages for over 50 non-overlapping generations (generation time: three weeks) with high census (over 1500 individuals).

The Drosophila Genetic Reference Panel (DGRP) lines were established from a natural population collected from Raleigh, North Carolina, USA, and inbred for 20 generations of full-sib mating, followed by random mating. This is a collection composed by approximately 200 isogenic and fully sequenced lines[66]. A subset of 75 lines from the DGRP were chosen to encompass the 72 different lineages identified according to the genetic distance between them (http://dgrp2.gnets.ncsu.edu/data.html and Lafuente et al, 2018 [68]). For the families comprising more than one line we based our choice on survival to *P. entomophila* infection data from M. Bou Sleiman et al., (2015)[69] (giving priority to lines with survival data), lifetime fecundity and lifespan data from Durham et al., (2014)[70] (again giving priority to lines with fecundity and lifespan data), and the presence (or absence) of *Wolbachia*. At the end we created a subset of DGRP encompassing as much as possible the full panel´s variation in survival, lifetime fecundity and lifespan and mostly infected with *Wolbachia*. This is because the majority of the DGRP stocks in our fly facility are infected with *Wolbachia* and we wanted to have a dataset as homogeneous as possible. It has been described that *Wolbachia* infection affects fitness traits, for example boosting fecundity and longevity[71], and confers resistance to infection with viruses[72]. Although that is true, Bou Sleiman et al. found that the presence of *Wolbachia* does not have an effect on *D. melanogaster* susceptibility to oral infection with *P. entomophila*[69]. In the laboratory these lines were kept with a small census, between 10 and 30 flies. A month before each line was tested, its individuals were transferred to bottles and were amplified to between 200 and 250 individuals and kept under the same conditions.

2.2 Bacterial infection

The Gram-negative bacterium *Pseudomonas entomophila* used contains a resistance marker to rifampicin (kind gift from Bruno Lemaitre). This is a wide spectrum antibiotic that inhibits transcription by blocking the bacterial RNA polymerase. The rifampicin resistance greatly reduces the risk of contamination and is essential when plating whole flies to count colony forming units (CFU) (see below). The bacterium was kept in 35% glycerol stocks (prepared from a single bacterial clone) at -80ºC. For each round of infection, the stocks were

used to streak a LB agar petri dish, supplemented with rifampicin (0.5 mg/mL). The plate was then kept at 30ºC for one or two days, allowing for the colonies to form and grow, and after that at 4ºC to be used at any time during the following week. From the petri dish, a single colony was inoculated in 5 mL of LB liquid medium and grown between 8 and 12 hours at 28ºC in an orbital shaker incubator. The 5 mL of the starter culture were poured into 500 mL of LB liquid medium and grown to saturation overnight in the same conditions. This *P. entomophila* suspension was centrifuged for 15 min at 15000 rpm and 4ºC, and its concentration adjusted to OD_{600} = 100 by diluting the pellet in LB liquid medium (to reach approx. 50% mortality, accordingly to previous unpublished work by Nuno Martins). The suspension was then mixed, in a ratio of 1:1, with a 5% sucrose solution, and food colouring was added to a concentration of 2%. A control solution was prepared by mixing equal parts of LB liquid medium with a 5% sucrose solution (see Appendix).

To infect flies, we used bottles containing a filter paper disk soaked in bacterial solution and humidified cotton covering the plug (in order to maintain the humidity during the infection period). In each bottle, approximately 80 previously separated female flies belonging to a single DGRP line (or the outbred population) were kept for 3 hours to feed on the *P. entomophila* suspension present in the filter paper (see Appendix). For each infection bottle, a control bottle was prepared in the same way, only replacing the bacterial solution for the control solution. After the infection period the flies were anesthetized with CO_2 for no more than five minutes, pooled in groups of 10, transferred to small vials under the same food and environmental conditions. The flies remained in the same vial for three days, until the end of the experiment. From each line, 8 individual flies were collected and kept on ice to measure the initial inoculum (see Appendix). Each round of experiments comprised 19 DGRP lines and the outbred population to a total of 140 vials (6 infection plus 1 control per line/population) containing 10 flies. The outbred population was tested at every experimental round while each DGRP line was tested in three replicates. While the majority of lines were tested in the same group for every replicate, between 2 and 3 lines per group were randomly selected from the pool, in order to control for random block effects.

2.3 Survival and Bacterial Loads

In this work, we observed and collected flies in order to characterize survival and bacterial load dynamics. The bacterial loads measured in this experiment were the initial inoculum, the bacterial load upon death (BLUD), and the set-point bacterial load (SPBL)[73]. Flies collected alive were anesthetized with light CO_2 as briefly as possible and kept on ice until plating. Every fly collected for bacterial load measurements was passed through one-minute serial washes in 70% ethanol, 50% bleach, and 70% ethanol and finally with Mili-Q water, in order to limit contamination by bacteria present on the outside of the fly. After the washing, individual flies were homogenized in 50 µL of sterile PBS with a Tissuelyser. The homogenate was then diluted to 1:10, 1:100, 1:1,000, 1:10,000 and 1:100,000 in PBS, ensuring that the number of bacterial colonies was within resolution limits. We then plated droplets of

4 μL from every dilution of each homogenate onto petri dishes with LB agar and rifampicin (0,05g/mL), so that each plate contained 40 droplets. The plates were incubated overnight at 30ºC and then kept at 4ºC until CFU counting, to estimate the number of live bacteria per fly (see Appendix).

To determine the size of the initial bacteria inoculum, 8 flies were collected immediately after the 3-hour period of infection and kept on ice for approximately 2 hours (average time necessary to prepare the rest of the experiment), until plating. Since each line was tested in three replicates, this makes a total of 24 individual flies measured for each inoculum per line (with exception of the outbred population, which was repeated every round of experiments). To estimate BLUD, infected hosts were checked every 30 min between 21 and 30 hours post infection (hpi) and dead flies collected and plated. We collected a maximum number of 8 flies per line each infection round, for a total of 24 individuals per line (again with exception of the outbred population, which was repeated every round of experiments). However, that number of flies could not be collected for every DGRP line used since mortality differs from line to line. For SPBL, 8 flies were collected at 72 hpi (at the end of the experiment) and immediately plated. For each line, a total of 24 flies was plated, except for the lines in which all individuals died before 72 hpi (and again the outbred population, for the reasons above).

Survival was measured by observing each vial and counting dead individuals every 24 hours from the start of infection for three days. Because flies were checked every half hour between 21 and 30 hpi for the BLUD, survival data was also gathered in these timepoints, creating an interval with higher resolution in the trajectory. As mentioned before, every round of experiments contained 60 infected individuals plus 10 controls per DGRP line, for a total of 180 infected flies plus 30 controls across all replicates (with exception of the outbred population, which was repeated every round of experiments).

2.4 Statistical Analysis

To compare survival across DGRP lines, the proportion of individuals surviving at day 10 after infection in each vial was first estimated using the Kaplan-Meier method. In each time point, individuals that either died from causes unrelated to bacterial infection (for example, that got stuck in the food) or escaped from vials were counted as censored observations. The same applied to individuals that were alive at the end of the experiment. If in any replicate abnormal mortality was observed in the control flies, the data would be censored from the final dataset (although this never happened). Subsequently, differences in survival between DGRP lines were tested using Cox regression models in order estimate hazard ratios. The hazard ratio is a comparison between the probability of events in a treatment group, compared to the probability of events in a control group. In this case, the hazard ratio can be defined as the relative risk of death. The models included the DGRP Line as a fixed factor, and Vial Replicate nested in Experimental Block and as random effects. To test for differences in pathogen loads, Linear models on the natural logarithm of bacterial counts were applied with the DGRP Line as a fixed

factor, and Vial Replicate nested in Experimental Block as random effects. These tests were done using the R libraries *lme4*, *coxme* and *survival*. We used Spearman correlations to assess the relationship between our measured traits (survival, initial inoculum, BLUD and SPBL). To assess differences between different survival categories (regarding initial inoculum, BLUD and SPBL) we used the non-parametric Kruskal Wallis test. All statistical analyses were done using R (v 3.6.2)[74].

3 Results and Discussion

3.1 Survival

Infection of different isogenic *D. melanogaster* lines with the bacterium *P. entomophila* generates a wide scope of outcomes, with survival at the population level ranging from 94% to 0% depending on the genetic background (Fig. 3.1 A, B). This amount of variation in infection susceptibility, observed in about 40% of the whole DGRP set, evidences how extremely diverse this phenotype is in natural populations. This strikingly high and reproducible variation in infection outcome of DGRP lines upon infection with *P. entomophila* had been previously described in the literature, with survival varying between 0% and 100%[69]. This means that with our subset we were able to capture almost all the variation previously described for this phenotype in the DGRP.

Vodovar et al., (2005)[58] found that *P. entomophila* infection causes severe damage to the gut of *D. melanogaster* larvae. They described that after 6 hours post infection (hpi) the mucus that protects the digestive epithelium was missing, and after 12 hpi the epithelial cells were either absent or abnormal. These findings suggest that there might be passage of bacteria from the gut to the body cavity, however they showed that the semipermeable peritrophic matrix that lines the gut was intact throughout the infection, and after 6 hpi they did not detect bacteria in the haemolymph of the infected flies. Therefore, the current model is that this bacterium is unable to cross to the body cavity and death occurs due to gut damage (caused by secreted virulence factors), that ultimately leads to shrinking and rupture – Gut Damage model. However, the evidence for the impossibility of *P. entomophila* crossing the gut are circumstantial and time restricted, and even the authors that described the process did not discard the possibility of bacteria reaching the haemolymph in earlier stages (and being unable to persist there) nor collected data for later stages of infection. Moreover, there is a technical constraint in trying to segregate both infections (oral and systemic) spatially without contaminating the samples. As such there is the possibility that death after oral infection with *P. entomophila* is the result of gut damage combined with a secondary systemic infection caused by a percentage of bacteria that are able to cross the gut epithelial barrier into the haemolymph. This hypothesis predicts that gut damage is a necessary condition for virulence, influencing the number of bacteria and timing of infection, and therefore the extent of damage and acuteness of the "secondary" systemic infection which will additively contribute to death – Systemic model. When a host is infected by a pathogen, survival as an outcome can only

happen by activating mechanisms that directly attack the pathogen, resistance, or mechanisms that limit the damage caused by infection, tolerance. Therefore, the variation in susceptibility between DGRP lines is a direct result of differences in these two traits. However, it is impossible to know the relative contribution of tolerance and resistance to the differences in susceptibility to infection taking into account only survival.

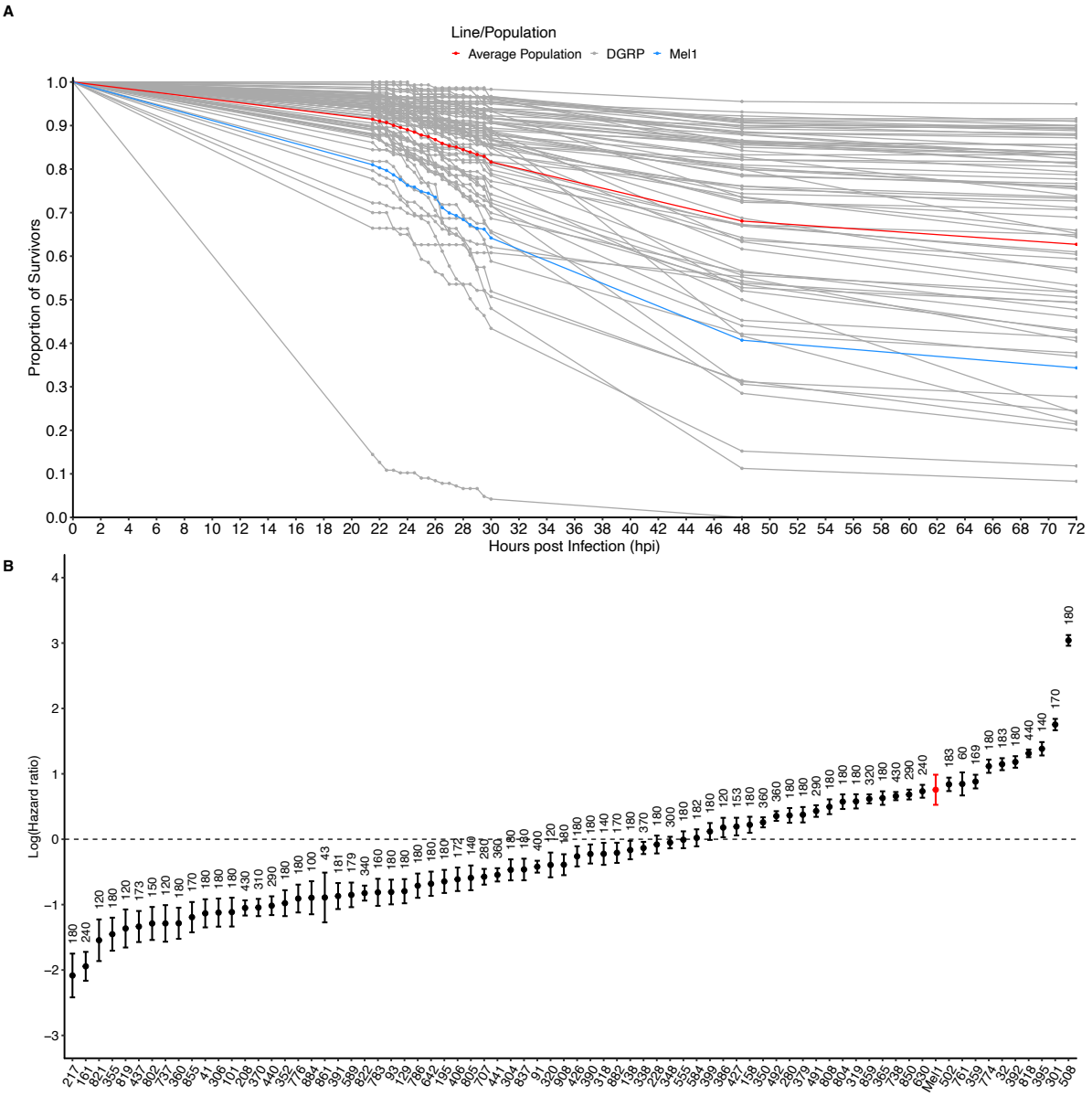


Fig. 3.1 – Oral infection of *Drosophila* Genetic Reference Panel lines with *P. entomophila* induces a wide spectrum of mortalities. (A) Survival curves for 75 selected DGRP lines and the outbred line, Mel1. The Y-axis represents the proportion of survivors and the X-axis the time post infection. Each grey curve represents a different line and each dot a timepoint in which survival was measured. The red curve corresponds to the average population (hypothetical line with the average survival of the 75 DGRP lines), and the blue curve corresponds to the outbred population, Mel1 **(B)** Hazard ratios of 75 selected DGRP lines, and the outbred population, Mel1. The Y-axis represents the hazard ratio (exponent of the coefficient in the Cox regression model) and the X-axis the line. The black dots represent the value of the hazard ratio and the bars represent the standard error. The dotted horizontal

line represents the baseline, the hypothetical average population. The sample size is represented above the error bars.

Mortality induced by *P. entomophila* infection starts before 21 hpi and increases in the vast majority of DGRP lines until 48 hpi. – with exception of DGRP-508 (in which 90% of the individuals die before 21 hpi). Because the complete DGRP contains the variation of a natural outbred population, the baseline used for calculating the hazard ratio in this work was a hypothetical average population comprising the empirical survival data of the 75 tested DGRP lines (Fig. 3.1 B). It can be noted that 59% of the tested DGRP have a lower hazard ratio than the average population (44 lines), while only 35% of the DGRP plus the outbred population, Mel1, have higher hazard ratio (26 DGRP and Mel1). Outlier analysis found DGRP-508, the line which has the highest probability of death, to be a conventional outlier (Fig. 3.2).

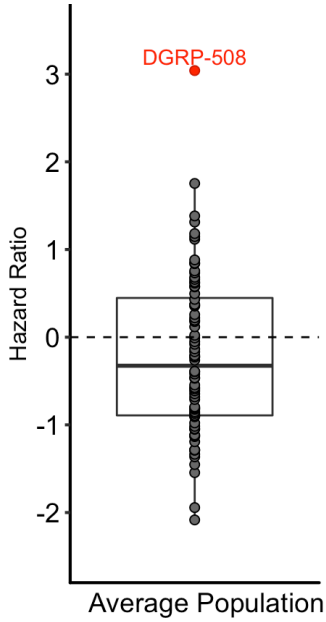


Fig. 3.2 – DGRP-508 is a conventional outlier, being extremely susceptible to infection. Boxplot showing the hazard ratio of the 75 selected DGRP lines, plus the outbred population, Mel1. Each coloured dot represents the hazard ratio of a single DGRP line. The box represents the interquartile range (the values between the upper and lower quartiles), the vertical black line the upper and lower whiskers and the horizontal black line the median. Conventional outliers are found to be above or below the whiskers. Y-axis represents hazard ratio and the X-axis the hypothetical average population (which contains all DGRP lines). The dotted horizontal line represents the baseline for the hazard ratio, the hypothetical average population.

Based on the mortality profile of the 75 DGRP lines we categorized them in different groups. The first step in doing this was comparing every DGRP line to a null model of constant mortality between 0 and 72 hpi (simplest model for mortality). Since the categorization is based on the survival dynamics, the first step was to consider all lines with survival above 85% to represent a separate category, consisting of lines with high survival (HS). For the categorization of the remaining DGRP lines, we divided the infection time in three different periods: 0-21.5

hpi, 21.5-48 hpi, and 48-72 hpi. For each of these periods we estimated the expected mortality of each line according to the null model of constant mortality. Then we compared the expected mortality with the absolute observed mortality in each timepoint. With this we assessed, for each line, if there were deviations from the model, creating three categories for each interval: L – lower observed mortality; E – equal observed mortality; H – higher observed mortality. We considered that only mortality values lower than half the expected or two times higher than the expected deviate from the model. The final category of each DGRP line is given by the combination of the three time intervals. The first letter of the category represents the time interval between 0-21.5 hpi, the second letter of the category represents the time interval between 21.5-48 hpi and the third letter of the category represents the time interval between 48-72 hpi.

The characterization of individual periods (Tables 1, 2) showed that 25% of the lines were categorized has having high survival (HS), in the time interval between 0 and 21.5 hpi 51% of the DGRP lines died as predicted in the model, in the time interval between 30-48 hpi 45% also died as predicted in the model, and in the time interval between 48-72 hpi 45% of the lines died less than predicted by the model. This means that, unlike what is predicted, in the data shows that mortality is lower the end of the experimental period. If we consider the Systemic model for mortality, this could be easily interpreted in the light of a typical bacterial growth curve: the first time period corresponds to the initial lag phase and exponential growth, the second period corresponds to the stationary phase where the number of bacteria is maximum, and the final period corresponds to the death phase of the bacterial growth curve. However, in an infection scenario the bacterial growth is constrained by the host internal environment and host mortality is not only dependent on bacterial load, meaning that the observed results could not be solely explained based on that. This becomes more evident if we consider the Gut Damage model where the bacterial load starts at the level of the stationary phase of the bacterial growth curve and decreases from that point onward. With this in mind, we need to consider that in the first interval the bacteria are establishing the infection and the virulence factors are still present in low concentrations. At the same time, the host only starts producing AMPs some hours after the beginning of the infection (lag varies depending on the AMP, with Toll-dependent AMPs expression peaking around 24 hpi and Imd-dependent AMPs expression peaking around 8 hpi[27]). The second time interval is where both virulence factors and host defence molecules reach their peak concentrations which is the point at which the host either controls the infection and starts clearing the pathogen or dies. Finally, the last interval is where the surviving hosts (given that only the ones who were able to previously control the infection reach this stage) clear the remaining bacteria, and the virulence factors are at its minimum. This over-simplification of a process as complex as infection, together with the bacterial growth dynamics, offer a possible explanation for the predominant observed pattern of lower mortality by the end of infection.

By putting together the information of all time periods, we characterized all DGRP lines according to their whole mortality profile (Tables 3.1 and 3.3). Looking at this characterization we can see that the two most predominant groups of lines correspond to the ones with high survival (HS), which represent 25% of all lines, and to the ones that follow the model of constant mortality in the first two periods of time and show lower mortality after that (EEL), which represent 21% of all lines (in accordance with the analysis of each time point separately).

These two categories (HS and EEL) together make up for 46% of the analysed lines and the remaining 54% fall into other six different categories. From these six categories it is worthwhile noting that 17% of the lines follow the model for constant mortality, and that DGRP-508 (which was identified as being a conventional outlier) fits in a category of its own (HEL). This is the only line in the whole dataset that shows higher mortality than predicted between 0-21.5 hpi which could be explained by having extremely low resistance and/or tolerance to this bacterium.

Table 3.1 – Categorization of the 75 DGRP lines according to their survival dynamics. In each time point, and for each individual DGRP, we compared the expected mortality (given by the null model of constant mortality) with the absolute observed mortality, considering that only values lower than half the expected or two times higher than the expected deviate from the model. Based on this deviation we created three categories for each interval: L – lower observed mortality (in green); E – equal observed mortality (in white); H – higher observed mortality (in red). The final category represents the combination of all three time intervals. The first letter of the category represents the time interval between 0-21.5 hpi, the second letter of the category represents the time interval between 21.5-48 hpi and the third letter of the category represents the time interval between 48-72 hpi. We considered lines with more than 85% survival to represent a separate category – high survival (HS, in blue).

DGRP	Expected mortality per period	Absolute mortality 0-21.5 hpi	Absolute mortality 21.5-48 hpi	Absolute mortality 48-72 hpi	Category
32	46	19	108	11	LHL
41	7	9	9	4	HS
91	45	16	62	57	LEE
93	10	14	11	4	EEL
101	7	5	14	2	HS
129	10	10	17	3	EEL
138	17	15	28	9	EEE
158	23	23	35	12	EEE
161	7	3	13	5	HS
195	11	14	18	1	EEL
208	26	11	46	22	LEE
217	3	1	7	1	HS
228	19	12	37	7	EEL
280	28	11	55	18	EEE
301	52	31	119	5	EHL
304	13	8	28	4	EHL
306	7	11	10	1	HS
318	12	17	17	3	EEL
319	30	20	65	6	EHL
320	10	4	20	5	LHE
338	53	40	92	26	EEE
348	48	18	96	29	LHE
350	65	37	127	30	EEL

352	9	1	24	1	HS
355	5	7	9	0	HS
359	34	21	74	7	EHL
360	6	9	5	4	HS
365	34	8	77	16	LHL
370	20	12	43	5	EHL
379	26	17	39	22	EEE
386	16	13	30	5	EEL
390	16	11	35	3	EHL
391	9	6	19	1	HS
392	48	11	117	15	LHL
395	37	47	49	14	EEL
399	21	28	31	4	EEL
406	11	10	14	8	EEE
426	15	9	26	9	EEE
427	20	8	42	9	LHL
437	6	6	12	0	HS
440	17	12	23	16	EEE
441	35	26	73	6	EHL
491	49	59	75	13	EEL
492	70	22	174	14	LHL
502	35	22	63	20	EEE
508	59	154	24	0	HEL
555	21	5	42	16	LEE
584	21	10	38	14	EEE
589	10	9	15	5	EEE
630	41	19	84	21	LHE
642	10	12	18	0	EEL
707	25	29	38	7	EEL
737	4	2	9	2	HS
738	109	51	164	111	LEE
761	11	18	8	8	EEE
774	37	50	48	12	EEL
776	8	5	16	2	HS
783	8	5	13	6	HS
786	10	10	16	4	EEL
802	5	1	12	3	HS
804	29	25	53	10	EEL
805	10	0	22	7	LHE
808	29	12	67	7	LHL
818	129	98	275	15	EHL
819	4	2	10	0	HS
821	3	0	10	0	HS

822	27	12	60	9	LHL
837	13	5	29	4	LHL
850	70	29	170	10	EHL
855	6	4	12	3	HS
859	83	12	174	63	LHE
861	2	2	3	2	HS
882	15	10	19	15	EEE
884	5	5	9	2	HS
908	14	11	27	3	EEL
Mel1	**218**	**190**	**400**	**63**	

Table 3.2 – Absolute number and frequency of DGRP line categories according the mortality profile for all time intervals. The category of each DGRP is given by the combination of three capital letters that represent the comparison between the expected and the observed mortality for each time interval: L – lower observed mortality; E – equal observed mortality; H – higher observed mortality. The first letter of the category represents the time interval between 0-21.5 hpi, the second letter of the category represents the time interval between 21.5-48 hpi and the third letter of the category represents the time interval between 48-72 hpi. We considered lines with more than 85% survival to represent a separate category – high survival (HS).

Category	N	Frequency (%)
HS	19	26
LEE	4	5
LHL	8	11
LHE	5	7
EEL	16	21
EEE	13	17
EHL	9	12
HEL	1	1
TOTAL	**75**	**100**

Table 3.3 – Absolute number and frequency of DGRP lines according their mortality profile for each time interval. For time interval (0-21.5 hpi, 21.5-30 hpi, and 30-48 hpi) "lower" comprises all lines with lower mortality than predicted, "equal" comprises all lines with no deviation from the null model, and "higher" comprises all lines with higher morality than predicted. "HS" represents all the lines categorized as having high survival (throughout the whole infection period).

Mortality 0-21.5 hpi	N	Frequency (%)
HS	19	25
Lower	17	23
Equal	38	51
Higher	1	1
TOTAL	75	100
Mortality 21.5-48 hpi	**N**	**Frequency (%)**
HS	19	25
Lower	0	0
Equal	34	45
Higher	22	29
TOTAL	75	100
Mortality 48-72 hpi	**N**	**Frequency (%)**
HS	19	25
Lower	34	45
Equal	22	29
Higher	0	0
TOTAL	75	100

3.2 Initial Inoculum

The initial inoculum is the number of bacteria that infects a fly. In a systemic infection, it is known that the size of the inoculum contributes to the infection dynamics and outcome: higher initial pathogen loads increase the probability of dying[75]. When infecting flies orally, although we cannot control for the inoculum as in a systemic infection (by injecting a fixed pathogen amount), we can control the food availability (in this case *ad libitum*) and ensure that the bacterial density in the food source is the same for all tested individuals. However, it remains impossible to account for variation in food intake between different DGRP lines. In fact, Garlapow et al. (2015)[76] quantified this variation among a set of DGRP lines, finding significant genetic variation for food consumption (in this work, ranging approximately from 0.5µl/Fly to 2.0µl/Fly depending on the DGRP-line). Because of this, we had to measure the initial inoculum immediately after infection to check for effects of this parameter in our

survival data. We could, however, get data for only 50 of the selected 75 DGRP lines. This was mainly due to some lines showing very small reproductive outputs, and, although we still measured the survival of those lines, we could not get enough individuals to collect for initial inoculum measurements (at least not without compromising other data that we prioritize, like survival and BLUD). Still, in those 50 DGRP lines, we found a high variation, with inoculum values ranging from 10^3 to $2{,}5x10^9$ bacteria (Fig. 3.3).

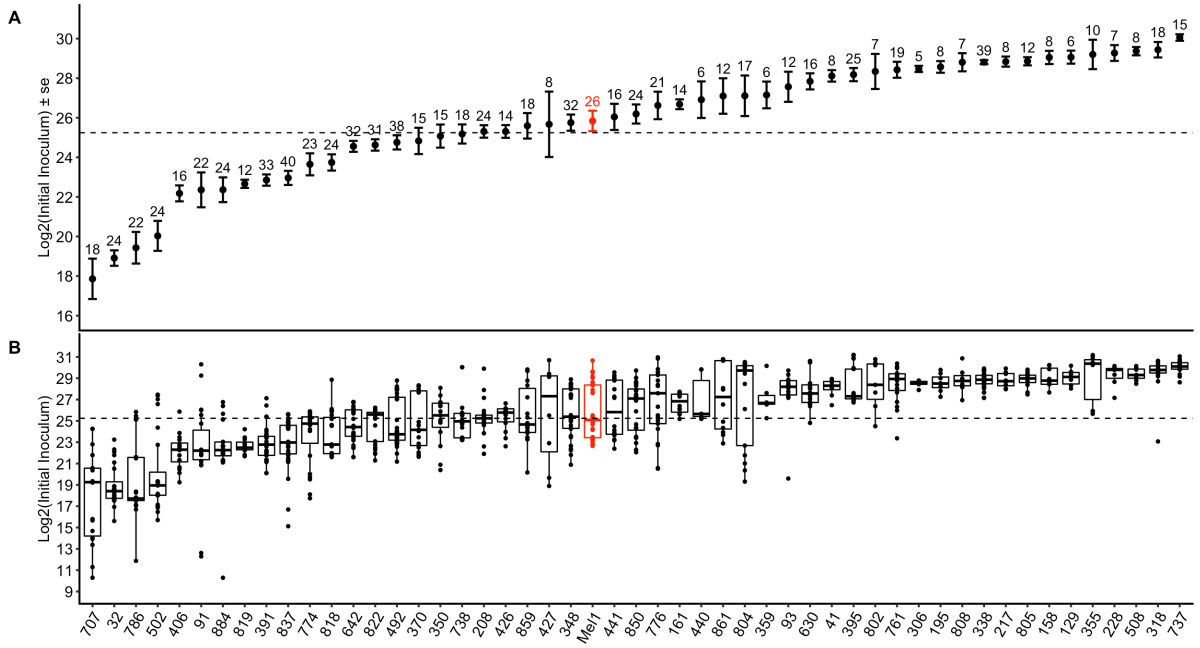


Fig. 3.3 – Feeding *P. entomophila* to *Drosophila* Genetic Reference Panel lines leads to a wide range of initial inoculums, with high variation within line. (A) Mean and Standard error of the Log_2 of the initial inoculum of 50 DGRP lines and the outbred line (Mel1). The Y-axis represents Log_2 of the initial inoculum and the X-axis the line. The black dots represent the value of the Log_2 of the initial inoculum and the bars represent the standard error. The horizontal dashed line represents the initial inoculum of the hypothetical average population created from the data for all DGRP lines. **(B)** Boxplot showing the initial inoculum of 47 DGRP lines and the outbred line, Mel1 (in red). Each dot represents a measure of Initial Inoculum taken from a single fly. The box represents the interquartile range (the values between the upper and lower quartiles), the vertical black lines the upper and lower whiskers and the horizontal black full line the median. The horizontal dashed line represents the minimum systemic bacterial load necessary for lethality. Y-axis represents Log_2 of the Initial Inoculum and the X-axis the Line.

Statistical analysis shows that, as expected, most of the variation found in this trait can be explained by differences between lines (p-value = $2.2x10^{-16}$). The number of bacteria found in flies after the infection is a significantly large amount of bacteria and indicates that even without starving the flies prior to infection, they don't seem to show aversion to the food containing *P. entomophila* (if that was the case, we would expect much less bacteria, or none at all, in the infected flies since they could have gone without eating for the 3-hour period of infection). However, to make such claim we would need to measure food intake in controlled conditions (without bacteria) and compare both. On the other hand the opposite can be argued,

as our data shows variation of six orders of magnitude while Garlapow et al. (2015)[76], measuring food intake without bacteria, found a four fold difference between the lines with lowest and highest food intake. Although the two sets of data are not directly comparable, this might indicate that there is a degree of avoidance to food containing *P. entomophila*.

In Fig. 3.3 we can observe that only 25% (13 lines) of the lines show values below the ones from the hypothetical average population while 61% (31 lines) of the lines show values above. We also did not find any correlation between initial inoculum and survival (Spearman's S = 0.024; p-value = 0.4979), which means that the differences shown in survival across the different DGRP lines are not, on average, a consequence of the variation in food intake (and consequently initial inoculum). Even if we account for eventual behavioural avoidance, it is not the mechanism that explains the majority of variation in our survival results. We also did not find any differences in initial inoculum between survival categories (Kruskal-Wallis chi-squared = 1.2828; df = 2; p-value = 0.5266), suggesting that initial inoculum does not explain the different survival trajectories found (Table 3.1).

To better interpret the data, we plotted the hazard ratio against the initial inoculum, dividing lines into four groups (Fig. 3.4). To do this we first separated lines according to the inoculum – those with higher inoculum than the average hypothetical population and those with lower inoculum than the average hypothetical population. Then we divided the lines according to the hazard ratio – lines with mortality above or below a reference mortality defined as the average of all 75 DGRP lines used. This way we can better visualize how the initial inoculum correlates with survival by comparing the four quadrants and consider the lines as divided between 4 categories. The lines on the top-left quadrant are lines with relatively low initial inoculum and survival. The ones on the top right show higher values for inoculum and lower survival. The bottom-left quadrant comprises lines with lower inoculum and higher survival. Finally, the bottom-right quadrant contains lines with higher inoculum and survival. As it can be observed in Fig. 3.4, there are lines on all four quadrants with the majority showing higher values for initial inoculum. This is likely to be a sampling artefact, since there is a group of four lines clearly showing a lower inoculum (DGRP-707, DGRP-32, DGRP-786 and DGRP-502), with one of them being an outlier (DGRP-707) (Fig. 3.5). However this might also be associated with avoidance behaviour, which was mentioned previously.

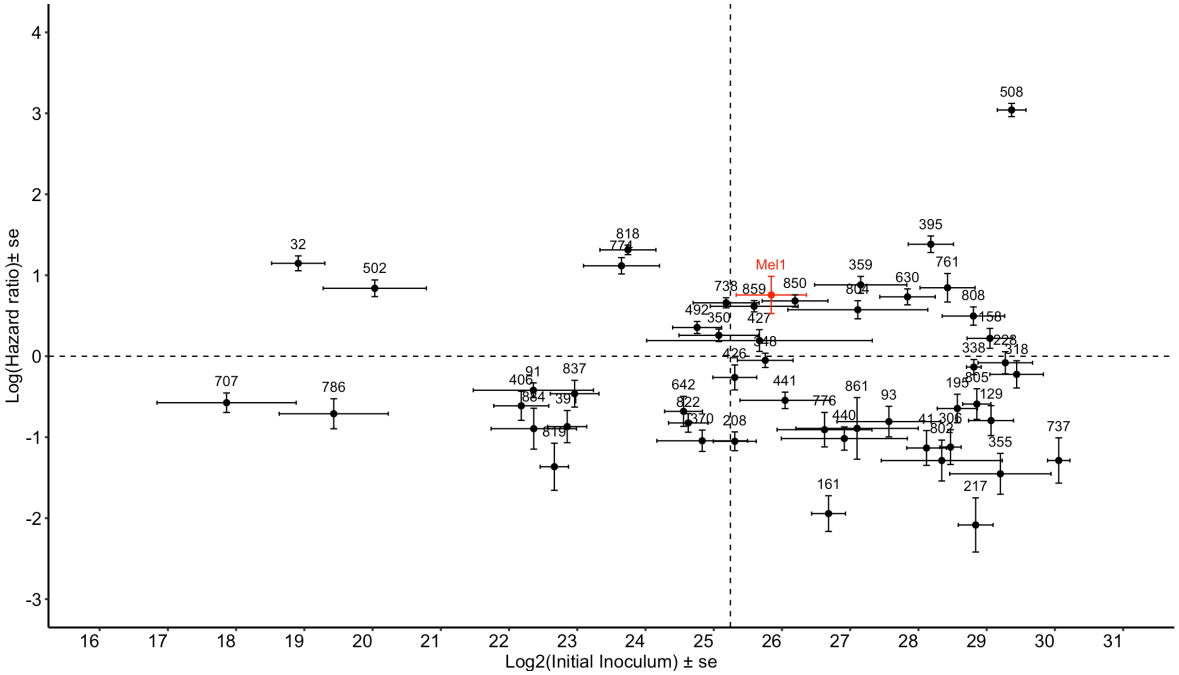


Fig. 3.4 – Initial Inoculum and Hazard Ratios for 50 lines of the *Drosophila* Genetic Reference Panel lines infected with *P. entomophila*. Mean and Standard error of the Log_2 of the Initial Inoculum of 50 DGRP lines and the outbred line (Mel1) plotted against the Hazard Ratio. The Y-axis represents the Hazard Ratio (exponent of the coefficient in the Cox regression model) and the X-axis Log_2 of the Initial Inoculum. The vertical and horizontal bars represent the Standard Error for the Log_2 of the Initial Inoculum and the Hazard Ratio, respectively. The horizontal and vertical dotted lines represent the baseline hazard ratio (given by the hypothetical average population), and the Initial Inoculum of the average hypothetical population, respectively.

Considering the Gut Damage model, the absence of correlation between the inoculum and survival suggests there is a threshold above which the initial bacterial load does not influence infection outcome. That is, below the threshold an increase in initial inoculum contributes to the overall damage sustained during infection, whereas above it any amount of bacteria already results in the maximum damage that can be dealt upon infection. An explanation might be that although the number of bacteria increases, there is a limit to how much bacteria can be in close contact with the epithelium (or the epithelial matrix), and therefore only that percentage of pathogen is contributing to the damage dealt. However, it does not seem like that is the case since *P. entomophila* has been described to damage the gut epithelium without contacting it by segregating toxins and other virulence factors that diffuse through the epithelial matrix[54]. In that perspective, a possible explanation for the existence of the inoculum threshold might be that although increasing the number of bacteria increases the amount of toxins they secrete, the epithelial membrane cannot diffuse at a faster rate. These two theories could potentially explain the existence of the initial inoculum threshold above which there is no correlation with survival, in the light of the Gut Damage model.

Considering the Systemic model, where mortality is the result of a secondary systemic infection caused by a fraction of bacteria that are able to cross the gut epithelial barrier into the haemolymph, the initial inoculum should play a major role in defining the timing and intensity of the secondary infection. That is, it should strongly correlate with the secondary systemic

initial inoculum that, in turn, determines infection outcome[75]. Although there are no data specifically supporting this scenario, we would expect that the inoculum would still correlate negatively with survival. This is because a higher number of bacteria inside the gut would logically translate into more bacteria crossing the gut-epithelial barrier (not taking into consideration differences in resistance and tolerance). However, this does not happen in our dataset, and, here again, it suggests there is a threshold above which the initial bacterial load does not influence infection outcome. One explanation would be again that an increase in the number of bacteria does not translate to an increase of the percentage of bacteria in close contact with the gut, and therefore an increase in overall epithelial damage.

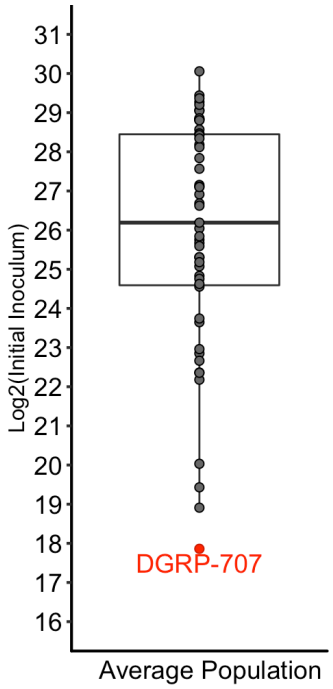


Fig. 3.5 – DGRP-707 is a conventional outlier, showing the lowest value for initial inoculum. Boxplot showing the average Initial Inoculum of 50 DGRP lines, plus the outbred population, Mel1. Each dot represents the average initial inoculum of a single DGRP line. The box represents the interquartile range (the values between the upper and lower quartiles), the vertical black line the upper and lower whiskers and the horizontal black line the median. Conventional outliers are found to be above or below the whiskers. Y-axis represents initial inoculum and the X-axis the hypothetical average population. The dotted horizontal line represents the baseline for the hazard ratio, the hypothetical average population.

3.3 Bacterial Load Upon Death

BLUD is the amount of bacteria a fly carries at the time of death. It is measured in the first 30 min upon death (before bacteria proliferate). Duneau et al. (2017)[73] described it as a measure of tolerance and not resistance that corresponds to the maximal bacterial load that the

host can sustain before dying from the infection. They also found that BLUD is constant over time and does not depend on initial infection dose nor eventual time to death. This tolerance measure was theorized in a context of systemic infection, which presents reversed bacterial load dynamics when compared to oral infection. In the first, the initial inoculum is low and increases over time, depending on the balance between bacterial growth and resistance mechanisms, whereas in the latter the initial inoculum is high and decreases over time, depending on resistance and defecation rate. It is likely that up until 21 hpi (time at which collection of flies for BLUD measurements begins), resistance mechanisms only play a small role in the reduction of bacterial load, with the majority of the decrease being due to defecation. The reason for this is that there is a delay until maximum AMP expression, with, for example, diptericin (very effective against Gram-negative bacteria) peaking systemically after 6 hpi[27] (Lemaitre and Hoffman, 2007), while experiments using food dyes revealed that food can travel the entire digestive tract in less than an hour[76]. Although there is no literature on the natural variation in defecation rate in *D. melanogaster*, Edgecomb et al. (1994)[77] developed a protocol to measure defecation in the fruit fly and found food nutritional value affected excretion. Cognigni et al. (2011)[78] corroborated previous results and found that intestinal transit is also affected by mating status. Looking at a set of nervous system Gal4 mutants they found differences in defecation dependent on a group of previously unknow epithelium-innervating neurons. Given this large number of variables affecting defecation rate, it is arguable that there is potential for high variation within this trait in natural populations of *D. melanogaster* and, therefore, in the DGRP. Another crucial difference between these two types of infection is the initial inoculum, which can be controlled for in a systemic but not in an oral scenario. Although flies were exposed to *ad libitum* food, and we ensured that the bacterial density in the available food source was equal and that the same number of flies were competing for the same amount of nutritional resources, we cannot control for differences in food intake between different DGRP lines (as mentioned before).

Considering the dynamics of an oral infection, the host can only survive if it is able to bring the bacterial load below the tolerance threshold. This is to say, whilst above this threshold, the fly can die with any amount of bacteria so that the determined BLUD will vary across lines. This means that lines could present higher values of BLUD for a number of reasons, not mutually exclusive:

1) because their initial inoculum was higher;
2) because they are less efficient at purging bacteria from their gut;
3) because they have less efficient resistance mechanisms but not (necessarily) because they are more tolerant.

This presents a big problem, because not only the differences in inoculum and gut clearance have a confounding effect in BLUD measurements, but also because the effect is different across lines. When looking at the Gut Damage model, we consider only an oral infection dynamic like mentioned above, with death being caused by unsustainable gut damage and eventually its rupture. Under this model BLUD data is hardly interpretable due to the previously referred confounding effects of food intake and gut clearance. These effects mean that the BLUD under the light of this model can never be interpreted as a measure of tolerance alone. Furthermore, having in mind that any bacterial load that is equal or superior to the

tolerance threshold can cause death, for each line the true BLUD value can be equal or inferior to the measurements, with no immediate way to remove the uncertainty.

However, that is not totally true when considering the Systemic model, where mortality is the result of a secondary systemic infection caused by a percentage of bacteria that are able to cross the gut epithelial barrier into the haemolymph. In this model flies suffer from gut damage, but it only plays a role as the definer of the timing and intensity (inoculum) of the secondary infection. Under this hypothesis the bacterial load progression from an oral infection can be regarded as the sum of two opposite dynamics: a dynamic derived from the bacteria in the gut, and an overlapping dynamic derived from the secondary systemic infection. Therefore, having two distinct bacterial dynamics determining the total (and measurable) load is problematic: the actual lethal systemic load is "hidden" under the much higher oral load and, thus, the determination of the BLUD is unachievable. However, this model is not susceptible to the variation in food intake or excretion like the previous one. Still, as in the Gut Damage model, the only information that can be immediately taken from BLUD measurements is that the true values are equal or lower to the measured ones.

When measuring BLUD in a set of DGRP lines, the immense variation in immune response comprises lines which did not present any mortality in the time interval in which flies were collected for measurements and others which did not reach enough mortality altogether to gather BLUD data. This means that after running a protocol through all DGRP lines, it needs to be optimized for different lines in order to collect the missing data. Unfortunately, that could not be done due to time constraints and at the end of the experiments we collected BLUD data for 47 DGRP lines with values ranging from $2x10^4$ to $1.3x10^8$ (Fig. 3.6 A, B). Statistical analysis show that the differences between line explain the variation with a p-value of $2.2x10^{-16}$. Again, the results will be presented in the light of the two aforementioned models for mortality (the Gut Damage model and the Systemic model). In light of the Gut Damage model, and according to the definition of BLUD, we can see a tolerance gradient in this DGRP subset (Fig. 3.6 A): lines with lower pathogen burden are less tolerant because a smaller amount of bacteria is enough to kill them, and increasing values mean increased tolerance. However, as previously mentioned, the applicability of this measure as a proxy for tolerance in a context of oral infection makes this interpretation uncertain due to the fact that a fly can die with any amount of bacteria equal or superior to the tolerance threshold, which means that lines with higher BLUD are either more tolerant, had higher inoculum or were less efficient in clearing their gut. This means that, as previously mentioned, inoculum and defecation rate (and partially resistance) have a strong confounding effect in this tolerance measurements, which might explain why some lines present high variance (Fig. 3.6 B).

The interpretation of the BLUD under the light of the Systemic model is more complex, since there are two opposite bacterial load dynamics and mortality is a direct cause of systemic damage. Another consideration is that data previously gathered in the lab by Joana Carvalho (unpublished work), using control backgrounds of knockout miRNA *D. melanogaster* lines, showed that the BLUD of systemically infected flies is in the order of 2^{20} (*circa* one million bacteria). This corresponds to the average amount of bacteria necessary to kill a fly systemically and, therefore, the expectation would be that BLUD is equal or above 1 million bacteria, but never below. Surprisingly, 25% of the lines show values below the aforementioned systemic lethal threshold of approximately 10^6 bacteria (Fig. 3.6 A),

suggesting these 12 lines die with less bacteria and are less tolerant than this baseline. As for the remaining lines we cannot take any conclusion regarding tolerance because although they are dying with pathogen load above 1 million bacteria, we do not know how many are present in the gut versus the body cavity. This means that, for example, a specific fly with a BLUD of 10 million bacteria could be more tolerant than the defined threshold if dying with 4 million bacteria in the system and 6 million in the gut. Reversely, this line would be less tolerant if dying with only 500 thousand bacteria in the system, and the remaining 9,5 million being inside the gut. With this example it becomes clear that the unknown relative contributions of each "bacterial location" to the total and measured bacterial load do not allow for conclusions to be drawn relative to the tolerance of lines with BLUD values above 1 million bacteria. In addition, even for lines with values below that threshold, we can only interpret their relative tolerance to a defined baseline. Finally, the confounding effect of the oral bacterial load would also explain the high variance present in some lines (Fig. 3.6 B).

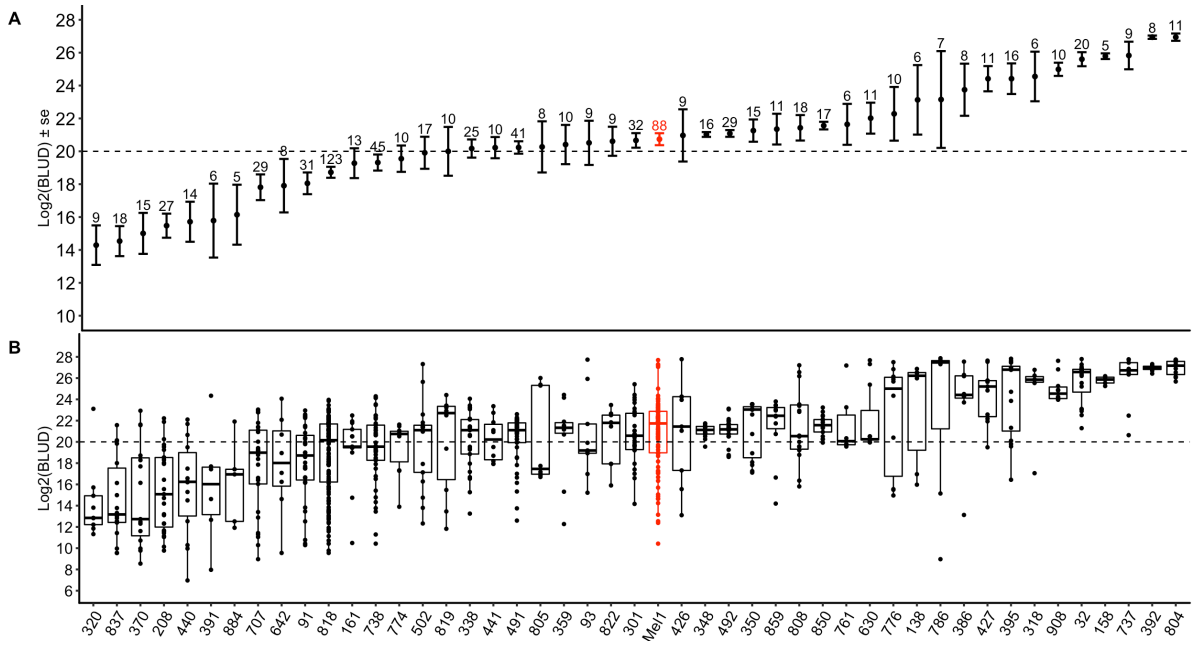


Fig. 3.6 – Feeding *P. entomophila* to *Drosophila* Genetic Reference Panel lines leads to a wide range of BLUD, with high variation within line. (A) Mean and standard error of the Log_2 of the BLUD of 47 DGRP lines and the outbred line, Mel1 (in red). The Y-axis represents Log_2 of the BLUD and the X-axis the Line. The black dots represent the value of the Log_2 of the BLUD and the bars represent the standard error. The horizontal dotted line represents the minimum systemic bacterial load necessary for lethality. **(B)** Boxplot showing the BLUD of 47 DGRP lines and the outbred line, Mel1 (in red). Each dot represents a measure of BLUD taken from a single fly. The box represents the interquartile range, the vertical black lines the upper and lower whiskers and the horizontal black full line the median. The horizontal dashed line represents the minimum systemic bacterial load necessary for lethality. Y-axis represents Log_2 of the BLUD and the X-axis the Line.

To better interpret the data, we plotted hazard ratio against the BLUD, clustering the lines in four different quadrants and categories (Fig. 3.7). To start, we separated the lines according to tolerance – lines with BLUD values above or below the tolerance threshold

(average amount of bacteria necessary to kill a fly systemically), as previously shown in Fig. 3.6 A and 3.6 B. Subsequently, we divided the DGRP lines according to the hazard ratio: lines with mortality above or below a reference mortality defined as the average of all 75 DGRP lines used. This allows us to better understand how tolerance may affect the ability to survive infection and to make relative comparisons between the different groups of lines. Although the tolerance threshold was defined based on a scenario of systemic infection, this value of one million bacteria can also be useful in interpreting our results under a Gut Damage model. Firstly, because we expect that more bacteria are needed to cause death only through gut damage rather than systemic infection, and therefore we can argue that lines dying with one million bacteria in their gut are certainly less tolerant than expected. Secondly, due to the fact that the lines distribute themselves according to a tolerance gradient (Fig. 3.6), any value of tolerance that we choose to separate them (the one million bacteria threshold included), will create a group that is relatively less tolerant and one that is relatively more tolerant (left and right quadrants of Fig. 3.7, respectively). Hence, although the two models present completely different dynamics, the relative conclusions regarding tolerance that can be extracted from Fig. 3.7 are the same for both. However, for the reasons aforementioned for the Gut Damage model the uncertainty around BLUD values does not allow conclusions to be drawn, and for the Systemic model the only conclusion is that lines with BLUD under 1 million bacteria are less tolerant than the reference extracted from systemic infection experiments.

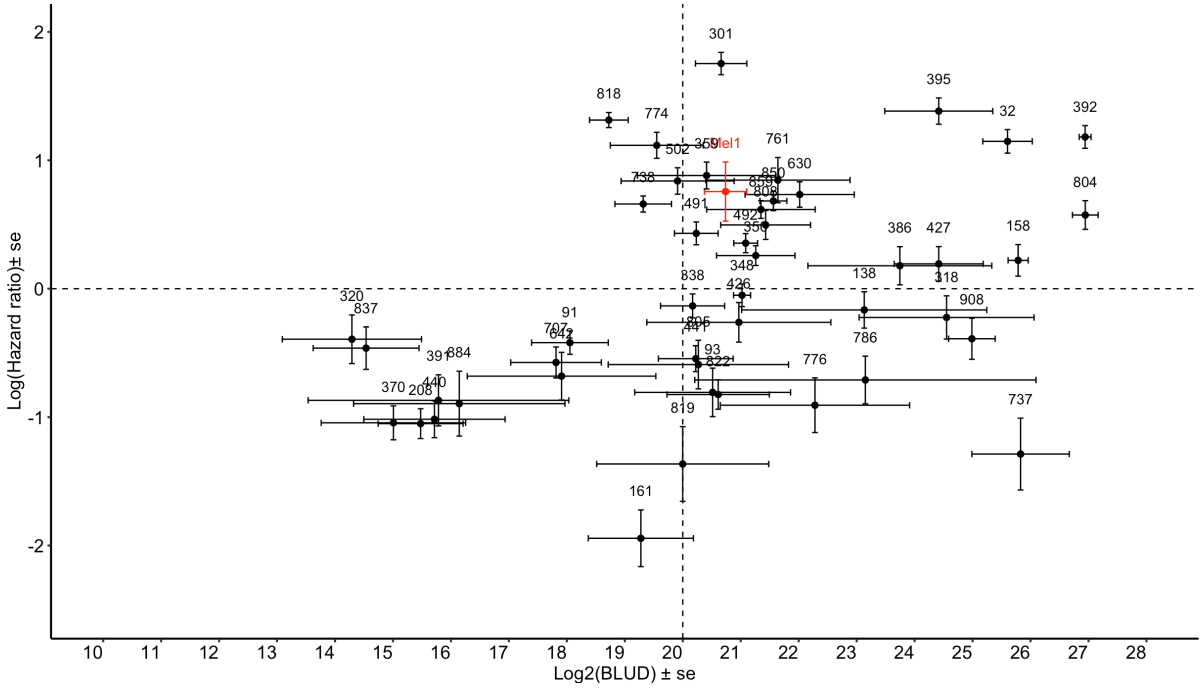


Fig. 3.7 – BLUD and Hazard Ratios for the 47 *Drosophila* Genetic Reference Panel lines infected with *P. entomophila*. Mean and standard error of the Log_2 of the BLUD of 47 DGRP lines and the outbred line, Mel1 (in red), plotted against the hazard ratio. The Y-axis represents the hazard ratio and the X-axis Log_2 of the BLUD. The vertical and horizontal bars represent the standard error for the Log_2 of the BLUD and the hazard ratio, respectively. The horizontal and vertical dotted lines represent the baseline hazard ratio (given by the hypothetical average population), and the minimum systemic bacterial load necessary for lethality, respectively.

With these limitations in mind, we will accept the premise that the BLUD values are representative of tolerance. In summary, DGRP lines on the left of the threshold (vertical dotted line) are less tolerant than the ones on the right, and lines above the baseline hazard ratio (horizontal dotted lines) are more susceptible to infection than the ones below. Because tolerance, by definition, is the mechanism through which a host limits health impact caused by a given pathogen burden, more tolerance will translate into lower disease susceptibility. This means the prediction would be for lines to be distributed between the bottom-right (higher tolerance and lower disease susceptibility) and the top-left quadrant (lower tolerance and higher disease susceptibility). Counter-intuitively, we observe that lines on the right side of the threshold (higher tolerance) are randomly distributed around the baseline hazard ratio (both higher and lower disease susceptibility), while lines on the left side of the threshold (lower tolerance) fall almost exclusively (with exception of DGRP-738 and DGRP-818 that fall very near the threshold) on the bottom quadrant (lower disease susceptibility). The absence of lines in the top-left quadrant may be due to sampling, since in this work we only used 75 out of approximately 200 DGRP lines, and from those only 47 have BLUD data. However, even if such phenotype exists, it should be very rare according to our data (only 4% of the lines fall in this category) and so it is not relevant enough to change the interpretation.

Statistical analysis also revealed that there is no correlation between BLUD and survival ($r=0.025$; p-value=0.08), nor between BLUD and survival categories (Kruskal-Wallis chi-squared = 1.2828; df = 2; p-value = 0.5266) (Table. 3.1). Both results are unexpected given that this parameter is a proxy for tolerance (higher BLUD means higher tolerance, and higher probability of survival). However, the absence of correlation might be because survival also depends on other factors like resistance, and due to experimental problems already mentioned (namely the confounding effects of oral infection).

When a fly is infected it can activate two types of mechanisms to defend itself and survive – tolerance and resistance. Therefore, mortality (here, the hazard ratio) is the result of both tolerance and resistance and cannot be predicted or interpreted considering just one of these mechanisms. Although it is already known that these mechanisms are not independent (nor totally dependent), a good (though rather simplistic) null model for the interpretation of our data is to treat them as fully independent mechanisms of immunity. Therefore, when we compare flies with similar tolerance, differences in disease susceptibility are due to different resistance. Going back to the 4 quadrants defined in Figure 10 under the null model of independence, we can conclude that lines on the bottom-left quadrant are less tolerant (Fig. 3.7 and 3.8), because they die with a lower bacterial load, but more resistant, because they are less susceptible to infection (lower hazard ratio). As previously mentioned, surviving to infection is due to the combination of only two types of mechanisms – tolerance and resistance – that are independent in our null model, meaning that lines that are less tolerant can only be less susceptible if they are more resistant (Fig. 3.7 and 3.8). Under the same logic, the top-left quadrant contains lines which are both less tolerant and resistant (Fig. 3.7 and 3.8), because they die with a lower bacterial load, and are more susceptible to infection. The lines present on the top-right quadrant are more tolerant, because they die with a higher bacterial load, but less resistant, because they are more susceptible to infection. Finally, the bottom-right quadrant lines are both more tolerant and resistant (Fig. 3.7 and 3.8), because they die with a higher bacterial load and would be less susceptible to infection.

Now that we defined the relative resistance and tolerance for all the four quadrants in (Fig. 3.7 and 3.8), we can try to elaborate on the explanation of the patterns observed. Going back to the absence of lines in the top-left quadrant, an alternative explanation for this may be not having alive individuals by the time of data collection, because these lines have both lower resistance and tolerance to bacterial infections compared to lines on the other quadrants. It is important to keep in mind that these conclusions are comparative, and we cannot make any statement about resistance or tolerance outside this framework (we cannot refer to neither of the traits in an absolute manner). Despite this, we know that BLUD could not be measured for lines that have no mechanisms of tolerance and resistance, since they would die before the data collection time window. This fits in what is observed in the figure, and, for example, a good candidate to be present in this group, would be the previously mentioned DGRP-508 which presents the highest hazard ratio in all the dataset, but did not have any surviving individuals by the time of data collection in any replicate. A more general conclusion would be that lines with extreme phenotypes (either high resistance and tolerance or the opposite) would not be present in the analysis because the individuals would all survive infection, or would all die before the time at which BLUD data is collected. With this in mind, we would not expect to have any line on the bottom-right quadrant (lines with high tolerant and high resistance) (Fig. 3.7). However, the presence of lines in that category might be due to the comparative nature of the interpretation. What this means is that although there are lines on the bottom-right quadrant, these are just more tolerant and resistant than the others, but do not represent the most extreme phenotypes. These arguments would explain why out of 75 DGRP lines selected for the study, BLUD measurements could only be taken for 47 of them (suggesting that the remaining 28 represent the most extreme phenotypes).

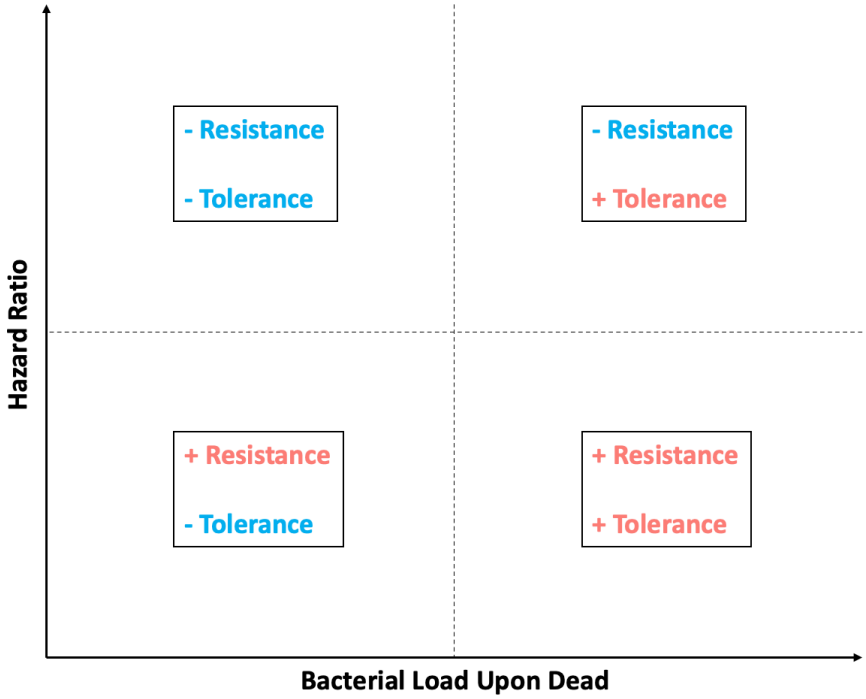


Fig. 3.8 – Relative resistance and tolerance phenotypes in each quadrant generated by plotting hazard ratio against BLUD. The Y-axis represents the hazard ratio and the X-axis Log_2 of the BLUD.

The horizontal and vertical dashed lines divide the plot in four quadrants. When plotting DGRP data each quadrant contains lines with specific relative phenotypes of resistance and tolerance. Higher relative phenotypes are represented in red and lower relative phenotypes are represented in blue.

Another explanation for the absence of lines that are both less tolerant and resistant could be that in the pool of natural variation there are no lines that fit in this description, since they would be subject of strong negative selection in the presence of pathogens (which in a natural environment should be the rule), due to not having an efficient mechanism to defend themselves. By the same reasoning, we would expect that the lines present on the bottom-right quadrant (both more tolerant and resistant) would take over and spread in the natural populations, due to being exposed to strong positive selection. The best explanation for maintaining variation in tolerance and resistance in natural populations, is the presence of trade-offs between tolerance/resistance and other fitness-related traits. If so, the increase in fitness that a host with high resistance and tolerance would enjoy during infection, may not compensate the reduction in other fitness-related traits (for example, reproduction). Moreover, although flies in nature are almost always in contact with pathogens, the selection for immune related traits depends on the virulence, hence varying over time (stronger or weaker depending on the infecting pathogen).

The interpretation of the results outside the null model of independence between tolerance and resistance is virtually impossible due to its comparative nature. That is, we can always classify lines according to being more or less tolerant/resistant than others, but comparisons with intermediate values of these traits are not possible. For example, we cannot say that a group of lines has a more intermediate level of resistance than another. With this in mind it is likely that our simplistic interpretation is far from the truth and that different combinations of resistance and tolerance levels can lead to the same infection outcome. Another important consideration to make is that it is known the existence of a trade-off between resistance and tolerance, easily seen in a phenomenon such as immunopathology. This means that an increase in resistance results in a reduction in tolerance, for example through the self-detrimental action of AMPs or ROS. Because of this, we would not expect to observe lines in the bottom-right quadrant (lines with higher tolerance and resistance), and makes us wonder if the presence of lines there is anything but an artifact of the experimental design. What this means is that, although natural selection would still work favourably towards both high resistance and tolerance, from a theoretical point of view it is not biologically possible to have both since they are negatively correlated[14]. Another interpretation would be that lines in the said quadrant although more tolerant and resistant than the others, might not have high enough levels of these traits (just higher than the others) to observe the trade-off.

A key aspect of the analysis of these data are the outliers, since these are the lines that differ significantly from the others, making them great candidates for further studies such as association analysis like GWAS, power permitting. Unfortunately, there are no conventional outliers for BLUD measurements (Fig. 3.9). In Fig. 3.7, even the outlier previously found in the survival analysis (DGRP-508) is not present due to the high mortality leading to the absence of individuals by the time of BLUD data collection. This is explained by the fact that BLUD outliers would be lines that are extremely tolerant and/or resistant (and would not show

mortality) or have extremely low tolerance and/or resistance (and would die before the data collection interval).

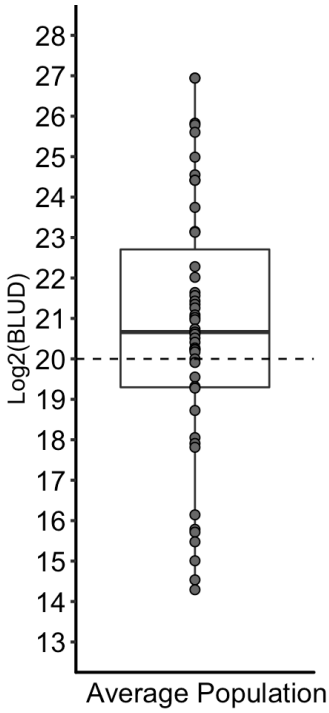


Fig. 3.9 – There are no conventional outliers for Bacterial Load Upon Death data. Boxplot showing the BLUD of 47 DGRP lines. Each dot represents the average BLUD of a single DGRP line. The box represents the interquartile range (the values between the upper and lower quartiles), the vertical black lines the upper and lower whiskers and the horizontal black line the median. Conventional outliers are found to be above or below the whiskers. Y-axis represents BLUD and the X-axis the hypothetical average population. The dashed horizontal line represents the baseline for the hazard ratio, the hypothetical average population.

Finally, we looked at the initial inoculum and BLUD together. In this analysis we looked at only 40 DGRP lines, namely the ones for which we have both initial inoculum and BLUD data. To do this we plotted the BLUD against the initial inoculum, again dividing the lines in four groups (Fig. 3.10). The first step was separating the lines according to the BLUD – lines with BLUD values above or below the tolerance threshold (average amount of bacteria necessary to kill a fly systemically). After we grouped the lines according to the initial inoculum – those with higher inoculum than the average hypothetical population and those with lower inoculum than the average hypothetical population. By separating the lines this way, we can have an idea of how the inoculum may affect the BLUD. However, because BLUD is a measure of tolerance (as discussed above), we do not expect it to be dependent or influenced by the initial inoculum. In fact, we already mentioned that Duneau et al. (2017)[73] found that, in systemic infection, BLUD is constant over time and does not depend on initial infection dose nor eventual time to death. Looking at Fig. 3.10, we can see that the lines on the top-left quadrant are lines with relatively low initial inoculum (in the context of this dataset) but high BLUD. The ones on the top right show higher values for both initial inoculum and

BLUD. The bottom-left quadrant comprises lines with lower inoculum and BLUD. Finally, the bottom-right quadrant contains lines with higher inoculum but lower BLUD. It can also be seen that the top quadrants are mainly empty. This would mean that there are almost no lines with BLUD values higher than the tolerance threshold, however, we have seen in previous results that it not the case. This is due to the fact that we couldn't collect data for initial inoculum for some of the lines we got BLUD measures for, and therefore they are not present in the figure.

The most important information we can take from Fig. 3.10 is that there is no correlation between initial inoculum and BLUD as expected (Spearman's S= -0.05; p-value = 0.65), nor between BLUD and survival categories (Kruskal-Wallis chi-squared = 1.8059; df = 2; p-value = 0.4054) (Table. 3.1). This supports the idea that the BLUD is a measure of tolerance, even in our experimental design (despite all the problems aforementioned). Considering the Gut Damage model, we would expect to find a positive correlation between inoculum and BLUD, since the flies ingest an amount of bacteria superior to the lethal dose during the infection time. This correlation could be eliminated if the differences between lines in resistance (local immunity in the gut) and defecation rates were great enough. However due to the high amount of bacteria ingested during the infection time we do not think differences in these parameters would be enough to completely eliminate the correlation between initial inoculum and BLUD. However, under the Systemic model, this correlation would not be seen. This is because in this model, the "real" initial inoculum is the secondary inoculum derived from the passage of a small amount of bacterial through the gut epithelial barrier. This means that from this point on, the lethal bacterial load dynamics would behave just as in a systemic infection and BLUD would depend exclusively on the tolerance of the fly. Even considering the previously mentioned problems of our experimental design, namely the fact that the measured BLUD is not the actual lethal load, but a mixture of the latter with an oral load prevenient from the oral infection, we would not expect a correlation. Even if we cannot measure the "real" lethal systemic load, this is the amount of bacteria responsible for killing the flies, which is mostly (but not totally) independent from the initial inoculum. Like previously mentioned, the initial inoculum in this scenario influences the timing and intensity (initial load) of the systemic infection, however, the bacterial growth and death of the fly depend solely on other parameters (like resistance and tolerance). Because of this, these data seem to agree with our Systemic model for *Drosophila* infection with *P. entomophila*.

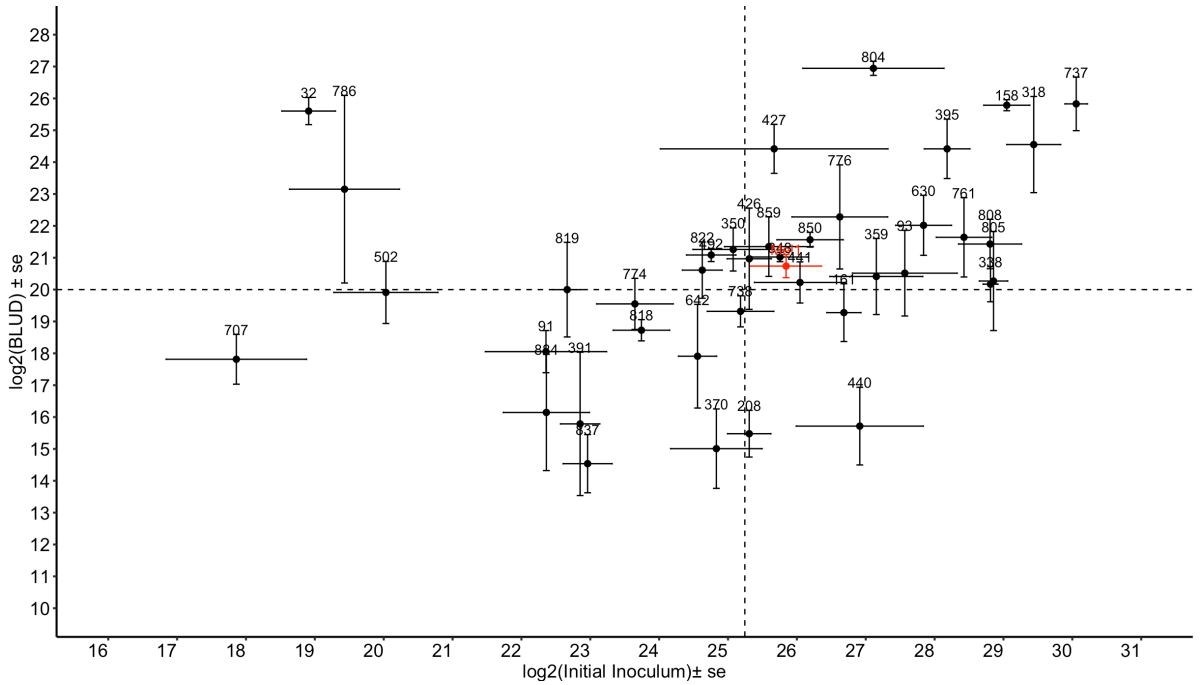


Fig. 3.10 – BLUD and initial inoculum for 40 *Drosophila* Genetic Reference Panel lines infected with *P. entomophila*. Mean and Standard error of the Log_2 of the initial inoculum of 47 DGRP lines and the outbred line, Mel1 (in red), plotted against the BLUD. The Y-axis represents the BLUD and the X-axis Log_2 of the initial inoculum. The vertical and horizontal bars represent the standard error for the Log_2 of the initial inoculum and the BLUD, respectively. The vertical and horizontal dotted lines represent the baseline hazard ratio (given by the hypothetical average population), and the initial inoculum of the average hypothetical population, respectively.

3.4 Set Point Bacterial Load

The Set Point Bacterial Load (SPBL) is a term first used by Duneau et al. (2017) to describe the chronic infection bacterial load. In that work they also describe the SPBL as being dependent on the initial inoculum. A chronic infection is a persistent infection that lasts for long periods (even after survival stabilizes) and happens when the immune system does not effectively clear the pathogen. A chronic infection can last for a set period until its eventually cleared or remain for the entire life of the host. To know for sure if a given DGRP line sustains a chronic infection, we would have to measure the bacterial load across time and see it stabilizes at some point. However, due to time constraints that was not possible in this work, and we will try to extrapolate from the SPBL data at 72 hpi. At this timepoint, survival was stable for at least 24 hours for most lines. Here again we could not get data for the whole DGRP subset, but only for 50 lines, due to some not surviving infection or having only few individuals surviving. We found SPBL values between 0 and $1{,}4x10^7$ (Fig. 3.11). Statistical analysis shows that the differences between line explain the variation with a p-value of $2.2x10^{-16}$. There is no bacterial threshold above which we can consider a fly to suffer from chronic infection, however we can see that the different DGRP lines segregate naturally into two groups: 13% above the SPBL of the average hypothetical population (30 lines) and 72% below (18 lines) (Fig. 3.11).

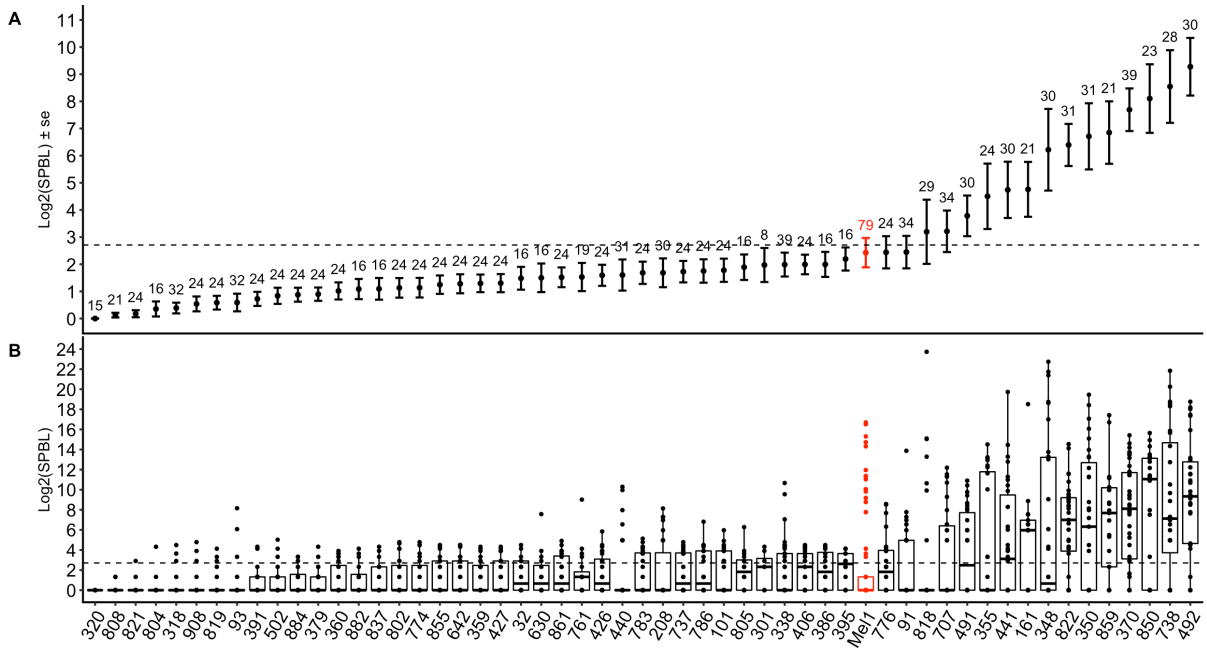


Fig. 3.11 – Feeding *P. entomophila* to *Drosophila* Genetic Reference Panel lines leads to chronic infection. (A) Mean and standard error of the Log_2 of the SPBL of 52 DGRP lines and the outbred line (Mel1). The Y-axis represents Log_2 of the SPBL and the X-axis the Line. The black dots represent the value of the Log_2 of the SPBL and the bars represent the standard error. The horizontal dotted line represents the SPBL of the hypothetical average population. **(B)** Boxplot showing the SPBL of 52 DGRP lines and the outbred line, Mel1 (in red). The box represents the Interquartile Range (the values between the upper and lower quartiles), the vertical black lines the upper and lower whiskers and the horizontal black full line the median. The horizontal dashed line represents the SPBL of the hypothetical average population. Y-axis represents Log_2 of the SPBL and the X-axis the Line.

Some DGRP-lines sustain bacterial loads in the order of few thousand bacteria which may indicate they are suffering from chronic infection. However, for all the lines, independently of where they stay relatively to the SPBL of the average hypothetical population, the bacterial load might not mean that they are sustaining a chronic infection, but that the clearance mechanisms are still having effect (and at later timepoints there would be no bacteria at all). However, one thing to keep in mind is the outlier analysis which found 8 conventional outliers (DGRP-492, DGRP-738, DGRP-850, DGRP-370, DGRP-859, DGRP-350, DGRP-822 and DGRP-348) (Fig. 3.12). All these lines show high bacterial load, and we may (based on the outlier analysis) consider that they are under the effect of a chronic infection. Of these lines, 5 show higher mortality than average (DGRP- 492, DGRP-738, DGRP 850, DGRP-859 and DGRP 350), two present lower mortality than average (DGRP-370 and DGRP-822) and one shows average mortality (DGRP-348).

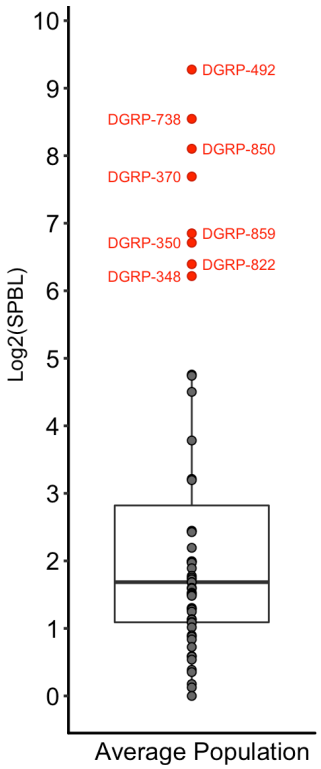


Fig. 3.12 – DGRP- 492, DGRP-738, DGRP 850, DGRP-859 and DGRP 350 are conventional outliers for Set point Bacterial Load data. Boxplot showing the SPBL of 52 DGRP lines, plus the outbred population, Mel1. Each coloured dot represents the SPBL of a single DGRP line. The colours represent the different DGRP lines. The box represents the Interquartile Range (the values between the upper and lower quartiles), the vertical black lines the upper and lower whiskers and the horizontal black line the median Conventional outliers are found to be above or below the whiskers. Y-axis represents SPBL and the X-axis the hypothetical average population.

For a host to sustain a persistent infection without dying, the bacterial load needs to be lower than its tolerance threshold (otherwise they would die). Because of this, the SPBL is informative about tolerance, meaning that lines with higher SPBL values show greater tolerance. Going back on the outlier analysis, because SPBL is representative of tolerance, and since higher tolerance means higher survival, the DGRP lines that show bellow average mortality might be the ones truly sustaining a chronic infection. Yet again, we cannot confirm this claim with only one measurement for SPBL. On the other hand, one line might be just as tolerant as another sustaining a chronic infection without having any bacteria in the body, if their resistance is higher and completely clears the pathogen. Hence, SPBL also gives information on the resistance of the host. Thus, we may extrapolate that the lines with SPBL values below the SPBL of the average hypothetical population are more resistant than the others due to clearing more bacteria by 72 hpi. However, this can only be true if there is no correlation between SPBL and initial inoculum, that is, if these lines possess less bacteria because they had less to start with. To investigate this, we plotted the SPBL against the initial inoculum, dividing lines into four groups (Fig. 3.13). To do this we first separated lines according to the inoculum – those with higher inoculum than the average hypothetical population and those with lower inoculum than the average hypothetical population. Then we divided the lines according to the SPBL – those with higher SPBL than the average hypothetical

population and those with lower. The lines on the top-left quadrant are lines with relatively low initial inoculum (in the context of this dataset) but high SPBL. The ones on the top right show higher values for inoculum and SPBL. The bottom-left quadrant comprises lines with lower inoculum and SPBL. Finally, the bottom-right quadrant contains lines with higher inoculum and lower SPBL.

We can see in the figure that there are lines present on all four quadrants of the graph and correlation analysis showed that there is no correlation between initial inoculum and SPBL ($r = 0.048$; p-value = 0.693), unlike what was found by Duneau et al. (2017). In their work, the authors inferred that a fly can only enter a chronic stage of infection if it is able to control the bacteria before it grows above the tolerance threshold (and kills the host). This way, the inoculum would have an impact, since higher inoculums would be closer to the tolerance threshold. In a context of oral infection, however, the expectation is different. Considering the Gut damage model, we would not expect a correlation between inoculum and SPBL, since in this feeding set up the initial inoculum is already for most cases higher than the tolerance threshold, and thus increasing it would not have any effect. In this scenario we can make conclusions relatively to the resistance of lines by looking at SPBL data (the ones on the left of Fig. 3.11 are more resistant than the ones on the right). Looking at the systemic model we could never see a correlation between initial inoculum and SPBL because the "real" systemic inoculum cannot be measured, only the one from the oral infection. In this case we cannot take conclusions relatively to resistance phenotypes by looking at SPBL data (because we are not sure whether there is a correlation between initial inoculum and SPBL or not).

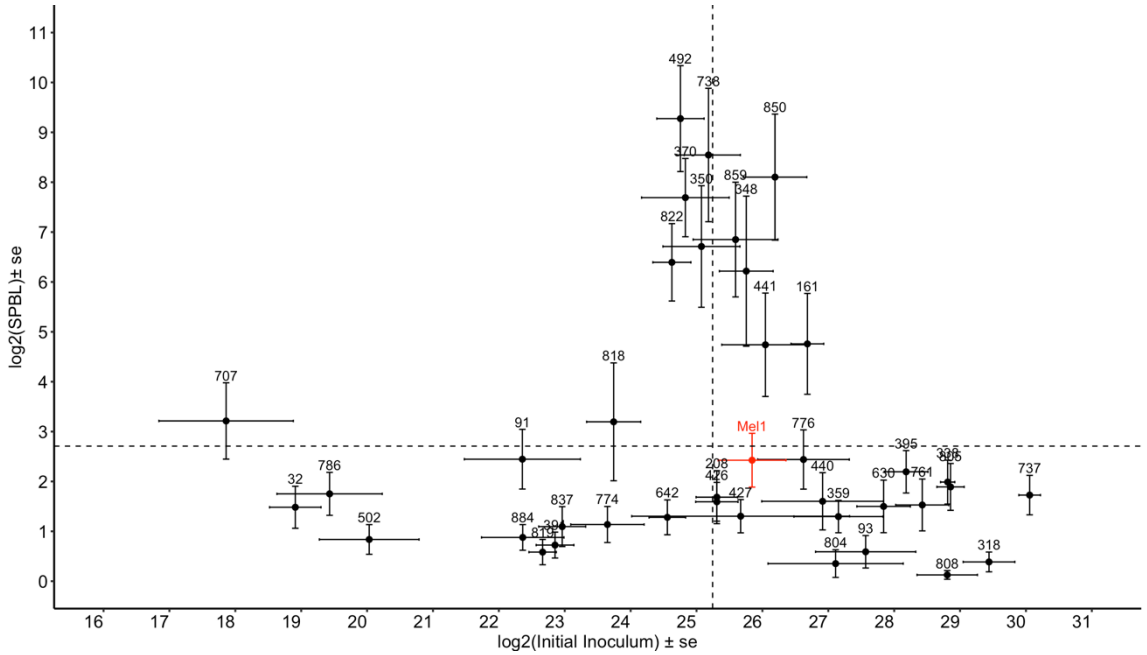


Fig. 3.13 – SPBL and initial inoculum for 52 *Drosophila* Genetic Reference Panel lines infected with *P. entomophila*. Mean and Standard error of the Log_2 of the SPBL of 52 DGRP lines and the outbred line, Mel1 (in red), plotted against the Initial inoculum. The Y-axis represents the SPBL (and the X-axis Log_2 of the initial inoculum. The vertical and horizontal bars represent the Standard Error for the Log_2 of the initial inoculum and the SPBL, respectively. The horizontal and vertical dotted lines

represent the SPBL of the hypothetical average population, and the initial inoculum of the hypothetical average population, respectively.

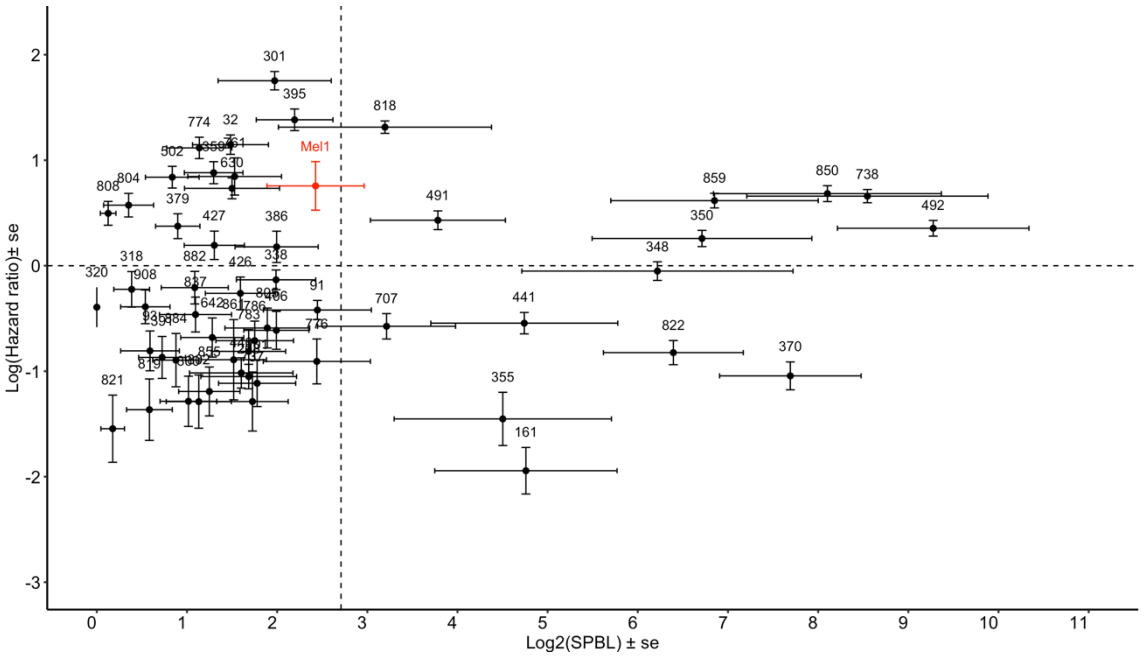


Fig. 3.14 – SPBL and Hazard Ratios for 52 *Drosophila* Genetic Reference Panel lines infected with *P. entomophila*. Mean and Standard error of the Log_2 of the SPBL of 52 DGRP lines and the outbred line (Mel1) plotted against the Hazard Ratio. The Y-axis represents the Hazard Ratio (exponent of the coefficient in the Cox regression model) and the X-axis Log_2 of the SPLB. The vertical and horizontal bars represent the Standard Error for the Log_2 of the SPBL and the Hazard Ratio, respectively. The horizontal and vertical dotted lines represent the baseline hazard ration (given by the hypothetical average population), and the SPBL of the hypothetical average population, respectively.

As said before, we can extract information regarding both tolerance and resistance by looking at SPBL. However, these parameters cannot be disentangled with only SPBL data, because one line might be just as tolerant as another sustaining a chronic infection without having any bacteria in the body, if their resistance is higher and completely clears the pathogen. Because of this, the resistance might in some cases "hide" the tolerance when looking at these data. However, one good way to try to disentangle this is to look at SPBL and survival data at the same time (Fig. 3.14). Here we once again separated lines first according to the hazard ratio – lines with mortality above or below a reference mortality defined as the average of all 75 DGRP lines used (even if for some of them there are no SPBL data). Then we divided the lines according to the SPBL – those with higher SPBL than the average hypothetical population and those with lower. The lines on the top-left quadrant are lines with low SPBL (in the context of this dataset) but high hazard ratio. The ones on the top right show higher values for both SPBL and hazard ratio. The bottom-left quadrant comprises lines with lower SPBL and hazard ratio. Finally, the bottom-right quadrant contains lines with higher SPBL and lower hazard ratio. The

first piece of information we can extract from the quadrants is that lines on the left are more resistant than lines on the right, because they cleared more bacteria. Due to the limitations of these data, we can only make comparisons regarding tolerance between quadrants by fixing the SPBL: comparing the right quadrants (the ones with higher SPBL) we can say that the bottom one contains lines with greater tolerance; equally when comparing the left quadrants (the ones with lower SPBL) we can infer that the bottom one contains lines with greater tolerance. In the figure we can also see that there are lines present on all quadrants and analysis found no correlation between SPBL and hazard ratio (r = 0.036; p-value = 0.764), nor between SPBL and the survival categorization (Kruskal-Wallis chi-squared = 1.8059, df = 2; p-value = 0.4054) (Table 3.1). The absence of correlation between hazard ratio and SPBL is not surprising since survival is an outcome of both resistance and tolerance and these parameters are entangled in the SPBL measurements.

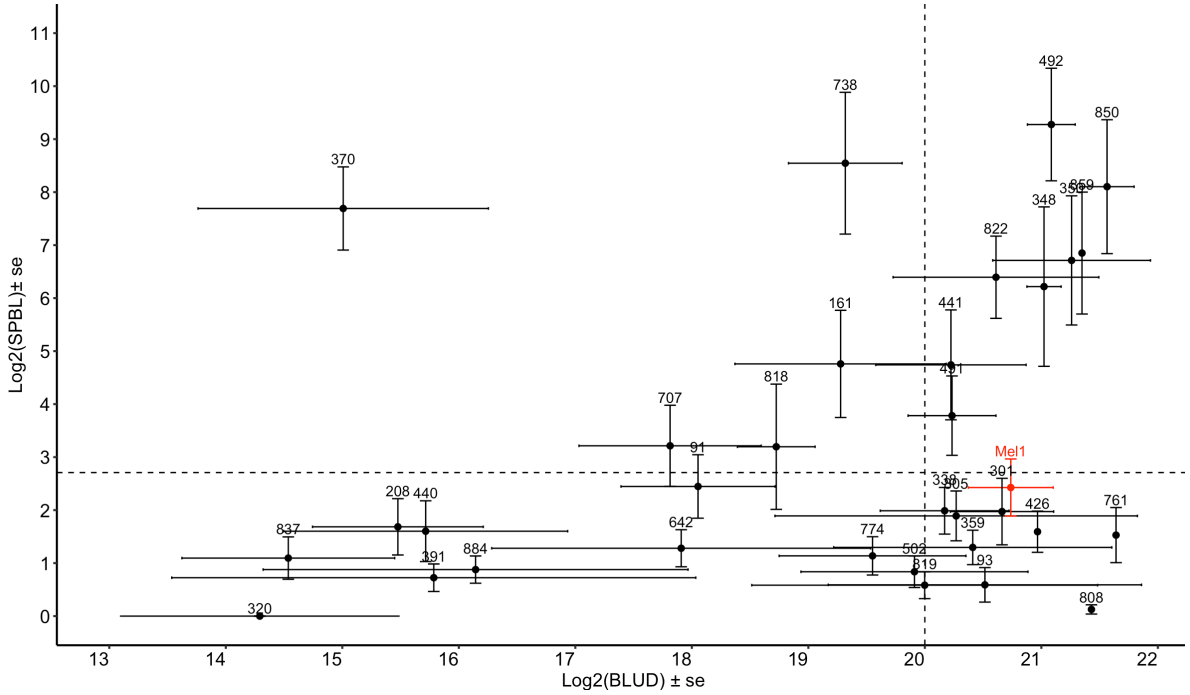


Fig. 3.15 – SPBL and BLUD for 52 *Drosophila* Genetic Reference Panel lines infected with *P. entomophila*. Mean and Standard error of the Log_2 of the BLUD of 52 DGRP lines and the outbred line (Mel1) plotted against the SPBL. The Y-axis represents the SPBL (exponent of the coefficient in the Cox regression model) and the X-axis Log_2 of the BLUD. The vertical and horizontal bars represent the Standard Error for the Log_2 of the BLUD and the SPBL, respectively. The horizontal and vertical dotted lines represent the SPBL of the hypothetical average population, and the minimum systemic bacterial load necessary for lethality, respectively.

We also wanted to see how BLUD and SPBL interact, since these measurements represent different types of tolerance. The BLUD, as previously discussed, represents the amount on bacteria necessary to kill a fly. The SPBL represents the amount of bacteria that a fly sustains at the end of the infection, without suffering mortality. To see this interaction we plotted the two variables against each other (Fig. 3.15). Again, we first separated lines according to the BLUD – those with BLUD values above or below the tolerance threshold

(average amount of bacteria necessary to kill a fly systemically). Then we divided the lines according to the SPBL – those with higher SPBL than the average hypothetical population and those with lower. The lines on the top-left quadrant are lines with relatively low BLUD but high SPBL. The ones on the top right show higher values for BLUD and SPBL. The bottom-left quadrant comprises lines with lower BLUD and SPBL. Finally, the bottom-right quadrant contains lines with higher BLUD and lower SPBL.

Because both BLUD and SPBL serve as proxy for tolerance, we wouldn't expect to find lines in the bottom-right nor top-left quadrant (higher BLUD and lower SPBL or lower BLUD and higher SPBL, respectively). However, this is likely because the two parameters are connected to different mechanisms of tolerance. This is logical given that BLUD is measured when host health is more affected by the infection (peak mortality) and SPBL is measured after survival stabilizes. Following this reasoning, we tested if there is any correlation between these two types of tolerance. We found no correlation ($r = 0.012$; p-value = 0.92) and concluded that the tolerance mechanisms behind these two bacterial loads are independent. However, due to the experimental problems around the BLUD (previously discussed) and SPBL measurements, conclusions cannot to be drawn under the light of any model (Gut Damage or Systemic).

4 Conclusion and Future Perspectives

Illness due to infection is a complex and dynamic process that depends not only on pathogen virulence, but also on host resistance and tolerance phenotypes. In this work we focused on how the host genotype can affect infection outcome by controlling the pathogen. The first logical step to this approach was to characterize survival to oral infection with *P. entomophila* in a diverse set of DGRP lines. With this, we were able to see that host genetic background alone is responsible for causing variable infection outcomes and survival dynamics over the course of infection. The differences observed in survival allowed us to characterize DGRP lines based on their mortality profile opening way for future association studies that may present candidate genes that intervene in the infection process. Conducting GWAS using hazard ratios or mortality after 3 days would identify genes that modulate both resistance and tolerance mechanisms. However, by using the survival categories, and comparing the results of both analyses, we expect to be able to further understand how resistance and tolerance can dictate the outcome of infection. This is because although both mechanisms affect infection outcome in terms of survival, they do it in very different ways, leading to a wide range of mortality dynamics. For example, some DGRP lines show a mortality profile that fits AMP expression patterns, suggesting that resistance mechanisms might be the driving force of their infection outcome[27].

After this, we measured bacterial loads such as initial inoculum, BLUD and SPBL in order to characterize how the host influences pathogen dynamics during infection, however these results were not clear enough to draw strong conclusions. One important misconception in our study design consisted in considering that BLUD would translate from systemic infection to oral infection in a similar and thus interpretable way. As previously discussed, this

is not the case and, in the future, we would need to gather more knowledge about the interaction between *P. entomophila* and *D. melanogaster* in an oral infection setting. Specifically, we would need to know if the bacteria are able to cross the flies' gut epithelial barrier into the haemolymph, establishing a secondary systemic infection. To do this we could try to separate both oral and systemic infections, by bleeding flies and check for bacteria in the haemolymph. Furthermore, we would need to adapt the bacterial load protocols to each individual DGRP line in order to collect as much data as possible (ideally completing the full DGRP set) and include more collection timepoints in the original protocols (for example, at 48 hpi and after 72 hpi). To adapt the protocol to each DGRP line we would first need to characterize their survival dynamics and identify, individually, the ideal timepoints and intervals to collect bacterial load data. This would generate a dataset consisting of about 200 DGRP lines, with information on survival, BLUD, and a full characterization of the bacterial load dynamics on living individuals (including initial inoculum and SPBL).

After doing this, we would be faced with one of two scenarios: 1) the Gut Damage model of mortality is correct; 2) the Systemic model of mortality is correct. In the first scenario, BLUD measurements would not be translatable to an oral infection scenario (as previously discussed), meaning that we would be left with a characterization of bacterial load dynamics on living flies. Performing GWAS using this dataset would allow us to identify candidate genes that modulate different resistance mechanisms, which are responsible for reducing the parasitic load. However, we would not be able to infer about tolerance using this dataset. In the second scenario, we would be able to use BLUD data to segregate resistance and tolerance by assessing the relative contributions to survival during infection. In this case, besides using the bacterial load dynamics on living flies to identify candidate genes that modulate resistance mechanisms, we would be able to place each DGRP in a space map using the tolerance and resistance phenotypes as coordinates (as we sketched in this work). However, to do this we would need to create a better model for the interaction between the two traits than assuming they are fully independent. The better we are able to generate this representative model, without compromising the phenotype identification, the clearer the results would be. Using this information to perform GWAS would allow us to identify genes that modulate resistance and tolerance, both dependently and independently.

Combining survival with bacterial load analysis we would be able to identify precise mechanisms through which the host genome can modulate infection dynamics. In this work we have taken the first step in creating a framework that leads to knowing how resistance and tolerance mechanisms interact with each other, which can then be applied to other organisms.

References

1. Dimijian, G. G. Evolving Together: The Biology of Symbiosis, Part 1. *Baylor Univ. Med. Cent. Proc.* **13**, 217a – 226 (2000).
2. Parmentier, E. & Michel, L. Boundary lines in symbiosis forms. *Symbiosis* **60**, 1–5 (2013).
3. Windsor, D. A. Most of the species on Earth are parasites. *Int. J. Parasitol.* **28**, 1939–1941 (1998).
4. Méthot, P. O. & Alizon, S. What is a pathogen? Toward a process view of host-parasite interactions. *Virulence* **5**, 775–785 (2014).
5. May, R. M. & Anderson, R. M. Epidemiology and genetics in the coevolution of parasites and hosts. *Proc. R. Soc. London - Biol. Sci.* **219**, 281–313 (1983).
6. Day, T. On the evolution of virulence and the relationship between various measures of mortality. *Proc. R. Soc. B Biol. Sci.* **269**, 1317–1323 (2002).
7. Agnew, P. Estimating virulence from relative survival. *bioRxiv* 1–27 (2019). doi:10.1101/530709
8. Read, A. F. The evolution of virulence. *Trends Microbiol.* **2**, 73–76 (1994).
9. Alizon, S., de Roode, J. C. & Michalakis, Y. Multiple infections and the evolution of virulence. *Ecol. Lett.* **16**, 556–567 (2013).
10. Hube, B. From commensal to pathogen: Stage- and tissue-specific gene expression of Candida albicans. *Curr. Opin. Microbiol.* **7**, 336–341 (2004).
11. Klemm, P., Hancock, V. & Schembri, M. A. Mellowing out: Adaptation to commensalism by Escherichia coli asymptomatic bacteriuria strain 83972. *Infect. Immun.* **75**, 3688–3695 (2007).
12. Armstrong, D. History of Opportunistic Infection in the Immunocompromised Host. *Clin. Infect. Dis.* **17**, S318–S321 (1993).
13. Miller, I. F., Jessica, C. & Metcalf, E. Evolving resistance to pathogens. *Science (80-.).* **363**, 1277–1278 (2019).
14. Ayres, J. S. & Schneider, D. S. Two ways to survive an infection: what resistance and tolerance can teach us about treatments for infectious diseases. *Nat Rev Immunol* **8**, 889–895 (2015).
15. Soares, M. P., Teixeira, L. & Moita, L. F. Disease tolerance and immunity in host protection against infection. *Nat. Rev. Immunol.* **17**, 83–96 (2017).
16. McCarville, J. L. & Ayres, J. S. Disease tolerance: concept and mechanisms. *Curr. Opin. Immunol.* **50**, 88–93 (2018).
17. Cooper, A. M. Cell-mediated immune responses in tuberculosis. *Annu. Rev. Immunol.* **27**, 393–422 (2009).
18. Graham, A. L., Allen, J. E. & Read, A. F. Evolutionary causes and consequences of immunopathology. *Annu. Rev. Ecol. Evol. Syst.* **36**, 373–397 (2005).
19. Long, G. H., Chan, B. H. K., Allen, J. E., Read, A. F. & Graham, A. L. Experimental manipulation of immune-mediated disease and its fitness costs for rodent malaria parasites. *BMC Evol. Biol.* **8**, 1–11 (2008).
20. Lambeth, J. D., Kawahara, T. & Diebold, B. Regulation of Nox and Duox enzymatic activity and expression. *Free Radic. Biol. Med.* **43**, 319–331 (2007).
21. Lambeth, J. D. Nox enzymes, ROS, and chronic disease: An example of antagonistic pleiotropy. *Free Radic. Biol. Med.* **43**, 332–347 (2007).
22. Libertucci, J. & Young, V. B. The role of the microbiota in infectious diseases. *Nat. Microbiol.* **4**, 35–45 (2019).

23. Wjst, M. Does Helicobacter pylori protect against asthma and allergy? *Gut* **57**, 1178–1179 (2008).
24. Martín, R., Miquel, S., Langella, P. & Bermúdez-Humarán, L. G. The role of metagenomics in understanding the human microbiome in health and disease. *Virulence* **5**, 413–423 (2014).
25. Hill, C., Ross, R. P., Stanton, C. & Toole, P. W. O. The Human Microbiome in Health and Disease Methods for Characterizing the Microbiota. 4–6 (2016).
26. Schmid-Hempel, P. Immune defence, parasite evasion strategies and their relevance for 'macroscopic phenomena' such as virulence. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 85–98 (2009).
27. Lemaitre, B. & Hoffmann, J. The host defense of Drosophila melanogaster. *Annu. Rev. Immunol.* **25**, 697–743 (2007).
28. Martins, N. E., Faria, V. G., Teixeira, L., Magalhães, S. & Sucena, É. Host Adaptation Is Contingent upon the Infection Route Taken by Pathogens. *PLoS Pathog.* **9**, (2013).
29. Vallet-Gely, I., Lemaitre, B. & Boccard, F. Bacterial strategies to overcome insect defences. *Nat. Rev. Microbiol.* **6**, 302–313 (2008).
30. Kylsten, P., Kimbrell, D. A., Daffre, S., Samakovlis, C. & Hultmark, D. The lysozyme locus in Drosophila melanogaster: different genes are expressed in midgut and salivary glands. *MGG Mol. Gen. Genet.* **232**, 335–343 (1992).
31. Dionne, M. S. & Schneider, D. S. Models of infectious diseases in the fruit fly Drosophila melanogaster. *DMM Dis. Model. Mech.* **1**, 43–49 (2008).
32. Reiter, L. T., Potocki, L., Chien, S., Gribskov, M. & Bier, E. A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster. *Genome Res.* **11**, 1114–1125 (2001).
33. Brennan, C. A. & Anderson, K. V. Drosophila: The genetics of innate immune recognition and response. *Annu. Rev. Immunol.* **22**, 457–483 (2004).
34. Leclerc, V. & Reichhart, J. M. The immune response of Drosophila melanogaster. *Immunol. Rev.* **198**, 59–71 (2004).
35. Tzou, P., De Gregorio, E. & Lemaitre, B. How Drosophila combats microbial infection: A model to study innate immunity and host-pathogen interactions. *Curr. Opin. Microbiol.* **5**, 102–110 (2002).
36. Hultmark, D. Drosophila immunity: Paths and patterns. *Curr. Opin. Immunol.* **15**, 12–19 (2003).
37. Lemaitre, B. & Miguel-Aliaga, I. The digestive tract of Drosophila melanogaster. *Annu. Rev. Genet.* **47**, 377–404 (2013).
38. Tzou, P. *et al.* Tissue-specific inducible expression of antimicrobial peptide genes in Drosophila surface epithelia. *Immunity* **13**, 737–748 (2000).
39. Bulet, P., Hetru, C., Dimarcq, J. L. & Hoffmann, D. Antimicrobial peptides in insects; structure and function. *Dev. Comp. Immunol.* **23**, 329–344 (1999).
40. Hanson, M. A. *et al.* Synergy and remarkable specificity of antimicrobial peptides in vivo using a systematic knockout approach. *Elife* **8**, 1–24 (2019).
41. Nunes, C., Sucena, É. & Koyama, T. Endocrine regulation of immunity in insects. *FEBS J.* 1–20 (2020). doi:10.1111/febs.15581
42. Ferrandon, D. *et al.* A drosomycin-GFP reporter transgene reveals a local immune response in Drosophila that is not dependent on the Toll pathway. *EMBO J.* **17**, 1217–1227 (1998).
43. Boutros, M., Agaisse, H. & Perrimon, N. Sequential activation of signaling pathways during innate immune responses in Drosophila. *Dev. Cell* **3**, 711–722 (2002).
44. Dostert, C. *et al.* The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of drosophila. *Nat. Immunol.* **6**, 946–953 (2005).

45. Liehl, P., Blight, M., Vodovar, N., Boccard, F. & Lemaitre, B. Prevalence of local immune response against oral infection in a Drosophila/Pseudomonas infection model. *PLoS Pathog.* **2**, 0551–0561 (2006).
46. Ryu, J. H. *et al.* An essential complementary role of NF-κB pathway to microbicidal oxidants in Drosophila gut immunity. *EMBO J.* **25**, 3693–3701 (2006).
47. Buchon, N. *et al.* Morphological and Molecular Characterization of Adult Midgut Compartmentalization in Drosophila. *Cell Rep.* **3**, 1725–1738 (2013).
48. Stoffolano, J. G. & Haselton, A. T. The adult dipteran crop: A unique and overlooked organ. *Annu. Rev. Entomol.* **58**, 205–225 (2013).
49. Wagner, C., Isermann, K., Fehrenbach, H. & Roeder, T. Molecular architecture of the fruit fly's airway epithelial immune system. *BMC Genomics* **9**, 1–12 (2008).
50. Ma, H. *et al.* 20-Hydroxyecdysone regulates the transcription of the lysozyme via Broad-Complex Z2 gene in silkworm, Bombyx mori. *Dev. Comp. Immunol.* **94**, 66–72 (2019).
51. Wang, J. L. *et al.* 20-hydroxyecdysone transcriptionally regulates humoral immunity in the fat body of Helicoverpa armigera. *Insect Mol. Biol.* **23**, 842–856 (2014).
52. Broderick, N. A. & Lemaitre, B. Gut-associated microbes of Drosophila melanogaster. *Gut Microbes* **3**, (2012).
53. Hegedus, D., Erlandson, M., Gillott, C. & Toprak, U. New insights into peritrophic matrix synthesis, architecture, and function. *Annu. Rev. Entomol.* **54**, 285–302 (2009).
54. Kuraishi, T., Binggeli, O., Opota, O., Buchon, N. & Lemaitre, B. Genetic evidence for a protective role of the peritrophic matrix against intestinal bacterial infection in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15966–15971 (2011).
55. Buchon, N., Broderick, N. A., Chakrabarti, S. & Lemaitre, B. Invasive and indigenous microbiota impact intestinal stem cell activity through multiple pathways in Drosophila. *Genes Dev.* **23**, 2333–2344 (2009).
56. Charroux, B. & Royet, J. Elimination of plasmatocytes by targeted apoptosis reveals their role in multiple aspects of the Drosophila immune response. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9797–9802 (2009).
57. Carvalho, G. B., Ja, W. W., Kapahi, P. & Benzer, S. Pitfalls of measuring feeding rate in the fruit fly Drosophila melanogaster [4]. *Nat. Methods* **5**, 214–215 (2008).
58. Vodovar, N. *et al.* Drosophila host defense after oral infection by an entomopathogenic Pseudomonas species. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 11414–11419 (2005).
59. Vodovar, N. *et al.* Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium Pseudomonas entomophila. *Nat. Biotechnol.* **24**, 673–679 (2006).
60. Buchon, N., Broderick, N. A., Poidevin, M., Pradervand, S. & Lemaitre, B. Drosophila Intestinal Response to Bacterial Infection: Activation of Host Defense and Stem Cell Proliferation. *Cell Host Microbe* **5**, 200–211 (2009).
61. Mulet, M., Gomila, M., Lemaitre, B., Lalucat, J. & García-Valdés, E. Taxonomic characterisation of Pseudomonas strain L48 and formal proposal of Pseudomonas entomophila sp. nov. *Syst. Appl. Microbiol.* **35**, 145–149 (2012).
62. Chakrabarti, S., Liehl, P., Buchon, N. & Lemaitre, B. Infection-induced host translational blockage inhibits immune responses and epithelial renewal in the Drosophila gut. *Cell Host Microbe* **12**, 60–70 (2012).
63. Amcheslavsky, A., Jiang, J. & Ip, Y. T. Tissue Damage-Induced Intestinal Stem Cell Division in Drosophila. *Cell Stem Cell* **4**, 49–61 (2009).
64. Ha, E. M. *et al.* An antioxidant system required for host protection against gut infection in Drosophila. *Dev. Cell* **8**, 125–132 (2005).
65. Richardson, L. A. Understanding Disease Tolerance and Resilience. *PLoS Biol.* **14**,

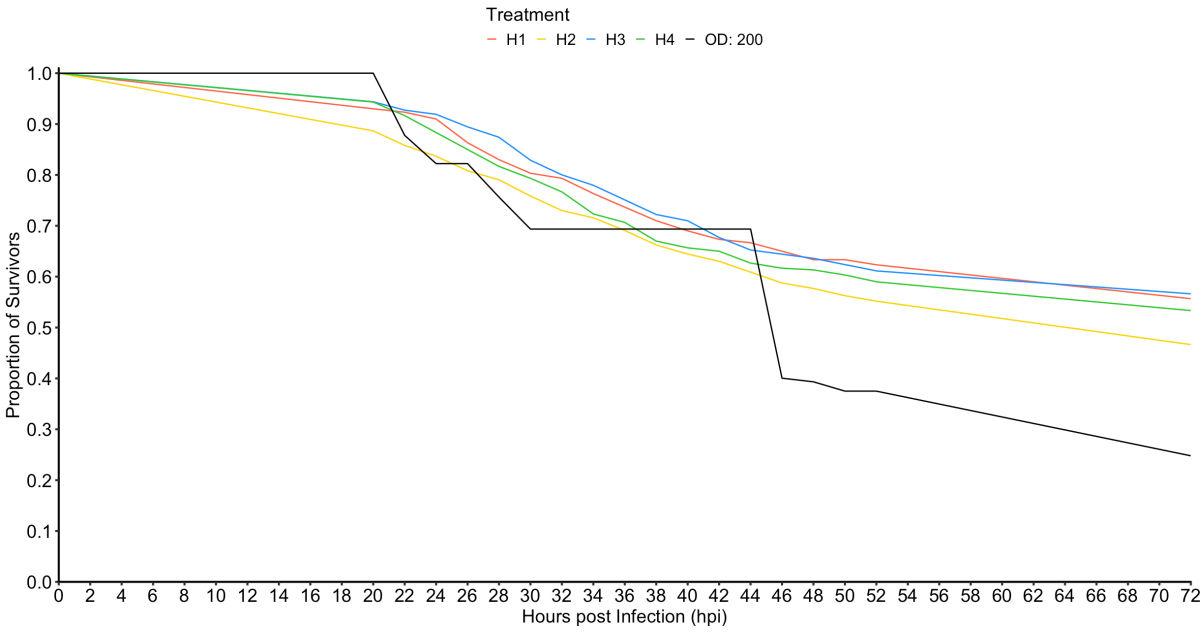
e1002513 (2016).
66. MacKay, T. F. C. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173–178 (2012).
67. Faria, V. G. & Sucena, É. From nature to the lab: Establishing Drosophila resources for evolutionary genetics. *Front. Ecol. Evol.* **5**, 1–9 (2017).
68. Lafuente, E., Duneau, D. & Beldade, P. Genetic basis of thermal plasticity variation in Drosophila melanogaster body size. *PLoS Genet.* **14**, 1–24 (2018).
69. Bou Sleiman, M. S. *et al.* Genetic, molecular and physiological basis of variation in Drosophila gut immunocompetence. *Nat. Commun.* **6**, 7829 (2015).
70. Durham, M. F., Magwire, M. M., Stone, E. A. & Leips, J. Genome-wide analysis in Drosophila reveals age-specific effects of SNPs on fitness traits. *Nat. Commun.* **5**, 1–8 (2014).
71. Fry, A. J., Palmer, M. R. & Rand, D. M. Variable fitness effects of Wolbachia infection in Drosophila melanogaster. *Heredity (Edinb).* **93**, 379–389 (2004).
72. Teixeira, L., Ferreira, Á. & Ashburner, M. The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster. *PLoS Biol.* **6**, 2753–2763 (2008).
73. Duneau, D. *et al.* Stochastic variation in the initial phase of bacterial infection predicts the probability of survival in D. melanogaster. *Elife* **6**, 1–23 (2017).
74. R Core Team. R: A Language and Environment for Statistical Computing. (2020).
75. Jane White, K. A. & Gilligan, C. A. The role of initial inoculum on epidemic dynamics. *J. Theor. Biol.* **242**, 670–682 (2006).
76. Garlapow, M. E., Huang, W., Yarboro, M. T., Peterson, K. R. & Mackay, T. F. C. Quantitative genetics of food intake in Drosophila melanogaster. *PLoS One* **10**, 1–25 (2015).
77. Edgecomb, R. S., Harth, C. E. & Schneiderman, A. M. Regulation of feeding behavior in adult Drosophila melanogaster varies with feeding regime and nutritional state. *J. Exp. Biol.* **197**, 215–235 (1994).
78. Cognigni, P., Bailey, A. P. & Miguel-Aliaga, I. Enteric neurons and systemic signals couple nutritional and reproductive status with intestinal homeostasis. *Cell Metab.* **13**, 92–104 (2011).

Appendix

Pilot Experiments

Bacterial Load Upon Death has to be measured on dead hosts within the timeframe of *P. entomophila* replication (about 30 min). Because of this, pilot experiments were conducted to select a time interval with maximum duration of 8 hours that showed the highest mortality, in order to collect as much BLUD data as possible (before plating). To this end, mortality of infected Mel1 flies every 2 hours between 20 and 54 hpi (Appendix 1). However, the pilot revealed a steady rate of mortality between 20 and 54 hpi, making it impossible to determine such tighter time-window. This led us to wonder if by increasing the bacterial concentration from OD_{600} = 100 to OD_{600} = 200 the mortality would show a peak instead of a steady rate. As such, on the second pilot we followed the mortality of flies infected with a bacterial concentration of OD_{600} = 200 every 2 hours between 20 and 30 hpi and 46 and 50 hpi (at which point we realized survival was stable) (Appendix 1). The result of the experiment was a steeper survival curve with steady mortality rate, and only about 25% of the infected individuals surviving at the end. Since by increasing the bacterial concentration we did not observe a peak in mortality, we decided to maintain the concentration of OD_{600} = 100 (to account for the high genetic variability found between DGRP lines) and chose the interval that was more convenient considering the protocol (since any interval with the same duration would lead to similar mortality).

Appendix 1 – Oral infection of Mel1 outbred flies with *P. entomophila* in a concentration of OD_{600} = 100 for 1, 2, 3 and 4 hours produces the same survival outcome. Mortality rate remained steady between 20 and 54 hpi. Increasing the bacterial concentration to OD_{600} = 200 leads to a steeper survival curve with steady mortality rate, and lower percentage of individuals surviving at the end (comparing to OD_{600} = 100). The Y-axis represents the proportion of survivors and the X-axis the time post infection. Each colour represents a different treatment: red for 1 hour of feeding with OD_{600} = 100, yellow for 2 hours of feeding with OD_{600} = 100, blue for 3 hours of feeding with OD_{600} = 100, green for 4 hours of feeding with OD_{600} = 100, and black for 3 hours of feeding with OD_{600} = 200.



Protocol for preparing a suspension of *Pseudomonas entomophila*

Beforehand:

Under sterility conditions:

1) Prepare a petri dish of *Pseudomonas entomophila* and incubate at 30ºC for one or two days;
2) Store the petri dish at 4ºC to be used later (up to one week).

Day 1:

Under sterility conditions:

1) Select a single colony of *P. entomophila* from the previously prepared petri dish and a liquid culture:
    a. with the help of a P200 touch a colony using a disposable tip;
    b. discard the tip into a tube containing 5mL of LB medium.
2) Incubate the starter colony at 28ºC for 9h with shaking;
3) Take the starter colony out of the shaker and pour the 5mL of LB into and 2L Erlenmeyer containing approx. 500 mL of LB;
4) Incubate the culture at 28ºC overnight with shaking.

Day 2:

1) Pellet the cells by centrifuging for 15 min at 15000 rpm and 4ºC;
2) Remove almost all the supernatant (do not remove all to avoid self-lysis) and resuspend the pellet in the remaining medium in a falcon tube (keep the suspension on ice from now on!);

Under sterility conditions:

3) Adjust the concentration of the bacteria suspension to OD_{600} = 100:
    a. prepare a dilution series of 10, 10^2 and 10^3 of the bacterial suspension;
    b. measure the optic density (OD_{600}) of the highest dilution, several times;
    c. calculate the average of the OD and multiply by the dilution factor; **
    d. use the formula $\boldsymbol{Ci\ x\ Vi = Cf\ x\ Vf}$ to adjust the OD;
    e. add LB to the suspension in order to obtain OD_{600} = 100;
    f. optionally add food colouring for a final concentration of 2%.

**If the suspension concentration is lower than 100, it must be centrifuged again and resuspended with less supernatant.

Protocol for oral infection with *Pseudomonas entomophila*

Prepare beforehand:

1. Suspension of *P. entomophila* at OD=10 (Protocol 1);
2. Solution of 5% sucrose;
3. Filter paper disks that fit inside small food vials and/or food bottles:
4. Separate the needed flies into fresh vials/bottles with food (use vials for less than 30 flies and bottles for more than 30).

Protocol:

1) Prepare a 1:1 bacteria/sucrose suspension, using a suspension of *P. entomophila* at OD=100 and a solution of 5% sucrose;
2) Add 2% food colouring to the suspension;
3) For each vial/bottle of flies to infect prepare an infection vial/bottle:
    a. cover the inside of the plug with humidified cotton;
    b. immerse a filter disk in the bacteria suspension (use a small petri dish to help in this step);
    c. place the disk in the bottom of the vial/bottle, completely covering the surface;
    d. make sure there are no droplets of suspension in the vial/bottle wall as flies will get stuck.
4) Flip the flies into the newly prepared infection vials/bottles without putting them to sleep and let them feed for 3h;
5) Flip/separate the infected flies into new vials/bottles.

Note: do not forget the control for the infection!!

Protocol for plating flies

Prepare beforehand:

1. 96-well plate (under sterility conditions):
    a. Fill each well with a glass pearl;
    b. Fill each well with 50 µL of 1x PBS.
2. 96-well plate for dilution (under sterility conditions):
    a. Fill each well with 90 µL

Protocol:

1) Wash the flies:

a. 70% ethanol for one minute;
b. 50% bleach for one minute;
c. 70% ethanol for one minute;
d. Water to wash the ethanol.

Under sterility conditions:

2) Place the flies in a filter paper for absorption of the remaining liquids;
3) Separate each fly to a single well in the 96-well plate;
4) Mush the flies using the TissueLyser;
5) Serially dilute the LB with bacteria in 96-well plates for dilution by adding 10 µL of the solution containing bacteria to the 90 µL of 1x PBS (prepared beforehand);
6) Plate 4 µL of each dilution in petri dishes with agar + rifampicin (0,05g/mL);

7) Incubate plates at 30^{o}C overnight;
8) Count the number of CFUs (colony forming units).