FUNCTIONAL AND POPULATION GENETICS OF *DROSOPHILA* INNATE IMMUNITY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Joo Hyun Im

May 2018

FUNCTIONAL AND POPULATION GENETICS OF DROSOPHILA INNATE IMMUNITY

Joo Hyun Im, Ph.D.

Cornell University 2018

Insects maintain a close relationship with microbes in the environment. This interaction has led to the development of the innate immune system, which is responsible for defending the host from a wide range of pathogens. As pathogens evolve to compromise host immunity, immune genes rapidly counter-evolve. In this dissertation, I present studies that answer how the innate immune system of *Drosophila* responds to a variety of pathogens and how host-pathogen interactions may have shaped the evolution of host cellular immunity. In Chapter 2, I explore how an arms race between *Drosophila* and its pathogens may have led to selective sweeps and adaptive evolution in the *Drosophila* cellular immune system. In Chapter 3, I present work on how host responses to infection with various bacteria are crucial in host survival and homeostasis. In Chapter 4, I investigate the role of genes that are expressed in a coordinated manner upon infection and identify key genes for future functional genetic studies. Altogether, these studies provide a clearer mechanistic and population genetic understanding of host-pathogen interactions and *Drosophila* immunity.

BIOGRAPHICAL SKETCH

Joo Hyun Im (임주현) was born and raised in Seoul, Korea. She graduated from Grinnell College with a Bachelor's of Arts degree in Biology and Global Development Studies in 2010. She then worked as a research technician in Dr. Rachel Brem's lab at University of California, Berkeley from 2010 to 2012. In 2013, Joo Hyun started a Ph.D. in Genetics, Genomics, and Development at Cornell University working with Dr. Brian Lazzaro.

To my family

ACKNOWLEDGEMENTS

I first thank my advisor, Brian Lazzaro, for his support, guidance, advice, and patience over the past five years. My thanks go to my committee members, Andrew Clark and Charles Danko for their advice and help along the way. I am also grateful for the guidance that Chip Aquadro, Mariana Wolfner, Philipp Messer, and Nicolas Buchon have offered me. I would like to extend my gratitude to past and present members of the Lazzaro lab for discussions and fun times: Rob Unckless, David Duneau, Moria Chambers, Susan Rottschaefer, Parvin Shahrestani, Ginny Howick, Robin Schwenke, Charlotte Renee, Gabe Fox, Vanika Gupta, Ashley Frank, Radhika Ravikumar, Jeremy McIntyre, Katie Gordon, and Katia Troha. Peter Schweitzer, Jennifer Mosher, and Anne Tate at the Sequencing Core helped me with generating RNA-seq libraries. Computing and statistical support from Qi Sun, Minghui Wang (CBSU), Lynn Johnson, and Françoise Vermeylen (CSCU) was instrumental in advancing my dissertation work. My work was supported by the R01 and R56 grants from the National Institutes of Health.

My friends in Ithaca truly made my stay here exceptional. I appreciate all the laughter and fun times I had with Adele Zhou, Andrea Slavney, Maria Sapar, Jackie Bubnell, Lauren Witter, Dan and Vilma Zinshteyn, Roman Spektor, Ian Rose, Ariel Chan, Julie Geyer, Rachel and David Hultengren, Jenny Chang, Writing group members, choir mates at VOICES, and many others. Friends elsewhere also have provided me with moral support throughout the graduate school, especially Seungmae Seo, Tiffany Hsu, Young Joo Yang, and Ojin Lee.

Lastly, I want to thank my family, and especially my spouse Karl Kremling, for the continued support and love.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Insects encounter a diverse range of microbes in the environment. Some grow and mature in decaying organic substances where microbes are prevalent and others are also vectors for microbes that cause diseases in animals and plants [1]. This interaction has necessitated a rise and development of mechanisms, which include a way to discriminate oneself from infectious non-self and target pathogens without damaging the host's own cells [2]. This first line of defense against a wide range of pathogens is called the innate immune system, which is conserved across all metazoan taxa [1, 3]. While the presence of any microbes would activate the innate immune response in general, the type of microbes that the host interacts with can induce specific physiological responses. Frequent host-pathogen interactions may also trigger an arms race that would shape and drive non-neutral evolution of the host immune system [4]. Indeed, immune genes are known to be rapidly evolving in humans, chimpanzees, and other organisms [5–8]. Therefore, in this dissertation, I present studies on how host innate immune system responds to a diverse range of pathogens and on how such host-pathogen interactions may have shaped the cellular immunity in *Drosophila*.

The fruit fly, *Drosophila melanogaster*, has been extensively used as a model to elucidate core mechanisms of innate immunity because flies lack adaptive immunity that can hinder studying innate immunity [2]. Using *Drosophila* infected with benign bacteria, previous research revealed the components of the signaling pathways important for immune defense and their functions [9–13]. Innate immunity consists of two main branches: humoral and

cellular responses. Bacterial presence triggers the activation of Toll and Imd pathways, two main arms of the humoral response, and then subsequent generation of antimicrobial peptides (AMPs) kills pathogens [1]. Lysine (Lys)-type peptidoglycans from Gram-positive bacteria induce Toll pathway while diaminopimelic acid (DAP)-type peptidoglycans from Gram-negative bacteria activate Imd pathway [14, 15]. Cellular antimicrobial response, on the other hand, mediates internalization and degradation of pathogens via phagocytosis and autophagy [1, 16].

While previous studies expanded our understanding of the basic principles of immune recognition and signaling, this knowledge of the immune response in *Drosophila* is based on studies that primarily used benign bacteria to stimulate the immune system [11, 12]. Highly virulent bacteria, regardless of their peptidoglycan structure, may release toxins and damage host tissue upon invasion [17, 18]. Thus, pathways that help repair tissue damage, clear bacterial toxins, or undergo metabolic re-wiring in addition to the canonical immune pathways would be crucial to the host's recovery and survival. While some genes in these pathways are activated upon infection, their roles during infection nor the regulator of those responses upon infection remain unclear [19]. The transcriptional complexity of the infection response that results in more nuanced host defense remains to be explored.

Immune genes in *Drosophila* have been subject to population genetic studies. Genes in the humoral Toll and Imd pathways are known to evolve adaptively at the amino acid sequence level in *Drosophila* [20–22]. Likewise, genes that encode opsonins and phagocytic receptors have shown evidence of adaptive evolution [8, 23–26]. On the other hand, most genes that encode pattern recognition

receptors that detect bacterial molecules have not shown evidence for adaptive evolution [8].

Bacteria are known to compromise the immune response, putting selective pressure on the host to counter-evolve [4]. Most population genetic studies on innate immunity up to date have focused on the humoral immune response genes and the phagocytosis receptor genes. The evolution of non-receptor phagocytosis genes and autophagy genes has not been extensively examined. Surveying the evolutionary patterns of all phagocytosis and autophagy genes in *Drosophila* would provide valuable information on how these genes evolved.

The overarching goal of this dissertation was to understand both functional and evolutionary aspects of the innate immunity in *Drosophila*. **In Chapter 2**, I hypothesized that host phagocytosis and autophagy genes may experience co-evolutionary pressure due to an arms race with pathogens and therefore may show a molecular evidence of adaptation. To understand the evolutionary patterns of the phagocytosis and autophagy genes, I performed population genetic analyses on these genes in *D. melanogaster* and *D. simulans*. I found signatures of recent and recurrent positive selection in several phagocytic recognition genes. In addition, glutamate receptor genes and several autophagy-related (*Atg*) genes showed signatures of recent selection in each species. Altogether, this chapter concludes that a dynamic conflict between *Drosophila* and its pathogens may shape the evolution of the host cellular immune system.

In Chapter 3, I explored host responses to infection with a range of bacteria and asked whether infection responses to various microbes are generic or specific and which physiological responses occur upon infection. To do

so, a collaborator and I performed RNA-seq to measure the *D. melanogaster* transcriptomic response to 10 bacteria that span the spectrum of virulence and did functional genetic experiments to uncover the role of a novel regulator of infection. We found that the host regulates distinct genes upon infection with each bacterium but also triggers a core transcriptional response beyond the canonical immune response to the majority of bacteria tested. In addition, we identified a transcription factor, *CrebA*, which modulates infection tolerance by controlling the increased cellular stress upon infection. In summary, this chapter underscores the importance of non-immune physiological responses upon infection and provides insights into how each bacterium may establish infection.

In Chapter 4, I asked which genes are responsible for general immune responses to a wide range of bacterial species and how such generality is coordinated among multiple genes. To this end, I used co-expression network analysis on the RNA-seq data from Chapter 2 to identify genes that are differentially expressed upon infection in an organized manner (modules) and key genes (hub genes) that likely make this coordination possible. I discovered four modules of genes with unique functional classifications, such as lipid and carbohydrate metabolism, protein secretion and transport, and AMP-based immune response. This chapter, thus, illustrates that changes in metabolism and secretion make up a crucial part of the response to infection and provides a set of candidate genes for future functional genetic studies.

In Chapter 5, I summarized my findings from the dissertation and proposed two areas of research that would further expand our understanding of the *Drosophila*-microbe interactions and the *Drosophila* innate immunity.

CHAPTER 2

POPULATION GENETIC ANALYSIS OF AUTOPHAGY AND PHAGOCYTOSIS GENES IN *DROSOPHILA MELANOGASTER* AND *D. SIMULANS*[0]

Author: Joo Hyun Im

2.1 Abstract

Autophagy and phagocytosis are cellular immune mechanisms for internalization and elimination of intracellular and extracellular pathogens. Some pathogens have evolved the ability to inhibit or manipulate these processes, raising the prospect of adaptive reciprocal co-evolution by the host. We performed population genetic analyses on phagocytosis and autophagy genes in *Drosophila melanogaster* and *D. simulans* to test for molecular evolutionary signatures of immune adaptation. We found signatures of recent and recurrent positive selection in several phagocytic recognition genes, consistent with previous reports. Two glutamate transporters involved in phagocytosis (*gb* and *polyph*) also show a pattern of elevated sequence divergence on the *D. simulans* lineage while three genes that facilitate autophagy (*Atg14*, *Atg16*, and *Atg8b*) display signatures of recent selection in each species. These results suggest that a dynamic conflict between *Drosophila* and its pathogens may shape the evolution of the host cellular immune system.

[0]Submitted to PLOS ONE

2.2 Introduction

Phagocytosis is a primary cellular immune process in *Drosophila* [1]. During phagocytosis, extracellular pathogens are recognized by opsonins and phagocytic receptors, engulfed at the host membrane, and then internalized and degraded in compartments called phagosomes [27] (Figure 2.1). Autophagy is an alternative cellular mechanism to remove intracellular pathogens [16]. During autophagy, intracelluar bacteria and viruses are encapsulated by isolation membranes called phagosphores, which then are nucleated and expanded to form autophagosomes to destroy the pathogen [28, 29] (Figure 2.1). Both phagosomes and autophagosomes are eventually fused with a lysosome to degrade internalized pathogens [30]. While autophagy and phagocytosis were previously thought to be distinct pathways, many autophagy proteins participate in the later stages of phagocytosis [30–32]. When phagocytosis fails to eliminate pathogens due to modification or damage to the phagosome by bacteria, autophagy works as a back-up process to overcome infection [33, 34].

Pathogens evolve to escape, resist, or compromise the host immunity [4]. Bacteria are known to inhibit or evade phagocytosis by preventing host opsonins and phagocytic receptors from binding to bacterial molecules [35, 36] and blocking host signaling pathways via effector proteins and toxins [37–39]. Similarly, bacteria and viruses also interfere with host autophagy by disrupting the signaling [40, 41] and blocking the production of reactive oxygen species (ROS) that sustains autophagy [42]. When a pathogen factor interferes with a critical host protein, the host protein can counter-evolve the bacterial hindrance via novel mutations. The change in the host may place renewed evolutionary pressure on the bacteria, and the process can repeat ad infinitum leading to a

dynamic co-evolutionary conflict, or arms race [43, 44]. Host alleles that resist or overcome pathogen interference mechanisms may be adaptive, potentially resulting in signatures of positive selection as beneficial alleles favored by natural selection rise to high frequency within a population. These alleles would exhibit a reduced level of polymorphism around the selected sites, and would accumulate a proportional excess of rare variants as the population recovers from a selective sweep [45]. Recurrent adaptive fixations over long evolutionary timescales could result in an elevated rate of amino acid divergence between species [46].

In this study, we used molecular population genetic analyses to test for adaptation in autophagy and phagocytosis genes. We analyzed published sequence data from *D. melanogaster* and *D. simulans* collected in Eastern Africa [47, 48]. We found that autophagy and phagocytosis genes as a whole showed an elevated level of homozygosity in both species. An opsonin and phagocytic receptors exhibit an excess of rare variants and an increased rate of non-synonymous substitution, suggesting that they are under positive selection. In addition, glutamate receptor genes and several *Atg* genes show signatures of non-neutral evolution.

2.3 Materials and Methods

Samples used in the population genetic analysis

For *D. melanogaster*, the sequenced genomes of 197 haploid embryo Siavonga lines from the *Drosophila* Genome Nexus Project 3 were used [48]. These lines represent a single ancestral population of *D. melanogaster* from Zambia. For *D.*

simulans, genome sequences from 20 isofemale lines, 10 collected in Madagascar and 10 collected in Kenya, were used [47]. In addition, a reference sequence of *D. yakuba* was used as an outgroup [49].

Curating genes of interest and control genes from the literature

A list of genes known to be involved in phagocytosis and autophagy was established by reviewing the primary literature (Table 2.4; Table 2.5). In addition to the known phagocytosis and autophagy genes in *Drosophila*, *Drosophila* homologs of phagocytosis or autophagy genes characterized in other organisms were included. Each gene was assigned to a functional category (e.g. Autophagy induction) (Table 2.4; Table 2.5; Figure 2.1). To test whether phagocytosis and autophagy pathways show different evolutionary patterns than canonical humoral pathways, known humoral immune signaling and recognition genes were used as a comparison. To control for effects of gene structure and chromosomal position, control genes were chosen to be similar to the focal immune genes in gene length (0.5-2x length of the coding region of the focal gene) and gene location (within 60kb ± of the start site of the focal gene). Three or four control genes that matched the criteria and had no annotated immune function were chosen for each focal gene. If a focal gene had fewer than three control genes that matched the criteria, it was removed from downstream analysis. Pairing focal and control genes controlled for potential effects of gene length and recombination rate. To ensure that focal and control genes indeed are matched for local recombination rate, the predicted recombination rates for all genes were examined using the *Drosophila melanogaster* Recombination Rate Calculator (RRC, version 2.3, [50]). This tool provides estimates of recombination rates based on the Marey

map approach [50] and direct measurement [51]. Both RRC and Comeron estimates showed that the correlation between recombination rates of focal and control genes in *D. melanogaster* is strong, confirming that focal and control genes truly share a similar recombination environment (R^2=0.994 based on RRC midpoint and R^2=0.890 based on Comeron midpoint, Figure 2.2). The presence of inversions can create strong haplotype structure and influence patterns of polymorphisms [52]. To minimize the effect of inversions, no focal gene or control gene used in the study was near the boundaries of known inversions. Populations collected in Africa sometimes contain genomic segments inferred to have recent cosmopolitan (non-African) ancestry [53]. To specifically analyze African genetic variation and eliminate the effect of relatedness among individual lines, genomic regions that were thought to have come from cosmopolitan (non-African) ancestry and that showed evidence of identity-by-descent in *D. melanogaster* were masked using Perl scripts obtained from http://www.johnpool.net/genomes.html.

Processing sequence data prior to population genetics analyses

Using custom scripts and bedtools [54], the coding sequence of each gene was extracted based on coordinates from the General Feature Format (GFF) file of the reference sequence for *D. melanogaster* (FlyBase release 5.25) and from the GFF file provided by the Thornton Lab GitHub (https://github.com/ThorntonLab) for *D. simulans*. When multiple isoforms were available for a gene, the longest sequence was chosen for downstream analysis. Custom filters were applied to exclude sequences with sites containing more than 10% missing data (noted as 'N's) and then alignment was performed using PRANK [55]. Genes with poor sequencing or alignment quality, with

large regions of gap, or with no annotation in any of the species were excluded from downstream analysis. To standardize the number of lines surveyed for each gene, 149 *D. melanogaster* lines and 14 *D. simulans* lines were randomly subsampled.

Surveying the level of polymorphisms and divergence

To evaluate the patterns of polymorphisms in each species, the following population genetic statistics were calculated: Watterson's θ (θ_w), Tajima's D (TajD), normalized Fay and Wu's H (nFWH), Ewens-Watterson test statistic (EW), and the compound statistics that combined Tajima's D, normalized Fay and Wu's H with Ewens-Watterson test statistic ($DHEW$). θ_w indicates the level of DNA sequence variation, and a reduction in sequence diversity can be due to a recent selective sweep [56]. Tajima's D test statistic compares the number of pairwise differences between individuals to the total number of segregating sites and detects the level of mutations of intermediate frequency relative to mutations that segregate at low frequencies [57]. The mutations that initially arise after variation is purged in a selective sweep will necessarily be at low frequency, skewing the site frequency spectrum and causing Tajima's D to become negative [46]. Negative Tajima's D values can also be caused by population expansion. Fay and Wu's H test statistic detects excess of high frequency alleles in the derived state, which is a signature of selective sweeps that is less likely under population expansion [58]. The normalized version of the H test was used to increase statistical power. The EW test statistic measures haplotype homozygosity by comparing the observed homozygosity to the expected homozygosity [59, 60]. A P-value greater than 0.95 suggests that allele frequencies are more unevenly distributed than the neutral expectation,

which suggests directional selection. The *DHEW* statistic is a composite statistic that provides more power to detect cases of positive selection [61, 62]. By combining Tajima's *D* and normalized Fay and Wu's *H* with Ewens-Watterson's expected homozygosity (*EW*), which is generally not sensitive to recombination, the *DHEW* overcomes the effect of recombination rates on the site frequency spectrum and makes the inference more conservative. The *DHEW* test first calculates required statistics separately and then combines the respective P-values into a vector to determine whether this vector deviates from the expectation under selective neutrality [61, 62]. A significant P-value points to positive selection. To calculate aforementioned statistics, the program DH was used [61–63]. Each test statistic was compared to 1) null distributions created using coalescent simulations with no recombination to obtain a P-value [62], and 2) the test statistics obtained from genomic control genes.

To evaluate the patterns of amino acid divergence, the ratio of non-synonymous to synonymous polymorphic sites in each species (Pn/Ps) and the ratio of non-synonymous to synonymous differences between the species (Dn/Ds) were calculated and the McDonald-Kreitman test (MK test) was performed [64] using a custom script. If the observed value of Pn/Ps is much different from Dn/Ds at a locus as determined by Fisher's exact test, that locus is rapidly diverging between the two species at the level of amino acid sequence, which is consistent with adaptive evolution. In addition, the Direction of Selection (DoS) coefficient was calculated [65]. DoS >0 indicates the presence of positive selection and DoS <0 indicates the presence of purifying selection.

Processing statistics and testing for significance

To assess whether phagocytosis and autophagy pathways as a whole show a departure from their control genes for various test statistics, the difference between the value of a given statistic in the focal gene and the median value of the statistic for the matched control genes, hereafter called a 'comparison score', was calculated for each focal-control pairing. The t-test then evaluated whether the mean of all comparison scores was significantly different from 0. If the confidence interval contains 0, the test provides no evidence for any statistically significant difference between the two groups as a whole. To identify individual genes that bear patterns of non-neutral sequence evolution in each species, the P-value of the compound test statistic *DHEW* was used. To correct for multiple testing, the p.adjust function in R based on the Benjamini and Hochberg method was implemented [66] to P-values of test statistics. To ensure that the pattern seen on focal genes is indeed due to a local selection, the comparison score was calculated for Tajima's *D*, Fay and Wu's *H*, and DoS for each focal-control pair and the scores were ranked from largest to smallest (listed as 'rank' column in Table 2.2 and Table 2.3). Randomly assigning a control gene to be "focal" and re-calculating the ranks of comparison scores for these statistics created a distribution of ranks. Then comparison scores from true coupling of focal and control genes were compared to this permutated distribution (listed as 'rank against null' column in Table 2.2 and Table 2.3). The more extreme this rank is, the stronger the confidence is that the focal gene differs from the control genes for a given statistic. We set a threshold for concluding evidence that a gene had experience a selective sweep as requiring both a significant P-value (<0.05) for DHEW statistic and comparison scores for *D*, *H*, or DoS ranking in the top 10%. Individual genes that have potentially undergone adaptive divergence on each lineage were identified as having a statistically significant (P <0.05) MK

test result and a Direction of Selection (DoS) coefficient is >0.

2.4 Results

Survey of autophagy and phagocytosis genes in *D. melanogaster*

Evidence of recent positive selection in autophagy and phagocytosis genes as a whole: To examine whether autophagy and phagocytosis genes exhibit any signature of recent selection, we calculated the following summary statistics for each focal gene and its corresponding control genes: Watterson's estimator (θ_w), Tajima's *D* (TajD), normalized Fay and Wu's *H* (nFWH), Ewens-Watterson's homozygosity (*EW*), and the compound test statistic *DHEW*. We first sought to determine whether the combined group of all autophagy and phagocytosis genes shows evidence of recent selection. To do so, we explored whether the population genetic statistics of these genes are statistically significantly different from their respective control genes by calculating the mean of comparison scores for each focal-control gene pairing for each statistic (Table 2.1). The only observed significant difference was in the *EW* statistic between autophagy and phagocytosis genes and their control genes in *D. melanogaster* (p= 0.022), pointing to the enrichment of low-frequency derived alleles with an increase in haplotype homozygosity in the focal genes. In order to find out whether this difference is indeed due to focal genes, we randomly assigned control genes to serve as "focal" genes and repeated the analysis. This approach removed this significant difference, indicating that the pattern of reduced polymorphism is unique to the autophagy and phagocytosis genes. Removing 10% of the genes with the lowest individual P-values and repeating the analysis still preserved

the pattern. Thus, this statistically significant difference between *EW* values of focal and control genes is a cumulative effect over the full set of autophagy and phagocytosis genes and is not driven by a few genes that are strongly divergent from the null expectation. Contrary to what was observed in autophagy and phagocytosis genes, sets of humoral signaling and recognition genes did not differ significantly from their control genes in any of the statistics.

Evidence of recent positive selection in individual autophagy and phagocytosis genes: We looked for signatures of recent selection in individual genes and a total of eight genes met the above criteria in *D. melanogaster* (Table 2.2): *Atg14*, *TepI*, *Atg16*, *emp*, *CG11975*, *Ykt6*, *Rac2*, and *pes*. No gene had a significant P-value after multiple testing correction. P-values listed below are uncorrected unless otherwise noted.

Out of these eight genes, the opsonin *TepI* [67] and CD36 scavenger receptors *emp* and *pes* [68–70] had been previously inferred to have undergone positive selection in *D. melanogaster* and/or have shown elevated rates of interspecific divergence in *Drosophila* [8, 24, 25, 71]. *Atg14* (*DHEW* p= 0.002) encodes an endoplasmic reticulum (ER) protein that helps autophagosomes nucleate while *Atg16* (*DHEW* p= 0.01) is involved in expanding phagosomes and autophagosomes for substrate uptake [72, 73]. The comparison scores for *D* and *H* test statistics for *Atg14* were the largest among 71 gene pairs surveyed and those of *Atg16* and its control genes were ranked in the top 10% (Table 2.2). These observations can be attributed to natural selection because effects of chromosomal position or demographic history would also have impacted the control genes. *Ykt6*, a SNARE protein involved in internalization of particles during phagocytosis and expansion of the

autophagosome membrane [74, 75], showed significant P-values for *DHEW* (p= 0.027), indicating reduced levels of variation. The *D* comparison score for *Ykt6* was the 5th largest, illustrating the enrichment of low-frequency variants specifically at this locus (Table 2.2). *CG11975* (*DHEW* p= 0.018) is a WD-repeat domain phosphoinositide-interacting (WIPI) 3/4-like gene involved in autophagic vacuole assembly while *Rac2* (*DHEW* p= 0.028) is a Rho-family GTPase that is essential for proper encapsulation of pathogens in *Drosophila* [76, 77]. The coding sequence of both genes showed patterns of non-neutral evolution as H comparison scores of these genes ranked top 10% (Table 2.2). In summary, an opsonin, two phagocytic receptors, two *Atg* genes and genes involved in the internalization step of phagocytosis showed indications of selective sweeps.

Evidence of recurrent adaptive evolution in individual autophagy and phagocytosis genes: For analyses of longer-term molecular evolution, we inferred the ancestral state of each substitution using *D. yakuba* and *D. simulans* as outgroups and assuming strict parsimony with no reverse or convergent mutation. We then compared polymorphism and divergence at synonymous and non-synonymous sites using the MK test [64] and calculated the Direction of Selection (DoS) on each gene (Table 2.3). Seven genes had significant MK test results, with one of them, *Ird1*, being significant after multiple testing correction (<0.05). However, the DoS value for each of these genes was negative, indicating that slightly deleterious mutations are segregating at these loci (Table 2.3) and providing no support for recurrent adaptive amino acid substitution.

Survey of autophagy and phagocytosis genes in *D. simulans*

Evidence of recent positive selection in autophagy and phagocytosis genes as a whole:

We tested whether phagocytosis and autophagy pathways as a whole show distinct patterns of sequence evolution than control genes in *D. simulans*. As was the case in *D. melanogaster*, the only significant difference between *D. simulans* phagocytosis and autophagy genes and their control genes was in the *EW* statistic (p= 0.002, Table 2.1). Randomly re-assigning a control gene to be "focal" eliminated this significant difference, suggesting that the observed pattern is attributable to phagocytosis and autophagy genes. Again, consistent with the patterns in *D. melanogaster*, the statistically significant difference between *EW* values of focal and control genes is not due to a few divergent genes, but rather is a pattern seen across all genes. Humoral immune signaling and recognition genes in *D. simulans* did not differ significantly from their control genes in any of the statistics.

Evidence of recent positive selection in individual autophagy and phagocytosis genes: Four genes met the criteria described in the Methods: *Rab1*, *Atg8b*, *Vamp7* and *Rac1*. The first three are involved in both phagocytosis and autophagy pathways, whereas *Rac1* is implicated only in phagocytosis. *Rab1* encodes a small guanosine triphosphate (GTP)-binding protein (GTPase) that regulates endocytosis and ER to Golgi trafficking and is also required for autophagosome formation in *Drosophila* and in mammalian cells [78]. Effector proteins from the bacteria *Shigella flexneri* and *Escherichia coli* inactivate host Rab1 protein to escape autophagy in mammalian cells [79]. *Rab1* showed a significant *DHEW* (p= 0.008) and its *H* comparison score was the 4th largest among the 68 gene pairs surveyed (Table 2.2). *Atg8b* is one of the two LC3 homologs that conjugate with phosphatidylethanolamine and localize to the autophagic membrane, creating the autophagosome [80, 81]. *Atg8b* showed an elevated level of low-frequency polymorphisms (*D*= -1.632, p= 0.041, *DHEW* p= 0.012)

and its comparison score for *H* was the largest in the 68-gene set. *Vamp7* (*D*= -1.959, p= 0.022, *DHEW* p= 0.020) is a SNARE protein that facilitates the fusion reaction between endomembrane organelles and phagosomes or autophagosomes [82] and its *D* and *H* comparison scores also ranked high (1st for *D* and 5th for *H*). Finally, *Rac1* is a GTPase involved in forming a phagocytic cup (*DHEW* p= 0.045) and its *D* comparison score (*D*= -1.930, p= 0.019) is ranked as the 4th largest contrast. Altogether, these results illustrate a strong enrichment in low-frequency variants at these loci.

Evidence of recurrent adaptive evolution in individual autophagy and phagocytosis genes: In order to test for recurrent adaptive amino acid substitution along the *D. simulans* lineage, we used sequences from *D. yakuba* and *D. melanogaster* to identify the ancestral state at each sequence position and compared polymorphism and divergence at synonymous and non-synonymous sites using the MK test. Five genes had significant MK test results before the FDR correction and positive DoS values (Table 2.3): *gb*, *polyph*, *pes*, *scb*, *Rbsn-5*. Only the P-value for *gb* remained significant after multiple test correction. *gb* and *polyph* encode glutamate transporters that regulate the extracellular glutamate levels in the nervous system [83] and control internal ROS and to promote phagosome maturation [84]. Both *gb* and *polyph* showed significant MK test results (*gb* MK corrected FET p=0.047, *polyph* MK FET p= 0.008). *pes* and *scb* are both phagocytic receptors [70, 85] that showed significant P-values (*pes* MK FET p= 0.002, *scb* MK FET p=0.008). Additionally, *Rbsn-5*, which facilitates phagosome maturation [27], showed a significant abundance of synonymous polymorphisms (MK FET p=0.01). We identified no autophagy genes evolving with an elevated rate of amino acid substitution along the *D. simulans* lineage.

2.5 Discussion

Dynamic conflict between hosts and pathogens may result in co-evolutionary adaptation in host genes, resulting in signatures of positive selection. Previous work to understand the molecular evolutionary patterns of immune genes in *Drosophila* has enriched our understanding of how the innate immune system has evolved. However, most population genetic studies on innate immunity have so far focused on the humoral immune response genes and phagocytic receptor genes, so the evolution of most of cellular immunity remains to be understood. In this study, we examined molecular evolutionary patterns of autophagy and phagocytosis genes in *D. melanogaster* and *D. simulans*. We found that phagocytosis and autophagy pathways as a whole showed an elevated level of haplotype homozygosity in both species, suggesting that genes in these pathways demonstrate small indications of adaptation that collectively result in a statistically measurable deviation from neutrality. Our results also agree with previous findings that phagocytic recognition genes evolve under positive selection. Additionally, we newly show that genes that were not previously tested as components of the immune response, including glutamate receptors and other internalization genes, also evolve non-neutrally.

Positive selection on recognition genes in phagocytosis

We found that genes encoding an opsonin (*TepI*) and scavenger receptors (*emp* and *pes*) that bind to pathogens evolve under positive selection in *D. melanogaster* (Table 2.2) and that the receptor genes *pes* and *scb* show elevated rates of amino acid evolution on the *D. simulans* lineage (Table 2.3). Opsonins and phagocytic receptors have been hypothesized to evolve under

host-pathogen co-evolution because they directly bind to, and interact with, molecules from pathogens in order to promote phagocytosis [8, 23–25, 86]. In *Anopheles* mosquitoes, *TepI* contains a thioester bond known to bind the microbial surface and promotes phagocytosis [87] and it is believed to act similarly in *Drosophila* [88]. Previous research also reported that phagocytic receptors, *emp*, *pes*, and *scb*, bore evidence of recent and adaptive evolution in *Drosophila* [8, 24, 71]. However, our data did not show an excess of nonsynonymous fixations between *D. melanogaster* and *D. simulans* in *TepII*, *TepIV* (opsonin), *Sr-CI* (scavenger receptor) and nimrod genes (*NimC1*, *NimB4*) as previously reported [8, 23, 24]. These discrepancies may come from differences in geographic locations from where samples were obtained, as different populations may be exposed to different types and frequencies of pathogens.

In the present study, phagocytosis genes and especially phagocytic recognition genes, make up a large proportion of the genes whose sequence changes appear to be most heavily shaped by adaptive evolution. In *D. melanogaster*, three of the eight genes that showed a signature of recent positive selection are phagocytosis recognition genes (*TepI*, *emp*, *pes*), and two are involved in the internalization step of phagocytosis (*Ykt6*, *Rac2*). In *D. simulans*, two of the five genes that showed elevated amino acid divergence are phagocytic receptors (*pes*, *scb*), while the other three genes facilitate phagosome maturation (*gb*, *polyph*, *Rbsn-5*). One possible explanation is that the pathogen molecules that phagocytosis receptors bind to may be diverse and variable in sequence and thus in structure. This diversity may prompt the selective pressure acting on the host genes to guide their evolution in a non-neutral way in order to effectively escape the interference. In short, our results confirm

previously reported signatures of positive selection in several phagocytic receptor genes.

Positive selection on glutamate transporters

We identified two glutamate transporter genes, *gb* and *polyph*, to have elevated amino acid sequence divergence on the *D. simulans* lineage (Table 2.3). *gb* encodes a glutamate transporter, which controls the extracellular glutamate levels in the nervous system [83]. When *gb* is mutated, glutamate level in the hemolymph is reduced, synthesis of glutathione (a major antioxidant) is disrupted, and the intracellular ROS is increased [89], leading to a failure in producing mature phagosomes and in phagocytosis of *Staphylococcus aureus* in *Drosophila* [84]. Similarly, *polyph*, another putative glutamate transporter gene, plays a role in regulating glutamate level and flies with a disrupted *polyph* exhibit decreased phagocytosis of *S. aureus* [84]. Although it is unknown whether these proteins physically interact with each other, both genes are expressed in *Drosophila* blood cells and share a function [84]. Due to this shared function and evolutionary pattern, it is tempting to speculate that positive selection may be acting on these proteins together. A host protein that evolves to escape pathogen interference may also evolve away from its native function within the host. Thus, compensatory mutation in interacting proteins that restore full function could become adaptive [90]. However, the effect of pleiotrophy may be possibly present as well.

It is interesting to note that evidence for accelerated rates of amino acid evolution was only present on the *D. simulans* lineage, but not on the *D. melanogaster* lineage. While they are closely related species, the locations in which they were collected, their ecology, and their demographic histories are

not identical [91]. It is possible that the pathogens that specifically interact with *D. simulans* might have caused this evolutionary pattern. Alternatively, a larger effective population size in *D. simulans* may have allowed more statistical power to detect adaptation in *D. simulans* [92].

Positive selection on *Atg* genes

The core autophagy machinery is conserved from yeast to higher eukaryotes and consists of *Atg* (Autophagy-related) genes [28]. Besides removal of pathogens, autophagy is also known to degrade damaged host proteins and organelles to recycle nutrients during stressful conditions, such as starvation [29]. Its other role in maintaining an organism's homeostasis could act as a constraint from selective pressure to adapt in the host and autophagy genes in general may be more likely to be under evolutionary constraint. However, we still identified signatures of recent positive selection on *Atg14* and *Atg16* in *D. melanogaster* and *Atg8b* in *D. simulans*.

Atg14 is conserved in both *Drosophila* and mammals and is involved in the nucleation of phagosphore membrane [16]. Upon infecting mouse and human cell lines, the intracellular bacterium *Brucella abortus* forms a *Brucella*-containing vacuole using the host proteins Atg14 and Beclin-1, the mammalian homolog of Atg6, in order to be trafficked to the ER where the bacterium proliferates [93]. Similar interactions with pathogens may lead to our observed evolutionary pattern at *Atg14* in *D. melanogaster*. *Atg16* may have experienced a strong selective pressure because it is responsible for removing both intra- and extracellular pathogens [73]. We additionally found that *Atg8b* and its paralog, *Atg8a*, show contrasting patterns between *D. melanogaster* and *D. simulans*. *Atg8b* is enriched with rare alleles in *D. simulans* but not in *D. melanogaster*,

while *Atg8a* showed a skew in its site frequency spectrum towards rare variants in *D. melanogaster* but not in *D. simulans*. During mammalian autophagy, the autophagic membrane is formed when cytosolic LC3, a homolog of *Atg8a* and *Atg8b*, is conjugated with phosphatidylethanolamine (Moy and Cherry 2013). Whether Atg8a and Atg8b directly interact with bacterial factors in *Drosophila* is unknown. However, previous work in human cell lines infected with a virulent *Legionella pneumophila* strain showed that the bacteria's RavZ effector cleaves the C-terminal end glycine residue (DESVYG to DESVY) of the host's LC3, thereby impairing the host's ability to assemble autophagosomes [40]. A similar arms race may be happening between *Atg8b* in *D. simulans* and a bacterial factor. Likewise, *D. melanogaster* may be using a different paralog, *Atg8a*, in order to respond to pathogen pressure. Since autophagy is involved in many physiological contexts, we cannot be certain that the source of evolutionary pressure driving positive selection is mediated by immune-related function. However, given the strong evidence that host-microbe interactions lead to co-adaption in many species, it is reasonable to speculate that an arms race may occur between host *Atg* genes and bacterial factors.

Soft sweeps and balancing selection

Our study mainly focused on detecting strong signatures of natural selection, as would be expected from hard selective sweeps favoring novel mutations [94], and on recurrent adaptive amino acid substitutions in a gene [64, 95]. These are the signatures expected from classic arms race models, which was our core biological hypothesis. We have less power to detect other forms of adaptation that might act on immune system genes, including soft sweeps of adaptation from standing genetic variation and balancing selection.

Soft sweeps refer to a mode of adaptation where multiple distinguishable adaptive alleles are present in the population at the same time [96]. Soft sweeps may be common in *D. melanogaster* [97, 98] but are not easily detected by frequency-based statistics such as Tajima's *D* or Fay and Wu's *H* because genetic diversity is not as severely reduced when the selected site is on multiple haplotype backgrounds or is not driven to complete fixation [99]. A previous study that looked for genes evolving via soft sweeps in *D. melanogaster* [98] showed evidence for soft sweeps in chromosomal regions that include *Vamp7* and *Sr-CII*, although the precise targets of these sweeps remain unidentified.

In the context of host-pathogen interactions, balancing selection can theoretically be generated if polymorphisms arise at co-evolving loci of both hosts and pathogens and two or more alleles are maintained at static intermediate or oscillating frequencies, or if alleles are costly in the absence of infection so cannot be driven to fixation [4, 7, 100, 101]. The genomic signatures of balancing or negative-frequency-dependent selection can be detected in organisms whose breeding structures or population sizes result in linkage disequilibrium that extends over long physical stretches of chromosomes, such as *Arabidopsis* [100, 102] and humans [103, 104]. However, it is much more difficult to detect balancing selection in organisms like *D. melanogaster* that have large population sizes and high rates of recombination [105, 106]. Therefore, we cannot rule out the possibility that some components of autophagy and phagocytosis systems in *Drosophila* may have evolved under undetected balancing selection.

Significance

We used a population genetic analysis to understand the molecular

evolutionary patterns of autophagy and phagocytosis genes in *D. melanogaster* and *D. simulans*. We found evidence of recent selection and an elevated rate of divergence in several genes. Our methods confirmed that phagocytic recognition genes might be evolving in a non-neutral way and revealed novel instances of positive selection in glutamate transporter genes and autophagy-related genes. These results indicate that a dynamic conflict between *Drosophila* and its pathogens may influence the molecular evolution of mechanisms that internalize and degrade pathogens, as has been previously shown for genes of the humoral immune response pathways.

Table 2.1: **Evaluation of polymorphism and divergence at the pathway level**

	D. melanogaster				
	Test statistic				
Class	**θ_w**	**TajD**	**nFWH**	**EW**	**DoS**
Autophagy & Phagocytosis	0.8409 (-2.3288, 4.0105) 0.5984	0.0470 (-0.0673, 0.1613) 0.4151	-0.0524 (-0.2487, 0.1439) 0.5961	**0.0497** **(0.0073, 0.0920)** **0.0221**	-0.0163 (-0.0578, 0.0251) 0.4348
Humoral Signaling & Recognition	-1.2716 (-4.0877, 1.5446) 0.3619	0.0305 (-0.2174, 0.2784) 0.8024	-0.2742 (-0.5973, 0.0490) 0.09298	0.0174 (-0.0456, 0.0804) 0.574	0.0310 (-0.0522, 0.1143) 0.4484

	D. simulans				
	Test statistic				
Class	**θ_w**	**TajD**	**nFWH**	**EW**	**DoS**
Autophagy & Phagocytosis	-0.2651 (-4.4637, 3.9333) 0.9	0.0232 (-0.1035, 0.1499) 0.7155	-0.0676 (-0.2071, 0.0719) 0.337	**0.0574** **(0.0211, 0.0937)** **0.002403**	-0.0218 (-0.0801, 0.0366) 0.4574
Humoral Signaling & Recognition	-1.711 (-6.1803, 2.7576) 0.4377	0.005 (-0.1762, 0.1853) 0.9591	-0.0771 (-0.3449, 0.1908) 0.5588	-0.0051 (-0.0256, 0.0154) 0.6126	0.0452 (-0.0898, 0.1801) 0.4919

θ_w, Watterson's θ; TajD, Tajima's D; nFWH, normalized Fay and Wu's H; EW, Ewens-Watterson statistic; DoS, Direction of Selection; For a given statistic, each value represents the mean of comparison scores for each focal-control gene pairing and the values in parenthesis are the 95% confidence intervals with an associated P-value below. Significant deviation is bolded.

Table 2.2: **Genes that show evidence of recent selection in *D. melanogaster* and in *D. simulans***

D. melanogaster

Gene ID	Gene Name	Function	θ_w	TajD	nFWH	EW	DHEW	DHEW corrected	TajD rank	TajD rank against null	nFWH rank	nFWH rank against null
FBgn0039636	*Atg14*	Autophagy: Nucleation	9.143	**-1.926**	**-1.898**	0.060	**0.002**	0.052	**1 (1.45%)**	**1 (0.40%)**	**3 (4.35%)**	**8 (3.19%)**
FBgn0041183	*TepI*	Phagocytosis: Recognition - opsonin	22.410	**-1.724**	-1.348	0.014	**0.003**	0.052	8 (11.59%)	42 (16.73%)	8 (11.59%)	**21 (8.37%)**
FBgn0039705	*Atg16*	Autophagy/Phagocytosis: Phagosome maturation; autophagosome expansion	5.916	-1.247	-1.715	0.130	**0.010**	0.058	55 (79.71%)	197 (78.49%)	**5 (7.25%)**	**12 (4.78%)**
FBgn0010435	*emp*	Phagocytosis: Recognition - receptor	4.482	-1.177	-1.544	0.202	**0.016**	0.058	63 (91.30%)	225 (89.64%)	**4 (5.80%)**	**9 (3.59%)**
FBgn0037648	*CG11975*	Autophagy: Nucleation	2.151	-1.432	-0.713	0.321	**0.018**	0.059	26 (37.68%)	125 (49.80%)	7 (10.14%)	**17 (6.77%)**
FBgn0260858	*Ykt6*	Autophagy/Phagocytosis: Phagosome internalization; autophagosome expansion/fusion with lysosome	2.869	**-2.151**	-0.617	0.552	**0.027**	0.069	**4 (5.80%)**	**18 (7.17%)**	31 (44.93%)	128 (51.00%)
FBgn0014011	*Rac2*	Phagocytosis: Internalization	5.378	-0.828	-1.506	0.045	**0.028**	0.069	41 (59.42%)	153 (60.96%)	**6 (8.70%)**	**14 (5.58%)**
FBgn0031969	*pes*	Phagocytosis: Recognition - receptor	15.418	-1.259	-0.196	0.014	**0.030**	0.069	**3 (4.35%)**	**15 (5.98%)**	52 (75.36%)	192 (76.49%)

D. simulans

Gene ID	Gene Name	Function	θw	TajD	nFWH	EW	DHEW	DHEW corrected	TajD rank	TajD rank against null	nFWH rank	nFWH rank against null
FBgn0191479	*Rab1*	Autophagy/Phagocytosis: Fusion with lysosome	1.572	-1.405	-0.868	0.367	**0.008**	0.315	8 (12.50%)	27 (11.49%)	**4 (6.25%)**	**7 (2.98%)**
FBgn0191695	*Atg8b*	Autophagy/Phagocytosis: Phagocytosis internalization; autophagosome expansion	2.830	-1.632	-0.883	0.286	**0.012**	0.315	7 (10.94%)	27 (11.49%)	**1 (1.56%)**	**5 (2.13%)**
FBgn0181823	*Vamp7*	Autophagy/Phagocytosis: Fusion with lysosome	1.887	-1.959	-0.760	0.439	**0.020**	0.315	**1 (1.56%)**	**2 (0.85%)**	**5 (7.81%)**	**7 (2.98%)**
FBgn0185198	*Rac1*	Phagocytosis: Internalization	2.830	-1.930	-0.302	0.235	**0.045**	0.480	**4 (6.25%)**	**15 (6.38%)**	17 (26.56%)	52 (22.13%)

θ_w, Watterson's θ; TajD, Tajima's *D*; nFWH, normalized Fay and Wu's *H*; EW, Ewens-Watterson statistic; DHEW, *DHEW* compound statistic P-value; DHEW corrected, *DHEW* compound statistic P-value with multiple testing correction. For each focal-control pair, comparison score was calculated for Tajima's *D* and Fay and Wu's *H* and the scores were ranked from largest to smallest in the 'rank' column. Random assignment of a control gene to be "focal" and re-calculating the ranks of comparison scores created a distribution of ranks. The ranks of comparison scores from true coupling of focal and control genes compared to this distribution are listed in the 'rank against null' column.

Table 2.3: **Genes that show evidence of amino acid divergence on the *D. melanogaster* lineage and on the *D. simulans* lineage**

D. melanogaster

Gene ID	Gene Name	Pn	Dn	Ps	Ds	MKcodons	Function	FETpval	FET corrected	DoS	DoS rank	DoS rank against null rank
FBgn0260935	*Ird1*	30	1	47	59	1119	Autophagy/Phagocytosis: Phagosome maturation; autophagosome nucleation	**>0.001**	**>0.001**	-0.373	17 (26.15%)	40 (16.81%)
FBgn0041100	*park*	7	1	10	20	388	Autophagy: Nucleation	**0.013**	0.267	-0.364	10 (15.38%)	24 (10.08%)
FBgn0039335	*Vps33B*	34	2	76	22	509	Autophagy/Phagocytosis: Fusion with lysosome	**0.023**	0.313	-0.226	27 (41.54%)	78 (32.77%)
FBgn0010395	*betaInt-nu*	13	5	18	28	618	Phagocytosis: Recognition - receptor	**0.026**	0.313	-0.268	35 (53.85%)	117 (49.16%)
FBgn0010435	*emp*	7	0	16	15	471	Phagocytosis: Recognition - receptor	**0.029**	0.336	-0.304	25 (38.46%)	77 (32.35%)
FBgn0261064	*Rbsn-5*	21	1	43	17	409	Phagocytosis: Phagosome maturation	**0.032**	0.347	-0.273	13 (20.00%)	25 (10.50%)
FBgn0041181	*TepIII*	95	5	192	27	1182	Phagocytosis: Recognition - opsonin	**0.046**	0.394	-0.175	12 (18.64%)	24 (10.08%)

D. simulans

Gene ID	Gene Name	Pn	Dn	Ps	Ds	MKcodons	Function	FETpval	FET corrected	DoS	DoS rank	DoS rank against null rank
FBgn0189637	***gb***	**4**	**7**	**53**	**8**	**447**	**Phagocytosis: Phagosome maturation**	**0.001**	**0.047**	**0.396**	**51 (100%)**	**179 (95.21%)**
FBgn0183718	*pes*	3	12	33	15	392	Phagocytosis: Recognition - receptor	**0.002**	0.065	0.361	41 (80.39%)	142 (75.53%)
FBgn0187055	*polyph*	3	4	47	5	320	Phagocytosis: Phagosome maturation	**0.008**	0.138	0.384	**47 (92.16%)**	**173 (92.02%)**
FBgn0182861	*scb*	11	12	75	21	814	Phagocytosis: Recognition - receptor	**0.008**	0.138	0.236	**49 (96.08%)**	**174 (92.55%)**
FBgn0045586	*Rbsn-5*	4	4	45	4	397	Phagocytosis: Phagosome maturation	**0.010**	0.142	0.418	**50 (98.04%)**	**178 (94.68%)**

Pn, the number of non-synonymous polymorphisms; Dn, the number of non-synonymous substitutions; Ps, the number of synonymous polymorphisms; Ds, the number of synonymous substitutions; MKcodons, the total number of codons subjected to the MK test; FETpval, P-value from Fisher's exact test; FET corrected, P-value from Fisher's exact test with multiple testing correction; DoS, direction of selection; For each focal-control pair, comparison score was calculated for DoS and the scores were ranked from largest to smallest in the 'rank' column. Random assignment of a control gene to be "focal" and re-calculating the ranks of comparison scores created a distribution of ranks. The ranks of comparison scores from true coupling of focal and control genes compared to this distribution are listed in the 'rank against null' column.

Table 2.4: **List of all genes in *D. melanogaster* surveyed for population genetic analysis**

Gene ID	Gene name	Function type	Function subtype	Matching control genes	Reference
FBgn0003231	*ref(2)P*	Autophagy	Expansion	FBgn0002044;FBgn0032787; FBgn0032814	Nagy et al., 2014; Polson et al., 2010; Nezis et al., 2008
FBgn0021796	*Tor*	Autophagy	Induction	FBgn0020367;FBgn0032465; FBgn0032467	Zirin and Perrimon 2010; Moy and Cherry, 2013; Hegedűs et al., 2014
FBgn0030960	*Atg101*	Autophagy	Induction	FBgn0001090;FBgn0030952; FBgn0030966	Zirin and Perrimon 2010; Nagy et al., 2014
FBgn0037363	*CG1347/Atg17*	Autophagy	Induction	FBgn0017550;FBgn0027951; FBgn0037364	Zirin and Perrimon 2010; Nagy et al., 2014
FBgn0261108	*Atg13*	Autophagy	Induction	FBgn0037573;FBgn0037578; FBgn0062412	Zirin and Perrimon 2010; Hegedűs et al., 2014; Moy and Cherry, 2013; Nagy et al., 2014
FBgn0037648	*CG11975*	Autophagy	Nucleation	FBgn0026380;FBgn0028708; FBgn0037654;FBgn0037656	Rusten and Simonsen 2008
FBgn0041100	*park*	Autophagy	Nucleation	FBgn0028427;FBgn0037051; FBgn0037059;FBgn0041607	Xie and Klionsky, 2007; Moy and Cherry, 2013; Polson et al., 2010
FBgn0039636	*CG11877/Atg14*	Autophagy	Nucleation	FBgn0014869;FBgn0039631; FBgn0039641;FBgn0041588	Florey and Overholtzer 2012; Moy and Cherry, 2013
FBgn0016700	*Rab1*	Autophagy/Phagocytosis	Fusion with lysosome	FBgn0038870;FBgn0038872; FBgn0038877;FBgn0038878	Dong et al., 2012; Flannagan et al., 2009;Winslow et al., 2010; Huang and Brumell, 2014
FBgn0027605	*Vps4*	Autophagy/Phagocytosis	Fusion with lysosome	FBgn0030873;FBgn0030877; FBgn0030881;FBgn0052557	Rusten et al., 2007; Flannagan et al., 2009
FBgn0033452	*CG1599/Vamp7*	Autophagy/Phagocytosis	Fusion with lysosome	FBgn0023175;FBgn0023180; FBgn0033454;FBgn0033458	Flannagan et al., 2009; Moreau et al., 2011; Takáts et al., 2013; Huang and Brumell, 2014
FBgn0037231	*Vps24*	Autophagy/Phagocytosis	Fusion with lysosome	FBgn0037228;FBgn0037230; FBgn0037235;FBgn0250821	Rusten et al., 2007; Flannagan et al., 2009
FBgn0039335	*CG5127/Vps33B*	Autophagy/Phagocytosis	Fusion with lysosome	FBgn0039341;FBgn0039343; FBgn0040212;FBgn0053494	Akbar et al., 2011; Takáts et al., 2014

FBgn0038539	*Atg8b*	Autophagy/Phagocytosis	Phagocytosis internalization; autophagosome expansion	FBgn0038536;FBgn0038540; FBgn0051251;FBgn0051360	Ichimura et al., 2000; Kirisako et al., 2000; Moy and Cherry, 2013; Meßling et al., 2017
FBgn0038973	*CG18594/Pebp1*	Autophagy/Phagocytosis	Phagocytosis internalization; autophagosome expansion	FBgn0038974;FBgn0038977; FBgn0038979;FBgn0085317	Reumer et al., 2009
FBgn0052672	*Atg8a*	Autophagy/Phagocytosis	Phagocytosis internalization; autophagosome expansion	FBgn0003204;FBgn0003360; FBgn0030228;FBgn0052675	Choy et al., 2012; Moy and Cherry, 2013; Meßling et al., 2017
FBgn0034110	*Atg9*	Autophagy/Phagocytosis	Phagocytosis internalization; autophagosome nucleation	FBgn0010611;FBgn0015925; FBgn0022070;FBgn0034118	Tung et al., 2010; Huang and Brummell, 2014
FBgn0034433	*EndoB*	Autophagy/Phagocytosis	Phagocytosis internalization; autophagosome nucleation	FBgn0034423;FBgn0034425; FBgn0034435;FBgn0034438	Takahashi et al., 2007; Flannagan et al., 2009; Zirin and Perrimon 2010
FBgn0260858	*Ykt6*	Autophagy/Phagocytosis	Phagosome internalization; autophagosome expansion/fusion with lysosome	FBgn0004404;FBgn0004922; FBgn0010292;FBgn0029980	Nair et al., 2011; Peltan et al., 2012
FBgn0029943	*Atg5*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0026144;FBgn0029941; FBgn0029944	Ogawa et al., 2005; Zirin and Perrimon 2010; Florey and Overholtzer, 2012
FBgn0031298	*Atg4a*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0015905;FBgn0031301; FBgn0031304;FBgn0031305	Zirin and Perrimon 2010; Florey and Overholtzer, 2012
FBgn0032935	*CG8678/Atg18b*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome	FBgn0000239;FBgn0010100; FBgn0026255;FBgn0026577	Moy and Cherry, 2013; Dooley et al., 2014

			expansion		
FBgn0034366	*Atg7*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0034363;FBgn0034367; FBgn0034372;FBgn0050122	Zirin and Perrimon 2010; Florey and Overholtzer, 2012
FBgn0035850	*CG7986/Atg18a*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0029113;FBgn0035851; FBgn0040290;FBgn0259916	Moy and Cherry, 2013; Dooley et al., 2014
FBgn0036255	*Atg12*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0011455;FBgn0036246; FBgn0036248;FBgn0036249	Zirin and Perrimon 2010; Moy and Cherry, 2013; Dooley et al., 2014
FBgn0036813	*CG6877/Atg3*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0036809;FBgn0036810; FBgn0036814	Zirin and Perrimon 2010; Florey and Overholtzer, 2012; Moy and Cherry, 2013
FBgn0039705	*CG31033/Atg16*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0000247;FBgn0039704; FBgn0039709;FBgn0039714	Fujita et al., 2008; Starr et al., 2012; Florey and Overholtzer, 2012
FBgn0010709	*Atg6*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome nucleation	FBgn0001230;FBgn0026576; FBgn0039140;FBgn0043455	Niu et al., 2008; Zirin and Perrimon 2010; Niu et al., 2012; Moy and Cherry, 2013
FBgn0015277	*Pi3K59F*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome nucleation	FBgn0015372;FBgn0034873; FBgn0034876;FBgn0034878	Juhász et al., 2008; Zirin and Perrimon 2010; Florey and Overholtzer, 2012
FBgn0030055	*CG12772*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome nucleation	FBgn0027330;FBgn0030060; FBgn0030061;FBgn0041629	Moy and Cherry, 2013; Martinez et al., 2015
FBgn0032499	*Uvrag*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome nucleation	FBgn0027586;FBgn0028406; FBgn0032494;FBgn0051728	Juhász et al., 2008; Zirin and Perrimon 2010; Moy and Cherry, 2013
FBgn0040491	*Buffy*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome nucleation	FBgn0033638;FBgn0033639; FBgn0033649;FBgn0050035	Bangs et al., 2000; Zirin and Perriomon 2010; Moy and Cherry, 2013

FBgn0260935	*Ird1*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome nucleation	FBgn0000412;FBgn0010803; FBgn0037664	Zirin and Perrimon 2010; Florey and Overholtzer, 2012; Moy and Cherry, 2013
FBgn0014010	*Rab5*	Autophagy/Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0028485;FBgn0031420; FBgn0031424	Florey and Overholtzer, 2012; Flannagan et al., 2009
FBgn0015795	*Rab7*	Autophagy/Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0039130;FBgn0039131; FBgn0039136;FBgn0039141	Flannagan et al., 2009; Huang and Brumell, 2014
FBgn0021814	*Vps28*	Autophagy/Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0033258;FBgn0033261; FBgn0033268	Rusten et al., 2007
FBgn0025806	*Rap2l*	Autophagy/Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0022708;FBgn0034957; FBgn0034962;FBgn0034974	Guo et al., 2010; Jean et al., 2015
FBgn0028741	*fab1*	Autophagy/Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0028743;FBgn0034266; FBgn0034267;FBgn0250851	Rusten et al., 2007; Flannagan et al., 2009
FBgn0010333	*Rac1*	Phagocytosis	Internalization	FBgn0029514;FBgn0035203; FBgn0035204;FBgn0035213	Sackton et al., 2007; Obbard et al., 2009
FBgn0010341	*Cdc42*	Phagocytosis	Internalization	FBgn0031057;FBgn0031058; FBgn0052528	Flannagan et al., 2011
FBgn0010348	*Arf79F*	Phagocytosis	Internalization	FBgn0037205;FBgn0044324; FBgn0053170	Flannagan et al., 2009
FBgn0014011	*Rac2*	Phagocytosis	Internalization	FBgn0035781;FBgn0035785; FBgn0035787;FBgn0052379	Sackton et al., 2007; Obbard et al., 2009; Williams et al., 2005
FBgn0024273	*WASp*	Phagocytosis	Internalization	FBgn0015589;FBgn0028373; FBgn0039600;FBgn0039602	Flannagan et al., 2011
FBgn0038320	*Sra-1*	Phagocytosis	Internalization	FBgn0026441;FBgn0026634; FBgn0038318;FBgn0038324	Peltan et al., 2012
FBgn0041781	*SCAR*	Phagocytosis	Internalization	FBgn0021761;FBgn0026147; FBgn0053129	Flannagan et al., 2011
FBgn0033572	*polyph*	Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0017414;FBgn0033566; FBgn0033569;FBgn0033574	Augustin et al., 2007; Gonzalez et al., 2013

FBgn0039487	*gb*	Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0029155;FBgn0039479; FBgn0039490	Augustin et al., 2007; Gonzalez et al., 2013
FBgn0039702	*Vps16B*	Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0002924;FBgn0039694; FBgn0039697;FBgn0039703	Akbar et al., 2011
FBgn0261064	*Rbsn-5*	Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0031997;FBgn0032003; FBgn0032004;FBgn0032005	Flannagan et al., 2009
FBgn0020240	*MCR*	Phagocytosis	Recognition-opsonin	FBgn0031985;FBgn0031988; FBgn0044323	Sackton et al., 2007
FBgn0041180	*TepIV*	Phagocytosis	Recognition-opsonin	FBgn0032796;FBgn0032798; FBgn0053116	Sackton et al., 2007; Obbard et al., 2009
FBgn0041181	*TepIII*	Phagocytosis	Recognition-opsonin	FBgn0031926;FBgn0031945; FBgn0031952	Sackton et al., 2007; Obbard et al., 2009
FBgn0041182	*TepII*	Phagocytosis	Recognition-opsonin	FBgn0031939;FBgn0031944; FBgn0031948	Sackton et al., 2007; Obbard et al., 2009
FBgn0041183	*TepI*	Phagocytosis	Recognition-opsonin	FBgn0000182;FBgn0028901; FBgn0028935	Sackton et al., 2007; Obbard et al., 2009
FBgn0003328	*scb*	Phagocytosis	Recognition-receptor	FBgn0004698;FBgn0033998; FBgn0034002;FBgn0259878	Nonaka et al., 2013
FBgn0010395	*betaInt-nu*	Phagocytosis	Recognition-receptor	FBgn0032911;FBgn0032914; FBgn0032915	Nonaka et al., 2013
FBgn0010435	*emp*	Phagocytosis	Recognition-receptor	FBgn0001147;FBgn0035086; FBgn0035091;FBgn0060296	Sackton et al., 2007
FBgn0014033	*SR-CI*	Phagocytosis	Recognition-receptor	FBgn0026394;FBgn0031609; FBgn0031611	Sackton et al., 2007
FBgn0015924	*crq*	Phagocytosis	Recognition-receptor	FBgn0015567;FBgn0023444; FBgn0027592;FBgn0031257	Sackton et al., 2007
FBgn0020377	*SR-CII*	Phagocytosis	Recognition-receptor	FBgn0025454;FBgn0033697; FBgn0033705	Sackton et al., 2007
FBgn0027594	*drpr*	Phagocytosis	Recognition-receptor	FBgn0013811;FBgn0035253; FBgn0250789	Sackton et al., 2007

FBgn0028542	*NimB4*	Phagocytosis	Recognition-receptor	FBgn0028519;FBgn0028872; FBgn0028904	Obbard et al., 2009
FBgn0028543	*NimB2*	Phagocytosis	Recognition-receptor	FBgn0012037;FBgn0025115; FBgn0028533;FBgn0032535	Obbard et al., 2009
FBgn0028936	*NimB5*	Phagocytosis	Recognition-receptor	FBgn0001961;FBgn0026373; FBgn0028430	Obbard et al., 2009
FBgn0031547	*SR-CIV*	Phagocytosis	Recognition-receptor	FBgn0031540;FBgn0031546; FBgn0031548;FBgn0051954	Sackton et al., 2007
FBgn0031969	*pes*	Phagocytosis	Recognition-receptor	FBgn0031972;FBgn0031976; FBgn0031981	Philips et al., 2005; Sackton et al., 2007
FBgn0035976	*PGRP-LC*	Phagocytosis	Recognition-receptor	FBgn0035978;FBgn0035983; FBgn0053700	Sackton et al., 2007; Obbard et al., 2009
FBgn0043576	*PGRP-SC1a*	Phagocytosis	Recognition-receptor	FBgn0033322;FBgn0050355; FBgn0050356	Sackton et al., 2007; Obbard et al., 2009
FBgn0259896	*NimC1*	Phagocytosis	Recognition-receptor	FBgn0001965;FBgn0028940; FBgn0032538	Sackton et al., 2007; Obbard et al., 2009
FBgn0040322	*GNBP2*	Humoral	Recognition	FBgn0036805;FBgn0052199; FBgn0259791	Sackton et al., 2007; Obbard et al., 2009
FBgn0043577	*PGRP-SB2*	Humoral	Recognition	FBgn0036654;FBgn0036655; FBgn0036656	Sackton et al., 2007; Obbard et al., 2009
FBgn0035977	*PGRP-LF*	Humoral	Recognition-IMD	FBgn0001227;FBgn0035985; FBgn0035987;FBgn0052039	Sackton et al., 2007; Obbard et al., 2009
FBgn0260458	*PGRP-LD*	Humoral	Recognition-IMD	FBgn0035630;FBgn0086694; FBgn0259167;FBgn0260936	Sackton et al., 2007; Obbard et al., 2009
FBgn0030310	*PGRP-SA*	Humoral	Recognition-Toll	FBgn0002723;FBgn0030311; FBgn0030314;FBgn0030316	Sackton et al., 2007; Obbard et al., 2009
FBgn0040321	*GNBP3*	Humoral	Recognition-Toll	FBgn0035941;FBgn0035943; FBgn0035951;FBgn0035955	Sackton et al., 2007; Obbard et al., 2009
FBgn0040323	*GNBP1*	Humoral	Recognition-Toll	FBgn0013717;FBgn0036806; FBgn0036807	Sackton et al., 2007; Obbard et al., 2009

FBgn0036978	*Toll-9*	Humoral	Signaling	FBgn0001247;FBgn0036967; FBgn0036974	Sackton et al., 2007; Obbard et al., 2009
FBgn0002962	*Nos*	Humoral	Signaling-IMD	FBgn0038681;FBgn0038682; FBgn0038686;FBgn0259704	Sackton et al., 2007
FBgn0014018	*Rel*	Humoral	Signaling-IMD	FBgn0015014;FBgn0037655; FBgn0037660;FBgn0260243	Sackton et al., 2007; Obbard et al., 2009
FBgn0020381	*Dredd*	Humoral	Signaling-IMD	FBgn0001341;FBgn0016038; FBgn0025634;FBgn0025640	Sackton et al., 2007; Obbard et al., 2009
FBgn0082598	*akirin*	Humoral	Signaling-IMD	FBgn0019662;FBgn0035761; FBgn0035763;FBgn0052380	Sackton et al., 2007; Obbard et al., 2009
FBgn0002930	*nec*	Humoral	Signaling-Toll	FBgn0015038;FBgn0015039; FBgn0024293;FBgn0033149	Obbard et al., 2009
FBgn0003495	*spz*	Humoral	Signaling-Toll	FBgn0039467;FBgn0039480; FBgn0039486	Sackton et al., 2007; Obbard et al., 2009
FBgn0010441	*pll*	Humoral	Signaling-Toll	FBgn0039501;FBgn0039508; FBgn0046247	Sackton et al., 2007; Obbard et al., 2009
FBgn0026318	*Traf6*	Humoral	Signaling-Toll	FBgn0029999;FBgn0030004; FBgn0030005;FBgn0030016	Sackton et al., 2007; Obbard et al., 2009
FBgn0026760	*Tehao*	Humoral	Signaling-Toll	FBgn0032515;FBgn0032517; FBgn0051847	Sackton et al., 2007; Obbard et al., 2009
FBgn0029512	*Aos1*	Humoral	Signaling-Toll	FBgn0010039;FBgn0010044; FBgn0038029;FBgn0038046	Sackton et al., 2007
FBgn0030051	*spirit*	Humoral	Signaling-Toll	FBgn0026679;FBgn0030048; FBgn0030054;FBgn0040929	Sackton et al., 2007; Obbard et al., 2009
FBgn0030774	*spheroide*	Humoral	Signaling-Toll	FBgn0030768;FBgn0030769; FBgn0030773;FBgn0030775	Sackton et al., 2007; Obbard et al., 2009
FBgn0030926	*psh*	Humoral	Signaling-Toll	FBgn0010194;FBgn0030925; FBgn0030927;FBgn0030936	Sackton et al., 2007; Obbard et al., 2009
FBgn0033402	*Myd88*	Humoral	Signaling-Toll	FBgn0002567;FBgn0026326; FBgn0083977	Sackton et al., 2007; Obbard et al., 2009

FBgn0039102	*SPE*	Humoral	Signaling-Toll	FBgn0039098;FBgn0039099; FBgn0040227;FBgn0051413	Sackton et al., 2007; Obbard et al., 2009
FBgn0039494	*grass*	Humoral	Signaling-Toll	FBgn0039488;FBgn0039489; FBgn0051065	Sackton et al., 2007; Obbard et al., 2009
FBgn0051217	*modSP*	Humoral	Signaling-Toll	FBgn0015520;FBgn0038447; FBgn0038449;FBgn0038452	Sackton et al., 2007
FBgn0022131	*aPKC*	Humoral	Signaling-Toll/Encapsulation	FBgn0027596;FBgn0029082; FBgn0033987;FBgn0033988	Obbard et al., 2009
FBgn0041205	*key*	Humoral	Signaling-Toll/IMD	FBgn0023181;FBgn0035057; FBgn0035065	Sackton et al., 2007; Obbard et al., 2009
FBgn0000250	*cact*	Humoral	Signaling-Humoral/Phagocytosis	FBgn0000307;FBgn0001994; FBgn0024734;FBgn0087041	Sackton et al., 2007; Obbard et al., 2009
FBgn0000229	*bsk*	Humoral	Signaling-JNK	FBgn0027491;FBgn0032210; FBgn0032213;FBgn0054043	Sackton et al., 2007; Obbard et al., 2009
FBgn0010303	*heph*	Humoral	Signaling-JNK	FBgn0030448;FBgn0052638; FBgn0052645;FBgn0250862	Sackton et al., 2007; Obbard et al., 2009

Table 2.5: **List of all genes in *D. simulans* surveyed for population genetic analysis**

Gene ID	Gene name	Function type	Function subtype	Matching control genes	Reference
FBgn0013883	*ref(2)P*	Autophagy	Expansion	FBgn0195564;FBgn0193178; FBgn0195578	Nagy et al., 2014; Polson et al., 2010; Nezis et al., 2008
FBgn0269678	*Atg101*	Autophagy	Induction	FBgn0187285;FBgn0188966; FBgn0187276	Zirin and Perrimon 2010; Nagy et al., 2014
FBgn0191299	*CG1347/Atg17*	Autophagy	Induction	FBgn0191092;FBgn0191087; FBgn0191083;FBgn0191091	Zirin and Perrimon 2010; Nagy et al., 2014
FBgn0190063	*Atg13*	Autophagy	Induction	FBgn0270869;FBgn0192334; FBgn0186814	Zirin and Perrimon 2010; Hegedűs et al., 2014; Moy and Cherry, 2013; Nagy et al., 2014
FBgn0190107	*CG11975*	Autophagy	Nucleation	FBgn0192290;FBgn0190114; FBgn0190111;FBgn0190112	Rusten and Simonsen 2008
FBgn0183875	*park*	Autophagy	Nucleation	FBgn0186615;FBgn0183880; FBgn0183872;FBgn0186608	Xie and Klionsky, 2007; Moy and Cherry, 2013; Polson et al., 2010
FBgn0192850	*CG11877/Atg14*	Autophagy	Nucleation	FBgn0270887;FBgn0192846; FBgn0268632;FBgn0186747	Florey and Overholtzer 2012; Moy and Cherry, 2013
FBgn0191479	*Rab1*	Autophagy/Phagocytosis	Fusion with lysosome	FBgn0190908;FBgn0190909; FBgn0191477;FBgn0191321	Dong et al., 2012; Flannagan et al., 2009;Winslow et al., 2010; Huang and Brumell, 2014
FBgn0188934	*Vps4*	Autophagy/Phagocytosis	Fusion with lysosome	FBgn0195821;FBgn0188931; FBgn0187322	Rusten et al., 2007; Flannagan et al., 2009
FBgn0181823	*CG1599/Vamp7*	Autophagy/Phagocytosis	Fusion with lysosome	FBgn0181821;FBgn0181822; FBgn0182455;FBgn0182458	Flannagan et al., 2009; Huang and Brumell, 2014
FBgn0191180	*Vps24*	Autophagy/Phagocytosis	Fusion with lysosome	FBgn0037228;FBgn0191203; FBgn0191209;FBgn0191184	Rusten et al., 2007; Flannagan et al., 2009
FBgn0192668	*CG5127*	Autophagy/Phagocytosis	Fusion with lysosome	FBgn0192672;FBgn0192674; FBgn0189729;FBgn0189727	Akbar et al., 2011; Takáts et al., 2014
FBgn0190125	*Vps16A*	Autophagy/Phagocytosis	Fusion with lysosome	FBgn0181965;FBgn0190122; FBgn0190124	Flannagan et al., 2012; Takáts et al., 2014
FBgn0191695	*Atg8b*	Autophagy/Phagocytosis	Phagocytosis internalization; autophagosome expansion	FBgn0190687;FBgn0190690; FBgn0186768;FBgn0186838	Ichimura et al., 2000; Kirisako et al., 2000; Moy and Cherry, 2013; Meßling et al., 2017

FBgn0192410	*CG18594/Pebp 1*	Autophagy/Phagocytosis	Phagocytosis internalization; autophagosome expansion	FBgn0189978;FBgn0189977; FBgn0183058;FBgn0189981	Reumer et al., 2009
FBgn0187657	*Atg8a*	Autophagy/Phagocytosis	Phagocytosis internalization; autophagosome expansion	FBgn0188563;FBgn0187655; FBgn0188564;FBgn0195877	Choy et al., 2012; Moy and Cherry, 2013; Meßling et al., 2017
FBgn0083470	*Atg9*	Autophagy/Phagocytosis	Phagocytosis internalization; autophagosome nucleation	FBgn0083433;FBgn0182931; FBgn0196829;FBgn0182941	Tung et al., 2010; Huang and Brummell, 2014
FBgn0196627	*EndoB*	Autophagy/Phagocytosis	Phagocytosis internalization; autophagosome nucleation	FBgn0183187;FBgn0183195; FBgn0196624	Takahashi et al., 2007; Flannagan et al., 2009; Zirin and Perrimon 2010
FBgn0188454	*Ykt6*	Autophagy/Phagocytosis	Phagosome internalization; autophagosome expansion/fusion with lysosome	FBgn0196085;FBgn0187778; FBgn0188449;FBgn0187773	Nair et al., 2011; Peltan et al., 2012
FBgn0083471	*Atg5*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0268718;FBgn0085967; FBgn0188425	Ogawa et al., 2005; Zirin and Perrimon 2010; Florey and Overholtzer, 2012
FBgn0194457	*Atg4a*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0194456;FBgn0194458; FBgn0194460;FBgn0194320	Zirin and Perrimon 2010; Florey and Overholtzer, 2012
FBgn0193067	*CG8678/Atg18b*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0193069;FBgn0195654; FBgn0195657;FBgn0195658	Moy and Cherry, 2013; Dooley et al., 2014
FBgn0196676	*Atg7*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0183146;FBgn0183148; FBgn0196669;FBgn0196674	Zirin and Perrimon 2010; Florey and Overholtzer, 2012

FBgn0185753	*CG7986/Atg18a*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0184741;FBgn0067157; FBgn0184737;FBgn0268599	Moy and Cherry, 2013; Dooley et al., 2014
FBgn0186045	*Atg12*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0184473;FBgn0186039; FBgn0269456;FBgn0186041	Zirin and Perrimon 2010; Moy and Cherry, 2013; Dooley et al., 2014
FBgn0186442	*CG6877/Atg3*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0184061;FBgn0184060; FBgn0184058	Zirin and Perrimon 2010; Florey and Overholtzer, 2012; Moy and Cherry, 2013
FBgn0189188	*CG31033/Atg16*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome expansion	FBgn0192904;FBgn0192905; FBgn0192908;FBgn0189150	Fujita et al., 2008; Starr et al., 2012; Florey and Overholtzer, 2012
FBgn0189857	*Atg6*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome nucleation	FBgn0067521;FBgn0192514; FBgn0189858;FBgn0192512	Niu et al., 2008; Zirin and Perrimon 2010; Niu et al., 2012; Moy and Cherry, 2013
FBgn0183516	*Pi3K59F*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome nucleation	FBgn0196352;FBgn0183510; FBgn0186872	Juhász et al., 2008; Zirin and Perrimon 2010; Florey and Overholtzer, 2012
FBgn0188490	*CG12772*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome nucleation	FBgn0187733;FBgn0187734; FBgn0188495;FBgn0188494	Moy and Cherry, 2013; Martinez et al., 2015
FBgn0195254	*Uvrag*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome nucleation	FBgn0195247;FBgn0193487; FBgn0195249;FBgn0195251	Juhász et al., 2008; Zirin and Perrimon 2010; Moy and Cherry, 2013
FBgn0182572	*Buffy*	Autophagy/Phagocytosis	Phagosome maturation; autophagosome nucleation	FBgn0197172;FBgn0197171; FBgn0182574;FBgn0182570	Bangs et al., 2000; Zirin and Perriomon 2010; Moy and Cherry, 2013
FBgn0194235	*Rab5*	Autophagy/Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0194538;FBgn0194540; FBgn0194544	Florey and Overholtzer, 2012; Flannagan et al., 2009
FBgn0192507	*Rab7*	Autophagy/Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0192502;FBgn0189863; FBgn0192506;FBgn0192511	Flannagan et al., 2009; Huang and Brumell, 2014
FBgn0186962	*Vps28*	Autophagy/Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0181967;FBgn0187033; FBgn0181960	Rusten et al., 2007

FBgn0183570	*Rap2l*	Autophagy/Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0183564;FBgn0183560; FBgn0183561;FBgn0187128	Guo et al., 2010; Jean et al., 2015
FBgn0196733	*fab1*	Autophagy/Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0183068;FBgn0183055; FBgn0183056;FBgn0196732	Rusten et al., 2007; Flannagan et al., 2009
FBgn0185198	*Rac1*	Phagocytosis	Internalization	FBgn0185324;FBgn0185316; FBgn0185197;FBgn0185192	Sackton et al., 2007; Obbard et al., 2009
FBgn0189009	*Cdc42*	Phagocytosis	Internalization	FBgn0187425;FBgn0189019; FBgn0195789	Flannagan et al., 2011
FBgn0183788	*Arf79F*	Phagocytosis	Internalization	FBgn0186733;FBgn0183787; FBgn0186728	Flannagan et al., 2009
FBgn0184778	*Rac2*	Phagocytosis	Internalization	FBgn0184779;FBgn0270100; FBgn0268325;FBgn0185709	Sackton et al., 2007; Obbard et al., 2009; Williams et al., 2005
FBgn0189560	*WASp*	Phagocytosis	Internalization	FBgn0189558;FBgn0189557; FBgn0192818;FBgn0192819	Flannagan et al., 2011
FBgn0190521	*Sra-1*	Phagocytosis	Internalization	FBgn0186830;FBgn0191830; FBgn0191836;FBgn0190525	Peltan et al., 2012
FBgn0195119	*SCAR*	Phagocytosis	Internalization	FBgn0193614;FBgn0195123; FBgn0195121	Flannagan et al., 2011
FBgn0189202	*Vps16B*	Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0189206;FBgn0189220; FBgn0189209;FBgn0189194	Akbar et al., 2011
FBgn0045586	*Rbsn-5*	Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0194890;FBgn0194895; FBgn0193834;FBgn0193833	Flannagan et al., 2009
FBgn0189637	*gb*	Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0192767;FBgn0192760; FBgn0269031	Augustin et al., 2007; Gonzalez et al., 2013
FBgn0187055	*polyph*	Phagocytosis	Phagosome maturation; fusion with lysosome	FBgn0186927;FBgn0187056; FBgn0197224	Augustin et al., 2007; Gonzalez et al., 2013
FBgn0193845	*MCR*	Phagocytosis	Recognition-opsonin	FBgn0193844;FBgn0194884; FBgn0194878	Sackton et al., 2007
FBgn0264798	*TepIV*	Phagocytosis	Recognition-opsonin	FBgn0193174;FBgn0195566; FBgn0195569	Sackton et al., 2007; Obbard et al., 2009
FBgn0268085	*TepIII*	Phagocytosis	Recognition-opsonin	FBgn0183765;FBgn0194865; FBgn0193868	Sackton et al., 2007; Obbard et al., 2009
FBgn0193881	*TepII*	Phagocytosis	Recognition-opsonin	FBgn0193876;FBgn0267689;	Sackton et al., 2007; Obbard et al., 2009

				FBgn0193871	
FBgn0182861	*scb*	Phagocytosis	Recognition-receptor	FBgn0182863;FBgn0196918; FBgn0021123;FBgn0196915	Nonaka et al., 2013
FBgn0269502	*betaInt-nu*	Phagocytosis	Recognition-receptor	FBgn0195638;FBgn0195645; FBgn0193078	Nonaka et al., 2013
FBgn0196216	*emp*	Phagocytosis	Recognition-receptor	FBgn0196214;FBgn0183652; FBgn0183654;FBgn0196225	Sackton et al., 2007
FBgn0044195	*SR-CI*	Phagocytosis	Recognition-receptor	FBgn0194116;FBgn0194101; FBgn0194677	Sackton et al., 2007
FBgn0064356	*crq*	Phagocytosis	Recognition-receptor	FBgn0194353;FBgn0194413; FBgn0194420;FBgn0194419	Sackton et al., 2007
FBgn0185159	*drpr*	Phagocytosis	Recognition-receptor	FBgn0185153;FBgn0185166; FBgn0185150	Sackton et al., 2007
FBgn0193438	*NimB4*	Phagocytosis	Recognition-receptor	FBgn0195313;FBgn0193433; FBgn0283651	Obbard et al., 2009
FBgn0193440	*NimB2*	Phagocytosis	Recognition-receptor	FBgn0262093;FBgn0069519; FBgn0195303	Obbard et al., 2009
FBgn0028936	*NimB5*	Phagocytosis	Recognition-receptor	FBgn0193452;FBgn0195296; FBgn0193437	Obbard et al., 2009
FBgn0067527	*SR-CIV*	Phagocytosis	Recognition-receptor	FBgn0194149;FBgn0194627; FBgn0194628;FBgn0194146	Sackton et al., 2007
FBgn0183718	*pes*	Phagocytosis	Recognition-receptor	FBgn0193854;FBgn0193849; FBgn0194880	Philips et al., 2005; Sackton et al., 2007
FBgn0182360	*PGRP-SC1a*	Phagocytosis	Recognition-receptor	FBgn0186937;FBgn0183669; FBgn0181923	Sackton et al., 2007; Obbard et al., 2009
FBgn0193436	*NimC1*	Phagocytosis	Recognition-receptor	FBgn0063991;FBgn0195309; FBgn0193442	Sackton et al., 2007; Obbard et al., 2009
FBgn0184066	*GNBP2*	Humoral	Recognition	FBgn0184062;FBgn0186433; FBgn0186434	Sackton et al., 2007; Obbard et al., 2009
FBgn0068647	*PGRP-SB2*	Humoral	Recognition	FBgn0186333;FBgn0186337; FBgn0184177	Sackton et al., 2007; Obbard et al., 2009
FBgn0067517	*PGRP-LA*	Humoral	Recognition-IMD	FBgn0185860;FBgn0184661; FBgn0185859	Sackton et al., 2007; Obbard et al., 2009

FBgn0185855	*PGRP-LF*	Humoral	Recognition-IMD	FBgn0185865;FBgn0185866; FBgn0185857	Sackton et al., 2007; Obbard et al., 2009
FBgn0185597	*PGRP-LD*	Humoral	Recognition-IMD	FBgn0184900;FBgn0189147; FBgn0185595;FBgn0185599	Sackton et al., 2007; Obbard et al., 2009
FBgn0067516	*PGRP-SA*	Humoral	Recognition-Toll	FBgn0188611;FBgn0188605; FBgn0268182;FBgn0268858	Sackton et al., 2007; Obbard et al., 2009
FBgn0185827	*GNBP3*	Humoral	Recognition-Toll	FBgn0185828;FBgn0184676; FBgn0185834;FBgn0184670	Sackton et al., 2007; Obbard et al., 2009
FBgn0067523	*GNBP1*	Humoral	Recognition-Toll	FBgn0184068;FBgn0184063; FBgn0186439	Sackton et al., 2007; Obbard et al., 2009
FBgn0183934	*Toll-9*	Humoral	Signaling	FBgn0183933;FBgn0186543; FBgn0186548	Sackton et al., 2007; Obbard et al., 2009
FBgn0025032	*Nos*	Humoral	Signaling-IMD	FBgn0190777;FBgn0191611; FBgn0191607;FBgn0191608	Sackton et al., 2007
FBgn0029245	*Rel*	Humoral	Signaling-IMD	FBgn0192287;FBgn0192289; FBgn0192284;FBgn0027801	Sackton et al., 2007; Obbard et al., 2009
FBgn0067524	*Dredd*	Humoral	Signaling-IMD	FBgn0085902;FBgn0188121; FBgn0188124;FBgn0188122	Sackton et al., 2007; Obbard et al., 2009
FBgn0185706	*akirin*	Humoral	Signaling-IMD	FBgn0184797;FBgn0184796; FBgn0184783	Sackton et al., 2007; Obbard et al., 2009
FBgn0067495	*nec*	Humoral	Signaling-Toll	FBgn0182047;FBgn0182046; FBgn0182044;FBgn0182232	Obbard et al., 2009
FBgn0067488	*spz*	Humoral	Signaling-Toll	FBgn0192755;FBgn0268111; FBgn0189638	Sackton et al., 2007; Obbard et al., 2009
FBgn0269378	*pll*	Humoral	Signaling-Toll	FBgn0186754;FBgn0186808; FBgn0186807	Sackton et al., 2007; Obbard et al., 2009
FBgn0187760	*Traf6*	Humoral	Signaling-Toll	FBgn0188459;FBgn0188461; FBgn0188463;FBgn0195904	Sackton et al., 2007; Obbard et al., 2009
FBgn0192033	*Aos1*	Humoral	Signaling-Toll	FBgn0190332;FBgn0190337; FBgn0190338;FBgn0192034	Sackton et al., 2007
FBgn0030774	*spheroide*	Humoral	Signaling-Toll	FBgn0187398;FBgn0269752; FBgn0187391;FBgn0187389	Sackton et al., 2007; Obbard et al., 2009
FBgn0195811	*psh*	Humoral	Signaling-Toll	FBgn0084925;FBgn0187297;	Sackton et al., 2007; Obbard et al., 2009

				FBgn0270551	
FBgn0192480	*SPE*	Humoral	Signaling-Toll	FBgn0192475;FBgn0192476; FBgn0192477;FBgn0192484	Sackton et al., 2007; Obbard et al., 2009
FBgn0186856	*grass*	Humoral	Signaling-Toll	FBgn0192763;FBgn0189636; FBgn0192765	Sackton et al., 2007; Obbard et al., 2009
FBgn0191746	*modSP*	Humoral	Signaling-Toll	FBgn0269837;FBgn0190619; FBgn0190621;FBgn0191749	Sackton et al., 2007
FBgn0196927	*aPKC*	Humoral	Signaling-Toll/Encapsulation	FBgn0196929;FBgn0182853; FBgn0182850;FBgn0182851	Obbard et al., 2009
FBgn0067525	*Dif*	Humoral	Signaling-Toll/IMD	FBgn0193286;FBgn0195463; FBgn0193292	Sackton et al., 2007; Obbard et al., 2009
FBgn0067496	*key*	Humoral	Signaling-Toll/IMD	FBgn0196230;FBgn0183633; FBgn0183636	Sackton et al., 2007; Obbard et al., 2009
FBgn0067503	*cact*	Humoral	Signaling-Humoral/Phagocytosis	FBgn0193330;FBgn0193342; FBgn0193344;FBgn0193331	Sackton et al., 2007; Obbard et al., 2009
FBgn0193697	*bsk*	Humoral	Signaling-JNK	FBgn0193685;FBgn0193691; FBgn0193686;FBgn0193689	Sackton et al., 2007; Obbard et al., 2009
FBgn0187559	*heph*	Humoral	Signaling-JNK	FBgn0195857;FBgn0187557; FBgn0188673	Sackton et al., 2007; Obbard et al., 2009

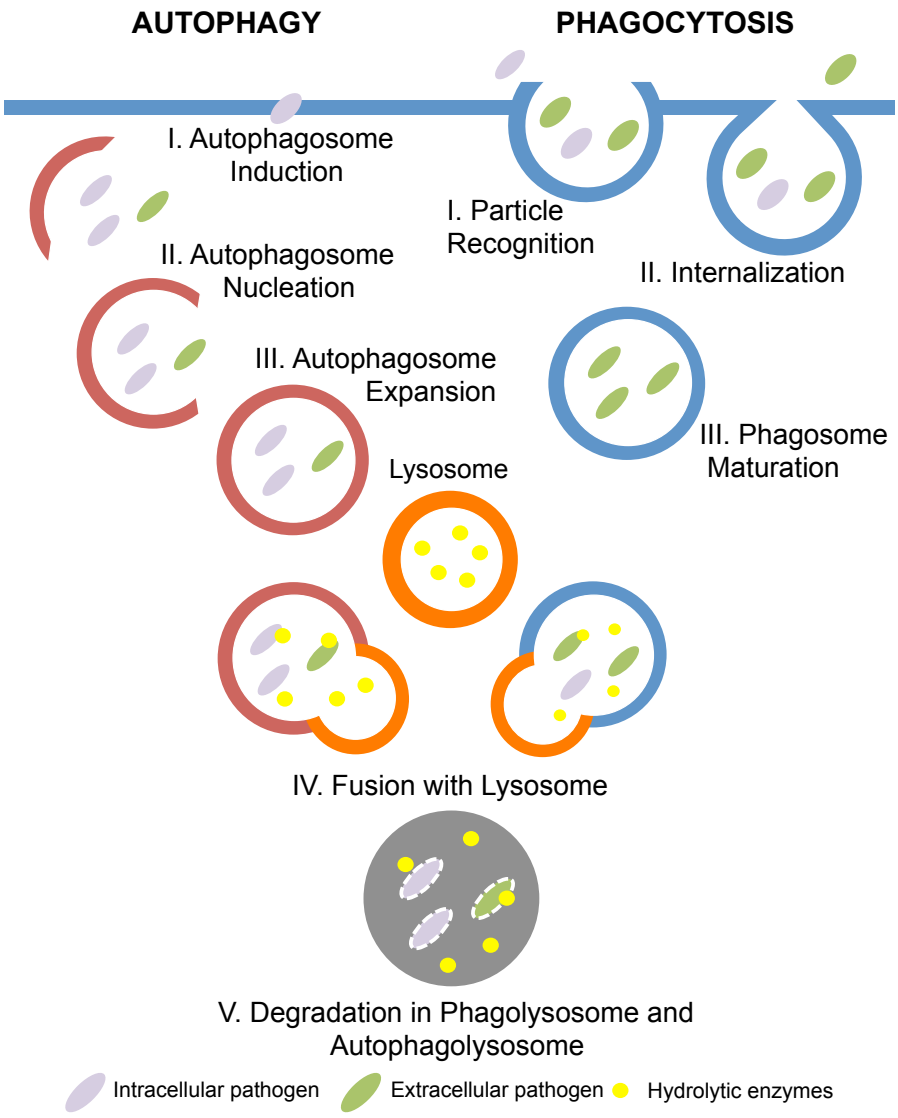


Figure 2.1: **Stages of autophagy and phagocytosis pathways.**

Genes in autophagy (red) and phagocytosis (blue) pathway function to recognize, internalize, and degrade cell debris and intracellular (purple) and extracellular pathogens (green). Organelles, such as phagosomes and autophagosomes, are form in the course of the process and are eventually fused with a lysosome full of hydrolytic enzymes (yellow) to degrade internalized pathogens.

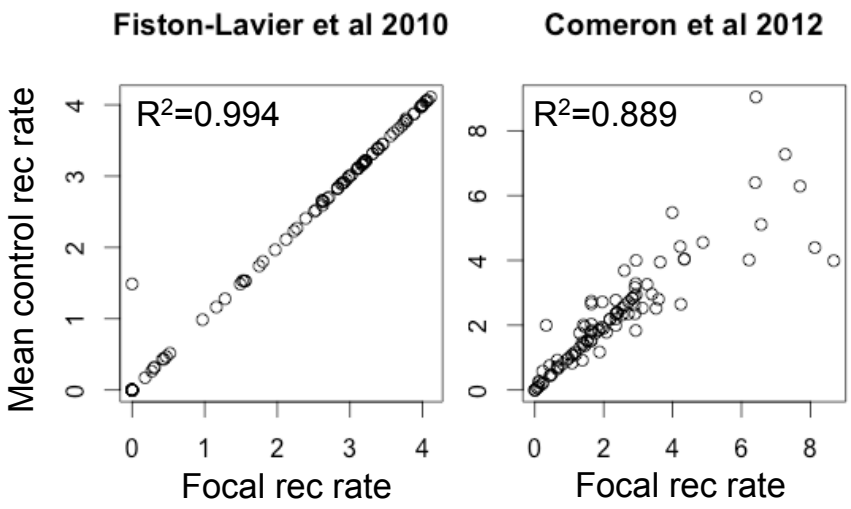


Figure 2.2: **Correlation between recombination rates of focal genes and recombination rates of corresponding control genes**

(L) Recombination rates of focal genes (x-axis) were plotted against mean recombination rates of respective control genes (y-axis) based on the study of [50]. (R) Recombination rates of focal genes (x-axis) were plotted against mean recombination rates of respective control genes (y-axis) based on the study of [51].

CHAPTER 3

COMPARATIVE TRANSCRIPTOMICS REVEALS *CREBA* AS A NOVEL REGULATOR OF INFECTION TOLERANCE IN *D. MELANOGASTER*[0]

Authors: Katia Troha*, Joo Hyun Im*, Jonathan Revah, Brian P. Lazzaro, Nicolas Buchon (*: co-first author)

3.1 Abstract

Host responses to infection encompass many processes in addition to activation of the immune system, including metabolic adaptations, stress responses, tissue repair, and other reactions. The response to bacterial infection in *Drosophila melanogaster* has been classically described in studies that focused on the immune response elicited by a small set of largely avirulent microbes. Thus, we have surprisingly limited knowledge of responses to infection that are outside the canonical immune response, of how the response to pathogenic infection differs from that to avirulent bacteria, or even of how generic the response to various microbes is and what regulates that core response. In this study, we addressed these questions by profiling the *D. melanogaster* transcriptomic response to 10 bacteria that span the spectrum of virulence. We found that each bacterium triggers a unique transcriptional response, with distinct genes making up to one third of the response elicited by highly virulent bacteria. We also identified a core set of 252 genes that are differentially expressed in response to the majority of bacteria tested. Among these, we determined

[0]Troha K, Im JH, Revah J, Lazzaro BP, Buchon N. Comparative transcriptomics reveals *CrebA* as a novel regulator of infection tolerance in *D. melanogaster*. *PLoS pathogens*. 2018 Feb;14(2):e1006847.

that the transcription factor *CrebA* is a novel regulator of infection tolerance. Knock-down of *CrebA* significantly increased mortality from microbial infection without any concomitant change in bacterial number. Upon infection, *CrebA* is upregulated by both the Toll and Imd pathways in the fat body, where it is required to induce the expression of secretory pathway genes. Loss of *CrebA* during infection triggered endoplasmic reticulum (ER) stress and activated the unfolded protein response (UPR), which contributed to infection-induced mortality. Altogether, our study reveals essential features of the response to bacterial infection and elucidates the function of a novel regulator of infection tolerance.

3.2 Author Summary

How does an organism survive infection? How generic or specific is the host response to diverse pathogens? To address these questions, we infected fruit flies with 10 different bacteria that vary in their ability to kill flies and measured changes in global gene expression. In general, we found that the host response is highly specific to individual bacteria. However, we also discovered a set of genes that changed expression in response to the majority of bacteria tested. Among these genes, we determined that the transcription factor *CrebA* is a novel regulator of the host response to infection. We found that upon infection, the immune system induces the expression of *CrebA*. *CrebA*-deficient flies are more likely to die from infection despite carrying the same number of bacteria as wildtype flies. *CrebA* is expressed in the fat body, an organ analogous to the mammalian liver and adipose tissues, where it regulates the transcription of multiple secretory pathway genes. Loss of *CrebA* during

infection triggers endoplasmic reticulum (ER) stress (a type of cellular stress), which is sufficient to sensitize flies to infection. These results suggest that the immune system can modulate host physiology to prevent the deleterious effect of infection-associated cellular stress.

3.3 Introduction

To combat infection, a host activates a combination of immune and physiological responses. While detection of microbial presence is sufficient to stimulate the innate immune response, physiological responses to infection occur as a consequence of microbial growth and virulence, and can therefore be very specific to the particular bacterium the host interacts with. Despite a growing body of literature on immunity, our knowledge of the different host processes that are activated or repressed in response to infection, and of how such responses contribute to host survival, remains limited. To identify new biological processes required to survive infection and to determine how specific or generic the immune and physiological responses to infection are, we surveyed changes in the transcriptome of *Drosophila melanogaster* in response to infection with 10 bacteria that span the spectrum of virulence.

Drosophila is a leading model system for studying how hosts respond to infection at the organismal level. To overcome infection, the fly relies on cellular and humoral innate immune responses. The cellular response consists of phagocytosis and encapsulation [107, 108]. The humoral response includes the pro-phenoloxidase cascade, which leads to the generation of reactive oxygen species and clotting, as well as the production of antimicrobial

peptides (AMPs) primarily by the fat body, an organ functionally analogous to the liver and adipose tissues of mammals [109–111]. In the early 2000s, microarray studies characterizing the transcriptional response to bacterial infection were conducted in *Drosophila* [11, 12, 112]. These experiments were based on infection with two non-pathogenic bacteria, *Micrococcus luteus* and *Escherichia coli*. This approach successfully identified a set of genes that are differentially expressed upon infection, which became known as the *Drosophila* Immune-Regulated Genes (DIRGs). A majority of the DIRGs were functionally assigned to specific aspects of the immune response phagocytosis, antimicrobial peptide synthesis, and production of reactive oxygen species among others [11]. These studies also confirmed that the Toll and Imd pathways are the major regulators of the immune response in *Drosophila*, and that both pathways direct expression of the majority of DIRGs [112]. In this model, the host response depends on the sensing of two microbe-associated molecular patterns (MAMPs): Lys-type peptidoglycan from Gram-positive bacteria, which activates the Toll pathway, and DAP-type peptidoglycan from Gram-negative bacteria, which induces the Imd pathway [14, 15, 113]. Upon activation, each pathway goes on to regulate a subset of DIRGs.

More recently, new findings have expanded our insight into the *Drosophila* response to infection. First, the Toll and Imd pathways can also be activated by virulence factors and damage-associated molecular patterns (DAMPs) [17, 18, 114–116]. Additionally, biological processes that would not be considered as classic immunological responses, such as tissue repair and regulation of metabolism, are clearly modulated by pathogenic infection [19, 117–119], and suggest that survival from pathogenic infections may require additional biological processes beyond those that are currently known [120].

In this study, we aimed to identify a comprehensive list of genes regulated by pathogenic and avirulent infections, and to determine what responses are general or specific to each infection. To that purpose, we used RNA-seq to profile the *D. melanogaster* transcriptomic response to systemic infection with 10 different species of bacteria that vary in their ability to grow within and kill the host. We found that each bacterium elicits a unique host transcriptional response. However, we also identified a small set of core genes that were differentially regulated by infection with the majority of microbes. These genes are involved in a variety of immune and non-immune functions, and a fraction of them remained highly expressed even after bacteria were cleared from the host. Among the core genes was *CrebA*, a Creb3-like transcription factor. *CrebA* expression is upregulated through both Toll and Imd signaling in the fat body following infection. Knockdown of *CrebA* significantly increased mortality from bacterial challenge but did not alter bacterial load, indicating that *CrebA* contributes to host tolerance of infection. *CrebA* regulates multiple genes involved in the secretory pathway, and the loss of *CrebA* triggered ER stress upon infection. This suggests that the *CrebA* tolerance phenotype may arise through protection from cellular stress during the rapid and dramatic response to infection.

3.4 Materials and Methods

Whole fly RNA-seq infections

Whole fly RNA-seq experiments were performed using wildtype strain Canton S flies. Flies were raised on standard yeast-cornmeal-sucrose medium

(50 g baker's yeast, 60 g cornmeal, 40 g sucrose, 7 g agar, 26.5 mL Moldex (10%), and 12 mL Acid Mix solution (4.2% phosphoric acid, 41.8% propionic acid) per 1L of deionized H_2O) at 24C and maintained at that temperature for the duration of the experiment. Individual males were infected with one of the ten experimental bacteria 5 to 8 days after eclosion from the pupal case. Control flies that were sterilely wounded or inoculated with heat-killed bacteria were handled equivalently. Flies were pin-pricked to generate septic injury. We standardized the initial inoculation dose across all bacteria to deliver 3,000 colony-forming units (CFU) per fly. The following bacteria (from overnight cultures) were used: *Micrococcus luteus* (A600 = 100), *Escherichia coli* (A600 = 100), *Serratia marcescens* Type (A600 = 1), *Ecc15* (A600 = 1), *Providencia rettgeri* (A600 = 1), *Enterococcus faecalis* (A600 = 1), *Staphylococcus aureus* (A600 = 1), *Providencia sneebia* (A600 = 1), *Serratia marcescens* Db11 (A600 = 1), and *Pseudomonas entomophila* (A600 = 1). Three sets of controls were included in the experiment: unchallenged and uninjured flies, sterilely wounded flies, and challenge with either heat-killed *P. rettgeri* or heat-killed *E. faecalis*. For every control and bacterial infection, with the exception of the 4 highly virulent infections, 20 flies were collected at 12 h, 36 h, and 132 h post-infection. For the 4 highly virulent bacteria, only the 12 h sample was collected because the majority of the flies had died before the later time points. Additionally, 20 unchallenged, uninjured flies were also collected at time 0 h as an extra control. Each sample of 20 flies was homogenized, and total RNA was isolated using a modified TRizol extraction protocol (Life Technologies). All experiments were done in triplicate. The same methodology was employed for the RNA-seq experiment focused specifically on the fat body.

3′-end RNA-seq library construction and sequencing

Following RNA extraction, the 3'end RNA-seq libraries were prepared using QuantSeq 3' mRNA-Seq Library Prep kit (Lexogen). The sample quality was evaluated before and after the library preparation using Fragment Analyzer (Advanced Analytical). Libraries were sequenced on two lanes of the Illumina Nextseq 500 platform using standard protocols for 75bp single-end read sequencing at the Cornell Life Sciences Sequencing Core.

Read processing, alignment, counts estimation, and PCA

On average, 6 million reads per sample were sequenced at their 3' termini. This is roughly equivalent in sensitivity to 20x coverage depth under a conventional random-priming RNA-seq method. Raw reads were first evaluated by fastqc for quality control (version 0.11.3) and were then trimmed using Trimmomatic version 0.32 [121]. Trimmed reads were mapped to the *D. melanogaster* reference transcriptome, which was constructed with the *D. melanogaster* reference genome (version 6.80) using STAR RNA-seq aligner (version 2.4.1a) [122]. Read depth at each transcript was then calculated using htseq (version 0.6.1) [123]. Principal Component Analysis and extraction of the PC1/PC2 genes were performed by custom R scripts (available upon request).

Differential expression, functional category and pathway, and transcription factor enrichment

The software edgeR version 3.10.5 was used to call the genes that are differentially expressed among treatments [124]. Nine samples of unchallenged flies matching the 3 different time points post-infection (12 h, 36 h, and 132 h) were collapsed into a single control once it was determined that their transcriptomic profiles were very similar. Library sizes were normalized using

a trimmed mean of M-values (TMM) approach implemented in edgeR. Genes with low counts (count-per-million < 1.2) were filtered out prior to differential expression analysis. Genes were considered statistically differentially expressed if they were differentially expressed between unchallenged condition and an infection condition of choice at the 5% false discovery rate (FDR). A fold-change cutoff was not applied to the data. Each gene was also evaluated for the number of infection conditions in which it was differentially regulated, where an infection condition refers to the transcriptomic profile in response to any of the live infections or controls at any point post-infection. Heatmaps were generated and clustering was performed using custom R scripts. Gene Ontology and KEGG pathway enrichment analysis was performed using the DAVID bioinformatics resource [125] and PANTHER [126]. The p-values from these analyses were corrected using the Benjamini and Hochberg procedure [66] with the FDR threshold set to 0.05. The search for putative transcription factor binding sites was performed using i-cisTarget under the default parameter values [127, 128].

Defining the kinetics of the transcriptional response to infection

For each gene under each infection condition, an expression path was assigned based on the series of inferred induction or repression of infection relative to unchallenged controls at each successive time point. Genes that were significantly induced or repressed at 12 or 36 h but then returned to basal expression levels were deemed to have "recovered". To quantify the degree of recovery for each gene, the level of fold change at 132 h after infection was compared to the fold change in expression at either 12 h or 36 h using custom R scripts (available upon request). The genes that were significantly induced (or

repressed) at 12 h and then significantly repressed (or induced) at 36 h relative to the unchallenged conditions (1% of the genes) were excluded, as were genes that never changed expression in any of the time points.

Fly strains and crosses for *CrebA* experiments

Subsequent to the initial RNA-seq experiment, genetic manipulations of *CrebA* expression were performed. Flies for all of these experiments were reared at 18°C or 24°C. The Rel^{E20} and spz^{rm7} stocks have been previously described [129, 130]. For manipulation of *CrebA* expression level, we used the *UAS/Gal4* gene expression system in combination with $Gal80^{ts}$ to restrict the expression of the constructs specifically to the adult stage. Male flies were collected 5 to 8 days after eclosion from the pupal case and then shifted to 29°C for an additional 8 days prior to any experiments. We used the following genotypes: 1. *c564-Gal4*; *tub-*$Gal80^{ts}$, *UAS-GFP* 2. *Lpp-Gal4*; *tub-*$Gal80^{ts}$, *UAS-GFP* 3. *c564-Gal4*; *tub-*$Gal80^{ts}$, *UAS-CrebA-IR* 4. *UAS-imd* 5. *UAS-spz** 6. *UAS-P35* 7. *UAS-Psn* (Bloomington 8305) 8. *UAS-BiP-IR* (Bloomington 32402) 9. *UAS-BiP* (Bloomington 5843) 10. TRiP control line *attP2* (Bloomington 36303) 11. TRiP control line *attP40* (Bloomington 36304) 12. $Xbp1_p$ > *dsRed* 13-15. *UAS-CrebA-IR* (Bloomington 42562 (A), 31900 (B) and 27648 (C)).

Survival Experiments

Infection was done via septic pinprick to the thorax. After inoculation, death was recorded daily, and flies were transferred to fresh vials every 3 days. All experiments were performed at least 3 times. Statistical significance was determined using a Log-rank (Mantel-Cox) test.

Quantification of bacterial CFUs

At specified time points following infection, flies were individually homogenized by bead beating in 500 l of sterile PBS using a tissue homogenizer (OPS Diagnostics). Dilutions of the homogenate were plated onto LB agar using a WASP II autoplate spiral plater (Microbiology International), incubated overnight at 29°C, and the CFUs were counted. All experiments were performed at least 3 times. Results were analyzed using a two-way (genotype and time) ANOVA in Prism (GraphPad Prism V7.0a, GraphPad Software, La Jolla, CA, USA).

RT-qPCR

For all experiments utilizing RT-qPCR, total RNA was extracted from pools of 20 flies using a standard TRIzol (Invitrogen) extraction. RNA samples were treated with DNase (Promega), and cDNA was generated using murine leukemia virus reverse transcriptase (MLV-RT) (Promega). qPCR was performed using the SSO Advanced SYBR green kit (Bio-Rad) in a Bio-Rad CFX-Connect instrument. Data represent the relative ratio between the Ct value of the target gene and that of the reference gene *RpL32* (also known as *Rp49*). Mean values of at least three biological replicates are represented ±SE. Data were normalized and then analyzed using an unpaired t-test in Prism (GraphPad Prism V7.0a; GraphPad Software, La Jolla, CA, USA). The primer sequences used in this study are available in Table 3.6.

Fat body imaging

In some experiments, fat bodies were visualized microscopically. For these experiments, *Drosophila* abdomens were dissected and fixed in a 4% paraformaldehyde in 1X PBS solution for 45 minutes and washed 3 times with

0.1% Triton-X in PBS. DNA was stained in 1:50,000 DAPI (Sigma-Aldrich) in PBS and 0.1% Triton-X for 45 minutes. Samples were then washed three times in PBS and mounted in antifadent medium (Citifluor AF1). Imaging was performed on a Zeiss LSM 700 fluorescent/confocal inverted microscope.

3.5 Results

Identification of bacteria with different virulence levels and peptidoglycan types

We began by assembling a panel of bacteria to probe the host response to infection. We selected bacteria that span the spectrum of virulence (from 0% to 100% mortality), focusing on microbes that are commonly used by the *D. melanogaster* research community and ensuring that we included bacteria with Lys-type or DAP-type peptidoglycan (PGN) in each virulence level. To assess the relative virulence of each bacterium, we measured host survival and bacterial load over time following infection (Figure 3.1A, Figure 3.2, and Figure 3.3). The bacteria with the lowest levels of virulence – *Escherichia coli* (*Ec*), *Micrococcus luteus* (*Ml*), and the Type strain of *Serratia marcescens* (*Sm*)–caused less than 10% mortality and did not grow past initial inoculum levels in the host. Bacteria exhibiting intermediate levels of virulence–*Pectinobacterium* (previously known as *Erwinia*) *carotovora 15* (*Ecc15*), *Providencia rettgeri* (*Pr*), and *Enterococcus faecalis* (*Ef*)–showed the ability to proliferate within the host and killed 15% to 55% of infected hosts. Highly virulent bacteria–*Staphylococcus aureus* (*Sa*), *Providencia sneebia* (*Ps*), *Serratia marcescens* strain Db11 (Db11), and *Pseudomonas entomophila* (*Pe*)–caused 100% mortality in less than 96 h (Figure

3.1A). *M. luteus*, *E. faecalis*, and *S. aureus* are Gram-positive bacteria (Lys-type PGN); all others are Gram-negative (DAP-type PGN).

Bacterial load time course experiments revealed differences between bacterial species in their ability to grow and persist within the host. For example, only *M. luteus* and *Ecc15* were eliminated from the host (i.e. their levels fall below our detection threshold of 30 CFU/fly) by 132 h post-infection. In the case of *Ecc15*, most but not all hosts were able to clear the infection (Figure 3.3A, F). Neither *E. coli* nor *S. marcescens* Type increased in density, but the bacteria persisted inside the host at 210 bacteria/fly even after 5 days of infection (Figure 3.3D-E). *P. rettgeri* and *E. faecalis* grew during the first 24 h of infection, killing a fraction of the hosts. The flies that survived these infections remained chronically infected with 2^{10} to 2^{13} bacteria per fly (Figure 3.3B, G) for at least 5.5 days. *P. entomophila*, *S. aureus*, *S. marcescens* Db11, and *P. sneebia* all grew monotonically in the host until death occurred (Figure 3.3C, H, I, J), causing complete mortality within 96 h (Figure 3.2E, J, K, L).

Having assembled our panel of bacteria, our next goal was to select relevant time points for transcriptomic analysis. Using our survival and bacterial load data, we identified three time points that are characteristic of different stages of infection: 12, 36, and 132 h. At 12 h post-infection, all flies remain alive, and they face the initial growth of microbes. Thirty-six hours represents an intermediate time point during infection, after the highly virulent bacteria have killed most or all flies and the moderately virulent bacteria have killed 15% to 55% of infected hosts. Finally, at 132 h post-infection (5.5 days), surviving flies are chronically infected with moderate to low levels of bacteria.

A diverse, partly specific *Drosophila* response to infection

To identify novel biological processes required to survive systemic infection, and to assess the level of specificity of the *Drosophila* response to microbes, we used RNA-seq to profile the *D. melanogaster* transcriptome after infection with each of our 10 experimental bacteria. We additionally included the following controls: unchallenged flies (UC), flies challenged with a sterile wound (SW), and flies inoculated with heat-killed *E. faecalis* (*Ef* HK) or heat-killed *P. rettgeri* (*Pr* HK). The purpose of the controls was to distinguish the response to live bacteria from that to aseptic injury and/or inert bacterial compounds (MAMPs) provided by the injection of dead bacteria. The expression value dataset for the entire experiment can be downloaded or accessed online in our associated database Flysick-seq (http://flysick.buchonlab.com)

We first determined the overall transcriptomic differences between flies infected by each of the 10 bacteria. Principal component analysis (PCA) showed that all three biological replicates clustered together, indicating good replicability of the response for each pathogen (illustrated in Figure 3.1B-C for the 12 h time point and Figure 3.4A for the full data set). In total, we identified 2,423 genes (13.7% of the genome) that were differentially expressed upon infection. Of these, 1,286 genes were upregulated and 1,290 genes were downregulated in response to at least one bacterial infection and time point (Figure 3.5 and Figure 3.4B). Out of the total number of genes differentially regulated by all 10 live infections, more genes were upregulated than downregulated; 6.1% of the 1,286 upregulated genes were induced in all bacterial infections, while only 0.6% of the 1,290 downregulated genes were repressed by all 10 bacteria (Figure 3.4B). We also determined that 51.1% of the downregulated genes were repressed in only one bacterial infection, while 38.6% of the upregulated genes were induced by a single bacterial condition

(Figure 3.4B). These data suggest that the host response to infection is highly specific to individual bacteria, but that there is also a core set of genes that are differentially expressed during most bacterial infections. Additionally, our data showed that downregulated genes tend to be unique to each infecting bacterium, perhaps reflecting the singular consequences of each infection to host physiology (Figure 3.4B).

In general, the largest number of differentially expressed genes was observed at 12 h post-infection. However, a substantial number of genes continued to be differentially regulated at 36 h and 132 h post-inoculation (Figure 3.5), presumably in part because the hosts continue to carry their bacterial infections at these later time points and/or because infection induces long-term changes in host physiology. Samples for the 36 h and 132 h time points were not available for infections with the highly virulent bacteria because they rapidly killed all their hosts. For the remaining infections, however, the number of upregulated genes at 12 h after infection was 1.6 times higher than the average number of genes that continued to be induced at 36 h and 132 h post-infection. Likewise, there were 2.8 times as many downregulated genes at 12 h post-infection than there were at later time points. These results demonstrate that the early transcriptional response to infection is larger than the sustained one, probably because the early response includes both an injury-induced transcriptional regulation and an aggressive initial immune response that is not yet tuned to bacterial titer or growth state within the host [131].

Major axes of variation in the transcriptional response to infection

We sought to investigate the source of differences in the host response to

various infections. We began by looking at the number of genes regulated by the host in response to each bacterium. The number of differentially regulated genes fluctuated considerably across bacterial infections (Figure 3.5). Flies inoculated with heat-killed *E. faecalis* and *P. rettgeri*, as well as flies challenged with avirulent bacteria, such as *E. coli* and *M. luteus*, induced the lowest number of genes. However, the number of genes regulated in the host did not directly correlate with the level of bacterial virulence. For example, despite the fact that both bacteria rapidly killed all flies, infection with *S. aureus* differentially regulated the expression of 1,193 genes, while *P. sneebia* infection altered the transcription of only 187 genes (Figure 3.5). In addition, there was a large variability in the number of genes regulated in response to different benign bacteria. Across all time points, *M. luteus* infection changed the expression of 794 genes, while *E. coli* infection affected only 446 genes (Figure 3.5). These results indicate that the breadth and the specificity of the host transcriptomic response is largely independent of virulence.

Next, we aimed to identify specific genes that underlie the transcriptomic differences in response to distinct infections. We focused on the first two principal components of our PCA analysis (Figure 3.1B), which respectively explain 34.0% and 27.2% of the variance in gene expression. We found that 73 of the top 100 genes contributing to the first principal component (PC1) and 75 of the 100 genes contributing most to the second principal component (PC2) are known targets of the Toll or Imd pathways (Figure 3.1C and Figure 3.6), confirming that these two pathways are key regulators of the specificity of the host response [112]. The genes that contributed most to PC1 included antimicrobial peptide genes (*Dpt*, *AttA*, *Drs*, and *Mtk*) as well as signaling components of the Toll (*Spz* and *PGRP-SA*) and Imd (*PGRP-LC*, *PGRP-SD*,

PGRP-LB, and *Rel*) pathways themselves (Fig 1C). Additionally, the expression of Turandot genes, stress peptides regulated by the JAK-STAT pathway, was strongly variable between infections, indicating that differential activation of the JAK/STAT pathway also contributes to PC1. Interestingly, metabolic genes involved in lipid synthesis (*ACC*), the Leloir pathway (*Galk*), and trehalose and glycogen synthesis (*Tps1*, *UGP*, and *Hex-C*) were downregulated to different levels depending on the infection, indicating that different bacteria alter host metabolism in unique ways. In general, PC1 appeared to reflect the transcriptional magnitude of the response to infection. Genes that contributed most to PC2 include target genes of the Toll pathway, including melanization and coagulation-related genes (*MP1* and *fondue*) (Figure 3.1C), as well as immune-induced proteins of the IM cluster. PC2 also included genes downregulated by infection that are involved in sugar digestion (i.e. the *Maltase* cluster), as well as P450 enzymes known for their functions in oxidoreduction reactions (i.e. *Cyp* genes). Flies infected with Gram-positive bacteria (Lys-type PGN) and Gram-negative bacteria (DAP-type PGN) were separated from each other on PC2, confirming that the type of bacterial peptidoglycan is a major parameter influencing the global response to infection (Figure 3.1B, Figure 3.4A) [112]. A heatmap showing the expression level of genes that contribute the most to each PC can be found in Figure 3.6.

Subsequently, we asked whether any differentially regulated genes were unique to a specific bacterial condition. We defined unique genes as those that significantly changed their expression in one and only one infection condition, regardless of time points, thus reflecting the response to a particular bacterium rather than temporal variations in the response to this bacterium. Without exception, we found that infection with each bacterium regulates an exclusive

set of genes. The number of uniquely regulated genes varied dramatically across bacterial infections (Figure 3.5). For instance, *P. sneebia* infection resulted in unique regulation of only 6 genes, whereas *S. aureus* infection exclusively regulated 336 genes. In order to determine what portion of the host response is specific to individual bacteria, we calculated the percentage of differentially expressed genes that were unique to each infection (Figure 3.4C). We found that this number also differs widely between bacteria. For instance, 20.1% of genes upregulated in response to *S. aureus* were exclusive to this infection, while only 7.1% of genes upregulated by *E. faecalis* infection were unique to this condition. Evaluating Gene Ontology (GO) terms associated with the genes uniquely altered by individual infections revealed bacteria-specific responses in some infection conditions (Figure 3.5). For example, *S. aureus* infection induced apoptosis-related genes and downregulated genes involved in glutathione and carboxylic acid metabolism. In contrast, infection with *P. entomophila* upregulated genes involved in epithelial cell proliferation and strongly decreased the expression of genes associated with cellular respiration and the electron transport chain. At the same time, infection by *P. rettgeri* specifically downregulated genes involved in the translation machinery (Figure 3.5). All the GO gene categories we identified are linked to stress responses that aim to maintain cell homeostasis (cell death and tissue repair) or metabolic homeostasis, suggesting that the unique physiological and virulence interactions of each bacterium with the host induce a specific set of organismal responses. Altogether, our results demonstrate that the host response to infection is shaped by a combination of immune potency, metabolic impact, and physiological alteration.

Identification of a core host response to infection

Next, we set out to identify the core set of genes that are regulated in response to most or all bacterial infections. We defined the core genes as those that are differentially expressed in response to 7 or more bacteria on at least one time point post-infection. We set the cutoff at 7 bacteria because we were concerned that requiring differential expression in response to all 10 infections would be overly restrictive. Specifically, we had reservations about the artificial omission of genes in cases where the bacteria are rapidly cleared from all or most hosts (e.g. *M. luteus* and *Ecc15*) and in cases where the bacterium might suppress or evade the canonical response (e.g. *P. sneebia*; [132]). Using these criteria, we identified a core response of 252 genes. This included 166 upregulated genes (Figure 3.7A and Table 3.1) and 86 downregulated genes (Figure 3.7B and Table 3.1). The set of core genes is fairly robust to the criteria for inclusion, decreasing only to 135 genes induced and 54 genes repressed when inclusion required differential expression in response to 8 of the bacterial conditions. Similarly, the numbers increased only to 216 genes induced and 136 repressed when inclusion was relaxed to 6 of the bacterial infections.

Within the core, 78 genes were also regulated in response to sterile wound alone or to challenge with heat-killed bacteria (Figure 3.7C). Most of the genes regulated by injury were also regulated by challenge with live or dead bacteria (96/114 genes), which is congruent with the fact that the infection method inherently inflicts injury. However, the core response to live infection was markedly distinct from the response to heat-killed bacteria. Of our core genes, 40% (105/252 genes) were differentially expressed in response to live infections but not in response to challenge with heat-killed bacteria. Moreover, we found 493 genes that were differentially regulated by treatment with heat-killed bacteria but were not part of the core response to live infection (Figure 3.7C).

Of those 493 genes, 164 were uniquely regulated in response to heat-killed bacteria and not in response to any live infection (Figure 3.8A). To determine whether genes exclusively regulated in response to heat-killed bacteria are simply artifacts of weak statistical detection, we relaxed the cutoff to a False Discovery Rate (FDR) <0.1 for classifying a gene as differentially expressed during infection. Even with this more lenient threshold, 61.6% of the 164 genes that were uniquely regulated in response to heat-killed bacteria were still not differentially regulated in response to any live infection. Our results, therefore, not only show that the response to live infections is fundamentally different from the biological challenges that simple injury and immune activation pose, but also demonstrate that challenge with dead bacteria induces a response that does not occur as a consequence of infection by live bacteria.

In 2001, a study identified a set of genes that are differentially expressed after infection with a combination of *E. coli* and *M. luteus* [11]. These genes became known as *Drosophila* Immune-Regulated Genes (DIRGs). We compared our set of 252 core response genes to the 381 DIRGs and found that only 84 of them were previously identified as DIRGs (Figure 3.8B). Intriguingly, the DIRGs identified in the previous study included 279 genes that were neither in our core response nor regulated by challenge with heat-killed bacteria (Figure 3.8B), and 246 of these DIRGs were not induced in the present study even by infection with *M. luteus* or *E. coli*(Figure 3.8C). These discrepancies may originate from differences in *Drosophila* genotype or rearing conditions, bacterial genotype, or experimental variation. Alternatively, they could imply that infection with a mixture of two bacteria can lead to the activation of a specific set of genes, different from each mono-microbial infection. When we compared our total number of differentially regulated genes (2,423) to the DIRGs, we found

that our study has identified 2,197 novel infection response genes, including 168 new core genes. Thus, our data offer a more comprehensive list of infection-responsive genes that is expanded both because of the sensitivity of RNA-seq technology over the previous microarrays and because of the broader diversity of bacteria used in our experiment.

To investigate the biological functions of our newly identified core response genes, we evaluated GO categories enriched in the core (Figure 3.7A-B). Upregulated core genes were primarily annotated with immune functions, such as Toll pathway and defense response to Gram-negative bacteria. This group also included genes involved in metabolism, including glycosaminoglycan metabolic process, carbohydrate metabolism, and metal ion transport. Additionally, core upregulated genes have a role in cellular and tissue processes, with genes acting in tissue repair, response to oxidative stress, cellular homeostasis, co-translational protein targeting to membrane, and protein targeting to ER (Figure 3.7A). The core downregulated genes were annotated with functions such as oxidation-reduction and starch and sucrose metabolism (Figure 3.7B). Core genes can be separated into two groups: genes regulated in response to live infections only and genes regulated in response to both live infections and heat-killed bacteria (Figure 3.7C). The 78 core genes that were also differentially expressed in the wound-only control and in the heat-killed bacteria control included genes coding for AMPs, PGRPs, Turandot (*Tot*) genes, and other classical targets of the Toll and Imd pathways [133]. Genes regulated only in response to live infection included key transcription factors of the immune system, such as *Rel* and *dl*, and were associated with biological processes such as metabolism, oxidation-reduction, regulation of iron ion transmembrane transport, and secretion. Altogether, these data indicate that

heat-killed bacteria mostly trigger classically defined immune responses, while live infections regulate a set of additional biological processes that presumably reflect physiological interactions between the host and invading pathogen. These processes, including metabolic rewiring, response to stress and damage, cellular translation, and secretion, could act as physiological adaptations or buffers to the stress and damage imposed by infection.

The hypothesis that *D. melanogaster* has a distinct response to infection by Gram-positive (Lys-type PGN) versus Gram-negative (DAP-type PGN) bacteria dominated the field for most of the 1990s and 2000s [1]. To address this hypothesis, we characterized the transcriptional response to Gram-positive versus Gram-negative bacterial infection in our study. We found that 662 genes are regulated only by infection with Gram-positive bacteria, 851 genes are regulated only by Gram-negative infection, and 1,063 genes are regulated by infections with bacteria of both Gram types (Figure 3.7D). Of the 662 genes exclusively regulated by Gram-positive bacteria, only 20 (*Cyp309a1*, *daw*, *CG31326*, etc.) are upregulated and 8 are downregulated by all three Gram-positive bacteria. Similarly, amongst genes regulated specifically by Gram-negative bacteria, only 1 gene is upregulated (*AttD*) and no genes are downregulated in response to all 7 Gram-negative bacteria. Our data suggest that the stereotypical response to Gram-negative infection also occurs as a consequence of Gram-positive infection, such that there is no large cohort of genes responding exclusively to Gram-negative infection. To confirm this, we performed RT-qPCR on *Dpt* and *Drs* transcripts as a proxy for activity of the Imd and Toll pathways, respectively [11]. We found that infection by most of our 10 bacteria induced both pathways, although to significantly different levels (Figure 3.9). Our results generally confirm the notion that the Toll pathway is

more responsive to infection with Gram-positive (Lys-type PGN) bacteria and the Imd pathway is more reactive to infection with Gram-negative (DAP-type PGN) bacteria, but also make clear that the differences in pathway activation are quantitative and not qualitative or binary.

Infection induces long-term changes in global host transcription

Since the bacteria belonging to the low and intermediate virulence categories do not kill all hosts, we followed the dynamics of gene expression in surviving hosts over several days. In particular, we aimed to contrast the sustained transcriptional response of flies that had cleared their infections to undetectable levels (i.e. after infection with *M. luteus* or *Ecc15*) to that of flies carrying chronic infections (i.e. *E. coli*, *S. marcescens* Type strain, *P. rettgeri* and *E. faecalis*). We hypothesized that persistent bacteria would continue to elicit a response from the host, which would be absent in flies that have cleared all bacteria. To test this idea, we determined whether genes "recover" from bacterial infection. We defined recovery in terms of gene expression: a gene that has recovered is differentially expressed at 12 and/or 36 h post-infection but returns to pre-infection levels by 132 h (5.5 days) after inoculation. We found that, on average, a minimum of 50% of the genes that are differentially regulated by each infection returned to basal levels by our last time point (Figure 3.10A), and this was the case even in hosts infected with persistent infections. The percentage of genes that fully recovered was substantially higher in moderately virulent infections (*P. rettgeri*: 79.8% and *E. faecalis*: 71.1%) than in benign infections (*E. coli*: 54.5% and *S. marcescens* Type: 55.5%), perhaps in part as a consequence of the higher number of genes induced upon infection with these bacteria (Figure 3.10A). Surprisingly, we also observed that only 56.5% and

58.7% of genes recovered in *M. luteus* and *Ecc15* infections, respectively, even though the majority of hosts (≥85%) survive these infections and the bacteria are eliminated (i.e. their levels fall below our detection threshold) within two days. These results demonstrate the complexities of the transcriptional response to infection. While there can be a substantial lingering transcriptional effect in flies that successfully cleared an infection, a subset of differentially regulated genes may return to basal levels even in chronically infected flies that continue to carry bacteria.

Next, we evaluated how the core upregulated and downregulated genes change in expression level over time (Figure 3.10B-C). We quantified the degree of recovery for each gene by comparing the fold change in expression at 132 h after infection to the fold change at either 12 h or 36 h, whichever was the highest if the gene was upregulated or the lowest if the gene was downregulated. In general, sterile wound and challenge with heat-killed bacteria resulted in the regulation of fewer core genes than live infection, and most of these genes recovered to pre-infection expression levels by 132 h post-challenge (Figure 3.10B-C). Core genes induced by *Ecc15* and *M. luteus* showed similar kinetics, and most genes had recovered or were on their way to recovery by 132 h, suggesting that the core response is not sustained in the absence of these bacteria. In contrast, core genes induced by *S. marcescens* Type, *P. rettgeri*, and *E. faecalis* did not recover as much, in agreement with the idea that infections with persistent bacteria continuously stimulate the core response. This paradigm was, however, not true for downregulated genes, as most downregulated genes did recover or were in the process of recovery by 132 h regardless of which bacteria was used for infection. Interestingly, we noticed that a group of genes did not recover at all in most conditions but continued to be upregulated over

time (boxed in Figure 3.10B). These included effector genes of the immune response (*AttD* and *Tep4*), regulators of iron homeostasis (*Tsf1* and *MtnD*), and negative regulators of the immune response (*PGRP-LB*, *nec*). Genes like *SPH93* and *su(r)* never returned to their basal expression levels in flies infected with *Ecc15* or *M. luteus*. Additionally, while the transcript levels of most antimicrobial peptide genes decreased over time, they never returned to basal, pre-infection levels, suggesting that the effect of infection lingers for several days after bacteria are eliminated.

CrebA is regulated in the fat body upon infection by the Toll and Imd pathways

Having identified a core transcriptional response to infection, we set out to find key regulators of that response. We used i-cisTarget to identify transcription factor binding motifs enriched in the regulatory regions of our core genes [127, 128]. Using this approach, we found enrichment in putative binding sites for Relish (Rel), Dif/Dorsal, Schnurri (Shn), CrebA, Atf6, Xbp1, and Tbp in the regulatory regions of upregulated core genes (Figure 3.11A and Table 3.2). Dif/Dorsal and Relish are the terminal transcription factors of the Toll and Imd pathways, respectively; therefore, finding enrichment for their predicted binding sites is in agreement with the central role that these pathways play in the immune response. Our data also agree with published reports showing that the TGF-beta pathway upstream of *shn* and the *Atf6* transcription factor are important to survive infection [134, 135]. Transcription factor binding site enrichment analysis of the repressed genes revealed putative binding sites for the Lola and GATA transcription factors (Table 3.3).

In addition to the i-cisTarget analysis, we searched for genes encoding

transcription factors within our list of core upregulated genes. We identified 3 transcription factors in the defined core that are upregulated themselves by infection: *Rel*, *dorsal*, and *CrebA* (Figure 3.11B). Although *Dif* is required to activate Toll pathway signaling in response to bacterial infection in *Drosophila* adults and *dorsal* is not [136], we surprisingly found that *Dif* is not significantly upregulated in response to any of the 10 bacteria tested. *CrebA* is the single *Drosophila* member of the Creb3-like family of transcription factors [137]. We found the predicted DNA motif bound by CrebA (TGCCACGT, see Figure 3.11A for position weight matrix [138]) in 71 genes upregulated by infection, including 18 upregulated core genes (Figure 3.12A). *CrebA* is itself significantly induced upon infection by all 10 bacteria (Figure 3.7A and Figure 3.11B). To validate our RNA-seq results on *CrebA* expression, we infected a new group of flies with *P. rettgeri* and *E. faecalis* and measured *CrebA* transcript levels at 12 h post-inoculation. In agreement with our RNA-seq data, we confirmed that *CrebA* expression is upregulated in response to infection with *P. rettgeri* (p=0.0026) and *E. faecalis* (p=0.0147) (Figure 3.12B). These results demonstrate that *CrebA* is a transcription factor induced by infection and is potentially a key regulator of the core response.

To identify the molecular mechanisms that control *CrebA* transcription in response to infection, we scanned 2 kb upstream and 2 kb downstream of the *CrebA* transcription start site for potential transcription factor binding sites using MatInspector (Genomatix) [139]. Within this region, we found an enrichment of putative binding sites corresponding to the transcription factors Dif/Dorsal and Relish. There were 12 predicted Relish binding sites, 16 predicted Dif binding sites, and 13 predicted Dorsal binding sites flanking the *CrebA* gene, suggesting that immune pathways may induce the expression

of *CrebA* (Figure 3.12C). To confirm regulation of *CrebA* by the Toll and Imd pathways, we quantified *CrebA* expression by RT-qPCR 12 h after infection with *P. rettgeri* and *E. faecalis* in wildtype (WT) flies, flies deficient for the Imd pathway (Rel^{E20}), and flies deficient for the Toll pathway (spz^{rm7}) (Figure 3.11C). *CrebA* expression was significantly reduced in both Rel^{E20} (p=0.0456 for *P. rettgeri* and p=0.0020 for *E. faecalis*) and spz^{rm7} (p=0.0118 for *P. rettgeri* and p=0.0026 for *E. faecalis*) mutants relative to wildtype controls, indicating that both the Imd and Toll pathways contribute to infection-induced *CrebA* upregulation. We then tested whether activation of the Imd or Toll pathway is sufficient to upregulate the level of *CrebA* expression in the absence of infection. Using the temperature-sensitive *UAS/Gal4/Gal80^{ts}* gene expression system to ubiquitously drive Imd or an active form of Spz (Spz*), we stimulated Imd and Toll pathway activity in adult flies [140, 141]. Transgenic activation of either the Imd or Toll pathway in the absence of infection was sufficient to significantly increase *CrebA* transcript levels in *D. melanogaster* adults (p=0.0114 for UAS-spz* and p=0.0062 for UAS-imd) (Figure 3.11C). Altogether, our results demonstrate that the Imd and Toll pathways are both necessary and sufficient to regulate *CrebA* transcription upon infection.

In order to identify the tissue(s) and/or organ(s) within the fly that upregulate *CrebA* expression upon bacterial challenge, we infected wildtype flies with *P. rettgeri* and dissected out the following tissues and body parts at 12 h post-infection: head, dorsal thorax (including wings and heart), ventral thorax (including legs), digestive tract (crop, midgut, and hindgut), Malpighian tubules, testes, and abdomen (abdominal fat body). The abdomen was the only tissue that exhibited significant upregulation of *CrebA* as determined by RT-qPCR (p=0.0315), suggesting that *CrebA* may be regulated in the fat body

upon infection (Fig 5D). We therefore knocked down *CrebA* expression by RNAi (via 3 independent RNAi constructs) using 2 separate fat body drivers, *c564-Gal4* and *Lpp-Gal4*, and quantified *CrebA* expression by RT-qPCR in whole flies 12 h after infection with *P. rettgeri*. The combination of 2 *CrebA* RNAi constructs (B and C) with the drivers fully prevented *CrebA* induction upon infection with *P. rettgeri*. In the case of the third RNAi construct (A), *CrebA* was significantly upregulated by infection with *P. rettgeri* (p=0.0002), but the induction was significantly lower (p=0.0442) than the expression level observed in infected wildtype samples (Figure 3.12D). These data indicate that the cells of the fat body represent the primary site of *CrebA* induction. In sum, our data suggest that the Toll and Imd pathways regulate the expression of *CrebA* in the fat body in response to infection.

***CrebA* is required to survive infection and promotes tolerance of infection**

We next asked whether *CrebA* is required for the host to survive infection. Since strong loss-of-function *CrebA* mutants are embryonic lethal, we tested the role of *CrebA* in response to infection by knocking it down in the fat body of adult flies using 3 independent RNAi constructs expressed under the control of the *c564-Gal4* driver (*Gal80^{ts}*; *c564-Gal4 > UAS-CrebA-IR*) and, separately, the *Lpp-Gal4* driver (*Gal80^{ts}*; *Lpp-Gal4 > UAS-CrebA-IR*) [142]. Because the *c564-Gal4* driver expresses strongly in both the fat body and hemocytes, we additionally tested the requirement for *CrebA* in the response to infection in hemocytes (*Hml-Gal4 > UAS-CrebA-IR*). All *CrebA* fat body knockdown flies exhibited increased susceptibility to systemic infection with *P. rettgeri* (p<0.0001) (Figure 3.13A and Fig S8A), while hemocyte-specific knockdown did not lead to any significant increase in mortality (Figure 3.14B). When *CrebA* was knocked down

in the fat body, nearly 100% of the flies died, and most of the death occurred during the first 24 h following infection. In contrast, almost 50% of control flies survived the infection for at least 7 days (Figure 3.13A and Figure 3.14A). To confirm that the survival phenotype observed in *CrebA* RNAi flies upon infection was solely due to loss of *CrebA* expression, we co-expressed a *CrebA* RNAi construct and a *CrebA* overexpression construct in flies (*Gal80^{ts}*; *c564-Gal4 > UAS-CrebA*, *UAS-CrebA-IR*) and infected them with *P. rettgeri*. We observed no significant difference between the survival of infected control flies and that of infected flies co-expressing both the RNAi and overexpression constructs, indicating that changes in *CrebA* expression are uniquely responsible for the lowered survival phenotype observed (Figure 3.13A). We also infected *CrebA* RNAi flies with *E. faecalis* and found that *CrebA* RNAi flies were remarkably more susceptible to infection when compared to control flies ($p<0.0001$) (Figure 3.13B). In addition, *CrebA* RNAi flies died at a significantly faster rate than control flies when inoculated with *P. sneebia* ($p<0.0001$) (Figure 3.13C). Finally, infection with *Ecc15*, *S. marcescens* Type, and *E. coli* also killed more flies with *CrebA* expression blocked in the fat body than controls ($p=0.0013$ for *Ecc15*, $p=0.0004$ for *S. marcescens* Type, and $p=0.0028$ for *E. coli*) (Figure 3.13D-F). None of these latter three infections were lethal to wildtype control flies, but approximately 30% of *CrebA*-deficient flies succumbed to infection. Collectively, our results demonstrate that *CrebA* is generally required to survive bacterial infection.

To test whether the *CrebA* survival phenotype is due to a failure to control bacterial proliferation (a resistance defect) or a decrease in the ability to withstand infection (a tolerance defect), we monitored bacterial load in individual *CrebA* RNAi and control flies following *P. rettgeri* infection [143]. We

focused our sampling on 1-2 h intervals over the first 24 h of infection, as this is the time when most of the *CrebA*-deficient flies succumbed. We did not find a significant difference in bacterial load between wildtype and *CrebA* knockdown flies at any measured time point (p=0.0664), indicating that *CrebA* RNAi flies are able to control bacterial load similarly to control flies (Figure 3.13G). To corroborate these results, we quantified bacterial load following infection with *P. rettgeri* in flies where *CrebA* was knocked down by a different RNAi construct and in flies co-expressing a *CrebA* RNAi construct and a *CrebA* overexpression construct (*Gal80^{ts}*; *c564-Gal4 > UAS-CrebA*, *UAS-CrebA-IR*). Again, we did not observe any significant difference in bacterial load between wildtype and *CrebA* knockdown flies (p=0.3208) or between wildtype and *CrebA* rescue flies (p=0.3030) (Figure 3.14C). To evaluate whether *CrebA* knockdown flies are less resistant to other pathogens, we measured bacterial load in individual flies following *E. faecalis* or *Ecc15* infection. In agreement with the results of our *P. rettgeri* experiments, we did not find a significant difference between wildtype and *CrebA*-deficient flies at the time points sampled (p=0.4204 for *E. faecalis* and p=0.7253 for *Ecc15*) (Figure 3.14D-E), suggesting that *CrebA* knockdown flies do not have a defect in resistance to infection.

We previously demonstrated that flies die at a stereotypical and narrowly distributed bacterial load, the bacterial load upon death (BLUD), which represents the maximum quantity of bacteria that a fly can sustain while alive [144]. We therefore sought to determine whether *CrebA* RNAi flies have a lower BLUD, which would indicate a reduced tolerance of infection. We quantified the bacterial load of individual flies within 15 minutes of their death and found that *CrebA* RNAi flies died carrying a significantly lower bacterial load than control flies (p<0.0001) (Figure 3.13H). These data demonstrate that while

CrebA-deficient flies control bacterial growth normally, they are more likely to die from infection, and they die at a lower bacterial load than wildtype flies. Therefore, the transcription factor *CrebA* acts to promote tolerance of infection.

Loss of CrebA alters the expression of secretory genes during infection

In order to identify the complete set of genes directly and indirectly regulated by *CrebA* upon infection, we performed RNA-seq on the fat bodies of wildtype flies and flies in which we knocked down *CrebA* in the fat body. We collected samples from both genotypes in unchallenged conditions and 12 h after infection with *P. rettgeri*. In total, we found that only 104 genes were downregulated in *CrebA* knockdown fat bodies compared to wildtype fat bodies following infection (Table 3.5). These genes were associated with GO categories such as protein targeting to the ER, signal peptide processing, protein localization to the ER, and antibacterial humoral responses. Antimicrobial peptide genes of the *Cecropin* gene family (*CecA1*, *CecA2*, *CecB*, and *CecC*) showed partially reduced induction when *CrebA* expression was disrupted. Nevertheless, they were still induced to extremely high levels (>200-fold) in *CrebA* knockdown fat bodies (Fig 7). Other antimicrobial peptide genes, such as *Dpt*, *Drs*, *Def*, and *AttC*, were expressed at similar levels in *CrebA* knockdown fat bodies compared to wildtype fat bodies, results corroborated by RT-qPCR analysis (Figure 3.16A-D). In contrast, a number of genes including sugar transporters and multiple lipases were upregulated upon infection in fat bodies deficient for *CrebA* but not in wildtype fat bodies. These data suggest that *CrebA* regulates immune, metabolic, and cellular functions during infection.

Previously, Fox and colleagues demonstrated that *CrebA* acts in the *Drosophila* embryo as a direct regulator of secretory capacity and is both

necessary and sufficient to activate the expression of many secretory pathway component genes [138]. We therefore asked whether *CrebA* controls secretion-related genes upon infection in the adult fat body. We found that the expression level of 32 secretion-related genes significantly increased upon infection with *P. rettgeri* in wildtype samples. However, the induction of these secretion-related genes was significantly lower ($p<0.05$) in *CrebA* RNAi fat body samples compared to wildtype fat body controls, a result that agrees with the findings of [138] (Figure 3.15 and Table 3.5). These 32 secretion-related genes we identified included core response genes that are central components of the cell's secretory machinery, including *TRAM*, *ergic53*, *Sec61β*, and *Spase25* (Figure 3.15). Using a separate set of samples from those of the RNA-seq, we further confirmed these findings by measuring *TRAM*, *ergic53*, *Sec61β*, and *Spase25* transcript levels by RT-qPCR in the fat bodies of flies infected with *P. rettgeri* at 12 h post-infection (Figure 3.16E-H). These four genes were significantly upregulated following infection with *P. rettgeri* in wildtype samples. However, we were not able to detect a significant increase in the levels of *TRAM*, *ergic53*, and *Sec61β* in *CrebA* RNAi fat bodies upon infection. The expression level of *Spase25* was significantly induced by infection with *P. rettgeri* even when *CrebA* expression was inhibited by RNAi in the fat body ($p<0.05$), but the induction was significantly lower ($p<0.001$) than the expression level observed in infected wildtype samples (Figure 3.16H). In sum, our data suggest that *CrebA* could act to regulate an increase in secretory capacity upon infection.

CrebA deficiency leads to ER stress upon bacterial challenge

Since our data suggested that *CrebA* may promote an increase in secretory capacity in the fat body upon infection, we hypothesized that loss of *CrebA*

expression could lead to altered protein secretion or defects in protein transport to the membrane. Accumulation of unfolded proteins or a decrease in protein secretion triggers endoplasmic reticulum (ER) stress, which in turn induces stereotypical pathways to limit the stress imposed on the cell. These pathways include IRE1α/XBP1-, PERK/ATF4-, and ATF6-mediated responses termed the unfolded protein response (UPR) [145, 146]. Upon sensing of ER stress, *Xbp1* mRNA undergoes alternative splicing via IRE1α; *Xbp1* splicing is thus considered to be a marker of ER stress and of the activation of UPR [147, 148]. To investigate whether loss of *CrebA* could trigger ER stress in fat body cells upon infection, we quantified the expression levels of both *Xbp1t* (total) and *Xbp1s* (spliced) in abdomens of wildtype and *CrebA*-knockdown flies under both unchallenged and infected conditions (Figure 3.17A-B). *Xbp1s* levels did not change upon infection in wildtype samples or differ between wildtype and *CrebA* RNAi samples in the absence of infection. However, *Xbp1s* levels spiked dramatically in *CrebA* RNAi fat body samples (p=0.0289) (Figure 3.17B) after infection, indicating that loss of *CrebA* upon bacterial challenge triggers ER stress in the fat body. Our data also revealed that *Xbp1t* expression was significantly higher in *CrebA* knockdown samples compared to wildtype samples following infection (p=0.0144) (Figure 3.17A). This result is in agreement with a previous study that suggested *Xbp1*s regulates *Xbp1* transcription [149]. To determine whether ER stress is induced in fat body cells directly, we labelled fat body cells in vivo by expressing a *dsRed* reporter under the control of the *Xbp1* regulatory sequence [150]. In agreement with our RT-qPCR experiments, we found that bacterial challenge did not induce dsRed expression in wildtype samples. However, infected *CrebA* RNAi fat body cells consistently expressed higher levels of *dsRed* compared to all other controls

(Figure 3.17C). These results demonstrate that *CrebA* expression prevents the occurrence of ER stress in the fat body upon infection.

We next asked whether the failure of *CrebA*-deficient flies to prevent ER stress following infection could explain their increased susceptibility to bacterial challenge. To test this, we genetically induced ER stress in *Drosophila* fat bodies either by overexpression of *Psn* (*Gal80^{ts}*; *c564-Gal4 > UAS-Psn*), which disrupts calcium homeostasis, or by knockdown of *BiP* (*Gal80^{ts}*; *c564-Gal4 > UAS-BiP-IR*), a regulatory protein of the unfolded protein response [151, 152]. Inducing ER stress in the fat body during infection made the flies more susceptible to *P. rettgeri* infection, phenocopying the result observed with *CrebA* knockdown flies ($p<0.0001$ for both constructs) (Figure 3.17D). Since the increased susceptibility of *CrebA* RNAi flies to infection stemmed from a tolerance defect (Figure 3.13G-H and Figure 3.14C-E), we sought to determine whether the increase in mortality observed in *BiP* RNAi and *Psn* overexpression flies following infection is also due to a tolerance deficiency. We monitored bacterial load in individual *BiP* RNAi and Psn overexpression flies following challenge with *P. rettgeri*. We did not observe a significant difference in bacterial load between wildtype and *BiP*-knockdown flies ($p=0.0624$) or between wildtype and *Psn* overexpression flies ($p=0.6462$) (Figure 3.18A). Quantification of bacterial load upon death (BLUD) following *P. rettgeri* infection in *BiP* RNAi flies showed that *BiP*-deficient flies perish carrying a significantly lower bacterial load than wildtype flies ($p<0.0001$) (Figure 3.18B). Altogether, our data indicate that induction of fat body ER stress during infection decreases fly survival by lowering host tolerance of infection.

Having demonstrated that *CrebA*-deficient flies experience fat body ER

stress upon bacterial challenge and that flies with genetically induced fat body ER stress display increased mortality without a concomitant change in bacterial load following infection, thus phenocopying *CrebA*-deficient flies, we subsequently asked whether alleviating ER stress in *CrebA*-deficient flies could rescue the *CrebA* survival phenotype. To test this, we overexpressed *BiP* in fat body cells in which *CrebA* was knocked down by RNAi (*Gal80^{ts}*; *c564-Gal4 > UAS-CrebA-IR*, *UAS-BiP*). Previous work has shown that overexpression of *BiP* can ameliorate ER stress [153]. While overexpression of *BiP* alone did not alter host survival during infection, expression of *BiP* in *CrebA* RNAi flies rescued fly survival upon challenge with *P. rettgeri* (Figure 3.17E). We observed no significant difference between the survival of infected control flies and that of infected flies co-expressing both the *CrebA* RNAi and *BiP* overexpression constructs (p=0.2786). These data indicate that reducing ER stress is sufficient to rescue the survival phenotype of *CrebA*-deficient flies during bacterial challenge. Excessive and prolonged ER stress can lead to apoptosis [154]. Therefore, we investigated whether *CrebA* RNAi flies are more susceptible to infection due to an increase in fat body cell apoptosis. We blocked apoptosis by overexpressing the apoptosis inhibitor P35 in the fat body of *CrebA*-knockdown flies (*Gal80^{ts}*; *c564-Gal4 > UAS-CrebA-IR*, *UAS-P35*) [155]. Expression of P35 in *CrebA* RNAi flies did not rescue the *CrebA* survival phenotype upon infection (Figure 3.18C), indicating that an increase in apoptosis is unlikely to explain the *CrebA* susceptibility defect. Collectively, our results show that *CrebA* is required in the fat body to prevent excessive and deleterious levels of ER stress upon infection.

3.6 Discussion

Bacteria trigger diverse and unique host responses

In this study, we have characterized the transcriptomic response of *Drosophila* to a wide range of bacterial infections. We found that the response to infection can involve up to 2,423 genes, or 13.7% of the genome. This is a considerably greater number of genes than what has been previously reported in similar transcriptomic studies [11, 12, 112]. As the response to infection was highly specific to each bacterium, the larger number of genes we identified is likely a consequence of having included more bacterial species in our experiment than previous studies. Likewise, we anticipate that future studies using different species of bacteria could further increase the number of genes found to be involved in the host response to infection. Our data clearly establish that while the core response to infection is narrow and conserved, every bacterium additionally triggers a very specific transcriptional response that reflects its unique interaction with host physiology.

At first, this high level of specificity may seem contrary to the traditional vision of the innate immune response. Early studies defined the innate immune system as generic, and the specificity of the *Drosophila* immune response was considered as a dichotomous activation of the Toll pathway by Gram-positive bacteria (Lys-type peptidoglycan) or the Imd pathway by Gram-negative bacteria (DAP-type peptidoglycan) [1, 111]. Our data show that the host response to infection goes beyond the activation of the Toll and Imd pathways, with each bacterium also modulating host cell biology, metabolism, and stress responses in a microbe-specific manner. Although we did find that the type

of bacterial peptidoglycan is a key factor shaping the response, we also found that each bacterium activates both the Toll and Imd pathways to quantitatively different levels, consistent with previous reports suggesting a much more complex coordination of the immune response [156–159]. Activation of the Toll and Imd pathways depends on recognition of microbe-associated molecular patterns (MAMPs) and detection of damage-associated molecular patterns (DAMPs), suggesting that virulent bacteria could activate the Toll and Imd pathways to a higher degree [17, 115]. However, we did not find a clear correlation between the virulence level of the bacterium or bacterial load sustained and the degree to which the canonical immune response is activated. In sum, our results support the notion that the response to infection comprises more than simple activation of immune functions, but instead is a function of precise physiological interactions between host and microbe.

Identification of a core response to infection

Although the response to infection appears to be largely specific, we identified a core set of genes that are regulated by infection with most bacteria. Induced genes include the classical targets of the Toll and Imd pathways, such as antimicrobial peptides and immune effectors (TEPs and IMs). However, genes involved in cell and tissue biology (translation, secretion, cell division) were also upregulated by the majority of infection conditions, possibly indicating a response to the stress imposed by infection. On the other hand, genes involved in metabolism (protease activity, oxidation-reduction, glucose metabolism, respiration), as well as digestive enzymes (e.g. the maltase cluster), were downregulated, suggesting a complete reshaping of host metabolism during infection [11]. It is tempting to speculate that the majority of core genes

that do not fall under the immunity category could be part of a tolerance core response. Although the subject of tolerance mechanisms has attracted a lot of interest in recent years, identifying the genes and processes that define tolerance has remained somewhat elusive [160, 161]. Further characterization of the core genes identified here may shed light on universal tolerance mechanisms.

The idea of a core response to infection has also been explored in other organisms. In *Caenorhabditis elegans*, for example, a study using four different pathogens to assay the transcriptional response to infection found that the core of the response included genes involved in proteolysis, cell death, and stress responses [162]. Comparative transcriptomics work in the honey bee, *Apis mellifera*, also revealed a core set of genes utilized in response to distinct pathogens, including genes involved in immunity, stress responses, and tissue repair [163]. In *Danio rerio*, immunity, metabolism, and cell killing have been implicated in host defense [164]. Collectively, these results and ours indicate that there is considerable overlap in the core response to infection across species, and that this consistency extends beyond classical immune sensing and signaling. Having a well-defined core response to infection in *Drosophila* will allow future studies to quantitatively assess differences in how distinct pathogens induce the core, as well as test the relative importance of various elements of the core in promoting resistance to and tolerance of infection.

A surprisingly high proportion (~40%) of the core response to infection was induced only by live microbes, but was not stimulated by challenge with heat-killed bacteria. One possibility is that MAMPs, such as peptidoglycan, are partially, if not fully, degraded at the sampled time points, obscuring our ability to appreciate the full extent of the response to MAMPs. An alternative

explanation is that almost half of the core response to infection is a reaction to microbial activity, rather than just to the presence of MAMPs. This latter model involves the detection of the host's own DAMPs upon infection [114]. For example, bacterial growth and secretion of toxins can inflict damage to host tissues, leading to the generation of DAMPs, such as actin, proteases, and elastases [17, 115, 165]. In turn, DAMPs can activate the Toll, Imd, and JAK-STAT pathways, which may trigger higher levels of signaling in these pathways beyond that which is induced by the detection of MAMPs [17, 18, 115, 165]. Higher degrees of activation in these pathways could then translate into the induction of a larger set of target genes, which could partially account for the 40% of core genes that are uniquely induced by live infections.

Bacterial infection triggers long term changes in host transcription

Interestingly, our study found that gene expression levels do not always reflect the changes in bacterial load during the course of infection. In chronically infected flies, we found that most genes downregulated at 12 h post-infection had returned to baseline expression levels by 132 h after infection. Likewise, many of the induced genes also decreased in expression or returned to basal levels even while flies still harbored bacteria. It is possible that the injury inflicted to systemically infected flies generates a complex early response, which is resolved at later time points. However, we note that injury alone did not generally trigger the downregulation of genes observed in live infections. An alternative explanation is that the bacteria have entered into a less aggressive state in the late stages of infection, persisting but with a reduced impact on the physiology of their host. Yet another hypothesis is that the host's initial response to infection is broad-spectrum and disproportionately strong, with

the proactive goal of suppressing all bacteria before they can establish a highly detrimental infection. In this scenario, a subdued infection can be controlled with more nuance at later stages [131]. Finally, it is also possible that the percentage of recovered genes following infection with moderately virulent bacteria is overestimated because the RNA-seq is performed on pools of flies that may have distinct individual fates upon infection, and therefore distinct transcriptional kinetics. We have previously shown that flies infected with these same bacteria either die with a high bacterial load or survive with a low-level, persistent infection [144]. The individual flies at the 12 h RNA-seq data point comprise flies destined for both outcomes, but only persistently infected flies are sampled at the 132 h time point after mortality has occurred. If flies fated to die induce genes that are not triggered in flies destined to survive, those genes may appear to be upregulated in the pooled 12 h RNA sample that contains a mix of flies destined for both outcomes. Likewise, those same genes will appear to have returned to baseline levels at the 132 h time point when just chronically infected flies are sampled, creating the false impression that they have recovered. Future work is required to evaluate these hypotheses and to provide insight into how the complex dynamics of gene expression relate to changes in pathogen burden [166].

We also observed seemingly long-term alterations to the transcription of some core response genes, even in the case of infections with bacteria, such as *M. luteus* and *Ecc15*, that are reduced to undetectable levels or cleared by the host. For example, the expression of several antimicrobial peptide genes (*Drs*, *Dro*, and *AttB*) as well as other effector molecules (*IM4* and *IM3*) never returned to basal levels, even multiple days after elimination of the infection. Such sustained reactions could provide long-lasting benefits in an environment

with high risk of infection. Moreover, it should perhaps be considered that the baseline expression levels of these genes in laboratory-reared *Drosophila* are artificially low because of aseptic maintenance conditions as compared to those in natural environments.

The transcription factor *CrebA* prevents infection-induced ER stress

Among our core response genes, we identified *CrebA* as a key transcription factor that promotes host tolerance to infection. CrebA is the single *Drosophila* member of the Creb3-like family of transcription factors, which includes five different proteins in mammals: Creb3/Luman, Creb3L1/Oasis, Creb3L2/BBF2H7, Creb3L3/CrebH, and Creb3L4/Creb4 [137]. A recent study demonstrated that *CrebA* is a master regulator of secretory capacity, capable of regulating the expression of the general machinery required in all cells for secretion [138]. *Drosophila CrebA* appears to have the same functional role as its mammalian counterparts. Exogenous expression of mammalian liver-specific *CrebH* caused upregulation of genes involved in secretory capacity and increased secretion of specific cargos [137]. Moreover, each of the five human CREB3 factors is capable of activating secretory pathway genes in *Drosophila*, dependent upon their shared ATB (Adjacent To bZip) domain [137]. In agreement with the function of CrebA and CREB3 proteins described in the literature, our study finds that *CrebA* regulates a rapid, infection-induced increase in the expression of secretory pathway genes in the fat body, an organ analogous to the liver and adipose tissues of mammals. Finally, it has been shown that proinflammatory cytokines act to increase the transcription of *CrebH*, and that CrebH becomes activated in response to ER stress [167]. Our data demonstrate that the two principal immune pathways in *Drosophila*,

the Toll and Imd pathways, upregulate the expression of *CrebA* in response to bacterial challenge and that loss of *CrebA* in the fat body triggers ER stress upon infection. Collectively, the functions of mammalian *CrebH* as a regulator of secretory homeostasis under stress bear a striking resemblance to the role that we have attributed to *Drosophila CrebA* after bacterial challenge, suggesting that *CrebH* could have a similar role in mammals during infection.

CREB proteins are activated by phosphorylation from diverse kinases, including PKA and Ca2+/calmodulin-dependent protein kinases on the Serine 133 residue [168]. CrebA does not contain a PKA consensus phosphorylation site, and its transcriptional activity is only slightly enhanced by cAMP [142]. Rather, we found that the Toll and Imd pathways are both necessary and sufficient to regulate *CrebA* expression in the fat body. Loss of *CrebA* leads to ER stress, further aggravating the physiological strains of infection. However, a lack of *CrebA* in unchallenged conditions does not lead to the induction of ER stress. We therefore propose a model in which the Toll and Imd pathways act early to upregulate *CrebA* in order to adapt the fat body cells for infection, thus preventing ER stress that would otherwise be triggered by the response to infection [169] (Figure 3.17F). This interpretation would suggest that immune activation generates a massive and rapid increase in translation [170] and secretion in response to infection, and thus triggers cellular stress in the fat body. In that context, the Toll and Imd pathways would proactively induce expression of *CrebA* to prevent some of the stress that comes from their own activation.

Lastly, *CrebA* knockdown flies are more likely to die from infection yet they show no increase in pathogen burden. This demonstrates that *CrebA* is required for tolerance of infection [171, 172]. Considering that ER stress is induced upon

infection in the absence of *CrebA*, our data suggest that *CrebA* is a tolerance gene that helps mitigate the stress imposed by the host response to infection. Fast induction of *CrebA* by the immune system upon infection can therefore be interpreted as an active tolerance mechanism that is generally required to survive bacterial infection.

Table 3.1: **Core upregulated and downregulated genes with functional annotations and the level of gene expression change at 12 h after infection with each bacterium (log2 fold change).**

Gene ID	Gene name	N sig	Functional categories	Direction
FBgn0004396	CrebA	10	Signaling, Stress response, Metabolism, Secretion	Up
FBgn0019925	Surf4	10	Stress response, Secretion, Translation control	Up
FBgn0020414	Idgf3	10	Stress response, Metabolism, Wound healing	Up
FBgn0028396	TotA	10	Antimicrobial response, Stress response	Up
FBgn0031701	TotM	10	Antimicrobial response, Stress response	Up
FBgn0032810	CG13077	10	Stress response, Metabolism, Metal ion homeostasis, Cell redox	Up
FBgn0033153	Gadd45	10	Signaling, Stress response, Cell cycle control, Tissue repair	Up
FBgn0034259	P32	10	Signaling, Stress response, Neuron-related, Tissue repair	Up
FBgn0037724	Fst	10	Signaling, Stress response, Tissue repair	Up
FBgn0038465	Irc	10	Stress response, Metabolism	Up
FBgn0039562	Gp93	10	Signaling, Stress response, Metabolism, Translation control	Up
FBgn0039666	Diedel	10	Signaling, Antimicrobial response, Stress response	Up
FBgn0044812	TotC	10	Antimicrobial response, Stress response	Up
FBgn0047135	CG32276	10	Signaling, Stress response, Metabolism, Secretion, Translation control	Up
FBgn0261560	Thor	10	Stress response, Metabolism, Cell growth regulation, Translation control	Up
FBgn0000276	CecA1	10	Antimicrobial response	Up
FBgn0000277	CecA2	10	Antimicrobial response	Up
FBgn0000278	CecB	10	Antimicrobial response	Up
FBgn0002930	nec	10	Antimicrobial response	Up
FBgn0004240	Dpt	10	Antimicrobial response	Up
FBgn0010381	Drs	10	Antimicrobial response	Up
FBgn0010385	Def	10	Antimicrobial response	Up
FBgn0010388	Dro	10	Antimicrobial response	Up
FBgn0012042	AttA	10	Antimicrobial response	Up
FBgn0014018	Rel	10	Antimicrobial response	Up
FBgn0014865	Mtk	10	Antimicrobial response	Up
FBgn0015221	Fer2LCH	10	Signaling, Metal ion homeostasis	Up

FBgn0022355	Tsf1	10	Signaling, Metal ion homeostasis	Up
FBgn0025583	IM2	10	Antimicrobial response	Up
FBgn0027584	CG4757	10	Metabolism	Up
FBgn0027929	NimB1	10	Antimicrobial response	Up
FBgn0030310	PGRP-SA	10	Antimicrobial response	Up
FBgn0031560	CG16713	10	Signaling	Up
FBgn0031561	CG16712	10	Signaling	Up
FBgn0032322	CG16743	10	Unknown	Up
FBgn0032638	SPH93	10	Signaling, Metabolism, Protease activity	Up
FBgn0032639	CG18563	10	Metabolism, Protease activity	Up
FBgn0033250	CG14762	10	Signaling, Neuron-related	Up
FBgn0034328	IM23	10	Antimicrobial response	Up
FBgn0034329	IM1	10	Antimicrobial response	Up
FBgn0034407	DptB	10	Antimicrobial response	Up
FBgn0034511	GNBP-like3	10	Antimicrobial response	Up
FBgn0034512	CG18067	10	Reproduction	Up
FBgn0034741	CG4269	10	Cell growth regulation	Up
FBgn0035022	CG11413	10	Unknown	Up
FBgn0035176	CG13905	10	Unknown	Up
FBgn0035343	CG16762	10	Unknown	Up
FBgn0035806	PGRP-SD	10	Antimicrobial response	Up
FBgn0037108	CG11306	10	Neuron-related, Tissue repair	Up
FBgn0037396	CG11459	10	Metabolism, Protease activity	Up
FBgn0038214	CG9616	10	Signaling	Up
FBgn0038250	CG3505	10	Metabolism, Protease activity	Up
FBgn0038299	Spn88Eb	10	Signaling, Metabolism, Protease activity	Up
FBgn0039564	CG5527	10	Signaling, Metabolism, Protease activity	Up
FBgn0039798	CG11313	10	Signaling, Metabolism, Protease activity	Up
FBgn0040340	TRAM	10	Signaling, Metabolism, Secretion	Up
FBgn0040653	IM4	10	Antimicrobial response	Up
FBgn0040734	CG15065	10	Unknown	Up
FBgn0040736	IM3	10	Antimicrobial response	Up
FBgn0041182	Tep2	10	Antimicrobial response	Up
FBgn0041579	AttC	10	Antimicrobial response	Up
FBgn0041581	AttB	10	Antimicrobial response	Up
FBgn0043575	PGRP-SC2	10	Antimicrobial response	Up
FBgn0043578	PGRP-SB1	10	Antimicrobial response	Up
FBgn0050026	CG30026	10	Unknown	Up

FBgn0052023	CG32023	10	Unknown	Up
FBgn0054003	NimB3	10	Antimicrobial response	Up
FBgn0054054	CG34054	10	Unknown	Up
FBgn0067905	IM14	10	Antimicrobial response	Up
FBgn0260474	CG30002	10	Signaling, Metabolism	Up
FBgn0261989	CG42807	10	Unknown	Up
FBgn0262838	CG43202	10	Antimicrobial response (putative AMP)	Up
FBgn0262881	CG43236	10	Antimicrobial response (putative AMP)	Up
FBgn0264541	CG43920	10	Antimicrobial response (putative AMP)	Up
FBgn0264979	CG4267	10	Metabolism	Up
FBgn0265048	cv-d	10	Signaling, Metabolism	Up
FBgn0265577	CR44404	10	Unknown	Up
FBgn0266405	CR45045	10	Unknown	Up
FBgn0010638	Sec61beta	9	Signaling, Stress response, Metabolism, Secretion, Neuron-related, Translation control	Up
FBgn0014002	Pdi	9	Signaling, Stress response, Secretion, Cell redox	Up
FBgn0025678	CaBP1	9	Signaling, Stress response, Cell redox	Up
FBgn0033129	Tsp42Eh	9	Signaling, Stress response, Tissue repair	Up
FBgn0036157	CG7560	9	Stress response, Metabolism, Cell redox	Up
FBgn0040285	Scamp	9	Signaling, Stress response, Metabolism, Neuron-related, Translation control	Up
FBgn0044810	TotX	9	Antimicrobial response, Stress response	Up
FBgn0052190	NUCB1	9	Signaling, Stress response, Metabolism, Secretion, Translation control	Up
FBgn0003961	Uro	9	Metabolism	Up
FBgn0023129	aay	9	Metabolism, Neuron-related	Up
FBgn0027095	Manf	9	Stress response, Metabolism, Neuron-related	Up
FBgn0030041	CG12116	9	Metabolism, Neuron-related	Up
FBgn0031937	CG13795	9	Neuron-related	Up
FBgn0032773	fon	9	Antimicrobial response	Up
FBgn0034761	CG4250	9	Signaling	Up
FBgn0035976	PGRP-LC	9	Antimicrobial response	Up

FBgn0037906	PGRP-LB	9	Antimicrobial response, Metabolism	Up
FBgn0038829	CG17271	9	Signaling, Reproduction	Up
FBgn0039022	CG4725	9	Metabolism, Protease activity	Up
FBgn0039023	CG4723	9	Metabolism, Protease activity	Up
FBgn0040503	CG7763	9	Signaling	Up
FBgn0040582	CG5791	9	Unknown	Up
FBgn0051092	LpR2	9	Signaling, Metabolism, Neuron-related	Up
FBgn0085455	CG34426	9	Metabolism	Up
FBgn0260632	dl	9	Antimicrobial response, Metabolism	Up
FBgn0260746	Ect3	9	Signaling, Metabolism	Up
FBgn0262481	CG43071	9	Antimicrobial response (putative AMP)	Up
FBgn0025700	CG5885	8	Signaling, Stress response, Metabolism, Secretion, Translation control	Up
FBgn0028327	l(1)G0320	8	Stress response, Metabolism, Secretion, Translation control	Up
FBgn0029688	lva	8	Stress response, Metabolism, Secretion, Translation control	Up
FBgn0033196	CG1358	8	Stress response, Metabolism, Secretion, Translation control	Up
FBgn0035623	mthl2	8	Signaling, Stress response, Metabolism	Up
FBgn0039629	CG11842	8	Stress response, Metabolism, Secretion, Protease activity	Up
FBgn0086357	Sec61alpha	8	Signaling, Stress response, Metabolism, Secretion, Translation control	Up
FBgn0250820	meigo	8	Stress response, Secretion, Neuron-related	Up
FBgn0000279	CecC	8	Antimicrobial response	Up
FBgn0011204	cue	8	Signaling, Metabolism	Up
FBgn0011672	Mvl	8	Signaling, Metal ion homeostasis	Up
FBgn0031397	CG15385	8	Metabolism	Up
FBgn0035909	ergic53	8	Signaling, Secretion	Up
FBgn0037819	CG14688	8	Metabolism	Up
FBgn0037875	ZnT86D	8	Metal ion homeostasis	Up
FBgn0041180	Tep4	8	Antimicrobial response	Up
FBgn0041710	yellow-f	8	Melanin production	Up
FBgn0046763	CG17278	8	Signaling	Up
FBgn0052185	edin	8	Antimicrobial response	Up

FBgn0052373	CG32373	8	Neuron-related	Up
FBgn0053192	MtnD	8	Signaling, Metal ion homeostasis	Up
FBgn0053303	CG33303	8	Metabolism, Secretion	Up
FBgn0062961	pncr016:2R	8	Unknown	Up
FBgn0082831	pps	8	Gene regulation	Up
FBgn0086450	su(r)	8	Metabolism, Cell redox	Up
FBgn0261258	rgn	8	Signaling, Tissue repair	Up
FBgn0262035	CG42846	8	Antimicrobial response (putative AMP)	Up
FBgn0262569	CG43109	8	Signaling	Up
FBgn0262794	CG43175	8	Unknown	Up
FBgn0264753	Rgk1	8	Neuron-related	Up
FBgn0000083	AnxB9	7	Stress response, Metabolism, Secretion, Translation control	Up
FBgn0010415	Sdc	7	Stress response, Metabolism, Neuron-related	Up
FBgn0011016	SsRbeta	7	Signaling, Stress response, Metabolism, Secretion, Translation control	Up
FBgn0011296	l(2)efl	7	Stress response, Metabolism, Translation control	Up
FBgn0025724	beta'COP	7	Stress response, Metabolism, Secretion, Translation control	Up
FBgn0032147	IP3K1	7	Signaling, Stress response	Up
FBgn0039593	CG9989	7	Stress Response	Up
FBgn0040070	Trx-2	7	Stress response, Metabolism	Up
FBgn0051216	Naam	7	Stress Response, Neuron-related	Up
FBgn0003495	spz	7	Antimicrobial response	Up
FBgn0025456	CREG	7	Metabolism, Cell redox	Up
FBgn0027329	Tcp-1zeta	7	Signaling, Metabolism	Up
FBgn0030305	CG1749	7	Metabolism	Up
FBgn0030306	Spase25	7	Metabolism, Secretion, Protease activity	Up
FBgn0031973	Spn28Dc	7	Wounding response	Up
FBgn0032472	CG9928	7	Unknown	Up
FBgn0032587	CG5953	7	Cell fate	Up
FBgn0032836	CG10680	7	Unknown	Up
FBgn0034290	CG5773	7	Unknown	Up
FBgn0034440	CG10073	7	Metabolism	Up
FBgn0035665	Jon65Aiii	7	Protease activity	Up
FBgn0036919	Grasp65	7	Unknown	Up
FBgn0038530	AttD	7	Antimicrobial response	Up
FBgn0039609	CG14529	7	Protease activity	Up

FBgn0050091	CG30091	7	Protease activity	Up
FBgn0052103	CG32103	7	Translation, Transmembrane transport	Up
FBgn0052284	CG32284	7	Metabolism	Up
FBgn0085244	CG34215	7	Unknown	Up
FBgn0085452	CG34423	7	Metabolism	Up
FBgn0259175	ome	7	Protease activity	Up
FBgn0263402	CG43448	7	Unknown	Up
FBgn0002563	Lsp1beta	10	Lateral inhibition	Down
FBgn0002565	Lsp2	10	Neuron-related	Down
FBgn0010053	Jheh1	10	Response to toxin, Metabolism	Down
FBgn0015570	alpha-Est2	10	Drug metabolism, Cell-cell adhesion, Neuron-related	Down
FBgn0034480	CG16898	10	Unknown	Down
FBgn0036837	CG18135	10	Metabolism, Reproduction	Down
FBgn0038865	CG10824	10	Unknown	Down
FBgn0039685	Obp99b	10	Olfactory behavior, Response to pheromone	Down
FBgn0015316	Try29F	9	Protease activity	Down
FBgn0028853	CG15263	9	Unknown	Down
FBgn0030593	CG9512	9	Oxidation-reduction, Metabolism	Down
FBgn0030607	dob	9	Metabolism	Down
FBgn0033543	CG12338	9	Oxidation-reduction, Metabolism	Down
FBgn0035228	CG12091	9	Protein dephosphorylation	Down
FBgn0035356	CG16986	9	Protein homotetramerization	Down
FBgn0036587	CG4950	9	Unknown	Down
FBgn0038702	CG3739	9	Protease activity	Down
FBgn0039777	Jon99Fii	9	Protease activity	Down
FBgn0040732	CG16926	9	Sleep	Down
FBgn0051104	CG31104	9	Unknown	Down
FBgn0053301	CG33301	9	Unknown	Down
FBgn0085195	CG34166	9	Unknown	Down
FBgn0003358	Jon99Ci	8	Protease activity	Down
FBgn0013772	Cyp6a8	8	Response to toxin/insecticide, Oxidation-reduction, Toxin/insecticide metabolic process	Down
FBgn0016675	Lectin-galC1	8	Cell-cell adhesion	Down
FBgn0024913	Actbeta	8	Cell growth and development	Down
FBgn0025454	Cyp6g1	8	Response to toxin/insecticide, Oxidation-reduction	Down
FBgn0026721	fat-spondin	8	Response to temperature	Down
FBgn0026755	Ugt37b1	8	Metabolism, Drug Metabolism	Down

FBgn0028473	Non1	8	Cell division	Down
FBgn0029529	CG13365	8	Unknown	Down
FBgn0029648	CG3603	8	Oxidation-reduction	Down
FBgn0030105	CG15369	8	Unknown	Down
FBgn0030362	regucalcin	8	Metabolism, Reproduction	Down
FBgn0031305	Iris	8	Sleep	Down
FBgn0031418	CG3609	8	Oxidation-reduction	Down
FBgn0031694	Cyp4ac2	8	Response to toxin/insecticide, Oxidation-reduction, Toxin/insecticide metabolic process	Down
FBgn0032285	CG17108	8	Unknown	Down
FBgn0032297	CG17124	8	Regulation of phosphorylation	Down
FBgn0035355	CG16985	8	Protein homotetramerization	Down
FBgn0036975	CG5618	8	Metabolism	Down
FBgn0037788	CG3940	8	Metabolism	Down
FBgn0038603	PKD	8	Metabolism, Signaling	Down
FBgn0038878	CG3301	8	Oxidation-reduction	Down
FBgn0038914	fit	8	Response to temperature	Down
FBgn0039298	to	8	Mating behavior, Feeding behavior	Down
FBgn0039801	Npc2h	8	Sterol transport, Intracellular cholesterol transport, Hemolymph coagulation	Down
FBgn0040096	lectin-33A	8	Carbohydrate-binding	Down
FBgn0040099	lectin-28C	8	Carbohydrate-binding	Down
FBgn0040349	CG3699	8	Oxidation-reduction	Down
FBgn0040931	CG9034	8	Unknown	Down
FBgn0063492	GstE8	8	Metabolism	Down
FBgn0261575	tobi	8	Metabolism	Down
FBgn0262353	CR43051	8	Unknown	Down
FBgn0002939	niXD	7	Phototransduction	Down
FBgn0014031	Spat	7	Metabolism	Down
FBgn0014469	Cyp4e2	7	Oxidation-reduction	Down
FBgn0015575	alpha-Est7	7	Drug metabolism, Cell-cell adhesion, Neuron-related	Down
FBgn0027560	Tps1	7	Metabolism	Down
FBgn0027657	glob1	7	Oxygen transport	Down
FBgn0029990	CG2233	7	Unknown	Down
FBgn0030999	Mur18B	7	Metabolism	Down
FBgn0031182	Cyp6t1	7	Oxidation-reduction	Down
FBgn0031515	CG9664	7	Olfactory behavior	Down
FBgn0032076	CG9510	7	Metabolism	Down
FBgn0032381	Mal-B1	7	Metabolism	Down

FBgn0033203	CG2070	7	Oxidation-reduction, Metabolism	Down
FBgn0034117	CG7997	7	Metabolism	Down
FBgn0034500	CG11200	7	Oxidation-reduction	Down
FBgn0035290	dsb	7	Phagosome maturation	Down
FBgn0035344	Cyp4d20	7	Oxidation-reduction	Down
FBgn0037090	Est-Q	7	Neuron-related	Down
FBgn0037678	CG16749	7	Protease activity	Down
FBgn0039118	CG10208	7	Unknown	Down
FBgn0039519	Cyp6a18	7	Oxidation-reduction	Down
FBgn0040260	Ugt36Bc	7	Metabolism, Drug Metabolism	Down
FBgn0040398	CG14629	7	Unknown	Down
FBgn0040813	Nplp2	7	Neuron-related	Down
FBgn0051075	CG31075	7	Oxidation-reduction, Metabolism	Down
FBgn0051548	CG31548	7	Oxidation-reduction	Down
FBgn0061356	CG18003	7	Oxidation-reduction	Down
FBgn0085360	CG34331	7	Unknown	Down
FBgn0250848	26-29-p	7	Protease activity	Down
FBgn0259715	CG42369	7	Unknown	Down
FBgn0261859	CG42788	7	Unknown	Down
FBgn0263120	Acsl	7	Metabolism, Neuron-related	Down

Table 3.2: **Putative transcription factor binding sites enriched in core upregulated genes as ascertained by i-cisTarget.**

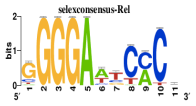
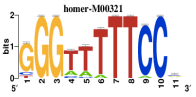
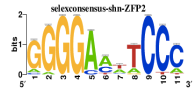
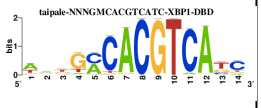
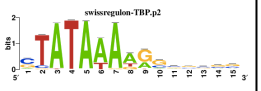
Gene ID	Gene name	Binding site	NES (Normalized enrichment score)	Putative/known function
FBgn0014018	*Rel*		10.45547	Immunity (Imd)
FBgn0260632	*dl*		7.46202	Immunity (Toll)
FBgn0003396	*shn*		5.47776	Development
FBgn0004396	*CrebA/Creb3-like*		5.68522	Secretion
FBgn0033010	*Atf6*		5.47312	Unknown
FBgn0021872	*Xbp1*		5.15217	Secretion (Unfolded Protein Response)
FBgn0003687	*Tbp*		4.14498	General

Table 3.3: **Putative transcription factor binding sites enriched in core downregulated genes as ascertained by i-cisTarget.**

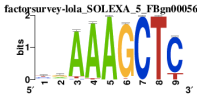
Gene ID	Gene name	Binding site	NES (Normalized enrichment score)	Putative/known function
FBgn0283521	*lola*		8.54478	Development; Reproduction; Metabolism
FBgn0001138	*grn/GATA1*		7.1363	Axon guidance; Development
FBgn0003117	*pnr*		6.11356	Development
FBgn0032223	*GATAd*		5.90999	General

Table 3.4: **Expression kinetics of core upregulated and downregulated genes.**

Upregulated

gene_id	gene_name	clean.prick	P.rett.heatl	E.fae.heatl	Ecc15	M.luteus	E.coli	S.marcesc	P.rettgeri	E.faecalis
FBgn0000276	CecA1	NA	0.782	NA	0.76	NA	0.522	-0.033	0.406	0.071
FBgn0000277	CecA2	0.488	1.007	0.959	0.662	0.704	0.409	-0.014	0.311	0.305
FBgn0000278	CecB	NA	0.97	NA	0.5	0.658	0.361	-0.049	0.313	0.225
FBgn0002930	nec	0.345	0.314	0.473	0.254	0.308	0.326	0.134	-0.533	0.395
FBgn0004240	Dpt	0.618	0.857	0.799	0.776	0.848	0.392	0.045	0.151	0.686
FBgn0004396	CrebA	NA	0.512	NA	0.624	0.559	0.512	0.241	0.649	0.594
FBgn0010381	Drs	0.674	0.963	0.818	0.711	0.341	0.614	0.086	0.237	0.338
FBgn0010385	Def	NA	1.04	NA	0.864	NA	0.614	-0.207	0.337	NA
FBgn0010388	Dro	0.589	0.8	0.817	0.665	0.633	0.266	0.022	0.12	0.516
FBgn0012042	AttA	0.434	0.798	0.725	0.662	0.501	0.29	0.049	0.207	0.316
FBgn0014018	Rel	NA	NA	NA	0.624	0.642	0.523	0.077	0.519	0.593
FBgn0014865	Mtk	0.585	0.866	0.864	0.821	0.569	0.329	0.046	0.096	0.445
FBgn0015221	Fer2LCH	NA	0.372	0.404	0.546	0.299	0.606	0.241	0.481	0.409
FBgn0019925	Surf4	NA	NA	NA	0.775	0.612	0.533	0.326	0.449	0.354
FBgn0020414	Idgf3	NA	0.968	NA	0.474	0.857	0.755	0.341	0.287	0.805
FBgn0022355	Tsf1	NA	0.124	0.472	-0.012	0.021	0.08	0.008	-0.158	0.093
FBgn0025583	IM2	0.576	0.568	0.822	0.766	0.279	0.439	0.324	0.239	0.338
FBgn0027584	CG4757	0.127	0.34	0.349	0.112	0.05	-0.018	-0.218	-0.326	0.08
FBgn0027929	NimB1	0.791	0.855	NA	1.174	0.7	1.071	0.583	0.882	0.695
FBgn0028396	TotA	0.588	0.937	0.822	0.749	0.859	0.389	0.541	0.599	0.684
FBgn0030310	PGRP-SA	0.723	0.628	0.77	0.557	0.474	0.657	0.236	0.346	0.304
FBgn0031560	CG16713	NA	NA	NA	0.241	0.269	0.127	-0.55	-0.085	0.257
FBgn0031561	CG16712	NA	0.492	NA	0.553	0.414	0.457	0.041	0.31	0.397
FBgn0031701	TotM	0.723	1.004	1.01	0.73	1.025	0.45	0.747	0.807	0.702
FBgn0032322	CG16743	0.726	0.554	NA	1.017	0.385	0.649	0.427	0.445	0.387
FBgn0032638	SPH93	0.182	0.243	-0.115	-0.353	-0.112	-0.122	-0.107	-0.559	-0.047
FBgn0032639	CG18563	NA	0.241	NA	-0.313	-0.148	-0.238	-0.113	-0.352	-0.134
FBgn0032810	CG13077	0.542	0.492	0.44	0.647	0.303	0.416	0.007	0.387	0.315

FBgn0033153	Gadd45	NA	NA	0.715	0.538	0.629	0.582	0.396	0.391	1.026
FBgn0033250	CG14762	0.862	0.676	0.739	0.737	0.365	0.714	0.016	0.162	0.501
FBgn0034259	P32	NA	0.728	NA	0.963	0.708	0.822	0.049	0.376	0.444
FBgn0034328	IM23	0.78	1.118	1.012	0.658	0.334	0.751	0.347	0.052	0.328
FBgn0034329	IM1	0.814	1.017	0.951	0.81	0.388	0.729	0.431	0.286	0.397
FBgn0034407	DptB	0.651	0.877	0.869	0.791	0.795	0.392	0.102	0.144	0.634
FBgn0034511	GNBP-like3	0.595	0.707	0.554	0.652	0.491	0.666	0.4	0.403	0.477
FBgn0034512	CG18067	0.647	0.668	0.721	0.447	0.206	0.505	0.603	0.189	0.335
FBgn0034741	CG4269	NA	NA	NA	0.507	0.895	0.69	0.439	0.671	1.32
FBgn0035022	CG11413	NA	1.178	2.733	0.81	1.621	1.039	1.093	0.82	0.824
FBgn0035176	CG13905	NA	NA	NA	0.393	NA	NA	1.059	1.148	0.825
FBgn0035343	CG16762	1.065	0.923	0.838	0.255	1.036	1.012	0.916	0.917	0.669
FBgn0035806	PGRP-SD	0.53	0.9	0.641	0.698	0.584	0.325	0.149	0.237	0.391
FBgn0037108	CG11306	NA	0.325	0.232	0.26	0.252	0.219	0.015	0.068	0.115
FBgn0037396	CG11459	0.964	1.092	1.105	1.307	0.548	0.676	0.863	0.575	0.495
FBgn0037724	Fst	1.289	NA	NA	0.906	1.083	1.149	1.065	1.026	0.817
FBgn0038214	CG9616	0.739	0.737	NA	0.793	0.139	0.93	0.113	0.323	0.166
FBgn0038250	CG3505	NA	NA	NA	0.471	0.364	0.479	0.071	0.449	0.373
FBgn0038299	Spn88Eb	0.627	0.548	0.367	0.527	0.207	0.478	-0.019	0.467	0.338
FBgn0038465	Irc	0.655	0.648	NA	0.691	0.506	0.564	0.188	0.278	0.484
FBgn0039562	Gp93	NA	NA	NA	NA	0.782	0.64	0.566	0.756	0.396
FBgn0039564	CG5527	NA	NA	NA	1.207	0.628	0.65	0.478	0.698	0.647
FBgn0039666	Diedel	0.855	0.882	1.099	0.577	1.036	0.755	0.821	0.754	0.505
FBgn0039798	CG11313	0.444	1.013	1.041	1.016	0.318	0.495	-0.11	-0.009	0.397
FBgn0040340	TRAM	NA	NA	NA	NA	0.891	0.651	0.504	0.337	0.711
FBgn0040653	IM4	0.618	0.626	0.703	0.638	0.356	0.553	0.318	0.093	0.372
FBgn0040734	CG15065	0.904	1.321	1.329	0.964	0.182	0.716	0.687	0.13	0.358
FBgn0040736	IM3	0.637	0.566	0.805	0.719	0.368	0.603	0.428	0.231	0.378
FBgn0041182	Tep2	0.869	0.965	0.824	0.78	0.568	0.732	0.635	0.694	0.498
FBgn0041579	AttC	0.575	0.804	0.776	0.799	0.632	0.267	0.017	0.139	0.464

FBgn0041581	AttB	0.453	0.816	0.671	0.643	0.618	0.301	0.024	0.183	0.37
FBgn0043575	PGRP-SC2	0.482	0.77	0.232	0.684	0.672	0.499	-0.178	0.239	0.341
FBgn0043578	PGRP-SB1	0.511	0.777	0.707	0.764	0.462	0.27	0.037	0.163	0.316
FBgn0044812	TotC	0.619	0.966	0.985	0.663	1.058	0.41	0.596	0.564	0.781
FBgn0047135	CG32276	NA	NA	NA	0.848	0.617	0.606	0.254	0.468	0.498
FBgn0050026	CG30026	NA	0.896	NA	0.564	0.473	0.638	0.163	0.224	0.477
FBgn0052023	CG32023	0.572	0.166	0.036	-0.01	0.21	0.159	0.362	0.283	0.135
FBgn0054003	NimB3	NA	NA	NA	0.942	0.779	0.906	0.668	0.381	0.78
FBgn0054054	CG34054	0.535	0.791	0.884	0.741	0.324	0.723	0.005	0.438	0.465
FBgn0067905	IM14	0.633	0.544	0.726	0.707	0.325	0.595	0.294	0.13	0.401
FBgn0260474	CG30002	0.46	0.673	0.591	0.542	0.198	0.389	0.068	-0.149	0.226
FBgn0261560	Thor	NA	0.289	0.846	0.864	0.943	1.558	0.955	0.893	1.141
FBgn0261989	CG42807	NA	0.7	NA	0.14	0.344	-0.037	0.06	0.291	-0.048
FBgn0262838	CG43202	0.619	0.69	0.788	0.697	0.426	0.694	0.282	0.522	0.464
FBgn0262881	CG43236	0.557	0.895	0.902	0.944	0.697	0.293	-0.05	0.205	0.449
FBgn0264541	CG43920	0.542	0.876	0.825	0.779	0.814	0.339	-0.038	0.155	0.502
FBgn0264979	CG4267	NA	NA	NA	0.668	0.429	0.796	0.026	0.549	0.595
FBgn0265048	cv-d	NA	0.481	0.364	0.441	0.138	0.311	0.034	0.148	0.339
FBgn0265577	CR44404	0.441	0.772	0.791	0.704	0.345	0.353	-0.039	0.161	0.236
FBgn0266405	CR45045	0.554	0.886	0.78	0.797	0.701	0.317	-0.027	0.13	0.53
FBgn0003961	Uro	NA	NA	0.614	NA	0.761	0.624	0.759	0.972	0.624
FBgn0010638	Sec61beta	NA	0.594	NA	0.705	0.676	0.633	0.503	0.45	0.579
FBgn0014002	Pdi	NA	0.592	0.223	NA	0.197	0.297	0.325	0.238	0.372
FBgn0023129	aay	NA	NA	NA	-0.103	NA	0.213	0.241	0.451	NA
FBgn0025678	CaBP1	NA	0.355	NA	NA	0.392	0.445	0.603	0.265	0.358
FBgn0027095	Manf	NA	NA	NA	0.917	0.485	0.386	0.455	0.443	0.523
FBgn0030041	CG12116	NA	NA	NA	NA	0.99	1.14	0.654	0.874	1.252
FBgn0031937	CG13795	NA	NA	NA	NA	-0.02	-0.057	-0.291	0.247	-0.128
FBgn0032773	fon	0.733	0.935	0.884	0.895	0.463	0.723	0.462	NA	0.589
FBgn0033129	Tsp42Eh	1.062	0.989	1.075	1.126	NA	0.879	0.655	0.829	1.022

FBgn0034761	CG4250	0.256	0.498	0.606	0.339	0.281	0.205	0.077	0.02	0.353
FBgn0035976	PGRP-LC	NA	0.218	-0.45	0.146	-0.695	0.216	-0.12	0.429	-0.719
FBgn0036157	CG7560	NA	0.698	0.419	0.517	0.064	0.068	-0.025	0.329	0.288
FBgn0037906	PGRP-LB	NA	NA	NA	0.319	-0.48	0.278	-0.081	0.207	NA
FBgn0038829	CG17271	NA	0.568	NA	0.522	0.308	0.543	0.509	0.343	0.437
FBgn0039022	CG4725	NA	1.178	NA	0.698	0.635	0.612	0.359	0.154	0.501
FBgn0039023	CG4723	0.647	0.908	NA	0.522	0.39	0.515	0.372	0.245	0.571
FBgn0040285	Scamp	NA	0.364	0.375	NA	-0.06	0.331	-0.288	0.343	0.05
FBgn0040503	CG7763	NA	0.314	0.135	0.603	-1.069	0.064	-0.551	-0.642	0.122
FBgn0040582	CG5791	0.536	0.608	0.798	0.321	0.13	0.442	0.477	-1.291	0.23
FBgn0044810	TotX	0.5	0.823	0.642	0.524	0.627	0.367	0.037	0.299	NA
FBgn0051092	LpR2	0.751	0.852	NA	1.051	0.579	0.717	0.225	0.144	0.576
FBgn0052190	NUCB1	NA	0.731	NA	0.514	0.444	0.654	0.399	0.338	0.471
FBgn0085455	CG34426	NA	NA	0.305	-0.496	0.3	-0.278	0.518	-0.041	0.182
FBgn0260632	dl	NA	NA	NA	NA	0.844	0.809	0.652	0.483	0.904
FBgn0260746	Ect3	NA	0.891	NA	0.843	-0.516	0.401	-0.245	0.357	-0.88
FBgn0262481	CG43071	0.485	0.95	NA	0.895	0.799	0.575	-0.149	0.24	0.864
FBgn0000279	CecC	NA	1.076	NA	0.611	NA	0.355	-0.052	0.306	NA
FBgn0011204	cue	NA	0.695	NA	1.121	0.594	0.847	0.347	0.121	0.69
FBgn0011672	Mvl	NA	NA	NA	NA	0.643	NA	0.237	0.492	0.531
FBgn0025700	CG5885	NA	NA	NA	NA	0.704	0.549	0.437	0.518	0.416
FBgn0028327	l(1)G0320	NA	NA	NA	NA	0.688	0.569	0.477	0.199	0.41
FBgn0029688	lva	NA	NA	NA	NA	0.303	0.246	0.273	0.472	0.241
FBgn0031397	CG15385	NA	NA	NA	NA	0.408	0.517	-0.198	0.395	-0.066
FBgn0033196	CG1358	NA	NA	NA	NA	0.473	0.596	-0.675	-0.206	0.578
FBgn0035623	mthl2	NA	NA	NA	NA	0.194	0.016	0.125	0.05	0.26
FBgn0035909	ergic53	NA	NA	NA	NA	0.545	0.465	0.348	0.22	0.463
FBgn0037819	CG14688	0.977	NA	1.098	1.289	1.184	1.138	1.438	0.969	1.245
FBgn0037875	ZnT86D	NA	NA	NA	NA	0.603	NA	0.451	0.685	0.692
FBgn0039629	CG11842	0.505	0.61	0.58	0.567	0.459	0.412	0.581	NA	0.422

FBgn0041180	Tep4	NA	NA	NA	NA	0.325	0.119	0.027	-0.115	0.076
FBgn0041710	yellow-f	NA	NA	NA	NA	0.362	0.67	-0.041	0.057	0.463
FBgn0046763	CG17278	NA	NA	NA	NA	0.822	0.619	0.399	0.874	0.772
FBgn0052185	edin	NA	NA	NA	0.611	NA	0.348	-0.047	0.365	NA
FBgn0052373	CG32373	NA	NA	NA	0.612	0.672	0.606	-0.11	0.369	NA
FBgn0053192	MtnD	0.483	NA	-0.232	-0.097	-0.259	0.408	0.311	0.318	0.68
FBgn0053303	CG33303	NA	0.219	0.089	NA	-0.123	0.361	0.378	0.331	0.272
FBgn0062961	pncr016:2R	0.551	1.045	NA	0.644	0.461	0.821	0.411	NA	0.485
FBgn0082831	pps	NA	0.557	NA	0.822	0.796	0.762	0.022	0.268	NA
FBgn0086357	Sec61alpha	NA	NA	NA	NA	0.534	0.414	0.463	0.301	0.199
FBgn0086450	su(r)	NA	0.337	-0.331	-1.473	NA	-0.267	0.331	-0.066	-0.037
FBgn0250820	meigo	NA	NA	NA	NA	0.791	0.784	0.813	0.547	0.601
FBgn0261258	rgn	NA	NA	NA	0.37	0.256	0.482	0.156	0.091	NA
FBgn0262035	CG42846	NA	0.22	-0.327	0.647	-0.001	-0.224	-0.391	0.089	-0.071
FBgn0262569	CG43109	0.502	1.089	NA	1.01	0.773	0.423	-0.015	0.12	NA
FBgn0262794	CG43175	NA	0.794	NA	0.557	NA	0.689	-0.04	0.222	NA
FBgn0264753	Rgk1	NA	NA	NA	NA	0.91	0.965	0.994	1.056	NA
FBgn0000083	AnxB9	NA	NA	0.8	NA	0.799	NA	0.716	0.626	0.911
FBgn0003495	spz	NA	NA	NA	NA	NA	0.837	0.477	0.697	NA
FBgn0010415	Sdc	NA	NA	NA	NA	0.356	0.631	0.072	0.42	NA
FBgn0011016	SsRbeta	NA	NA	NA	NA	0.79	0.456	0.407	NA	0.429
FBgn0011296	l(2)efl	1.34	NA	0.705	NA	NA	NA	0.752	1.087	0.68
FBgn0025456	CREG	NA	NA	NA	NA	0.634	NA	0.54	NA	0.771
FBgn0025724	beta'COP	NA	NA	NA	NA	0.355	NA	0.187	0.193	0.261
FBgn0027329	Tcp-1zeta	NA	NA	0.138	0.431	-0.043	-0.251	0.287	0.493	-0.068
FBgn0030305	CG1749	NA	NA	NA	NA	1.024	NA	0.612	0.643	0.619
FBgn0030306	Spase25	NA	NA	NA	NA	0.768	NA	0.383	0.35	0.435
FBgn0031973	Spn28Dc	NA	NA	NA	NA	0.672	NA	NA	0.808	0.753
FBgn0032147	IP3K1	NA	NA	NA	NA	0.057	-0.267	NA	0.277	0.275
FBgn0032472	CG9928	NA	NA	NA	0.659	0.237	0.086	0.038	NA	0.253

FBgn0032587	CG5953	NA	NA	NA	0.607	0.914	NA	NA	0.914	1.279
FBgn0032836	CG10680	NA	NA	NA	NA	0.45	NA	0.371	0.424	0.718
FBgn0034290	CG5773	NA	NA	NA	NA	1.031	NA	0.051	0.567	1.12
FBgn0034440	CG10073	NA	0.776	NA	0.857	0.358	NA	-0.466	-0.11	0.006
FBgn0035665	Jon65Aiii	NA	1.042	NA	0.985	1.323	NA	1.377	NA	1.162
FBgn0036919	Grasp65	NA	NA	NA	NA	NA	0.356	0.28	0.538	0.511
FBgn0038530	AttD	NA	NA	NA	0.393	NA	0.162	-0.333	0.287	NA
FBgn0039593	CG9989	NA	0.684	NA	NA	-0.22	-0.067	-0.742	-0.011	NA
FBgn0039609	CG14529	NA	0.975	NA	1.185	1.16	1.073	0.094	0.345	NA
FBgn0040070	Trx-2	1.02	NA	0.666	0.852	0.517	NA	NA	0.631	0.836
FBgn0050091	CG30091	NA	NA	NA	NA	0.068	-0.21	-0.366	-0.579	0.059
FBgn0051216	Naam	0.964	NA	NA	0.994	NA	NA	NA	0.369	0.886
FBgn0052103	CG32103	NA	NA	NA	NA	0.396	NA	0.64	0.112	NA
FBgn0052284	CG32284	NA	NA	NA	NA	NA	NA	NA	NA	NA
FBgn0085244	CG34215	NA	NA	NA	0.598	-0.365	-0.556	-1.163	0.009	-0.75
FBgn0085452	CG34423	0.344	0.206	0.508	0.341	0.513	0.598	0.041	-0.01	0.451
FBgn0259175	ome	NA	NA	NA	NA	0.498	NA	-0.12	0.364	0.361
FBgn0263402	CG43448	NA	1.651	1.674	1.646	0.973	0.922	0.184	0.781	0.737

Downregulated

gene_id	gene_name	clean.prick	P.rett.heatk	E.fae.heatk	Ecc15	M.luteus	E.coli	S.marcesce	P.rettgeri	E.faecalis
FBgn0002563	Lsp1beta	1.153	1.269	1.295	1.322	1.415	1.223	1.052	0.874	1.484
FBgn0002565	Lsp2	1.294	1.241	1.595	1.331	1.057	1.171	1.069	0.836	1.227
FBgn0010053	Jheh1	NA	NA	NA	0.839	0.703	0.929	0.563	0.798	0.721
FBgn0015570	alpha-Est2	NA	NA	NA	0.5	0.891	0.983	0.516	0.729	0.831
FBgn0034480	CG16898	1.238	1.064	1.285	1.317	1.041	1.051	0.832	0.869	0.916
FBgn0036837	CG18135	0.863	1.361	1.149	1.037	1.069	1.067	0.655	0.675	1.14
FBgn0038865	CG10824	1.056	1.007	0.885	0.932	1.109	1.193	0.91	0.808	1.065
FBgn0039685	Obp99b	1.186	NA	1.294	1.393	1.012	1.08	1.078	0.744	1.193
FBgn0015316	Try29F	0.751	0.727	0.55	0.752	0.583	0.497	0.607	0.375	0.624

FBgn0028853	CG15263	1.375	1.385	1.546	1.308	1.15	1.223	1.215	1.18	1.285
FBgn0030593	CG9512	NA	0.561	NA	1.34	1.276	1.006	0.701	1.061	0.938
FBgn0030607	dob	NA	1.374	NA	NA	1.332	1.019	0.695	0.787	0.965
FBgn0033543	CG12338	NA	NA	NA	0.838	0.995	0.726	0.722	0.744	0.721
FBgn0035228	CG12091	NA	0.817	0.727	0.922	0.872	0.788	0.522	0.805	0.75
FBgn0035356	CG16986	NA	NA	NA	1.697	1.374	1.721	1.242	1.176	1.365
FBgn0036587	CG4950	NA	NA	NA	NA	1.37	1.052	0.887	0.792	0.953
FBgn0038702	CG3739	NA	1.423	NA	NA	1.202	NA	1.303	1.694	NA
FBgn0039777	Jon99Fii	NA	NA	NA	NA	0.379	0.473	-0.019	0.536	0.362
FBgn0040732	CG16926	NA	NA	0.267	0.187	0.647	0.658	0.209	0.688	0.624
FBgn0051104	CG31104	NA	NA	NA	NA	1.369	1.599	1.32	1.422	1.436
FBgn0053301	CG33301	1.066	NA	NA	1.007	0.66	NA	0.554	0.511	0.865
FBgn0085195	CG34166	NA	NA	NA	1.612	1.824	NA	1.289	1.232	1.799
FBgn0003358	Jon99Ci	NA	NA	2.851	3.477	1.446	1.779	1.288	1.473	1.491
FBgn0013772	Cyp6a8	NA	NA	NA	NA	0.522	0.951	0.472	0.726	0.888
FBgn0016675	Lectin-galC1	NA	NA	0.714	0.696	0.907	0.891	0.873	0.579	0.868
FBgn0024913	Actbeta	NA	NA	NA	NA	NA	1.378	1.061	1.153	1.176
FBgn0025454	Cyp6g1	NA	NA	NA	NA	1.372	1.547	1.226	1.013	1.445
FBgn0026721	fat-spondin	NA	NA	NA	NA	1.272	1.226	1.101	0.989	1.237
FBgn0026755	Ugt37b1	NA	NA	2.942	NA	0.989	NA	1.172	1.014	NA
FBgn0028473	Non1	NA	NA	0.744	1.056	0.752	0.527	0.691	0.493	0.59
FBgn0029529	CG13365	NA	NA	0.57	NA	0.468	NA	0.098	0.63	0.722
FBgn0029648	CG3603	NA	NA	NA	1.282	1.036	1.439	NA	1.172	1.29
FBgn0030105	CG15369	1.163	NA	NA	NA	0.877	0.978	1.307	1.238	1.257
FBgn0030362	regucalcin	NA	NA	NA	NA	NA	1.017	1.148	0.562	NA
FBgn0031305	Iris	NA	NA	NA	NA	NA	1.37	0.992	1.188	NA
FBgn0031418	CG3609	NA	NA	NA	3.319	1.466	NA	1.49	1.549	1.441
FBgn0031694	Cyp4ac2	NA	NA	NA	NA	1.218	1.209	0.922	0.967	1.108
FBgn0032285	CG17108	NA	NA	NA	NA	0.893	0.85	NA	1.057	0.926
FBgn0032297	CG17124	NA	NA	NA	NA	0.632	0.75	0.785	0.807	0.784

FBgn0035355	CG16985	NA	1.128	NA	NA	1.291	NA	1.011	1.258	1.3
FBgn0036975	CG5618	NA	NA	NA	NA	0.114	NA	NA	0.637	0.156
FBgn0037788	CG3940	NA	NA	NA	NA	1.82	NA	1.135	1.115	1.755
FBgn0038603	PKD	NA	NA	0.584	NA	NA	0.045	0.026	0.245	0.394
FBgn0038878	CG3301	NA	NA	NA	NA	1.283	1.425	1.352	1.271	1.293
FBgn0038914	fit	NA	NA	NA	NA	1.12	NA	1.406	NA	1.226
FBgn0039298	to	NA	NA	NA	NA	NA	NA	0.976	0.929	NA
FBgn0039801	Npc2h	NA	NA	NA	NA	1.051	NA	0.655	1.062	1.131
FBgn0040096	lectin-33A	NA	NA	NA	NA	1.122	1.155	NA	0.758	0.719
FBgn0040099	lectin-28C	NA	NA	NA	1.32	1.595	NA	1.124	0.946	NA
FBgn0040349	CG3699	NA	NA	NA	NA	1.354	1.509	1.045	1.208	1.306
FBgn0040931	CG9034	NA	NA	NA	0.568	0.49	0.409	0.236	0.613	0.289
FBgn0063492	GstE8	NA	1.084	1.475	1.551	1.606	1.427	1.014	1.469	1.047
FBgn0261575	tobi	1.462	1.28	NA	1.409	1.321	1.477	NA	1.068	1.339
FBgn0262353	CR43051	0.863	NA	0.702	NA	0.898	0.652	0.614	0.85	0.655
FBgn0002939	ninaD	NA	NA	NA	NA	0.827	NA	0.763	0.855	NA
FBgn0014031	Spat	NA	NA	NA	NA	1.5	NA	1.435	NA	1.278
FBgn0014469	Cyp4e2	NA	NA	NA	NA	1.202	NA	1.189	1.225	NA
FBgn0015575	alpha-Est7	NA	NA	NA	NA	1.567	NA	1.161	1.136	NA
FBgn0027560	Tps1	NA	NA	NA	NA	0.794	0.836	NA	0.874	0.802
FBgn0027657	glob1	NA	NA	NA	NA	1.489	NA	1.151	1.214	NA
FBgn0029990	CG2233	NA	NA	NA	NA	NA	NA	NA	NA	NA
FBgn0030999	Mur18B	NA	NA	NA	NA	1.454	1.366	1.384	1.391	1.297
FBgn0031182	Cyp6t1	NA	NA	NA	NA	1.199	NA	1.064	0.871	1.103
FBgn0031515	CG9664	0.814	NA	NA	NA	NA	0.724	0.502	0.933	0.759
FBgn0032076	CG9510	NA	NA	NA	2.714	1.683	NA	1.393	1.339	1.72
FBgn0032381	Mal-B1	1.006	1.016	0.999	1.233	NA	NA	1.104	0.758	0.892
FBgn0033203	CG2070	0.979	NA	NA	NA	NA	NA	0.906	0.896	1.042
FBgn0034117	CG7997	NA	NA	NA	NA	NA	NA	1.962	1.522	NA
FBgn0034500	CG11200	NA	NA	NA	NA	0.132	NA	0.505	0.725	0.708

FBgn0035290	dsb	NA	NA	NA	NA	NA	NA	0.985	1.014	1.419
FBgn0035344	Cyp4d20	NA	0.638	NA	NA	0.3	NA	-0.158	0.694	0.702
FBgn0037090	Est-Q	NA	NA	NA	NA	1.265	NA	0.582	0.71	1.101
FBgn0037678	CG16749	NA	NA	NA	NA	1.165	NA	0.797	0.789	0.94
FBgn0039118	CG10208	NA	0.522	NA	1.147	NA	NA	0.588	1.077	0.682
FBgn0039519	Cyp6a18	NA	NA	NA	NA	1.056	1.204	1.014	NA	1.207
FBgn0040260	Ugt36Bc	NA	NA	NA	NA	0.972	NA	NA	0.786	0.986
FBgn0040398	CG14629	NA	NA	NA	NA	NA	0.695	0.798	0.889	0.973
FBgn0040813	Nplp2	NA	NA	NA	NA	1.218	NA	0.835	NA	1.003
FBgn0051075	CG31075	NA	NA	1.421	NA	1.186	1.485	NA	1.021	1.447
FBgn0051548	CG31548	NA	NA	NA	NA	1.264	NA	0.943	0.929	0.927
FBgn0061356	CG18003	NA	NA	NA	NA	1.422	NA	1.206	1.139	1.272
FBgn0085360	CG34331	NA	NA	NA	NA	NA	NA	1.192	1.086	NA
FBgn0250848	26-29-p	NA	NA	NA	NA	NA	0.837	0.801	0.984	1.098
FBgn0259715	CG42369	NA	NA	NA	NA	1.146	NA	1.267	1.026	1.007
FBgn0261859	CG42788	NA	NA	NA	NA	NA	NA	1.314	NA	NA
FBgn0263120	Acsl	NA	NA	NA	NA	1.279	NA	0.749	0.886	1.276

Table 3.5: **List of genes differentially regulated between infected wildtype fat bodies and infected *CrebA* RNAi fat bodies.**

Gene_ID	Gene_name	logFC	PValue
FBgn0263219	Dscam4	7.99361527	9.44E-13
FBgn0264267	CR43767	5.60073366	5.52E-09
FBgn0033395	Cyp4p2	3.92315248	3.18E-07
FBgn0259973	Sfp79B	3.7794489	8.38E-06
FBgn0029681	CG15239	3.73524247	5.12E-06
FBgn0266791	CR45256	3.46817129	0.00050812
FBgn0264273	Sema2b	3.43874687	1.22E-08
FBgn0051832	CG31832	3.22252935	5.09E-19
FBgn0032187	CG4839	3.12057412	0.00047183
FBgn0053530	Acp53C14c	2.88086072	0.00106268
FBgn0050289	CG30289	2.85217816	0.0017243
FBgn0038136	CG8774	2.72661318	0.00713496
FBgn0266826	CR45288	2.64608441	0.0019038
FBgn0028987	Spn28F	2.60097068	0.00284775
FBgn0266785	CR45250	2.58838535	0.0099632
FBgn0264377	CG43829	2.54433377	0.00108534
FBgn0266363	CG45011	2.52916364	0.01275338
FBgn0266764	CR45229	2.52357707	0.01120932
FBgn0259961	CG42471	2.5192592	0.00311642
FBgn0262547	CG43101	2.4625578	0.00726729
FBgn0004414	msopa	2.46171901	0.00369568
FBgn0261853	CG42782	2.37103677	0.00190364
FBgn0264480	CR43888	2.34261206	0.00317983
FBgn0263326	CG43407	2.33176903	0.01621205
FBgn0000094	Anp	2.31252988	0.00421161
FBgn0031800	CG9497	2.30245388	0.00996125
FBgn0250841	CG17242	2.29254481	0.00636804
FBgn0259963	Sfp33A2	2.26801051	0.00610041
FBgn0264513	CR43912	2.24409681	0.01946213
FBgn0259998	CG17571	2.23786758	0.00824333
FBgn0083936	Acp54A1	2.23605103	0.00610086
FBgn0043530	Obp51a	2.23328909	0.00363459
FBgn0053290	CG33290	2.22979935	0.00616952
FBgn0266052	CR44817	2.22563665	0.01999656
FBgn0028532	CG7968	2.2155433	0.01239822
FBgn0032122	CG31883	2.1887618	0.01756501
FBgn0265349	Sfp33A4	2.15824606	0.01563712
FBgn0028415	Met75Cb	2.14272507	0.00081167
FBgn0083965	CG34129	2.12874771	0.00744724
FBgn0036320	CG10943	2.12753318	0.0005072
FBgn0036970	Spn77Bc	2.10941725	0.00340984

FBgn0039783	PH4alphaNE	2.10496927	0.01618653
FBgn0262880	CG43235	2.09289727	0.00016884
FBgn0036969	Spn77Bb	2.07334221	0.00450136
FBgn0265270	CR43239	2.06704103	0.01368412
FBgn0267752	CR46083	2.06183698	0.04787933
FBgn0259965	Sfp35C	2.05775557	0.02577572
FBgn0011559	Acp36DE	2.05510425	0.00890127
FBgn0259964	Sfp33A3	2.04893	0.02426642
FBgn0033145	CG12828	2.04060369	0.01080062
FBgn0013772	Cyp6a8	2.0388519	0.00142528
FBgn0054034	CG34034	2.03449276	0.0044088
FBgn0259970	Sfp70A4	2.02647496	0.03611706
FBgn0033980	Cyp6a20	2.00930526	3.67E-07
FBgn0035670	CG10472	1.99302002	0.00279894
FBgn0011669	Mst57Db	1.98972475	0.01580251
FBgn0263086	CG43354	1.98668404	0.03039716
FBgn0036186	CG6071	1.98568265	0.01382
FBgn0264083	CR43752	1.97450759	0.0311182
FBgn0052054	CG32054	1.95965046	0.00016451
FBgn0267562	CR45902	1.95480005	0.02431557
FBgn0262143	CG42869	1.94358589	0.02837176
FBgn0263410	CR43456	1.93953336	0.03224349
FBgn0038701	CG18493	1.93621344	0.02083677
FBgn0052133	Ptip	1.91158873	0.00586254
FBgn0262099	CG42852	1.90388892	0.02030904
FBgn0262007	CG42825	1.90366693	0.01307084
FBgn0033353	CG13749	1.90045737	4.02E-05
FBgn0015586	Acp76A	1.87475309	0.02740945
FBgn0011670	Mst57Dc	1.86528102	0.03156867
FBgn0267607	CR45945	1.86196504	0.02031229
FBgn0051872	CG31872	1.85945339	0.01004239
FBgn0003965	v	1.85299015	7.35E-09
FBgn0266364	CG45012	1.84788369	0.02506934
FBgn0265264	CG17097	1.84421738	0.03018801
FBgn0028986	Spn38F	1.84101785	0.01679454
FBgn0262999	CG43307	1.83772318	0.03166277
FBgn0261058	Sfp38D	1.82570265	0.03015455
FBgn0027864	Ogg1	1.81965034	0.02171597
FBgn0011668	Mst57Da	1.81337376	0.02438391
FBgn0263625	CR43634	1.80954864	0.03255988
FBgn0050395	CG30395	1.80698342	0.00683531
FBgn0265750	CR44557	1.79732274	0.04736652

FBgn0259960	CG42470	1.79029593	0.01950752
FBgn0267477	CR45827	1.78920268	0.02079538
FBgn0263249	CG43392	1.78134467	0.01516014
FBgn0002855	Acp26Aa	1.77699062	0.02500844
FBgn0039598	aqrs	1.77627275	0.03263824
FBgn0259969	Sfp65A	1.77466449	0.04019545
FBgn0047334	BG642312	1.75376331	0.04791808
FBgn0259954	CG42464	1.7520542	0.04894757
FBgn0261061	Sfp96F	1.75194341	0.03623002
FBgn0266365	CR45013	1.74827684	0.03416785
FBgn0262623	CG43147	1.74496289	0.03224302
FBgn0264330	CG43789	1.74173063	0.04312089
FBgn0085325	CG34296	1.73142667	0.00182375
FBgn0264726	CR43994	1.71921442	0.04513472
FBgn0040959	Peritrophin-1	1.71079854	0.04138477
FBgn0032505	CG16826	1.70890008	0.04660038
FBgn0267543	CR45883	1.70790266	0.04162866
FBgn0002863	Acp95EF	1.70710541	0.03378234
FBgn0024997	CG2681	1.70047528	0.01929836
FBgn0259956	Sfp24C1	1.69563733	0.03948303
FBgn0262814	CG43185	1.69530337	0.031541
FBgn0265538	CG44388	1.6897293	0.02665867
FBgn0265788	CR44577	1.68723076	0.03371341
FBgn0020509	Acp62F	1.6849548	0.03880673
FBgn0262365	CG43063	1.68031871	0.02754795
FBgn0001112	Gld	1.66666461	0.00082037
FBgn0034195	Spn53F	1.65886506	0.02642401
FBgn0052382	sphinx2	1.64861938	0.04352826
FBgn0037004	CG17637	1.64674025	0.01245782
FBgn0263597	Acp98AB	1.64529859	0.03907647
FBgn0265087	CR44198	1.63560916	0.04682007
FBgn0262580	CG43120	1.63090446	0.00775615
FBgn0267347	squ	1.61641227	0.0004929
FBgn0038919	CG17843	1.6044584	0.04653589
FBgn0265541	CR44391	1.58770004	0.00470576
FBgn0036154	CG6168	1.58377073	0.02827306
FBgn0261060	Sfp93F	1.56683635	0.04167311
FBgn0051021	CG31021	1.56345816	0.03813632
FBgn0264705	CR43974	1.5273898	0.03536743
FBgn0043825	CG18284	1.52507554	0.04223911
FBgn0036285	toe	1.51825957	0.02136438
FBgn0265267	CG18258	1.51788875	0.04825228

FBgn0015584	Acp53Ea	1.51056925	0.03867817
FBgn0083966	CG34130	1.50156437	0.0434997
FBgn0039521	CG5402	1.48657205	0.01213887
FBgn0051901	Mur29B	1.48290658	0.04970717
FBgn0051148	Gba1a	1.48178401	0.04649538
FBgn0033978	Cyp6a23	1.45303274	0.00021668
FBgn0033271	CG8708	1.44443873	0.02216619
FBgn0034605	CG15661	1.42416978	0.00032753
FBgn0054051	CG34051	1.42190892	0.04218579
FBgn0259972	Sfp77F	1.41474314	0.03427898
FBgn0033110	CG9447	1.40703755	0.03056373
FBgn0261632	CR42715	1.39682488	0.0275977
FBgn0051198	CG31198	1.38877502	0.02351657
FBgn0265268	CG18234	1.34628154	0.04625014
FBgn0250835	CG15394	1.3284032	0.01112293
FBgn0264344	CG43800	1.31180565	0.01365352
FBgn0034225	veil	1.30259425	0.01246367
FBgn0013771	Cyp6a9	1.29458631	0.00137802
FBgn0033657	Sln	1.27196405	0.00540754
FBgn0025682	scf	1.25541637	0.00035495
FBgn0050104	NT5E-2	1.23012107	0.00014125
FBgn0050083	CG30083	1.2054618	0.01107287
FBgn0031549	Spindly	1.16746269	0.04738924
FBgn0053319	CR33319	1.09833431	0.03409399
FBgn0034883	Eglp2	1.08904171	0.01100601
FBgn0031484	CG3165	1.07720081	0.01078613
FBgn0052373	CG32373	1.04467162	0.02930071
FBgn0014877	Roe1	1.03319563	0.00216824
FBgn0027507	CG1344	0.97717912	0.00568354
FBgn0040364	CG11378	0.94101992	0.00266848
FBgn0037090	Est-Q	0.94083607	0.01427731
FBgn0027070	CG17322	0.90375558	0.01902833
FBgn0035083	Tina-1	0.9001793	0.007735
FBgn0004428	LysE	0.8916926	0.03574229
FBgn0035983	CG4080	0.84826419	0.0170651
FBgn0263019	CR43314	0.81301681	0.04349902
FBgn0033710	CG17739	0.8112402	0.0090497
FBgn0000053	ade3	0.80423493	0.01593956
FBgn0266000	CG44774	0.80137569	0.0070228
FBgn0267408	AOX1	0.74102428	0.04157973
FBgn0261068	LSm7	0.73831994	0.02443489
FBgn0037140	SLC22A	0.72933982	0.03866343

FBgn0030828	CG5162	0.71911343	0.01968358
FBgn0051704	CG31704	0.65523935	0.03897263
FBgn0027348	bgm	0.64967483	0.03957772
FBgn0036945	Ssk	0.6467746	0.04661582
FBgn0034796	CG3700	0.64302518	0.03556398
FBgn0030060	CG2004	0.62498142	0.03322182
FBgn0039798	CG11313	-10.725224	2.45E-59
FBgn0050091	CG30091	-5.3724481	7.88E-19
FBgn0267780	CR46111	-2.9803512	0.00192536
FBgn0053926	CG33926	-2.7614848	1.75E-16
FBgn0085312	CG34283	-2.641471	0.0005549
FBgn0043791	phu	-2.2407856	0.03419548
FBgn0053301	CG33301	-1.9728586	0.01249276
FBgn0033302	Cyp6a14	-1.949839	0.01350904
FBgn0266780	CR45245	-1.8848654	0.00627935
FBgn0000281	Cec2	-1.8715493	0.00587289
FBgn0040340	TRAM	-1.7328284	1.38E-05
FBgn0040096	lectin-33A	-1.6466774	0.00760249
FBgn0032382	Mal-B2	-1.5835733	0.00023698
FBgn0044011	Spn43Ad	-1.5784571	0.01947329
FBgn0033742	CG8550	-1.5660656	0.00010335
FBgn0037396	CG11459	-1.5097897	0.02411658
FBgn0030305	Uba5	-1.4670665	3.50E-05
FBgn0050098	CG30098	-1.4518477	0.00092196
FBgn0033691	CG8860	-1.3121796	2.15E-05
FBgn0004396	CrebA	-1.3048033	0.00014486
FBgn0031974	CG12560	-1.2842987	0.01260504
FBgn0267348	LanB2	-1.2765688	0.01838718
FBgn0262801	twr	-1.2029658	9.56E-05
FBgn0031260	Spp	-1.1995563	9.89E-05
FBgn0027578	CG14526	-1.1697711	0.00042811
FBgn0000280	Cec-Psi1	-1.1577651	0.04984833
FBgn0032603	CG17928	-1.1515815	0.00016736
FBgn0035121	Tudor-SN	-1.1379836	0.00043833
FBgn0028425	JhI-21	-1.107097	0.03440936
FBgn0000278	CecB	-1.1052833	0.00572199
FBgn0051205	CG31205	-1.0952852	0.00363779
FBgn0262003	CG42821	-1.0933058	0.03131128
FBgn0000279	CecC	-1.0665644	0.00586704
FBgn0000276	CecA1	-1.0655582	0.00037077
FBgn0047135	CG32276	-1.061416	0.00187247
FBgn0037612	CG8112	-1.0593977	0.01353603

FBgn0027584	CG4757	-1.0404357	0.01514874
FBgn0262574	CG43114	-1.0384957	0.00247177
FBgn0035401	CG1291	-1.0282816	0.00286542
FBgn0030465	CG15743	-1.0001639	0.01194242
FBgn0052282	Drsl4	-0.9977753	0.00120346
FBgn0025700	CG5885	-0.9881127	0.00319958
FBgn0010391	Gtp-bp	-0.9805509	0.0012244
FBgn0036919	Grasp65	-0.9721802	0.0041467
FBgn0259238	CG42336	-0.9642657	0.00334445
FBgn0034247	CG6484	-0.9614627	0.00325837
FBgn0034870	CG13559	-0.9586236	0.04787821
FBgn0262563	CG43103	-0.9558805	0.00207527
FBgn0040582	CG5791	-0.904621	0.00194576
FBgn0032283	CG7296	-0.9039512	0.0057235
FBgn0061492	loj	-0.9018181	0.0028757
FBgn0000277	CecA2	-0.8977052	0.00263845
FBgn0032635	CG15141	-0.8932044	0.03932665
FBgn0035089	Phk-3	-0.887065	0.0238418
FBgn0002868	MtnA	-0.8838031	0.0020876
FBgn0085320	CG34291	-0.8687764	0.03547258
FBgn0033170	sPLA2	-0.8645498	0.02687871
FBgn0037875	ZnT86D	-0.8621191	0.00783166
FBgn0031373	CG15358	-0.8604612	0.00669744
FBgn0020655	ArfGAP1	-0.8597089	0.01413026
FBgn0030306	Spase25	-0.858964	0.00676121
FBgn0031558	CG16704	-0.8539217	0.01577015
FBgn0035206	CG9186	-0.853033	0.00667117
FBgn0015714	Cyp6a17	-0.8508776	0.03403179
FBgn0038810	Srp72	-0.8422575	0.01121997
FBgn0085453	CG34424	-0.8378751	0.00681615
FBgn0085220	Ufm1	-0.8332888	0.01096128
FBgn0053120	CG33120	-0.8310345	0.02497035
FBgn0038829	CG17271	-0.8199988	0.00703744
FBgn0283461	Drs	-0.8190181	0.04298454
FBgn0034733	CG4752	-0.8149853	0.01768145
FBgn0029167	Hml	-0.8118492	0.02492457
FBgn0014469	Cyp4e2	-0.804871	0.02365659
FBgn0086347	Myo31DF	-0.798515	0.02991197
FBgn0010638	Sec61beta	-0.7964556	0.01265049
FBgn0022359	Sodh-2	-0.7946655	0.01908288
FBgn0045866	bai	-0.7687571	0.00856612
FBgn0034061	Ufc1	-0.7591955	0.02091232

FBgn0015372	RabX1	-0.7563842	0.02021024
FBgn0263200	Galt	-0.7493597	0.01626709
FBgn0033339	Sec31	-0.728818	0.0212302
FBgn0030990	CG7556	-0.7280977	0.02425194
FBgn0040260	Ugt36Bc	-0.7260759	0.04752511
FBgn0043783	CG32444	-0.7164299	0.03260969
FBgn0035771	Sec63	-0.7155946	0.02141891
FBgn0011509	SrpRbeta	-0.7151056	0.03918473
FBgn0051777	CG31777	-0.7101747	0.02781726
FBgn0034354	GstE11	-0.7065813	0.02894605
FBgn0021795	Tapdelta	-0.7003481	0.02612695
FBgn0032322	CG16743	-0.69775	0.02136475
FBgn0015298	Srp19	-0.6967046	0.03281301
FBgn0027835	Dp1	-0.6927219	0.02351938
FBgn0034511	GNBP-like3	-0.6899137	0.02842507
FBgn0067905	IM14	-0.6846708	0.02609561
FBgn0040623	Spase12	-0.6839987	0.04793993
FBgn0283427	FASN1	-0.6671901	0.02596683
FBgn0016675	Lectin-galC1	-0.6580763	0.04256131
FBgn0035909	ergic53	-0.6576962	0.03597916
FBgn0263083	CG43351	-0.6567207	0.02509999
FBgn0012042	AttA	-0.6210145	0.04539943
FBgn0031305	Iris	-0.6095503	0.04077
FBgn0001114	Glt	-0.6011398	0.04640275
FBgn0283509	Phm	-0.5977902	0.04815624
FBgn0050026	CG30026	-0.5952026	0.04901786

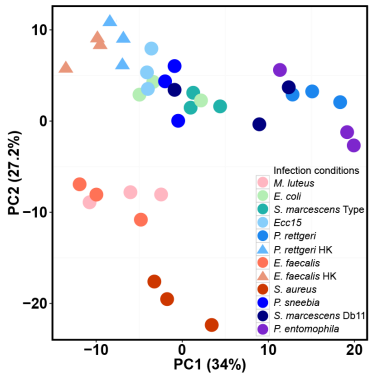
Table 3.6: **List of primers used in RT-qPCR experiments.**

Gene	Primer Sequences	
	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
AttC	TGCCCGATTGGACCTAAGC	GCGTATGGGTTTTGGTCAGTTC
CrebA	GACGGAGCACTCCTACAGTCT	GAAATGGCGGGAAAGCACTC
Def	GTTCTTCGTTCTCGTGG	CTTTGAACCCCTTGGC
Dpt	GCTGCGCAATCGCTTCTACT	TGGTGGAGTGGGCTTCATG
Drs	CGTGAGAACCTTTTCCAATATGATG	TCCCAGGACCACCAGCAT
ergic53	CTTCTGGTACACCACGGAAAAG	TCGAAGGAATCGAACATGATGG
RpL32	GACGCTTCAAGGGACAGTATCTG	AAACGCGGTTCTGCATGAG
Sec61beta	GGTGGCATGTGGCGTTTCTA	GAAGACGGAAGCGATGAAGAG
Spase25	AAGGATGAAAAGTCACAGCAAGG	CCAAGCATGGGCCATTATAGC
TRAM	AAGAACCCGCCTATCCTGAG	GATGTAGGTGTATGGCTTGCC
Xbp1s	CCGAACTGAAGCAGCAACAGC	GTATACCCTGCGGCAGATCC
Xbp1t	TCTAACCTGGGAGGAGAAAG	GTCCAGCTTGTGGTTCTTG

A

Bacteria	Gram	PGN	Isolated from *Dmel*	Survival at 7 days (%)	LT_{50} (hours) (95%CI)	Virulence level
E. coli	-	DAP-type	No	98.3	N/A	*
M. luteus	+	Lys-type	No	96.7	N/A	*
S. marcescens Type	-	DAP-type	No	94.9	N/A	*
Ecc15	-	DAP-type	No	85.0	N/A	**
P. rettgeri	-	DAP-type	Yes	48.3	N/A	**
E. faecalis	+	Lys-type	Yes	45.0	N/A	**
S. aureus	+	Lys-type	No	0.0	31.3 (27.2 – 35.4)	***
P. sneebia	-	DAP-type	Yes	0.0	31.2 (27.8 – 34.7)	***
S. marcescens Db11	-	DAP-type	Yes	0.0	13.4 (12.8 – 14.8)	***
P. entomophila	-	DAP-type	Yes	0.0	13.7 (12.7 – 15.8)	***

B



C

PC	Gene name	Function	SW	Pr HK	Ef HK	MI	Ef	Sa	Ecc	Ec	Ps	Sm	Db11	Pr	Pe
PC1	PGRP-LC	Imd signaling	1.1	1.4	1.3	1.3	1.2	1.3	1.6	1.7	1.9	1.9	2.5	4.4	4.9
PC1	Rel	Imd signaling	1.2	1.3	1.1	1.8	1.8	4.5	1.5	1.5	2.2	2.4	4.5	7.3	8.1
PC1	PGRP-LB	Imd signaling	1.3	1.3	1.1	1.4	1.3	2.5	2.1	2.8	4.3	4.3	8.3	15.9	13.9
PC1	PGRP-SD	Imd signaling	4.2	6.1	4.4	7.2	4.3	6.5	9.1	8.7	9.4	11.8	11.7	20.2	27.5
PC1	AttA	AMP	7.2	10.7	6.5	29.7	16.7	51.2	24.0	35.5	42.7	61.3	65.2	100	100
PC1	Dpt	AMP	35.3	85.1	40.0	52.2	23.3	23.2	100	100	100	100	100	100	100
PC1	Drs	AMP	9.5	11.9	7.5	54.6	42.6	65.6	16.5	27.3	14.9	26.1	20.2	29.9	40.9
PC1	Mtk	AMP	29.0	66.7	34.9	69.5	50.1	72.6	85.5	91.7	69.9	100	92.9	100	100
PC1	Spz	Toll signaling	1.1	1.2	1.3	1.4	1.3	1.7	1.3	1.5	1.4	1.6	1.7	2.9	4.4
PC1	Spirit	Toll signaling	1.0	1.3	1.1	1.7	1.4	2.6	1.1	1.1	-1.0	1.5	1.6	2.9	4.5
PC1	PGRP-SA	Toll signaling	2.1	2.3	1.8	2.9	2.0	3.3	2.7	3.7	2.4	4.3	3.6	4.8	7.6
PC1	Spn88Eb	Toll signaling	1.8	2.2	1.9	4.6	6.0	13.0	1.8	2.9	2.8	3.3	4.5	5.7	6.7
PC1	TotA	stress peptide (JAK)	24.4	72.1	20.6	10.9	4.1	8.9	13.2	43.3	8.3	18.5	18.8	23.7	20.5
PC1	TotC	stress peptide (JAK)	52.2	100	35.3	13.7	4.3	10.6	19.9	69.0	11.6	28.2	33.9	40.6	28.8
PC1	TotM	stress peptide (JAK)	79.8	100	35.2	16.7	6.7	15.3	19.2	99.3	14.0	45.3	35.6	65.8	59.4
PC1	ACC	acetyl-coA-carboxylase	-1.0	-1.0	-1.0	-1.2	-1.2	-1.8	1.2	1.0	-1.2	-1.1	-1.4	-1.9	-2.0
PC1	UGP	glycogen synthesis	-1.2	-1.0	1.0	-1.8	-1.3	-2.5	-1.1	-1.1	-1.3	-1.3	-1.9	-2.1	-2.9
PC1	Galk	galactokinase	-1.3	-1.3	-1.0	-1.3	-1.3	-2.3	-1.3	-1.3	-1.3	-1.2	-2.2	-2.4	-2.8
PC1	Tps1	trehalose 6P synthase	-1.1	-1.1	-1.2	-1.7	-1.9	-2.6	-1.1	-1.5	-1.3	-1.3	-2.0	-3.3	-3.6
PC1	Hex-C	hexokinase	-1.2	-1.4	-1.3	-1.4	-1.6	-2.7	-1.2	-1.3	-1.6	-1.5	-2.2	-2.9	-3.7
PC1	FASN1	Fatty acid synthase	1.0	1.0	1.2	-1.8	-1.4	-2.5	1.1	-1.2	-1.1	-1.4	-2.0	-3.5	-3.3
		Fold regulation vs unchallenged							< -4	< -2	NS	> 2	> 4	> 8	> 12

PC	Gene name	Function	SW	Pr HK	Ef HK	MI	Ef	Sa	Ecc	Ec	Ps	Sm	Db11	Pr	Pe
PC2	Sp7	Toll signaling	-1.2	-1.1	-1.8	1.2	-1.0	2.0	1.1	-1.2	-1.6	-1.1	1.4	1.6	1.8
PC2	cactus	Toll signaling	1.1	1.2	1.0	1.6	1.8	3.8	1.3	1.2	1.1	1.2	1.3	1.4	1.3
PC2	nec	Toll signaling	1.4	1.5	1.7	2.8	3.4	6.2	1.4	1.5	1.3	1.3	1.6	1.2	1.3
PC2	SPE	Toll signaling	1.3	1.2	1.2	2.5	2.8	3.7	1.5	1.5	1.3	1.5	1.3	1.3	1.2
PC2	fon	coagulation	1.9	1.6	1.8	2.8	2.9	2.5	2.0	2.2	1.8	1.8	1.7	-1.1	1.2
PC2	MP1	melanization	1.6	1.0	-1.1	1.5	2.6	4.9	1.5	1.3	1.4	1.4	-1.2	1.7	1.9
PC1/2	IM2	AMP	3.4	2.1	2.7	12.1	12.1	15.6	2.2	4.1	3.4	3.0	2.8	1.3	2.2
PC1/2	IM3	AMP	3.4	2.3	2.8	14.0	13.9	19.2	2.4	4.5	3.5	3.4	2.9	1.5	2.7
PC1/2	IM1	AMP	4.7	3.5	3.9	33.3	29.9	39.8	4.8	9.6	5.4	6.3	5.2	3.5	6.2
PC1/2	IM23	AMP	4.9	3.5	4.1	46.6	39.0	61.1	5.1	8.7	5.7	6.0	6.1	3.4	6.7
PC2	nimB3	phago	1.7	1.9	1.5	4.8	3.8	5.3	2.3	2.7	2.0	2.3	2.1	1.9	3.2
PC2	Mal-A1	Digestion	-1.3	-1.3	-1.1	-1.5	-1.8	-3.3	-1.3	-1.5	-1.2	-1.1	-2.2	-1.2	-1.2
PC2	Mal-A8	Digestion	-1.4	-1.3	-1.2	-1.4	-1.9	-3.5	-1.3	-1.7	-1.2	-1.5	-2.1	-1.4	-1.2
PC2	Mal-A7	Digestion	-1.6	-1.3	-1.1	-1.6	-1.8	-3.6	-1.3	-1.6	-1.2	-1.4	-2.0	-1.4	-1.4
PC2	Mal-A4	Digestion	-1.3	-1.3	-1.4	-2.1	-2.2	-4.0	-1.4	-1.9	-1.2	-1.6	-1.4	-1.2	-1.1
PC2	Gs1	AA biosynthesis	1.3	1.4	1.4	1.0	-1.0	-1.7	1.6	1.3	1.5	1.5	1.2	-1.1	-1.1
PC2	Spat	AA biosynthesis	-1.2	-1.2	-1.2	-2.0	-2.0	-3.9	-1.1	-1.4	-1.2	-1.5	-2.0	-2.7	-3.3
PC2	Cyp6w1	P450	1.7	1.3	1.5	-2.5	-1.3	-2.3	1.4	-1.2	1.7	-1.4	1.8	1.2	-1.2
PC2	Cyp6g1	P450	-1.3	-1.2	-1.1	-2.0	-1.9	-3.6	-1.3	-1.7	-1.5	-1.9	-2.3	-2.1	-2.4
PC2	Cyp4ac1	P450	1.0	1.0	1.0	-1.6	-1.5	-3.4	1.2	-1.1	-1.2	-1.1	-1.2	-1.1	-1.1
PC2	Cyp309a1	P450	-1.3	-1.1	-1.3	-1.9	-2.9	-6.1	1.1	1.0	1.2	1.0	1.0	1.1	1.3

Figure 3.1: **Major parameters influencing the global response to infection.**

(A) List of bacteria used in the RNA-seq experiment, including Gram classification, type of bacterial peptidoglycan (PGN), source of isolate, percent survival at 7 days post-infection, median lethal time (LT_{50}) for each bacterium, and assignment into broad virulence categories. (B) PCA plot showing the first two principal components of the 12 h dataset. Red and orange (warm) colors indicate infections with Lys-type PGN bacteria, while green, blue and purple (cool) colors indicate infections with DAP-type PGN bacteria. Circles indicate infection with live bacteria and triangles denote inoculation with heat-killed bacteria. (C) Genes that contribute the most to PC1 (left column), PC2 (right column), or both PCs (right column) are presented with their associated level of expression change (fold change) at 12 h post-infection. Warm colors indicate the degree of transcriptional induction, while cool colors show the extent of transcriptional downregulation.

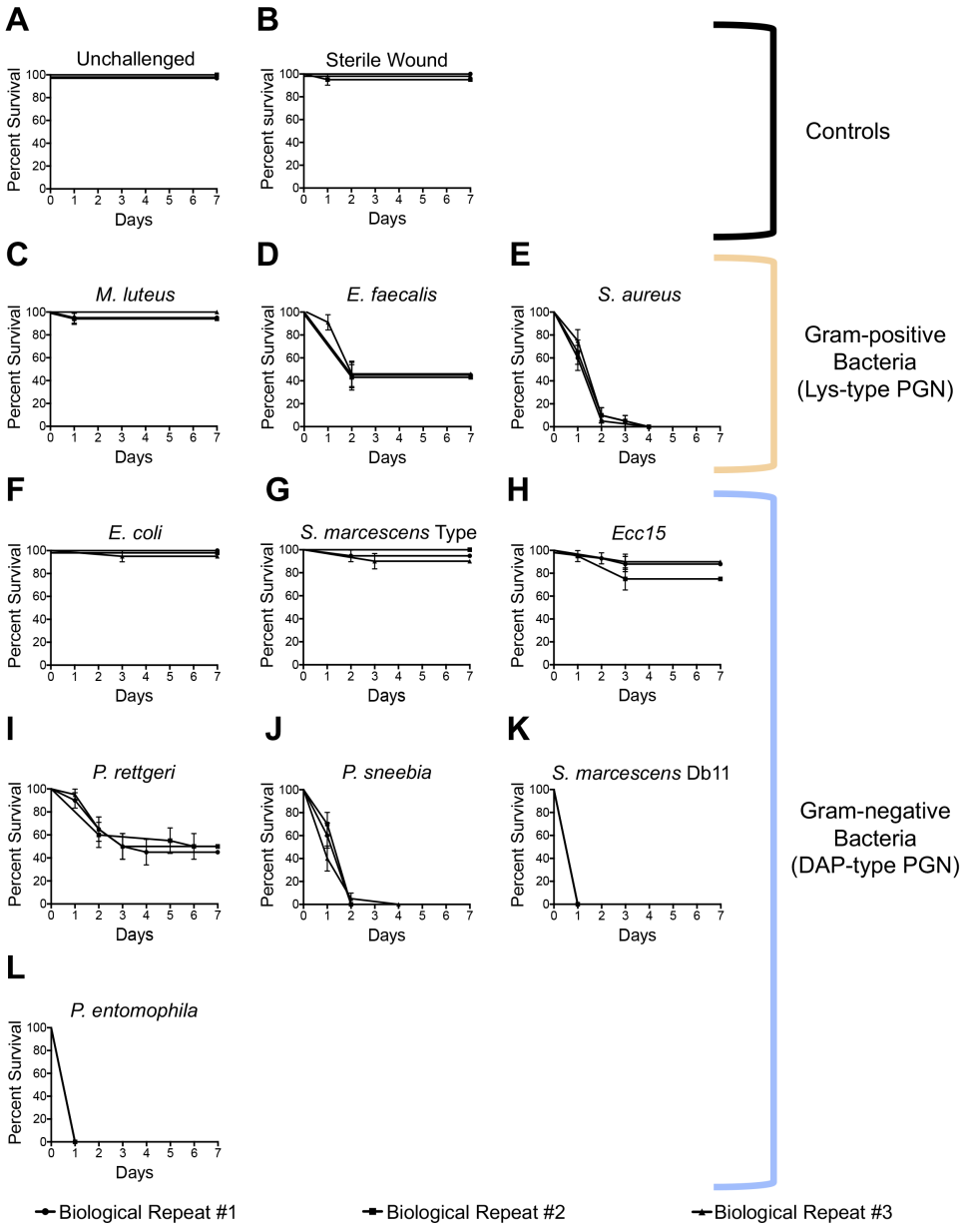


Figure 3.2: **10 bacteria cause different mortalities in *Drosophila*.**

Survival curves (in %) over time of control and infected Canton S flies. Three biological replicates are graphed independently for each condition. Treatments are as follows: (A) Unchallenged. (B) Sterile wound. (C) *Micrococcus luteus*.

(D) *Enterococcus faecalis*. (E) *Staphylococcus aureus*. (F) *Escherichia coli*. (G) *Serratia marcescens* Type strain. (H) *Pectinobacterium* (formerly *Erwinia*) *carotovora Ecc15*. (I) *Providencia rettgeri*. (J) *Providencia sneebia*. (K) *Serratia marcescens* strain Db11. (L) *Pseudomonas entomophila*

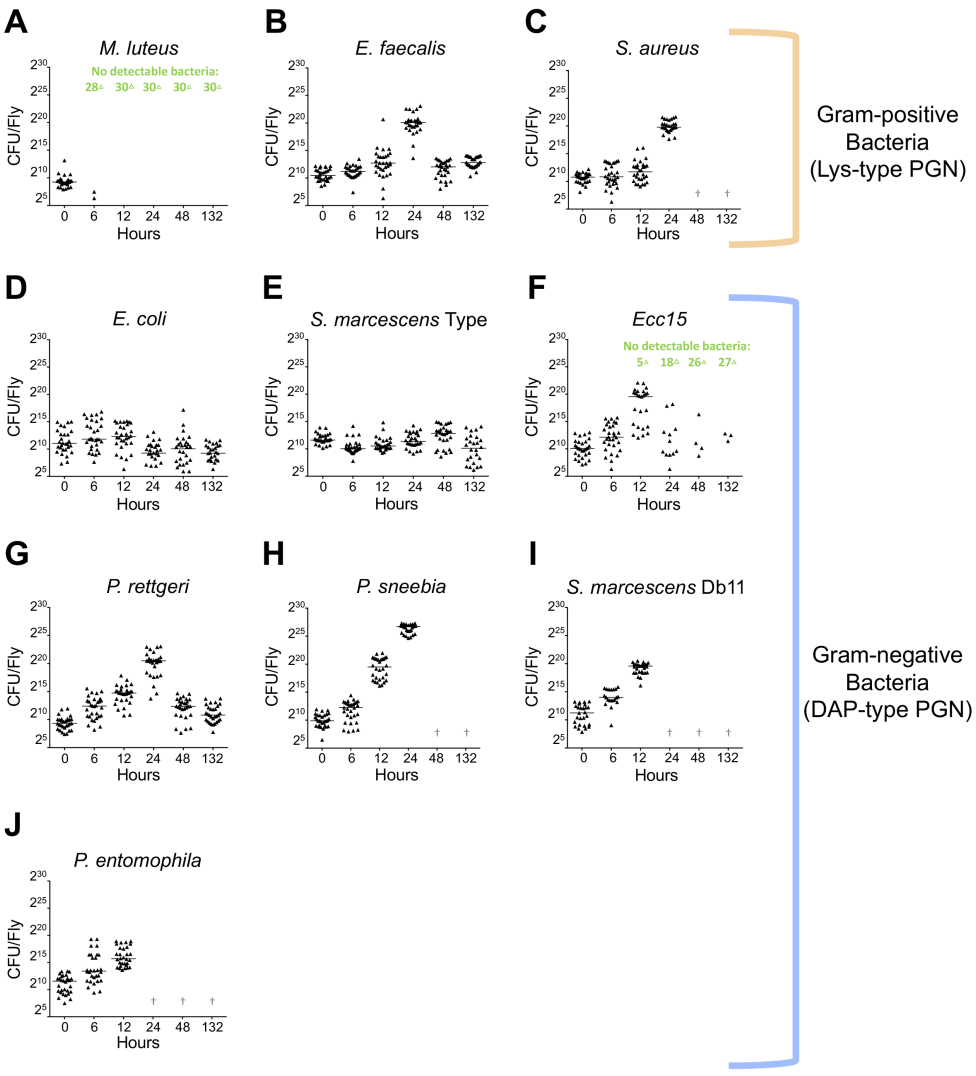


Figure 3.3: **10 bacteria differ in their rate of growth in the fly.**

Bacterial load time courses of infected Canton S flies over 132 h following infection. Three biological repeats are graphed together, with each triangle representing the bacterial burden in an individually sampled fly.

(A) *M. luteus*. (B) *E. faecalis*. (C) *S. aureus*. (D) *E. coli*. (E) *S. marcescens* Type. (F) *Ecc15*. (G) *P. rettgeri*. (H) *P. sneebia*. (I) *S. marcescens* Db11. (J) *P. entomophila*. The symbol †denotes no flies were sampled because most, if not all, flies had succumbed by that time point. A number followed by the symbol Δ indicates the number of flies found to have no bacteria (flies that carry undetectable levels of bacteria or that have cleared the infection) at the specified time point.

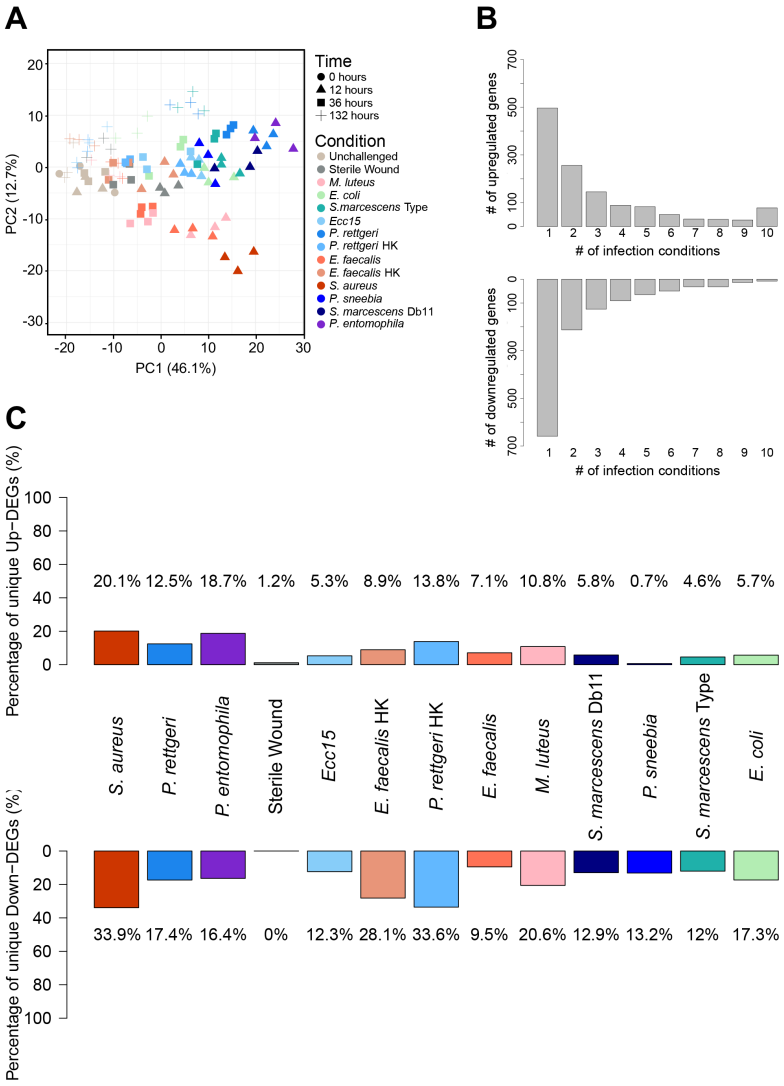


Figure 3.4: **Each bacterial infection induces a unique host response.**

(A) PCA plot showing the first two principal components of the entire dataset. Pink, orange, and red (warm) colors show infections with Gram-positive (Lys-type PGN) bacteria, while green, blue, and purple (cool) colors denote infections with Gram-negative (DAP-type PGN) bacteria. HK indicates stimulation with heat-killed bacteria.

(B) Histogram of differentially upregulated (top) or downregulated (bottom) genes by the number of infection conditions in which a given gene was differentially expressed. (C) Percentage of genes that are uniquely upregulated (top) or downregulated (bottom) by each infection.

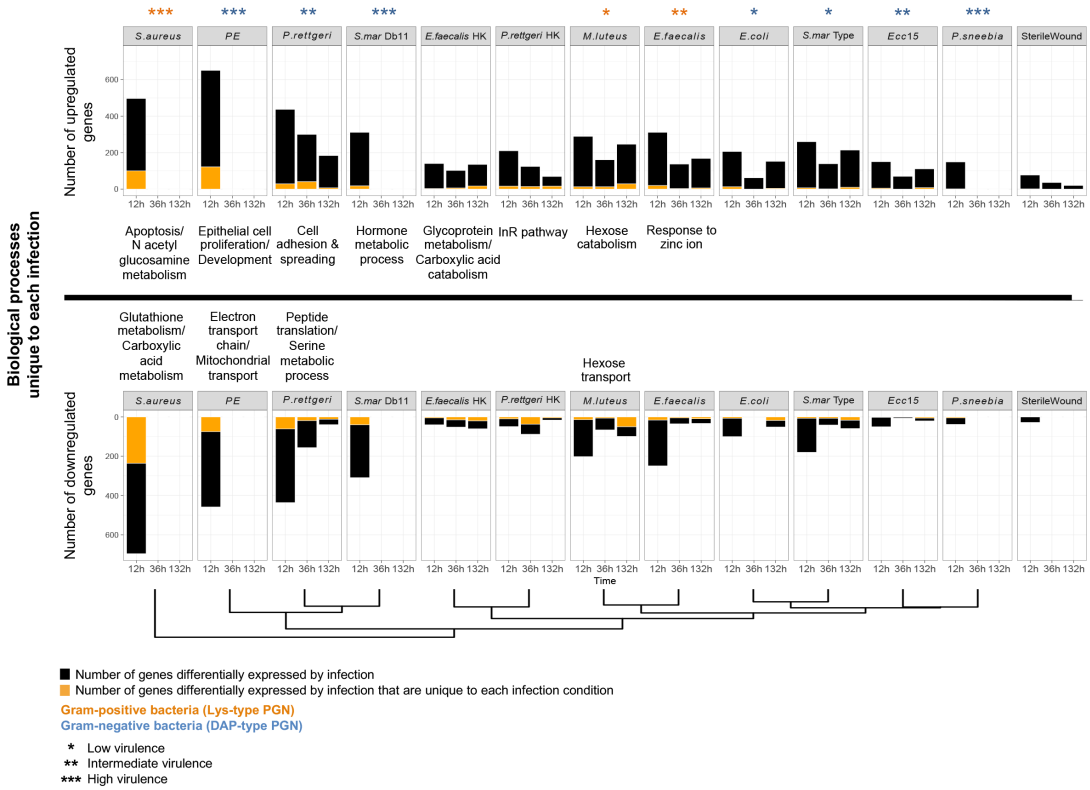


Figure 3.5: **The host response to bacterial infection is diverse.**

Number of genes differentially regulated by each infection (black bars) and number of genes differentially expressed by infection that are unique to each individual bacterium (orange bars). This information is listed separately by time point for each infection. Upregulated and downregulated genes are above and below the horizontal demarcation line, respectively. The biological processes regulated exclusively in response to each individual bacterium, if any, are listed adjacent to each bacterium. These biological processes are also separated by the horizontal demarcation line depending on whether they are upregulated (above) or downregulated (below) by each infection. The relative level of virulence of each bacterium is indicated by the number of stars: *Low virulence, **Intermediate virulence, ***High virulence. The type of bacterial peptidoglycan is indicated by the color of the stars: orange (bacteria with Lys-type PGN) and blue (bacteria with DAP-type PGN). The clustering of infection conditions (shown at the bottom of the graph) is based on the similarities of the expression patterns measured at 12 h.

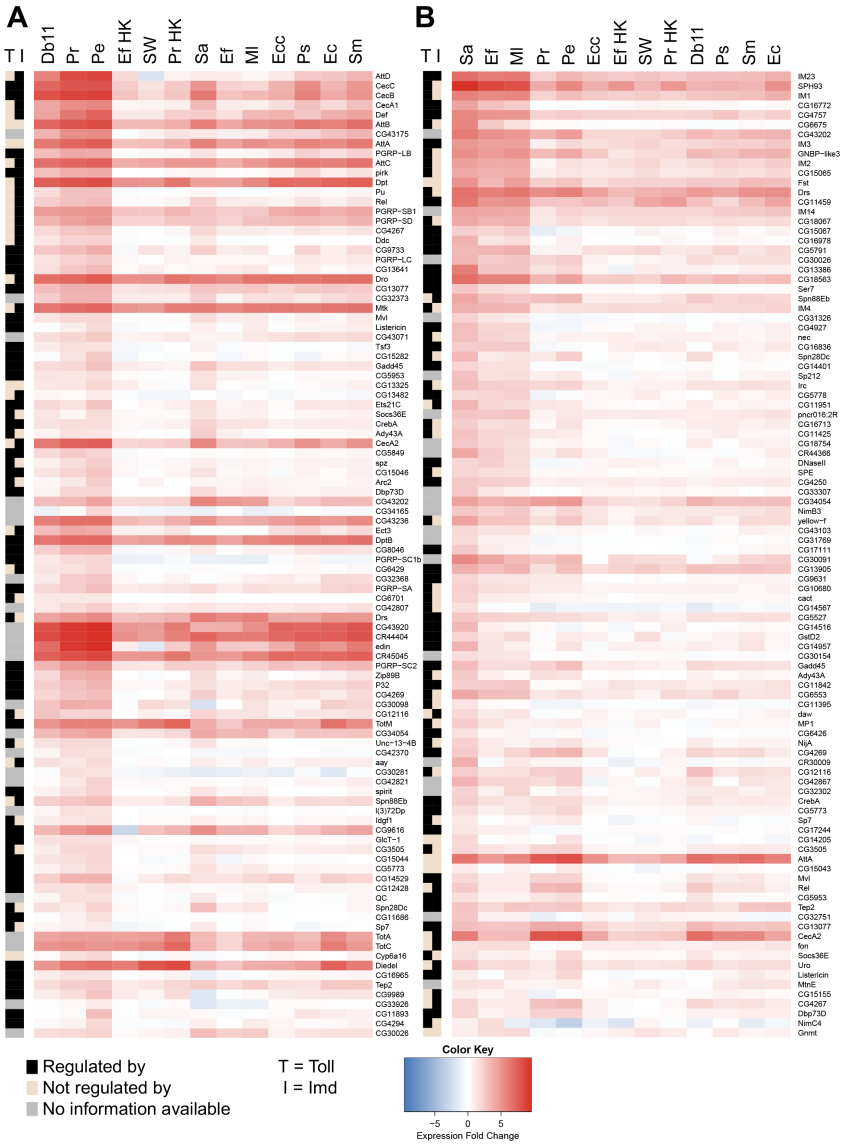


Figure 3.6: **Individual infections differ in their ability to induce the Toll and Imd pathways and reshape host metabolism.**

Heatmap (log2 fold change) of top 100 genes that contribute the most to PC1 (A) and of top 100 genes that most contribute to PC2 (B). A gene was deemed to be regulated by the Toll or Imd pathways if the absence of the key genes in each pathway (*spz* for Toll and *Rel* for Imd) changed the expression level of said gene by 20% or more compared to the expression level of the gene in wildtype, as previously reported in [112].

A color scale on the left side of each heatmap indicates whether each gene is regulated by Toll (T) or Imd (I). In the first column (T), genes regulated by the Toll pathway are marked in black, while genes not regulated by Toll are marked in beige. Similarly, in the second column (I), genes regulated by Imd are marked in black, while genes not regulated by Imd are marked in beige.

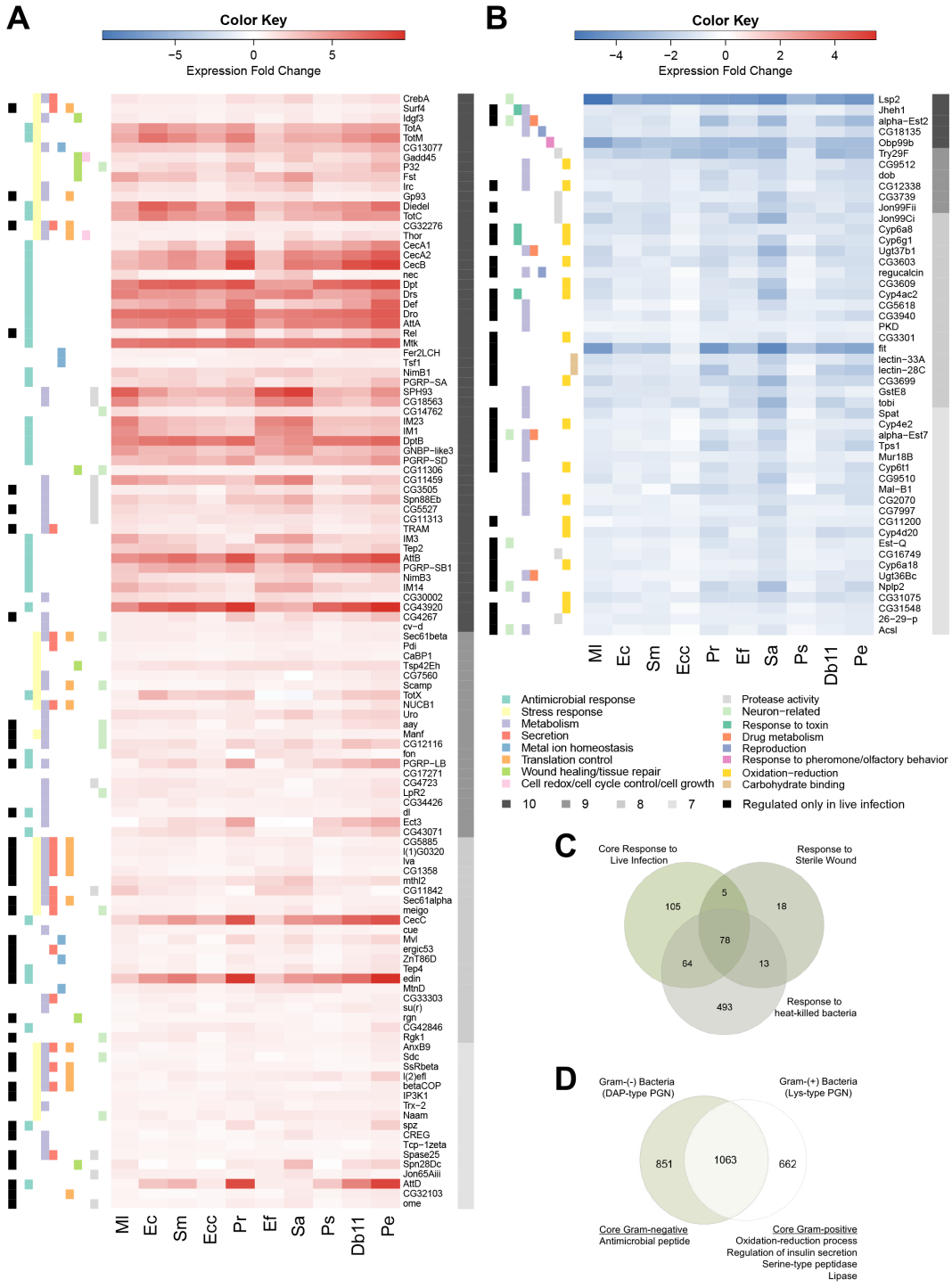


Figure 3.7: **Systemic infection triggers a core host response.**

Heatmap showing the expression level (log2 fold change) of selected core upregulated (A) and downregulated (B) genes for all 10 bacteria.

Core genes are differentially expressed in response to infection by 7 or more bacteria. A gray scale on the right side of each heatmap indicates the number of bacterial infections that significantly change the expression of a given gene (dark gray = 10, medium gray = 9, light gray = 8, and very light gray = 7). A color scale on the left side of each heatmap denotes the functional categories that each gene belongs to, and the legend for each color is listed at the bottom of the graph. (C) Venn diagram showing the intersection of core genes, sterile wound genes, and genes differentially regulated by challenge with heat-killed bacteria. (D) Venn diagram indicating the number of genes differentially regulated only in response to infection with Lys-type peptidoglycan (PGN) bacteria (right), DAP-type PGN bacteria (left), and those genes differentially expressed by challenge with both types of bacteria. GO terms associated with genes exclusively regulated by infection with Lys-type PGN bacteria or DAP-type PGN bacteria are listed.

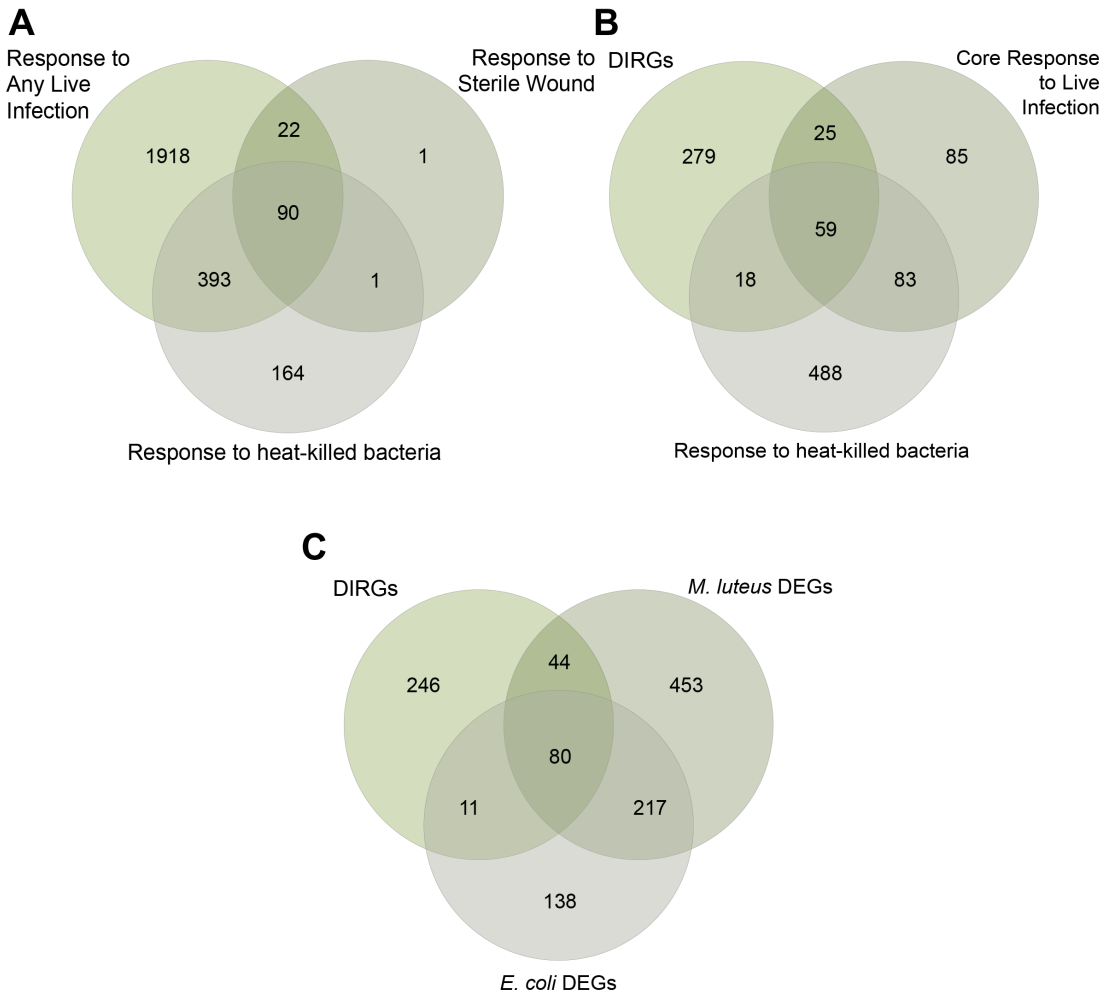


Figure 3.8: **Overlap between genes responding to live infection, sterile wound, heat-killed bacteria, and the DIRGs.**

(A) Venn diagram showing the intersection between genes that are differentially regulated in response to at least one live infection, sterile wound, and challenge with heat-killed bacteria. (B) Venn diagram depicting the overlap between the previously described *Drosophila* Immune-Regulated Genes (DIRGs) [6], core genes differentially regulated in response to live infection, and genes differentially expressed in response to heat-killed bacteria. (C) Venn diagram illustrating the overlay between genes differentially regulated in response to *M. luteus* infection and, separately, *E. coli* infection in the present study and the DIRGs, which were previously identified from infection with a mixed cocktail of *E. coli* and *M. luteus* [11].

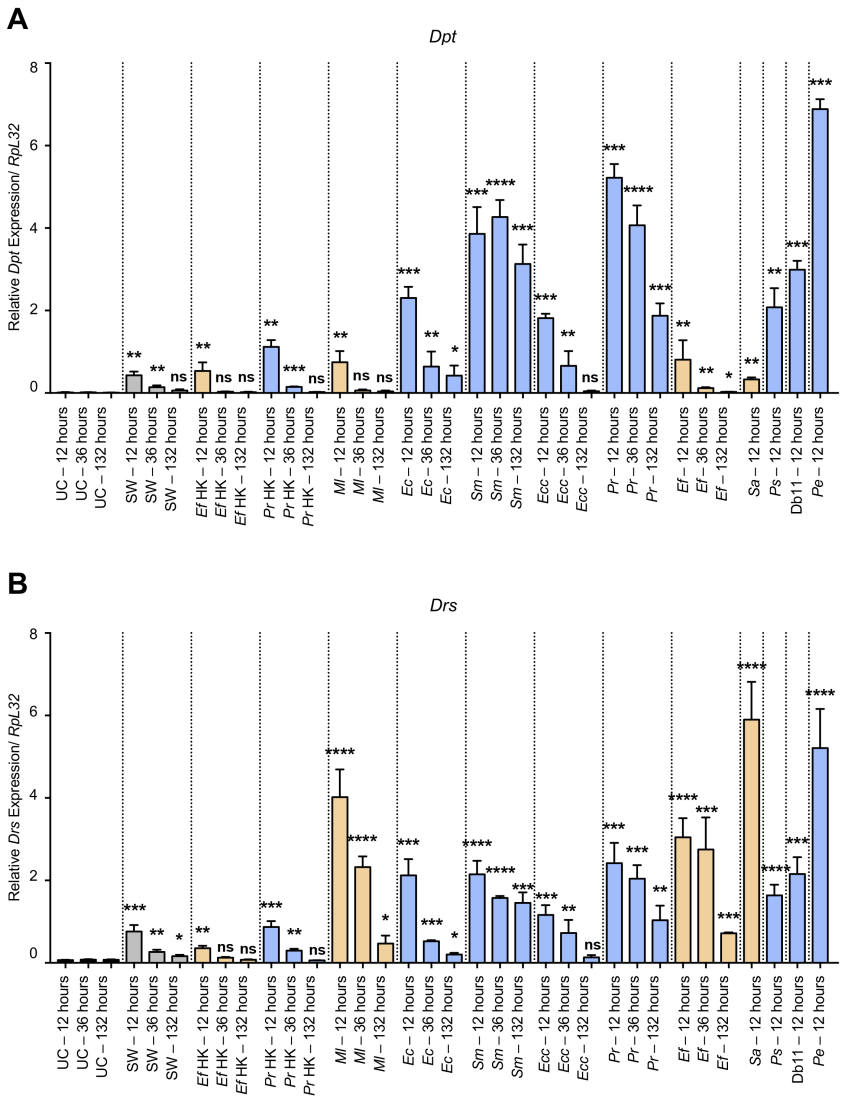


Figure 3.9: **10 bacteria induce different expression levels of antimicrobial peptide genes.**

RT-qPCR measuring (A) *Diptericin* and (B) *Drosomycin* expression levels in control and infected Canton S flies at 12, 36, and 132 h post-infection. These samples are separate biological replicates, distinct from those used in the RNA-seq experiment. Mean values of three biological repeats are represented ±SE. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001 in a Student's t-test.

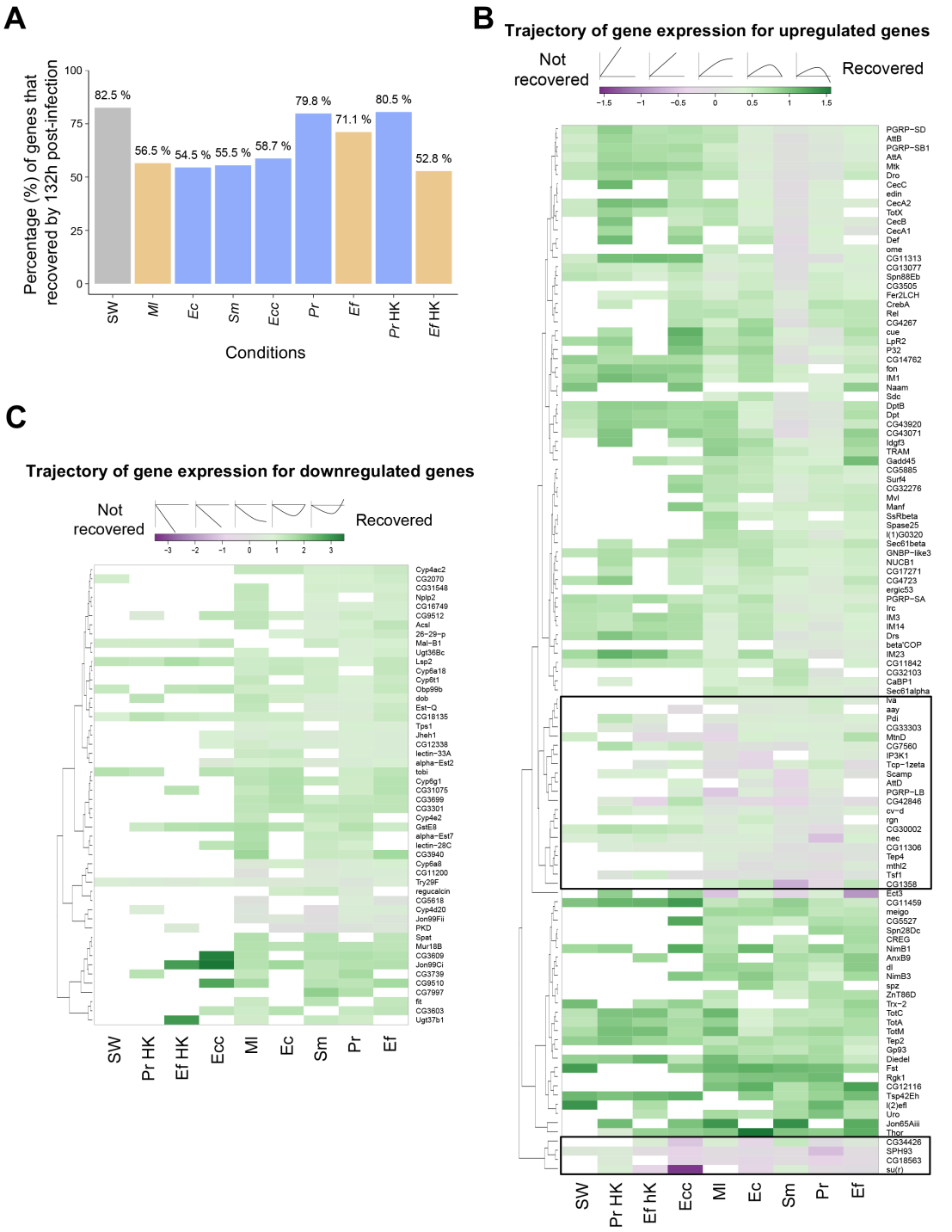


Figure 3.10: **Bacterial infection elicits long-term changes in global host transcription.**

(A) Percentage of genes found to be differentially expressed at 12 and/or 36 h post-infection in a given condition that returned to basal levels of expression (recovery) by 132 h post-infection.

(B) Gene expression trajectory of core upregulated genes. By 132 h post-infection, the expression level of core upregulated genes continued to increase (purple), plateaued (gray), or returned to basal, pre-infection levels (green) as indicated by the graphic above the color key. A black box encloses genes that did not recover in most infections (as denoted by the purple color). (C) Trajectory of gene expression for core downregulated genes. By the last time point (132 h), the transcript levels of core downregulated genes continued to decrease (purple), plateaued (gray), or returned to basal expression levels (green) as illustrated on the graphic above the color key. Genes that were not differentially regulated by a given condition are marked in white.

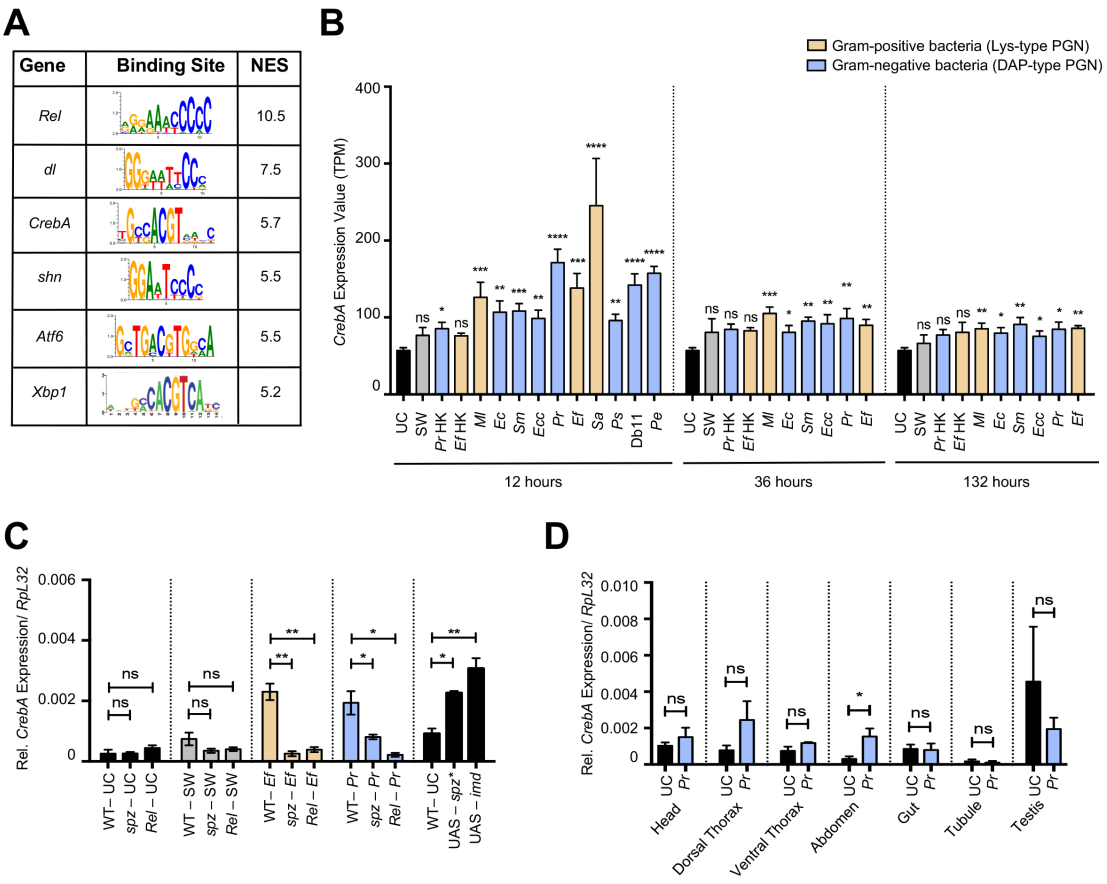


Figure 3.11: ***CrebA* is a core transcription factor regulated by Toll and Imd in the fat body.**

(A) Subset of transcription factors whose predicted binding sites are enriched in the promoter regions of core upregulated genes. The table includes the transcription factors' gene symbols, consensus binding sites, and their normalized enrichment scores (NES), which indicate the degree to which a binding site is overrepresented at the top of a ranked list of binding sites. (B) RNA-seq expression values in TPM (transcripts per million) of *CrebA* at 12, 36, and 132 h after infection with all 10 bacteria. (C) RT-qPCR of *CrebA* levels in Rel^{E20} and spz^{rm7} mutants and wildtype flies following: no challenge (UC), sterile wound (SW), infection with *E. faecalis* (*Ef*), and infection with *P. rettgeri* (*Pr*). In the last histogram, WT indicates wildtype flies given no challenge, UAS-spz* denotes *CrebA* expression in the absence of challenge when an activated form of Spz is ubiquitously overexpressed, and UAS-imd shows *CrebA* expression in flies that constitutively overexpress Imd in the absence of challenge.

(D) RT-qPCR of *CrebA* levels in dissected organs and body parts (head, dorsal thorax, ventral thorax, abdomen, gut, Malpighian tubule, and testis) following infection with *P. rettgeri*. Mean values of at least three biological replicates are represented ±SE. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001 in a Student's t-test.

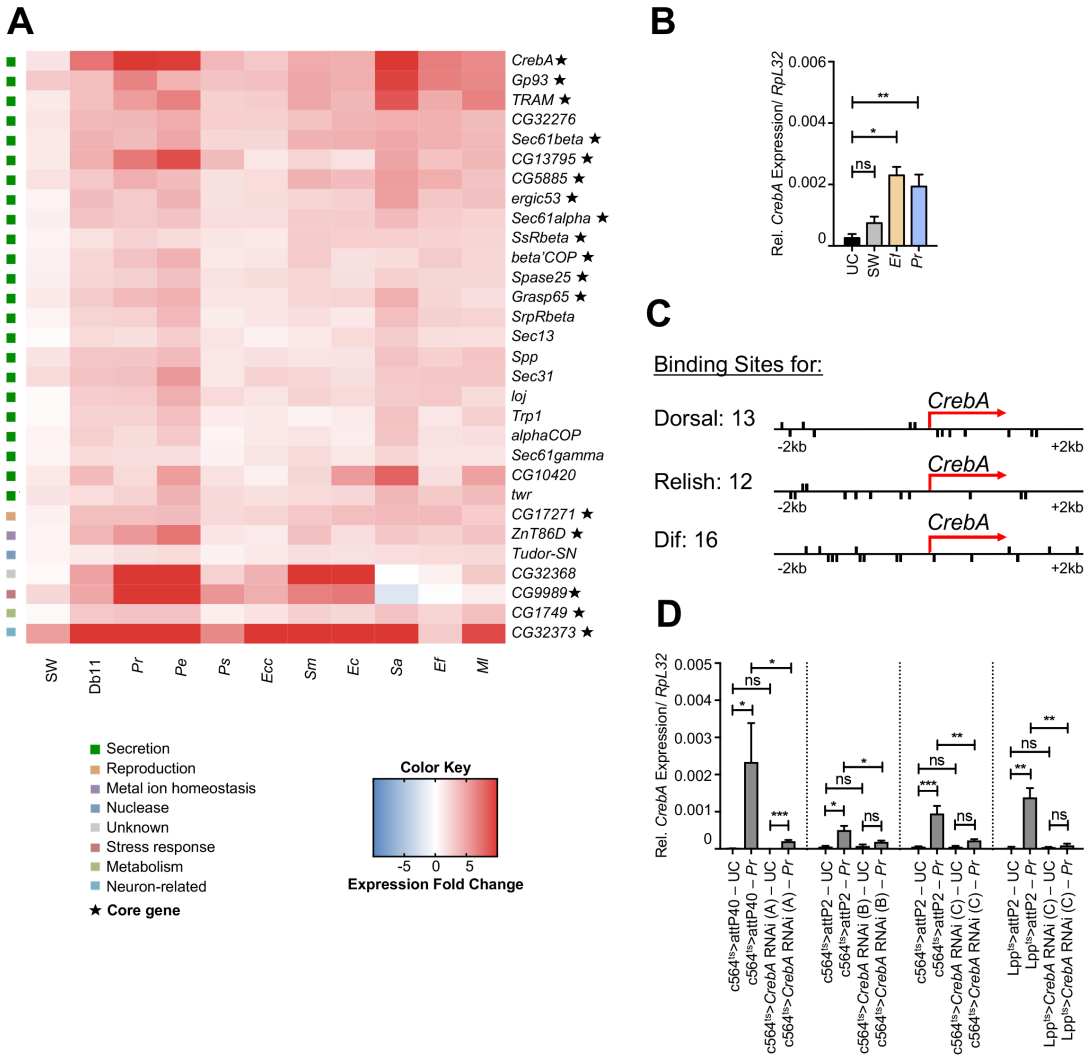


Figure 3.12: ***CrebA* is a target of Toll and Imd and a putative regulator of core genes in the fat body.**

(A) Heatmap showing the expression levels (log2 fold change) of a select group of putative *CrebA* target genes found to be significantly upregulated by infection. Core genes (marked by a star) and their functions are highlighted. (B) RT-qPCR validation of *CrebA* induction levels 12 h after infection with *P. rettgeri* (*Pr*) and *E. faecalis* (*Ef*) using samples distinct from those used in the RNA-seq. (C) Schematic of predicted Dif, Dorsal, and Relish binding sites on the *CrebA* promoter region (+/-2kb from the start site). (D) Whole fly RT-qPCR of flies with *CrebA* knockdown in the fat body following infection with *P. rettgeri*. *CrebA* RNAi (A), (B), and (C) denote three distinct RNAi constructs used to target *CrebA* mRNA. Mean values of three or more repeats are represented ±SE. *p<0.05 **p<0.01 ***p<0.001 in a Student's t-test.

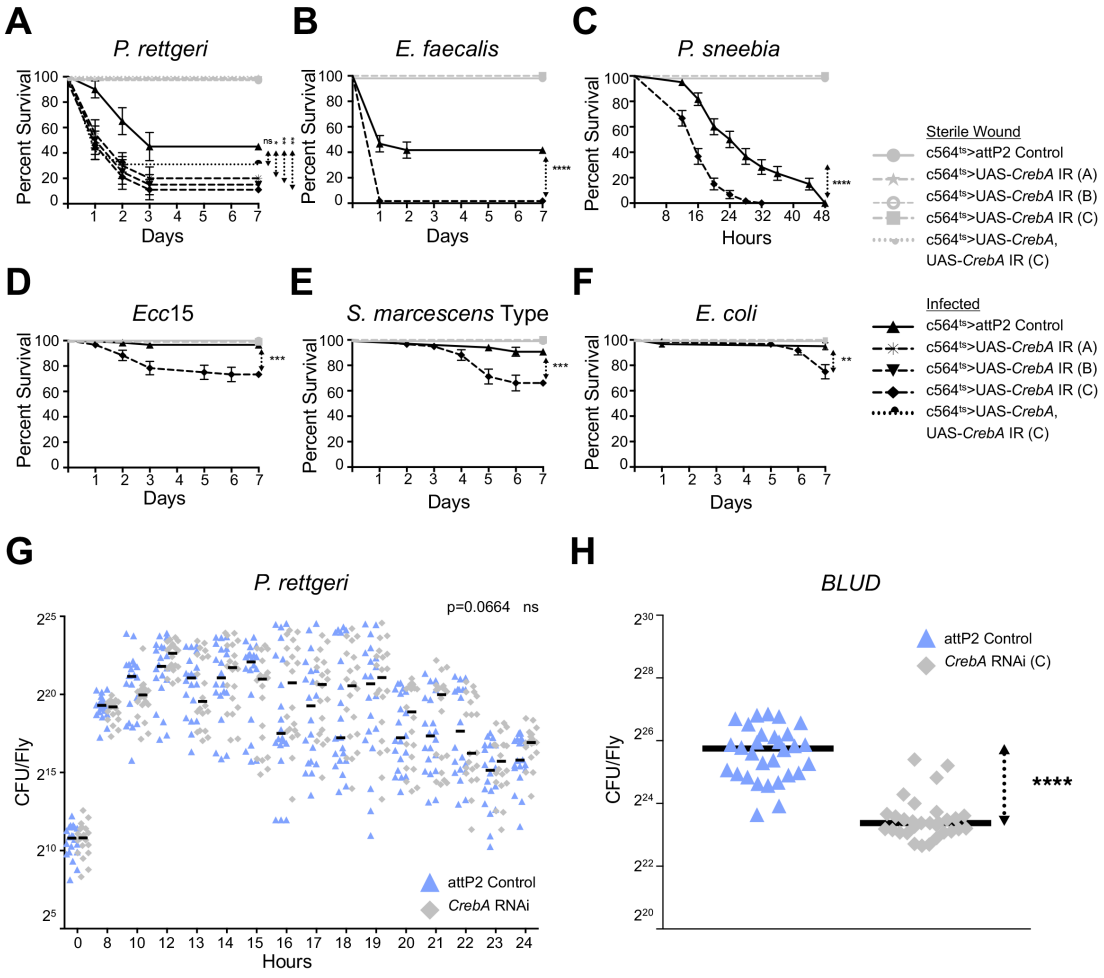


Figure 3.13: ***CrebA* promotes infection tolerance.**

Survival curves over 7 days (or 48 h in the case of *P. sneebia*) following infection of flies whose expression of *CrebA* is blocked with RNAi. *UAS-CrebA IR* (A), (B), and (C) indicate three distinct RNAi constructs that target *CrebA* transcripts. *UAS-CrebA, UAS-CrebA IR* refers to flies simultaneously co-expressing a *CrebA* RNAi and a *CrebA* overexpression construct. *attP2* is the background genotype control, in which *CrebA* is fully expressed. Knockdowns were driven in the fat body and hemocytes using a conditional c564 temperature sensitive driver. The curves represent the average percent survival ±SE of three biological replicates. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001 in a Log-rank test. Infections were performed with (A) *P. rettgeri*. (B) *E. faecalis*. (C) *P. sneebia*. (D) *Ecc15*. (E) *S. marcescens* Type. (F) *E. coli*. (G) Bacterial load time course of control flies and flies expressing *CrebA* RNAi in the fat body following infection with *P. rettgeri*.

(H) *P. rettgeri* bacterial load upon death (BLUD) of wildtype controls and flies with *CrebA* expression knocked down by RNAi in the fat body. Three repeats are graphed together, with each symbol representing an individual fly's number of colony forming units (CFU). Horizontal lines represent median values for each condition. ****p<0.0001 in a Student's t-test.

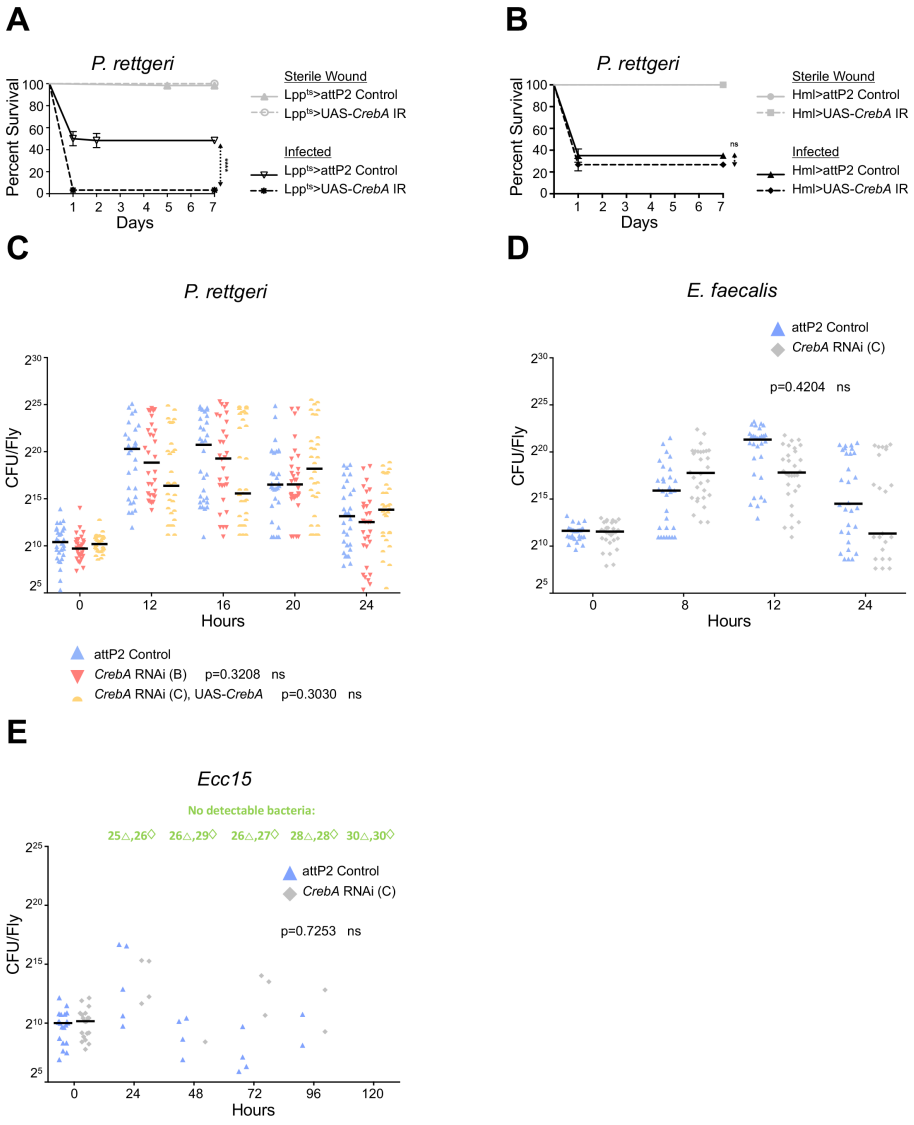


Figure 3.14: ***CrebA* RNAi flies do not carry a larger bacterial burden than wildtype flies upon infection.**

(A) Survival curves over 7 days following *P. rettgeri* infection of flies whose expression of *CrebA* is blocked with RNAi specifically in the fat body with a second driver, *Lpp-Gal4* (*Gal80^{ts}; Lpp-Gal4 > UAS-CrebA-IR*). *attP2* is the background genotype control, in which *CrebA* is fully expressed.

(B) Survival of unchallenged and infected (*P. rettgeri*) control flies and flies expressing *CrebA* RNAi in hemocytes only (*Hml-Gal4* driver). The curves represent the average percent survival ±SE of three biological replicates. ****p<0.0001 in a Log-rank test. (C) Bacterial load time course of control flies, flies expressing a separate *CrebA* RNAi construct (construct B), and flies simultaneously co-expressing a *CrebA* RNAi and a *CrebA* overexpression construct in the fat body following infection with *P. rettgeri*. Bacterial load time course of *CrebA* knockdown and control flies after infection with (D) *E. faecalis* and (E) *Ecc15*. Three repeats are graphed together, with each symbol representing an individual fly's number of colony forming units (CFU). Horizontal lines represent median values for each condition. A number followed by the symbol Δ (*attP2* control flies) or the symbol ◇ (*CrebA* RNAi flies) indicates the number of flies found to have no bacteria (flies that carry undetectable levels of bacteria or that have cleared the infection) at the specified time point.

Gene Name	Function	*P. rettgeri* / UC WT	*P. rettgeri* / UC *CrebA* RNAi
SECRETORY PATHWAY			
ArfGAP1	Secretion	1.8	1.2
bai	Secretion	1.9	1.3
CG32276★	Secretion	3.0	2.0
CG5885★	Secretion	3.1	1.6
CG8860	Secretion	2.2	1.1
CrebA★	Secretion	7.3	4.1
ergic53★	Secretion	2.4	1.8
Grasp65★	Secretion	2.6	1.4
Gtp-bp	Secretion	2.3	1.2
loj	Secretion	2.5	2.1
Sec31	Secretion	1.5	1.3
Sec61beta★	Secretion	2.6	1.8
Sec63	Secretion	2.0	1.2
Spase12	Secretion	2.3	1.5
Spase25★	Secretion	2.6	1.5
Spp	Secretion	2.9	1.4
Srp19	Secretion	1.5	1.1
Srp72	Secretion	1.5	0.8
SrpRbeta	Secretion	1.7	1.1
Tapdelta	Secretion	2.4	1.7
TRAM★	Secretion	4.6	1.7
twr	Secretion	3.3	1.7
UNKNOWN			
CG30026★	Unknown	6.0	5.4
CG5791★	Unknown	1.3	1.0

Gene Name	Function	*P. rettgeri* / UC WT	*P. rettgeri* / UC *CrebA* RNAi
IMMUNITY			
CecA1★	Antimicrobial peptide	183.5	87.8
CecA2★	Antimicrobial peptide	422.5	352.0
CecB★	Antimicrobial peptide	1285.0	281.6
CecC★	Antimicrobial peptide	402.4	127.4
CG11313★	Coagulation	3.3	0.6
CG43114	Defense Response	0.7	0.4
Drs★	Antimicrobial peptide	77.5	50.7
GNBP-like3★	Antimicrobial response	3.3	2.8
IM14★	Antimicrobial response	2.2	1.7
METABOLISM			
SLC22A	Sugar transporter-like	3.9	6.2
CG32054	Sugar transporter-like	2.4	23.2
CG42825	Sugar transporter-like	0.7	3.9
Gld	Glucose dehydrogenase	0.7	2.9
CG17097	Lipase	0.9	9.1
CG18258	Lipase	1.0	6.3
CG18284	Lipase	1.1	11.3
CG31872	Lipid metabolic process	0.9	8.9
Uba5★	Mo-molybdopterin cofactor biosynthetic process	2.4	1.3
PROTEOLYSIS			
CG30091★	Serine-type endopeptidase activity	3.3	Not expressed
METAL ION HOMEOSTASIS			
ZnT86D★	Zinc ion transmembrane transporter activity	2.7	1.7
REPRODUCTION			
CG17271★	Multicellular organism reproduction	1.9	1.2

Color Legend: ≥1.5 ≥2 ≥3 ≥5 ≥10 ≥100

★ Core gene

Figure 3.15: ***CrebA* regulates the expression of secretory pathway genes upon infection.**

Select list of 45 genes whose expression significantly changes in infected *CrebA* RNAi fat body samples compared to infected control samples. Gene symbols, functions, and fold enrichment of expression with infection (*P. rettgeri*/unchallenged) are indicated. Core genes are highlighted with a star symbol.

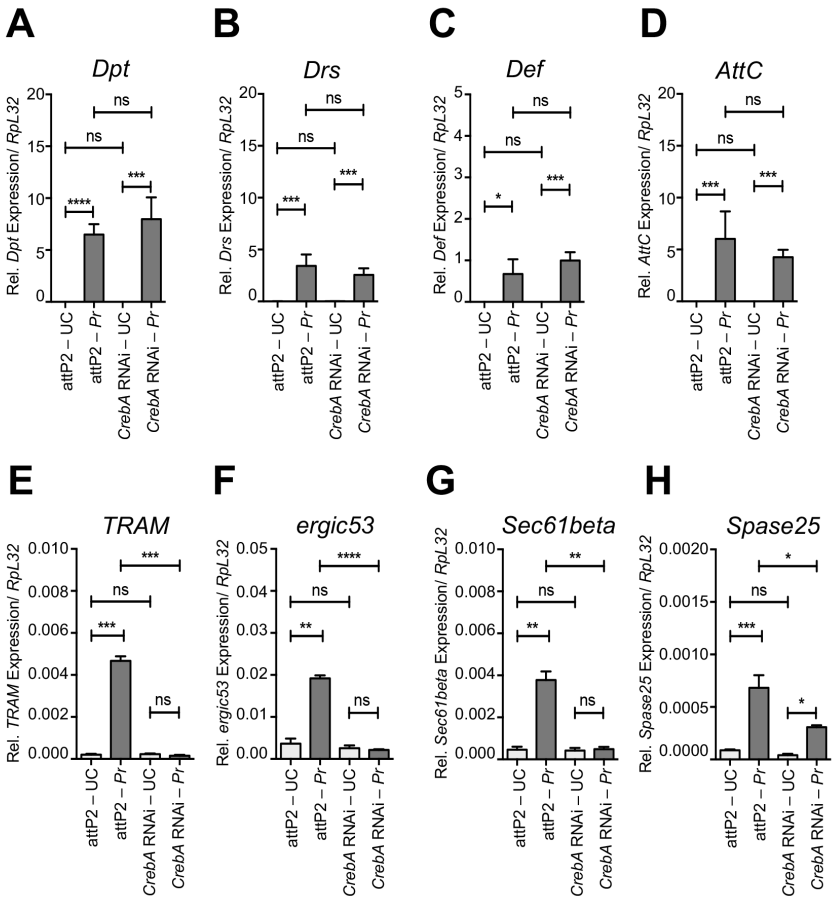


Figure 3.16: ***CrebA* regulates secretory capacity during infection.**

Expression level of predicted *CrebA* target genes in unchallenged (UC) or infected (Pr) conditions in control (*attP2*) and CrebA RNAi fat body samples. Assayed genes encoding antimicrobial peptides are (A) *Diptericin*, (B) *Drosomycin*, (C) *Defensin*, and (D) *Attacin C*. Surveyed genes encoding secretory factors are (E) *TRAM*, (F) *ergic53*, (G) *Sec61beta*, and (H) *Spase25*. Mean values of three biological replicates are represented ±SE. *p<0.05 **p<0.01 ***p<0.001 ****p<0.0001 in a Student's t-test.

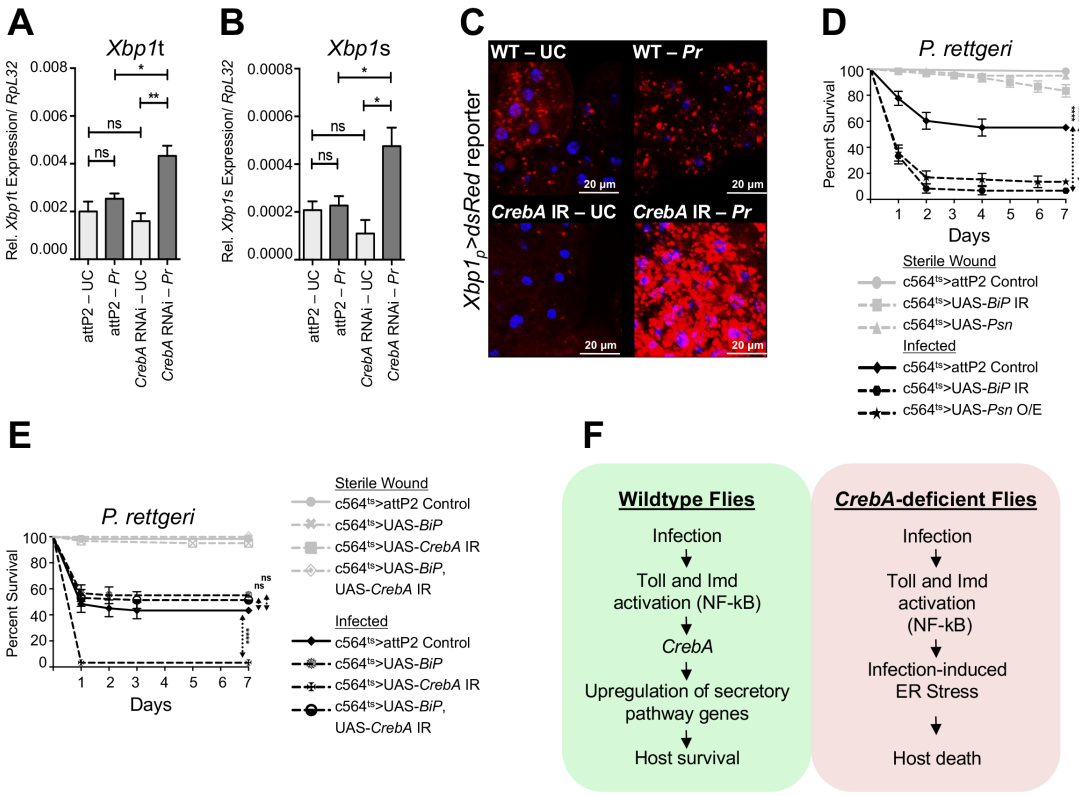


Figure 3.17: **Loss of *CrebA* triggers ER stress during infection.**

RT-qPCR of *Xbp1t* (unspliced, inactive) (A) and *Xbp1s* (spliced, active) (B) levels in the fat bodies of *CrebA* RNAi and wildtype flies in unchallenged and infected (*P. rettgeri*) conditions. An increase in the spliced form of *Xbp1* (*Xbp1s*) is a sign of ER stress. Mean values of three repeats are represented ±SE. *p<0.05 **p<0.01 in a Student's t-test. (C) Fat bodies from the *Xbp1p>dsRed* reporter crossed to *CrebA* RNAi or wildtype flies in unchallenged (UC) and infected conditions (Pr). (D) Survival curves of flies with genetically-induced ER stress (*Psn* overexpression or *BiP* RNAi) in the fat body in unchallenged and infected conditions. (E) Survival curves over 7 days of flies co-expressing both the *CrebA* RNAi and *BiP* overexpression constructs in their fat bodies following infection with *P. rettgeri*. The curves represent the average percent survival ±SE of three biological replicates. ****p<0.0001 in a Log-rank test. (F) Upon infection, activation of the Toll and Imd pathways in the fat body transcriptionally upregulates the expression of the transcription factor *CrebA*. In turn, *CrebA* upregulates the expression of secretory pathway genes. In absence of *CrebA*, a failure to upregulate secretion machinery genes leads to infection-induced ER stress, followed by host death.

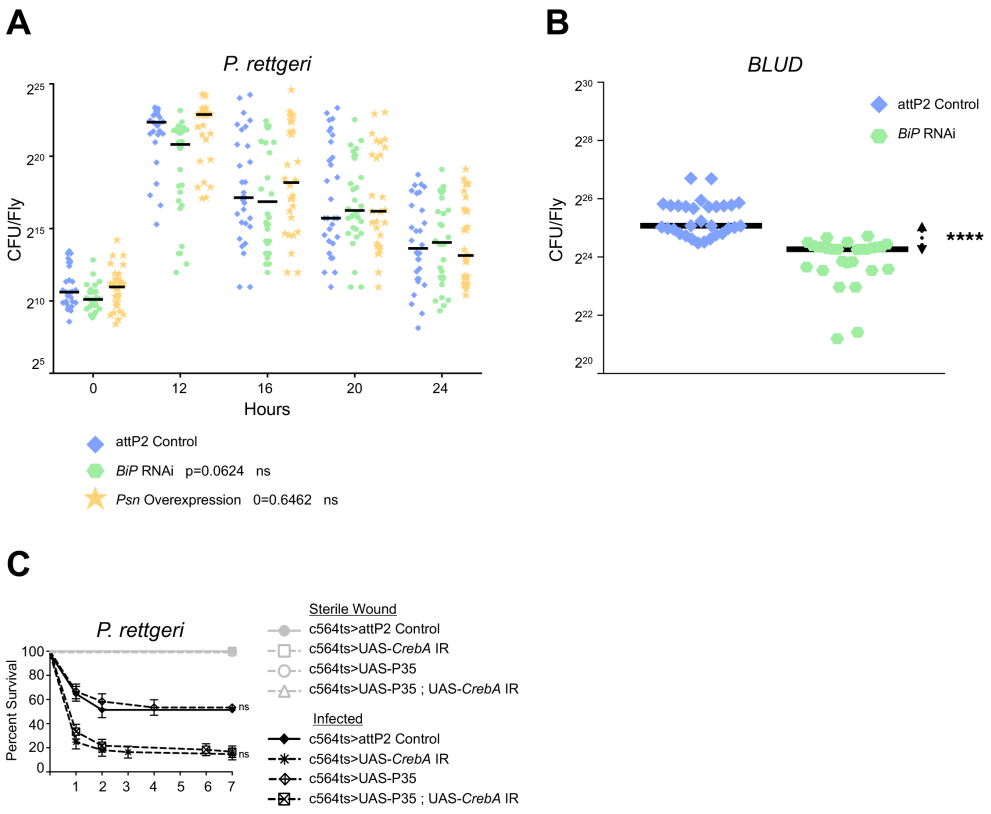


Figure 3.18: ***CrebA* expression prevents ER stress upon infection.**

(A) Bacterial load time course of control flies and flies expressing *BiP* RNAi or *Psn* overexpression in the fat body following infection with *P. rettgeri*. (B) Bacterial load upon death (BLUD) following *P. rettgeri* infection of wildtype controls and flies with *BiP* expression knocked down by RNAi in the fat body. Three repeats are graphed together, with each symbol representing an individual fly's number of colony forming units (CFU). Horizontal lines represent median values for each condition. ****p<0.0001 in a Student's t-test. (C) Survival curves of flies co-expressing *CrebA* RNAi and the apoptosis inhibitor P35 in fat body cells. The curves represent the average percent survival ±SE of three biological replicates.

CHAPTER 4

CO-EXPRESSION NETWORK ANALYSIS ILLUSTRATES KEY PHYSIOLOGICAL RESPONSES TO INFECTION IN *D. MELANOGASTER*

Author: Joo Hyun Im

4.1 Abstract

Insects have a defense mechanism that recognizes and removes a variety of pathogens. How is the generality of this infection response achieved? Looking at changes in RNA level upon infection can reveal the genes that become active or repressed in this process and identify the underlying structure of gene networks that illustrate how these genes interact with each other. In this study, I applied co-expression network analysis method on the transcriptome of *Drosophila melanogaster* after infections with 10 different microbes to determine genes that are co-expressed post-infection. This resulted in four modules of genes associated with lipid and carbohydrate metabolic process, protein secretion and transport, and antimicrobial peptide (AMP)-based immune response. This study illustrates that post-infection responses include an induction of the canonical immune response as well as changes in metabolism and management of cellular stress and provides candidate genes for molecular genetic studies.

4.2 Introduction

Infection alters host physiology, disrupts host homeostasis, and causes tissue damage and potentially death. The innate immune system acts as the first line of defense against infections with a wide range of pathogens and is highly conserved across animals [1]. Within hours of infection, a robust transcriptional response activates production of antimicrobial peptides (AMPs) and defense molecules [1, 111]. Measuring the transcriptional response to infection has identified candidate genes and pathways important for host defense [11, 12, 112, 117, 133, 173]. Early transcriptome profiling of *D. melanogaster* with benign *Escherichia coli* and *Micrococcus luteus* for infection discovered the *Drosophila* Immune-Regulated Genes (DIRGs), a set of genes that are significantly differentially regulated upon infection and are relevant to host defense (production of AMPs, microbial recognition, phagocytosis), regulation of iron metabolism, and generation of reactive oxygen species [11]. Recent work examining the *D. melanogaster* transcriptomic response to phenotypically diverse natural pathogens showed that a set of core genes are differentially expressed upon the majority of infections and are enriched with functional categories that are beyond canonical immune response, such as stress response, translation control, wound repair, and secretion [173].

Work up to date only has focused on understanding how individual genes are regulated upon a specific infection and failed to look at how these genes as a group change expression against diverse bacterial challenges. Upon stimuli, a group of genes in the same biological process or involved in multiple signaling pathways can be co-regulated. Identifying genes that are differentially expressed between two conditions using a simple pairwise comparison ignores

these functional interactions between genes. In addition, this method is limited in inferring the function of unannotated genes if no contextual information, such as binding partner, functional domain, or pathway membership, is available. Co-expression network analysis overcomes these limitations. Based on the idea of guilt by association, this approach assumes that genes with similar expression profiles across multiple samples are also likely to be a part of the same biological process and thus to be regulated by the common factor [174, 175]. It constructs an undirected network of genes based on the correlations of genes' expression levels. Next, the method groups genes into transcription modules, collections of genes that are regulated together and likely to be associated with the same biological processes [176, 177]. The ultimate output is a graph with nodes (genes) and edges (correlations in expression values between genes). From these network modules, one can ask which genes have the greatest number of connections with other genes in the module and thus are likely to be functionally important for the phenotype in response to a given stimulus (called hub genes) [178–180]. Secondary analyses can then reveal which biological processes are most commonly represented and which transcription factor's binding sites are most enriched [177, 181]. Co-expression network analysis has been fruitful in constructing gene networks underlying phenotypes and in inferring functions of uncharacterized genes in *Arabidopsis*, *Salmonella*, zebrafish, and human [182–185].

To identify a network of genes that regulate the general infection response to a wide range of bacteria, I used co-expression network analysis on the RNA-seq data of *D. melanogaster* upon infection with ten bacteria that span the spectrum of virulence [173]. This resulted in four transcription modules associated with cellular lipid and carbohydrate metabolic process, protein secretion regulated

by *CrebA*, AMP-based immune response, and an unknown function. Our approach illustrates that changes in metabolism and protein secretion are a big part of the general response to infection in addition to the canonical immune response and lays a foundation for molecular genetic studies of hub and otherwise crucial yet uncharacterized genes.

4.3 Materials and Methods

Data acquisition

The RNA-seq data feature experiments designed to measure transcriptional changes over time in the whole body of adult fruit flies upon infection with the following ten different live bacteria: *Micrococcus luteus*, *Escherechia coli*, Type strain of *Serratia marcescens*, *Erwinia caratovora Ecc15*, *Providencia rettgeri*, *Enterococcus faecalis*, *Staphylococcus aureus*, *Providencia sneebia*, *Serratia marcescens* strain Db11, and *Pseudomonas entomophila* [173] (sequence accession number: SRP127794). Samples infected with each bacterium were collected at 12 h, 36 h, and 132 h post-infection, except for four pathogenic infections (*S. aureus*, *P. rettgeri*, *P. entomophila*, and *S. marcescens* Db11). Samples infected with four highly virulent pathogens were collected only at 12 h for since most flies had died before the subsequent sampling time points. For each sample, three replicates were generated. Details of infection conditions, data generation, and data processing can be found in [173].

Co-expression network analysis

Raw RNA-seq reads were aligned, were processed into count data, and

then were normalized as described in [173]. Genes with low expression (count-per-million <1.2) were removed. Only genes that were significantly differentially expressed in at least one live infection regardless of time points compared to the no-infection condition were selected to make sure that they were biologically relevant. The modified version of the weighted gene coexpression network analysis (WGCNA) was applied to the RNA-seq data using the WGCNA R package [186, 187]. The data were log2 transformed after adding 1 to the count values to remove any entries with 0 and then, for a given pair of genes, a similarity score that combines the Pearson correlation and Euclidean distance was calculated. To filter out potentially false-positive correlations, the similarity matrix was then converted into an adjacency matrix using 'adjacency.fromSimilarity' function with the power = 5. To detect co-expression modules in the network and to cut branches, hierarchical clustering function 'hclust' and the cutreeDynamicTree function based on a dynamic branch cut algorithm were applied [188]. This resulted in modules containing genes whose expression levels are positively correlated with each other. Using igraph R package (v.1.0.1), the network information was exported as a graphml file [189].

Visualization of the network and identification of features in the network

To visualize the co-expression modules, Cytoscape v3.5.1 was used [190]. A network consists of nodes and edges, the former representing genes and the latter demonstrating the correlations of expression values between connected genes. The network analysis resulted in total 63 clusters with two or more nodes with the mean number of nodes across clusters (size) being 12 genes (Figure 4.5). Distribution of module sizes is skewed towards left indicating

that there are a few modules with a large number of nodes and many others with a few nodes (Figure 4.5). To only focus on the modules that are likely to carry meaningful biological information, a co-expression module was subsequently defined as a group of genes that contains minimum 12 genes as members and that share the biological function. The following network statistics for each co-expression module and its gene members were calculated using Cytoscape's NetworkAnalyzer plugin [191] with an assumption that edges are undirected: node degree, betweenness centrality, and network centralization. The 'node degree', also known as 'node connectivity', refers to the number of edges linked to a given node. A node with the most connections or the highest 'node degree' was deemed as a hub node because it best represents the behavior of a given module [192]. Hub genes are expected to be more functionally relevant or essential in responding to the conditions under which the data were collected [178]. Thus, hubs genes are predicted to play a crucial role in the response to infection, and/or may be involved in several different functional pathways that become all differentially regulated as a part of the general response to infections. The 'betweenness centrality' ($C_b(n)$) parameter measures how central a given node is to the module and is calculated the following way: $C_b(n) = \sum_{s \neq n \neq t} (\sigma_{st}(n) / \sigma_{st})$, where s, n, and t are distinct nodes, σ_{st} indicates the number of shortest paths from s to t, and $\sigma_{st}(n)$ is the number of shortest paths from s to t via n. It is used to identify bottleneck nodes that are shortest-path connectors in the network, some of which can be hub genes [177, 193, 194]. A high score is assigned to nodes that have a high degree of intersections, bridging sub-networks [193, 194]. Biologically, this parameter is useful in finding additional functionally relevant genes that do not necessarily have the largest number of connections.

These shortest-path connectors determine how robust the network is and disruption in their activity may result in functional consequences. The 'network centralization' parameter measures the distribution of connections evaluating whether the network has a star-like topology or the nodes in the network have relatively the same level of connectivity and is calculated the following way: $Centralization = \frac{n}{n-2}\left(\frac{\max(k)}{n-1} - Density\right) \approx \frac{\max(k)}{n} - Density.$ where $Density = \frac{\sum_i \sum_{j \neq i} a_{ij}}{n(n-1)}$, n is the number of all nodes, k is the total number of edges, and a_{ij} is the number of connections between node i and node j. Network modules with a central node connected to the rest of the nodes with only one link (a star topology) would have this value close to 1, while the parameter value close to 0 represents decentralized networks with multiple sub-modules, which are highly interconnected regions of a network [192]. If the network consists of multiple sub-modules, it might biologically mean that these sub-modules are functionally redundant for the same biological process or each of them belongs to distinct pathways that are together activated upon infection.

To identify sub-modules in a large module, Cytoscape's ClusterViz plugin with the Fast AGglomerate algorithm for mining functional modules based on the Edge Clustering coefficients (FAG-EC) clustering algorithm [195] was used with default options. To identify putative transcription factor binding sites enriched in the genes of a given module, the i-cisTarget tool [127] with default parameter values was used. Only transcription factors with normalized enrichment score (NES) >5, which indicates the degree to which a binding site is overrepresented at the top of a ranked list of binding sites from the Jaspar database, were selected. To find overrepresented gene ontology (GO) categories and pathway annotations in a module, the PANTHER tool [126]

was used . To predict potential interactions between user-provided lncRNAs and mRNAs, the LncTar tool [196] was used. For a given lncRNA and mRNA pair, the tool calculates the minimum binding free energy (deltaG) based on thermodynamics that maximizes the number of base pairing and then normalizes deltaG by length of both RNAs, which results in the ndG score. Pairs of interactions with a ndG ≤-0.1 were selected as suggested. To survey known functional domains and find enriched binding sites of unannotated genes, Interpro (http://www.ebi.ac.uk/interpro/, [197]) was used on the amino acid sequence of each gene and the Analysis of Motif Enrichment (v 4.12.0, http://meme-suite.org/tools/ame, [198]) was applied on the 2kb upstream and downstream of each gene's sequence.

4.4 Results and Discussion

Canonical immune pathways as well as processes involved in host metabolism and cellular/tissue biology (transcription, translation, secretion, and cell division) become differentially regulated against a bacterial challenge [11, 12, 112, 117, 133, 173]. In doing so, expression of a set of genes can be coordinated to achieve host defense. To identify a gene network underlying the generality of infection response, I have used the dataset that characterized the transcriptomic response of *D. melanogaster* to a wide range of bacterial infections. The analysis resulted in total 63 groups with two or more nodes with the mean number of nodes across groups being 12 genes. Total of four modules had more than 12 genes with distinct functional categories (Table 4.1).

Module 1 reflects changes in lipid and carbohydrate metabolism genes upon

infection

Module 1 consists of 380 genes and is the largest (Figure 4.1, Table 4.1). It is a set of genes that show reduced expression at 12 hours post-infection and then recover their expression level to the pre-infection level over time. Its network centralization value is 0.103, indicating that the network is decentralized potentially with multiple sub-modules. Indeed, ClusterViz results showed that Module 1 contains total 33 sub-modules with two largest sub-modules that make up more than 50% of the genes (first: 147 genes, second: 79 genes (Figure 4.1). Assessing GO terms for each of these two sub-modules showed that the only GO term associated with the second sub-module - glycogen metabolic process - is shared by the first sub-module, making these sub-modules functionally not distinguishable. Thus, we looked at genes in Module 1 as a whole. GO terms enriched in Module 1 underscore metabolism (Table 4.2): lipid metabolic process, carbohydrate metabolic process (monosaccharide/disaccharide/glycogen metabolic process), vitamin biosynthetic process, and sulfur compound metabolic process.

Lipids are storage of fatty acids and are accumulated in the form of lipid droplets in insect fat body cells upon feeding and then released for movement and development [199]. Recently, several studies found a direct link between lipid droplets and immune defense. In mammals, activated Toll-like receptors (TLR) trigger production of lipid droplets in infected tissues and prompt recruitment of signaling molecules to lipid droplets which then eventually help produce interferons against virus [200, 201]. In mosquitoes, *Aedes aegypti*, activation of Toll and Imd immune pathways by a challenge with Gram-negative bacteria and virus results in an increase in the amount of lipid

droplets [202]. In *Drosophila*, histones bound to lipid droplets kill bacteria in embryos and potentially in adults [203].

Lipid metabolic process is linked to the largest number of genes (33 genes) in this module. Other genes associated with generating and storing lipid droplets, such as lipid storage droplet gene *Lsd-2* and its binding partner *alpha-Est7* [204, 205], are also present. It is plausible that these genes are eventually upregulated after a bacterial attack to increase the quantity of lipid/lipid droplets to generate immunity modulators as lipid droplets do in mammals [206]. The hub gene, NADH-Cytochrome B5 reductase (*Cyb5r*, also known as *CG5946*), is also associated with the regulation of lipid storage. As an enzyme highly expressed in the adult fat body, it is conserved in *Drosophila*, other insects, as well as in mammals. Overexpression of *Cyb5r* in flies extends fly life span and increases long-chain fatty acids [207]. Overexpression of *Cyb5r* in mice liver also upregulates genes involved in lipid biosynthetic process and its overexpression in primary culture of adult mouse hepatocytes increases de novo lipogenesis [207]. In short, gene associated with lipid metabolism may be co-regulated and *Cyb5r* may play an integral role in generating more lipids upon infection.

Genes in Module 1 are also associated with carbohydrate metabolic process, which is important for growth, development and homeostasis of animals [208]. This process is drastically shifted during infection. For instance, flies infected with *Mycobacterium marinum* exhibit downregulation of genes involved in glycogen synthesis and degradation as well as triclyceride synthesis [117]. Upon infection with the intracellular pathogen *Listeria monocytogenes*, flies experience downregulation of glycolysis genes and lose glycogen stores [209]. Genes involved in sugar digestion (i.e. the *Maltase*

cluster) are also downregulated by infection [173]. Our results demonstrate global suppression of various carbohydrate metabolic processes, such as monosaccharide (glycolysis, pentose-phosphate shunt), disaccharide, and polysaccharide (glycogen) metabolism in addition to carbohydrate transport. These data suggest that shift in carbohydrate metabolism coincides with immune defense.

Fifty-three out of 86 core downregulated genes (61.6%), those that were repressed upon infection with most of the pathogens used in [173], are also present in this module. The putative transcription factor binding sites most common in these core downregulated genes (pnr/GATAa, lola, GATAe, and grn/GATAc) are also enriched in Module 1 (Table 4.5). GATA transcription factors are important for the development and maintenance of adult intestinal tissues and pnr/GATAa and lola regulate glucose metabolic process [210, 211]. In addition, GATA motifs are enriched in AMP promoters and are important for induction of AMPs [212].

A connection between lipid and carbohydrate metabolism and immunity is only beginning to be explored. Infection with *Listeria* leads to reduction in fatty acids and glycogen storage in fruit flies [209]. A transcription factor MEF2 is known to facilitate generation of lipogenic and glycogenic enzymes in healthy flies and then works as a switch to promote transcription of AMPs upon infection [119]. Our results point to lipid and carbohydrate metabolic genes that are initially downregulated together and then recover their expression or are induced post-infection. This global downregulation immediately after infection may occur in order for the host to reduce the availability of free nutrients for pathogens. Future work in understanding the connection between changes in

lipid and carbohydrate metabolism upon induction of immune response would reveal how host physiology is altered post-infection.

Module 2 suggests changes in protein secretion and transport upon infection

Module 2 contains 160 genes that are induced upon infection but to a different degree depending on the type of infections (Table 4.1, Figure 4.2). The most common predicted binding sites are for transcription factor Atf6, Atf3, Creb3/CrebA, cnc, and kay (Table 4.5), some of which are crucial in host survival upon infection [135, 138, 173]. Previously, we have shown that *CrebA* is upregulated upon bacterial infection to potentially regulate protein secretion in cells in order to prevent infection-induced ER stress [173]. Indeed, *CrebA* is present in the module as well as 34 genes that are targeted by *CrebA* (21.3% of the genes in this module, indicated by yellow in Figure 4.2). These results demonstrate that *CrebA* may play a crucial role in this module.

The network centralization score (the degree of how centralized the network is) of Module 2 is low (0.069), indicating that the module may be decentralized with multiple sub-modules (Table 4.1). Indeed, Module 2 features a bipartite structure with a sole connection between *CrebA* and *jdp*. The betweenness centrality measure that demonstrates how central a given node is in a module is the highest for *CrebA* (0.511) and *jdp* (0.456), confirming its role in bridging sub-modules. Genes that belong to the same sub-module as *CrebA* are not enriched for any biological processes or signaling pathways, potentially due to the small number of genes (55 genes) used for input. One of them is a hub gene of Module 2, *CG17271*, which is associated with signaling, reproduction, and metal ion binding, although its precise role in *Drosophila* is unknown. Its human ortholog, MCFD2, is known to form a complex with LMAN1 (a human

ortholog of *ergic53*), a key component of the secretory machinery, to assemble a cargo receptor for transport of proteins between ER and Golgi [213].

jdp belongs to the Heat Shock Protein 40 family and is associated with unfolded protein binding. It is induced in all infections though statistically significantly so only in pathogenic infections (*P. rettgeri*, *S. aureus*, *P. entomophila*) [173]. The precise function of *jdp* is unknown, but *jdp* is not a target of *CrebA* because the expression of *jdp* did not change in *CrebA* RNAi knockdown flies compared to wildtype flies [173]. Genes that belong to the same sub-cluster as *jdp* are overrepresented for GO terms, exocytosis and intracellular protein transport (Table 4.3). These results suggest that *jdp* may play a role in *CrebA*-initiated secretion machinery once it becomes active upon infection.

Altogether, these results suggest that the expression of genes associated with protein secretion change together upon various bacterial challenges. What is the relationship between protein secretion and infection? Infection increases cellular stress of the host since cells have to rapidly translate and deliver necessary protein cargos, such as AMPs, to the right compartments [170]. Thus, infection may activate *CrebA*-mediated secretory machinery as a whole as previously speculated [173].

Module 3 indicates no common biological process

Module 3 comprises of 38 genes that are in general downregulated in 36h or 132h post-infection although these changes are not statistically significant in many conditions. Most of these, including the hub gene, are not characterized (Figure 4.3). Thus, no GO category or pathway is associated with this module. Those genes that are annotated are associated with functional categories such

as proteolysis, glycogen metabolic process, protein transport, phospholipid metabolic process, and response to stimulus. The putative binding sites most commonly present in Module 3 are for transcription factors zen and CG11839 (Table 4.5). zen is a Hox-like homeobox transcription factor involved in epithelium development and dorsal/ventral axis specification while CG11839 belongs to the spliceosome complex C and is involved in mitotic cell cycle [214, 215]. Neither of them has been associated with the immune response. The Module is not partitioned into multiple sub-modules as indicated by the network centralization score (0.26) (Table 4.1).

Module 4 of immune genes provides insight into function of uncharacterized genes

Module 4 includes 14 genes that are induced upon all ten bacterial infections with a higher expression level shown in infections with Gram-negative bacteria than those with Gram-positive bacteria (Figure 4.4, [173]). Known immune genes, such as AMPs and peptidoglycan recognition receptor proteins (PGRPs), make up this module (Table 4.4). While two AMP genes, *AttB* and *DptB*, have the highest node degree (=7) of connections and are known immune genes, other non-hub genes are equally essential in the immune defense. Its network centralization value is 0.321, illustrating that the distribution of connectivity among nodes in this module is more even relative to that of other modules. This suggests that there is no strong hub gene in this module.

AMPs are known to bind to membrane bilayers of bacteria to form pores or inhibit enzymatic activity of bacteria [216]. They are typically small (<10 kDa) and cationic proteins and are made in the fat body [1]. Module 4 features two uncharacterized genes (*CG43236* and *CG43920*) that are more highly induced

in pathogenic infections than in benign infections and the level of expression does not recover back to the basal level long after any bacterial challenge (*CG43236*) or upon infections with bacteria whose presence persists in infected flies even multiple days after infection (*CG43920*) [173]. To understand the chemical properties of the proteins that these genes encode, I calculated the molecular weight, theoretical isoelectric point (pI), and predicted charge at pH using Protein Calculator (v3.4, http://protcalc.sourceforge.net/). *CG43236* and *CG43920* both encode small proteins (5.7 kDa and 6.2 kDa respectively) that are predicted to be cationic since the pI for each protein is bigger than the predicted charge at pH (10.28 vs 2.9 for *CG43236* and 8.27 vs 1.4 for *CG43920*). These results suggest that they may function as AMPs or else other molecules important for response to a bacterial challenge. Surveying known functional domain in the coding sequence and enriched binding sites in the 2kb up- and downstream of these genes resulted in no information.

Long non-coding RNAs (lncRNAs) are transcripts longer than 200 bp that do not encode proteins and are known to regulate transcription, post-transcription, and translation [217]. lncRNAs are also generated post-infection and are involved in regulating the transcription of immune genes. In human macrophages, lincRNA-Cox2 and THRIL regulate genes involved in the inflammatory process by interacting with the heterogeneous nuclear ribonucleoproteins that help transcribe genes [218, 219]. However, no insect lncRNAs have been implicated in the immune response to date. Module 4 contains two uncharacterized long non-coding RNA genes, *CR44404* and *CR45045*. An independent dense time-course of D. melanogaster transcriptome upon stimulation with lipopolysaccharides (LPS) also showed that the expression patterns of *CR44404* and *CR45045* mirror those of AMPs in the whole

fly, confirming that this observation is robust (Personal communication with Florencia Schlamp). Prediction by the LncTar tool to determine whether these lncRNAs may physically interact with AMP transcripts that are induced in the majority of infections from [173] demonstrated that CR44404 is predicted to interact with AttB, AttA, CG43236, CG43920, CecA2, DptA, Dro (all in Module 4) and CecA1. Likewise, CR45045 may be interacting with AttB, AttA, DptB, CG43920, Mtk, Dro (all in Module 4), CecA1, and AttD. In short, these results suggest that these lncRNAs and CG genes may play a direct role in the immune response.

Table 4.1: **Network statistics of each module**

Module	# of genes	Hub gene[a]	Node degree of a hub gene[b]	Network centralization[c]
1	380	*Cyb5r*	49	0.103
2	160	*CG17271*	14	0.069
3	38	*CG9970*	13	0.26
4	14	*AttB, DptB*	7	0.321

a: a gene with the most connections
b: the number of edges linked to the hub node
c: the degree to evaluate whether the network is star-shaped or decentralized

Table 4.2: **Selected Gene Ontology categories enriched in Module 1**

	Drosophila melanogaster (REF)	Client Text Box Input (▽ Hierarchy NEW!)					
PANTHER GO-Slim Biological Process	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
disaccharide metabolic process	2	2	.05	43.67	+	2.91E-03	2.27E-02
↳carbohydrate metabolic process	232	34	5.31	6.40	+	1.39E-16	3.13E-14
↳primary metabolic process	2916	115	66.77	1.72	+	9.81E-10	4.43E-08
↳metabolic process	3558	133	81.47	1.63	+	4.90E-10	2.77E-08
pentose-phosphate shunt	6	4	.14	29.12	+	4.65E-05	7.01E-04
↳monosaccharide metabolic process	62	16	1.42	11.27	+	1.34E-11	1.01E-09
fatty acid beta-oxidation	23	7	.53	13.29	+	3.45E-06	7.79E-05
↳fatty acid metabolic process	106	13	2.43	5.36	+	2.57E-06	6.46E-05
↳lipid metabolic process	262	33	6.00	5.50	+	2.06E-14	2.32E-12
vitamin biosynthetic process	10	3	.23	13.10	+	2.69E-03	2.34E-02
↳vitamin metabolic process	15	3	.34	8.73	+	7.06E-03	4.43E-02
glycolysis	21	6	.48	12.48	+	2.39E-05	4.15E-04
↳generation of precursor metabolites and energy	186	16	4.26	3.76	+	1.27E-05	2.39E-04
glycogen metabolic process	36	9	.82	10.92	+	5.49E-07	1.55E-05
↳polysaccharide metabolic process	58	12	1.33	9.04	+	4.28E-08	1.38E-06
gluconeogenesis	13	3	.30	10.08	+	5.01E-03	3.54E-02
carbohydrate transport	30	6	.69	8.73	+	1.33E-04	1.87E-03
cellular amino acid biosynthetic process	47	8	1.08	7.43	+	2.81E-05	4.54E-04
lipid transport	58	7	1.33	5.27	+	6.03E-04	7.17E-03
steroid metabolic process	57	6	1.31	4.60	+	2.79E-03	2.34E-02
sulfur compound metabolic process	77	8	1.76	4.54	+	6.22E-04	6.70E-03

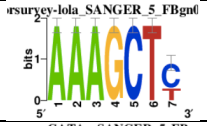
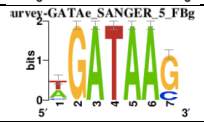
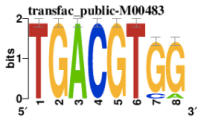
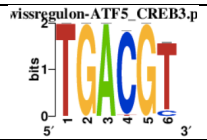
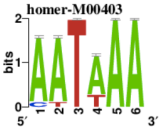
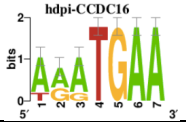
Table 4.3: **Selected Gene Ontology categories enriched in Module 2**

	Drosophila melanogaster (REF)	Client Text Box Input (Hierarchy NEW!)					
PANTHER GO-Slim Biological Process	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
exocytosis	82	8	.84	9.52	+	3.54E-06	1.33E-04
↳vesicle-mediated transport	335	18	3.43	5.24	+	1.79E-08	1.35E-06
↳transport	1012	29	10.37	2.80	+	4.87E-07	2.75E-05
↳localization	1108	30	11.36	2.64	+	9.54E-07	4.31E-05
intracellular protein transport	411	21	4.21	4.99	+	2.43E-09	2.74E-07
↳protein transport	427	22	4.38	5.03	+	8.34E-10	1.89E-07
protein localization	318	13	3.26	3.99	+	3.20E-05	1.03E-03

Table 4.4: **Selected Gene Ontology categories enriched in Module 4**

	Drosophila melanogaster (REF)	Client Text Box Input (Hierarchy NEW!)					
PANTHER GO-Slim Biological Process	#	#	expected	Fold Enrichment	+/-	raw P value	FDR
defense response to bacterium	11	4	.01	> 100	+	1.91E-10	4.31E-08
response to biotic stimulus	12	4	.01	> 100	+	2.54E-10	2.87E-08
immune response	33	4	.02	> 100	+	9.16E-09	6.90E-07
↳immune system process	106	4	.08	51.91	+	7.81E-07	3.53E-05
response to abiotic stimulus	51	4	.04	> 100	+	4.70E-08	2.66E-06
response to stress	282	4	.20	19.51	+	3.47E-05	1.31E-03

Table 4.5: **Enrichment of putative transcription factor binding sites in each module**

Module	Gene ID	Gene name	Binding site	NES (Normalized enrichment score)
1	FBgn0003117	*pnr/GATAa*		6.1721
1	FBgn0283521	*lola*		6.02408
1	FBgn0038391	*GATAe*		5.4965
1	FBgn0001138	*grn/GATAc*		5.20903
2	FBgn0033010	*Atf6*		5.76397
2	FBgn0004396	*Creb3/CrebA*		5.71747
2	FBgn0028550	*Atf3*		5.61357
2	FBgn0262975	*cnc*		5.35670
2	FBgn0001297	*kay*		5.16958
3	FBgn0004053	*zen*		5.45949
3	FBgn0039271	*CG11839*		5.10028

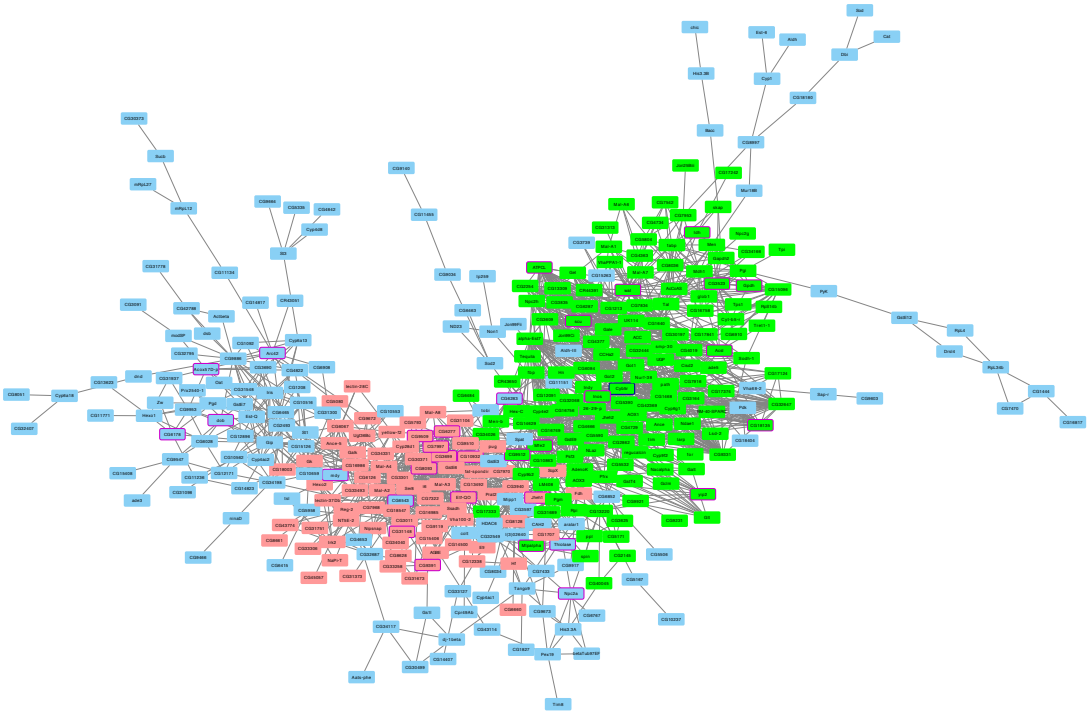


Figure 4.1: **Coordinated changes in metabolism genes upon infection.**

The network module was plotted in an organic format. Each box represents a node (gene) in the module. The box with a thick blue outline marks the hub gene. Green boxes represent 147 genes in sub-module 1 while peach boxes indicate 79 genes in sub-module 2. Boxes with purple outline represent genes associated with lipid metabolism and lipid droplets.

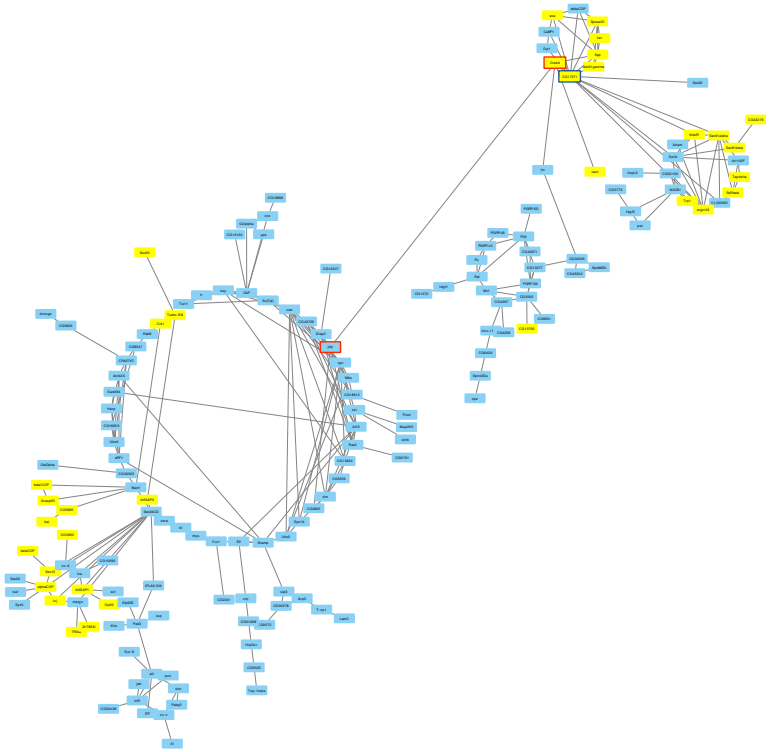


Figure 4.2: ***CrebA*-mediated secretory machinery undergoes changes upon infection.**

The network module was plotted in a circular format to emphasize two sub-modules. Each box represents a node (gene) in the module and the box with a thick blue outline marks the hub gene. The boxes with a thick red outline marks two genes, *CrebA* and *jdp*, which connect two sub-modules. The yellow boxes represent genes regulated by *CrebA*.

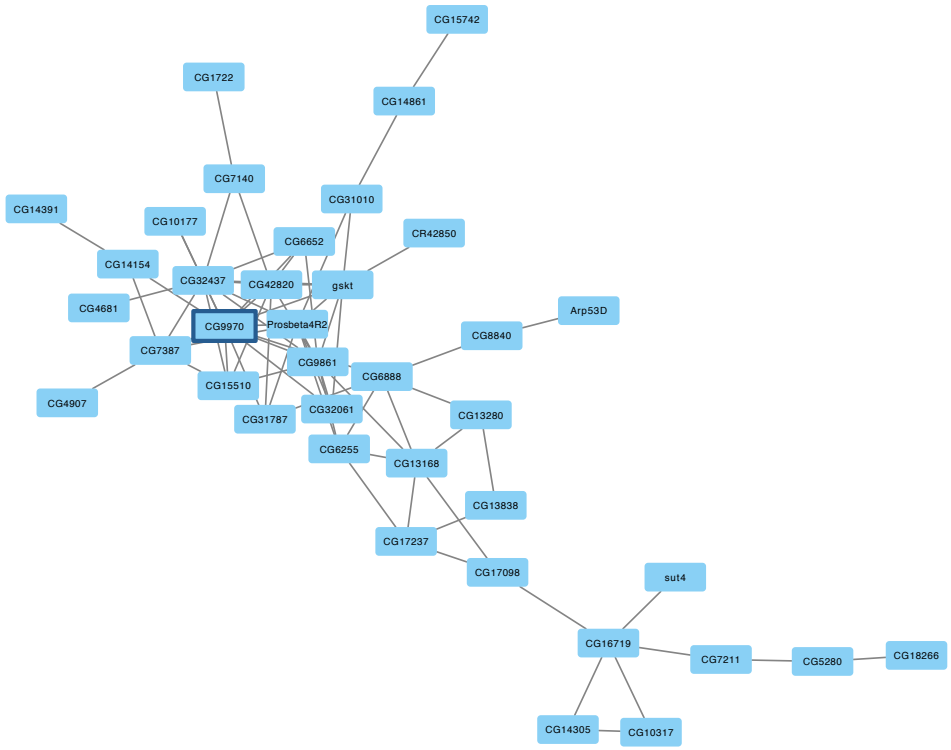


Figure 4.3: **Uncharacterized genes are clustered.**

The network module was plotted in an organic format. Each box represents a node (gene) in the module. The box with a thick blue outline marks the hub genes.

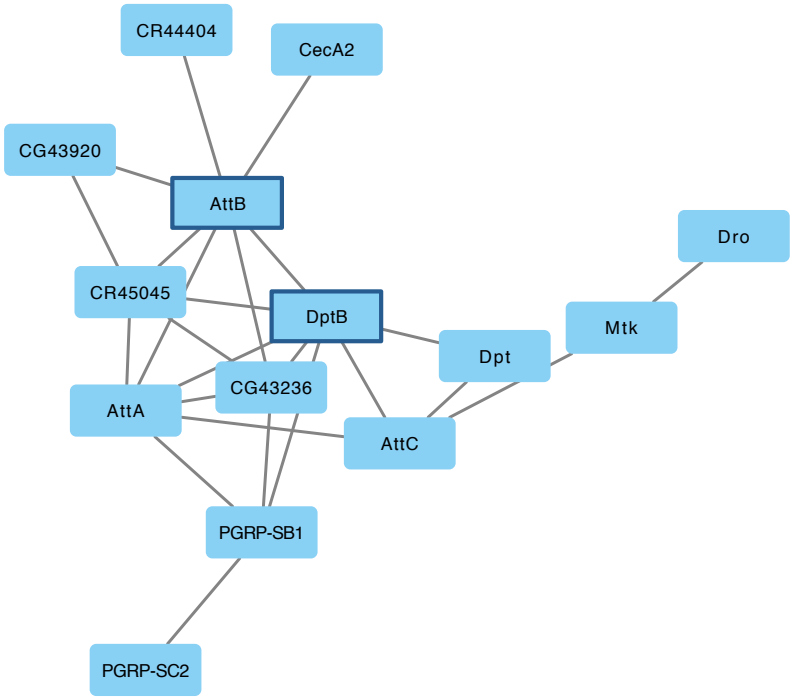


Figure 4.4: **Clustering of immune genes reveals potential function of uncharacterized genes.**

The network module was plotted in an organic format. Each box represents a node (gene) in the module. The box with a thick blue outline marks the hub genes.

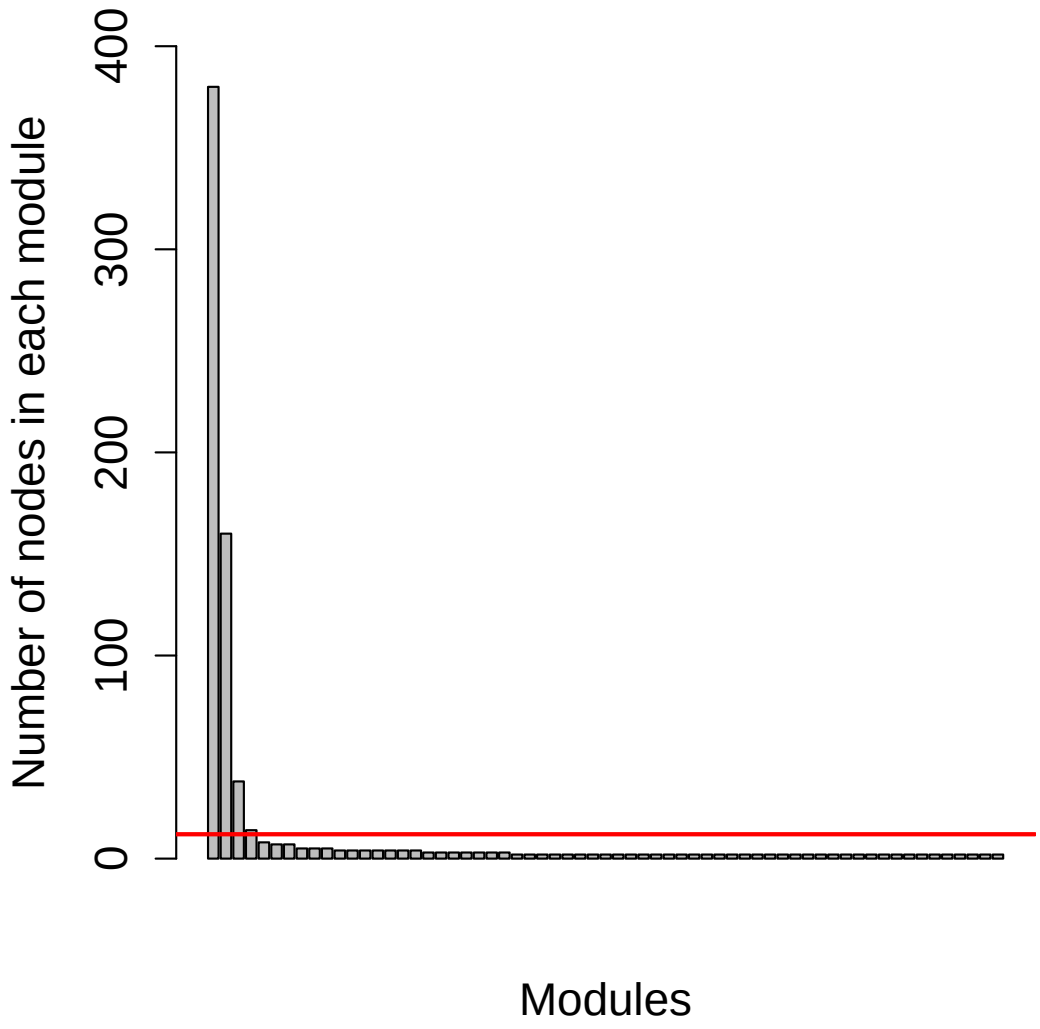


Figure 4.5: **Distribution of the size of modules.**

X-axis indicates each module and Y-axis indicates the number of genes in each module. Red line indicates the cutoff used to define co-expression modules for subsequent studies (12 genes).

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

In this dissertation, I presented research on how host-pathogen interactions may drive adaptation in host phagocytosis and autophagy genes in *Drosophila* and on how non-immune physiological responses upon infection, such as changes in metabolism and protein secretion and transport, play a crucial part in the *D. melanogaster* response to infection. These studies provide a clearer mechanistic and population genetic understanding of host-microbe interactions and establish a ground for future research in *Drosophila* immunity.

In Chapter 1, I performed population genetic analyses on phagocytosis and autophagy genes of *D. melanogaster* and *D. simulans* to evaluate their evolutionary patterns. I found signatures of recent and recurrent positive selection in several phagocytic recognition genes and reported that glutamate receptor genes and several autophagy-related (*Atg*) genes show signatures of recent selection in each species. Chapter 1 illustrates that an arms race between *Drosophila* and its pathogens may shape the evolution of the host cellular immunity.

In Chapter 2, together with a collaborator, I performed RNA-seq to measure the *D. melanogaster* transcriptomic response to 10 bacteria that span the spectrum of virulence and conducted functional genetic experiments to understand the role of a transcription factor, *CrebA*. We found that while the host regulates unique sets of genes upon infection with each bacterium, it also triggers a core transcriptional response and then identified that *CrebA* modulates infection tolerance. Chapter 2 emphasizes how host physiological responses to infection beyond the typical immune response are crucial in host

survival and homeostasis and provides a foundation for studying how each bacterium may establish infection.

In Chapter 3, I conducted co-expression network analysis on the RNA-seq data from Chapter 2 to identify gene modules that regulate the response to infection in an coordinated manner and their hub genes. This analysis resulted in total four modules of genes associated with lipid and carbohydrate metabolic process, protein secretion and transport, and AMP-based immune response. Chapter 3 illustrates that changes in metabolism and protein secretion play a crucial role in the response to infection and provides a basis for future functional genetic studies.

While much work has been done on examining the immune defense of *Drosophila*, little is known about how bacteria establish infections in the host, and specifically how differently they respond to the host immune defense. Variabilities in bacterial traits like growth rate, types of toxins, adhesins and other effector proteins secreted, and ability to form a biofilm and to establish an intracellular growth all contribute to how infection proceeds [132, 173, 220]. Co-infections with multiple pathogen species at once or one after another are common in the wild and may lead to synergistic or antagonistic interactions among bacteria or compromise the host immune response, which results in a different health outcome from that of single infections [221]. My work could be further expanded to directly compare the pathology of bacterial species that produce different survival outcomes in flies and to observe how these bacteria may behave when they are simultaneously present in the host. For instance, closely related Gram-negative bacteria, *Providencia sneebia* and *P. rettgeri*, are pathogenic to flies to a different extent since *P. sneebia* is more virulent than

P. rettgeri; infecting flies with only *P. sneebia* or with both bacteria indicated that *P. sneebia* might suppress or evade the canonical immune response [132, 173]. In order to further explore the pathology of these bacteria, I would infect flies with each bacterium or both and perform RNA-seq on bacterial RNA isolated from infected flies to identify genes differentially expressed between *P. sneebia* and *P. rettgeri*. Subsequently, I would create knockout strains of bacteria that lack those differentially expressed genes, infect the flies again, and observe how such changes may alter the infection outcome for hosts. Ultimately, a dual-seq approach over multiple time points would be useful in evaluating the response from both parties and infer host-pathogen interactions at the transcriptomics level [222]. This method has characterized host-pathogen interactions like a plant-fungi relationship [223, 224], mice-pathogen interaction [225], mosquito-worm interactions [226], and fly-*Wolbachia* system [227] but has not been used to elucidate the relationship between fly and non-symbiotic bacteria. Thus, this method would be useful in assessing the host-microbe interactions from bacteria's perspective, complementing work done in Chapter 3 and 4.

The other under-explored area of research is to predict host resistance to pathogens based on the basal gene expression of the hosts. Genetic variations lead to variability in how flies survive infection [160, 228, 229]. Yet whether variations in the constitutive expression among individuals also explain an individual's ability to overcome infection is unknown. To explore this idea, I would first measure basal gene expression across many different individual lines from *Drosophila* Genetics Reference Panel (DGRP) [230, 231] or ideally from an outbred population to avoid inbreeding depression, and the percentage of flies surviving 5 days after *P. rettgeri* infection. I would then group

individual lines into three categories based on their survival (resistant, medium, susceptible) using a cutoff determined by a pilot experiment and identify genes that are constitutively differentially expressed between resistant and susceptible individual lines pre-infection. However, differences in phenotypes are not necessarily due to differential expression of these genes. Differential expression analysis is limited to only distinguishing between already disparate groups and cannot classify groups without prior knowledge [232]. Thus, I would build a prediction model based on support vector machines (SVMs) using these differentially regulated genes as inputs to identify biomarkers that explain this difference in phenotype [232, 233]. Such model would evaluate individual transcripts and rank them based on their ability to distinguish between two states, thereby further selecting genes that may truly exert influence on the phenotype [232]. To find out whether these biomarkers are effective in predicting the infection outcome in general, I would then apply the model on the gene expression data of other populations with known survival data. No work has been done in *Drosophila* immunity so far to build and test a framework to predict susceptibility to bacterial infection based on constitutive gene expression data. With the reduction in sequencing costs and increasing interests in generating RNA-seq data using automatized machines, such large-scale dataset will soon become available to pursue these questions.

BIBLIOGRAPHY

[1] B. Lemaitre and J. Hoffmann. The host defense of *Drosophila melanogaster*. *Annu Rev Immunol*, 25:697–743, 2007.

[2] J. A. Hoffmann and J. M. Reichhart. *Drosophila* innate immunity: an evolutionary perspective. *Nat Immunol*, 3(2):121–6, 2002.

[3] R. Medzhitov, P. Preston-Hurlburt, and Jr. Janeway, C. A. A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature*, 388(6640):394–7, 1997.

[4] M. E. Woolhouse, J. P. Webster, E. Domingo, B. Charlesworth, and B. R. Levin. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat Genet*, 32(4):569–77, 2002.

[5] R. Nielsen. Molecular signatures of natural selection. *Annu Rev Genet*, 39: 197–218, 2005.

[6] R. Nielsen, I. Hellmann, M. Hubisz, C. Bustamante, and A. G. Clark. Recent and ongoing selection in the human genome. *Nat Rev Genet*, 8 (11):857–68, 2007.

[7] E. Decaestecker, S. Gaba, J. A. Raeymaekers, R. Stoks, L. Van Kerckhoven, D. Ebert, and L. De Meester. Host-parasite 'Red Queen' dynamics archived in pond sediment. *Nature*, 450(7171):870–3, 2007.

[8] T. B. Sackton, B. P. Lazzaro, T. A. Schlenke, J. D. Evans, D. Hultmark, and A. G. Clark. Dynamic evolution of the innate immune system in *Drosophila*. *Nat Genet*, 39(12):1461–8, 2007.

[9] B. Lemaitre, E. Kromer-Metzger, L. Michaut, E. Nicolas, M. Meister, P. Georgel, J. M. Reichhart, and J. A. Hoffmann. A recessive mutation, immune deficiency (*imd*), defines two distinct control pathways in the *Drosophila* host defense. *Proc Natl Acad Sci U S A*, 92(21):9465–9, 1995.

[10] B. Lemaitre, E. Nicolas, L. Michaut, J. M. Reichhart, and J. A. Hoffmann. The dorsoventral regulatory gene cassette *spatzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults. *Cell*, 86(6):973–83, 1996.

[11] E. De Gregorio, P. T. Spellman, G. M. Rubin, and B. Lemaitre. Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc Natl Acad Sci U S A*, 98(22):12590–5, 2001.

[12] P. Irving, L. Troxler, T. S. Heuer, M. Belvin, C. Kopczynski, J. M. Reichhart, J. A. Hoffmann, and C. Hetru. A genome-wide analysis of immune responses in *Drosophila*. *Proc Natl Acad Sci U S A*, 98(26):15119–24, 2001.

[13] S. Rutschmann, A. Kilinc, and D. Ferrandon. Cutting edge: the toll pathway is required for resistance to Gram-positive bacterial infections in *Drosophila*. *J Immunol*, 168(4):1542–6, 2002.

[14] V. Gobert, M. Gottar, A. A. Matskevich, S. Rutschmann, J. Royet, M. Belvin, J. A. Hoffmann, and D. Ferrandon. Dual activation of the *Drosophila* toll pathway by two pattern recognition receptors. *Science*, 302 (5653):2126–30, 2003.

[15] F. Leulier, C. Parquet, S. Pili-Floury, J. H. Ryu, M. Caroff, W. J. Lee, D. Mengin-Lecreulx, and B. Lemaitre. The *Drosophila* immune system

detects bacteria through specific peptidoglycan recognition. *Nat Immunol*, 4(5):478–84, 2003.

[16] R. H. Moy and S. Cherry. Antimicrobial autophagy: a conserved innate immune response in *Drosophila*. *J Innate Immun*, 5(5):444–55, 2013.

[17] M. Gottar, V. Gobert, A. A. Matskevich, J. M. Reichhart, C. Wang, T. M. Butt, M. Belvin, J. A. Hoffmann, and D. Ferrandon. Dual detection of fungal infections in *Drosophila* via recognition of glucans and sensing of virulence factors. *Cell*, 127(7):1425–37, 2006.

[18] M. Ming, F. Obata, E. Kuranaga, and M. Miura. Persephone/Spatzle pathogen sensors mediate the activation of Toll receptor signaling in response to endogenous danger signals in apoptosis-deficient *Drosophila*. *J Biol Chem*, 289(11):7558–68, 2014.

[19] N. Buchon, N. A. Broderick, M. Poidevin, S. Pradervand, and B. Lemaitre. *Drosophila* intestinal response to bacterial infection: activation of host defense and stem cell proliferation. *Cell Host Microbe*, 5(2):200–11, 2009.

[20] D. J. Begun and P. Whitley. Adaptive evolution of *relish*, a *Drosophila* NF-kappaB/IkappaB protein. *Genetics*, 154(3):1231–8, 2000.

[21] T. A. Schlenke and D. J. Begun. Natural selection drives *Drosophila* immune system evolution. *Genetics*, 164(4):1471–80, 2003.

[22] F. M. Jiggins and K. W. Kim. A screen for immunity genes evolving under positive selection in *Drosophila*. *J Evol Biol*, 20(3):965–70, 2007.

[23] B. P. Lazzaro. Elevated polymorphism and divergence in the class C scavenger receptors of *Drosophila melanogaster* and *D. simulans*. *Genetics*, 169(4):2023–34, 2005.

[24] F. M. Jiggins and K. W. Kim. Contrasting evolutionary patterns in *Drosophila* immune receptors. *J Mol Evol*, 63(6):769–80, 2006.

[25] L. Salazar-Jaramillo, A. Paspati, L. van de Zande, C. J. Vermeulen, T. Schwander, and B. Wertheim. Evolution of a cellular immune response in *Drosophila*: a phenotypic and genomic comparative analysis. *Genome Biol Evol*, 6(2):273–89, 2014.

[26] A. M. Early, J. R. Arguello, M. Cardoso-Moreira, S. Gottipati, J. K. Grenier, and A. G. Clark. Survey of global genetic diversity within the *Drosophila* immune system. *Genetics*, 205(1):353–366, 2017.

[27] R. S. Flannagan, G. Cosio, and S. Grinstein. Antimicrobial mechanisms of phagocytes and bacterial evasion strategies. *Nat Rev Microbiol*, 7(5): 355–66, 2009.

[28] J. Zirin and N. Perrimon. *Drosophila* as a model system to study autophagy. *Semin Immunopathol*, 32(4):363–72, 2010.

[29] J. Huang and J. H. Brumell. Bacteria-autophagy interplay: a battle for survival. *Nat Rev Microbiol*, 12(2):101–14, 2014.

[30] K. Cadwell and J. A. Philips. Autophagy meets phagocytosis. *Immunity*, 39(3):425–7, 2013.

[31] M. A. Sanjuan, C. P. Dillon, S. W. Tait, S. Moshiach, F. Dorsey, S. Connell, M. Komatsu, K. Tanaka, J. L. Cleveland, S. Withoff, and D. R. Green. Toll-like receptor signalling in macrophages links the autophagy pathway to phagocytosis. *Nature*, 450(7173):1253–7, 2007.

[32] W. Shui, L. Sheu, J. Liu, B. Smart, C. J. Petzold, T. Y. Hsieh, A. Pitcher, J. D. Keasling, and C. R. Bertozzi. Membrane proteomics of phagosomes

suggests a connection to autophagy. *Proc Natl Acad Sci U S A*, 105(44): 16952–7, 2008.

[33] M. G. Gutierrez, S. S. Master, S. B. Singh, G. A. Taylor, M. I. Colombo, and V. Deretic. Autophagy is a defense mechanism inhibiting BCG and *Mycobacterium tuberculosis* survival in infected macrophages. *Cell*, 119(6): 753–66, 2004.

[34] M. A. Sanjuan and D. R. Green. Eating for good health: linking autophagy and phagocytosis in host defense. *Autophagy*, 4(5):607–11, 2008.

[35] V. A. Fischetti. Streptococcal M protein: molecular design and biological behavior. *Clin Microbiol Rev*, 2(3):285–314, 1989.

[36] B. China, M. P. Sory, B. T. N'Guyen, M. De Bruyere, and G. R. Cornelis. Role of the YadA protein in prevention of opsonization of *Yersinia enterocolitica* by C3b molecules. *Infect Immun*, 61(8):3129–36, 1993.

[37] K. E. Beauregard, K. D. Lee, R. J. Collier, and J. A. Swanson. pH-dependent perforation of macrophage phagosomes by listeriolysin O from Listeria monocytogenes. *J Exp Med*, 186(7):1159–63, 1997.

[38] N. Ramarao, S. D. Gray-Owen, S. Backert, and T. F. Meyer. *Helicobacter pylori* inhibits phagocytosis by professional phagocytes involving type IV secretion components. *Mol Microbiol*, 37(6):1389–404, 2000.

[39] K. Pethe, D. L. Swenson, S. Alonso, J. Anderson, C. Wang, and D. G. Russell. Isolation of *Mycobacterium tuberculosis* mutants defective in the arrest of phagosome maturation. *Proc Natl Acad Sci U S A*, 101(37): 13642–7, 2004.

[40] A. Choy, J. Dancourt, B. Mugo, T. J. O'Connor, R. R. Isberg, T. J. Melia, and C. R. Roy. The *Legionella* effector RavZ inhibits host autophagy through irreversible Atg8 deconjugation. *Science*, 338(6110):1072–6, 2012.

[41] M. Chaumorcel, M. Lussignol, L. Mouna, Y. Cavignac, K. Fahie, J. Cotte-Laffitte, A. Geballe, W. Brune, I. Beau, P. Codogno, and A. Esclatine. The human cytomegalovirus protein TRS1 inhibits autophagy via its interaction with Beclin 1. *J Virol*, 86(5):2571–84, 2012.

[42] D. M. Shin, B. Y. Jeon, H. M. Lee, H. S. Jin, J. M. Yuk, C. H. Song, S. H. Lee, Z. W. Lee, S. N. Cho, J. M. Kim, R. L. Friedman, and E. K. Jo. *Mycobacterium tuberculosis* Eis regulates autophagy, inflammation, and cell death through redox-dependent signaling. *PLoS Pathog*, 6(12):e1001230, 2010.

[43] L. Van Valen. A new evolutionary law. *Evolutionary theory*, 1:1–30, 1973.

[44] R. Dawkins and J. R. Krebs. Arms races between and within species. *Proc R Soc Lond B Biol Sci*, 205(1161):489–511, 1979.

[45] S. Casillas and A. Barbadilla. Molecular population genetics. *Genetics*, 205 (3):1003–1035, 2017.

[46] J. J. Vitti, S. R. Grossman, and P. C. Sabeti. Detecting natural selection in genomic data. *Annu Rev Genet*, 47:97–120, 2013.

[47] R. L. Rogers, J. M. Cridland, L. Shao, T. T. Hu, P. Andolfatto, and K. R. Thornton. Landscape of standing variation for tandem duplications in *Drosophila yakuba* and *Drosophila simulans*. *Mol Biol Evol*, 31(7):1750–66, 2014.

[48] J. B. Lack, C. M. Cardeno, M. W. Crepeau, W. Taylor, R. B. Corbett-Detig, K. A. Stevens, C. H. Langley, and J. E. Pool. The *Drosophila* genome nexus:

A population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population. *Genetics*, 2015.

[49] Consortium Drosophila 12 Genomes, A. G. Clark, M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver, T. A. Markow, T. C. Kaufman, M. Kellis, W. Gelbart, V. N. Iyer, D. A. Pollard, T. B. Sackton, A. M. Larracuente, N. D. Singh, J. P. Abad, D. N. Abt, B. Adryan, M. Aguade, H. Akashi, W. W. Anderson, C. F. Aquadro, D. H. Ardell, R. Arguello, C. G. Artieri, D. A. Barbash, D. Barker, P. Barsanti, P. Batterham, S. Batzoglou, D. Begun, A. Bhutkar, E. Blanco, S. A. Bosak, R. K. Bradley, A. D. Brand, M. R. Brent, A. N. Brooks, R. H. Brown, R. K. Butlin, C. Caggese, B. R. Calvi, A. Bernardo de Carvalho, A. Caspi, S. Castrezana, S. E. Celniker, J. L. Chang, C. Chapple, S. Chatterji, A. Chinwalla, A. Civetta, S. W. Clifton, J. M. Comeron, J. C. Costello, J. A. Coyne, J. Daub, R. G. David, A. L. Delcher, K. Delehaunty, C. B. Do, H. Ebling, K. Edwards, T. Eickbush, J. D. Evans, A. Filipski, S. Findeiss, E. Freyhult, L. Fulton, R. Fulton, A. C. Garcia, A. Gardiner, D. A. Garfield, B. E. Garvin, G. Gibson, D. Gilbert, S. Gnerre, J. Godfrey, R. Good, V. Gotea, B. Gravely, A. J. Greenberg, S. Griffiths-Jones, S. Gross, R. Guigo, E. A. Gustafson, W. Haerty, M. W. Hahn, D. L. Halligan, A. L. Halpern, G. M. Halter, M. V. Han, A. Heger, L. Hillier, A. S. Hinrichs, I. Holmes, R. A. Hoskins, M. J. Hubisz, D. Hultmark, M. A. Huntley, D. B. Jaffe, et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, 450(7167):203–18, 2007.

[50] A. S. Fiston-Lavier, N. D. Singh, M. Lipatov, and D. A. Petrov. *Drosophila melanogaster* recombination rate calculator. *Gene*, 463(1-2):18–20, 2010.

[51] J. M. Comeron, R. Ratnappan, and S. Bailin. The many landscapes of

recombination in *Drosophila melanogaster*. *PLoS Genet*, 8(10):e1002905, 2012.

[52] R. B. Corbett-Detig and D. L. Hartl. Population genomics of inversion polymorphisms in *Drosophila melanogaster*. *PLoS Genet*, 8(12):e1003056, 2012.

[53] J. E. Pool, R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno, M. W. Crepeau, P. Duchen, J. J. Emerson, P. Saelao, D. J. Begun, and C. H. Langley. Population genomics of sub-saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genet*, 8(12):e1003080, 2012.

[54] A. R. Quinlan and I. M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–2, 2010.

[55] A. Loytynoja and N. Goldman. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A*, 102(30): 10557–62, 2005.

[56] G. A. Watterson. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, 7(2):256–276, 1975.

[57] F. Tajima. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3):585–95, 1989.

[58] J. C. Fay, G. J. Wyckoff, and C. I. Wu. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature*, 415 (6875):1024–6, 2002.

[59] W. J. Ewens. Concepts of substitutional load in finite populations. *Theoretical Population Biology*, 3(2):153–161, 1972.

[60] G. A. Watterson. Reversibility and the age of an allele. II. two-allele models, with selection and mutation. *Theor Popul Biol*, 12(2):179–96, 1977.

[61] K. Zeng, S. Mano, S. Shi, and C. I. Wu. Comparisons of site- and haplotype-frequency methods for detecting positive selection. *Mol Biol Evol*, 24(7):1562–74, 2007.

[62] K. Zeng, S. Shi, and C. I. Wu. Compound tests for the detection of hitchhiking under positive selection. *Mol Biol Evol*, 24(8):1898–908, 2007.

[63] K. Zeng, Y. X. Fu, S. Shi, and C. I. Wu. Statistical tests for detecting positive selection by utilizing high-frequency variants. *Genetics*, 174(3): 1431–9, 2006.

[64] J. H. McDonald and M. Kreitman. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, 351(6328):652–4, 1991.

[65] N. Stoletzki and A. Eyre-Walker. Estimation of the neutrality index. *Mol Biol Evol*, 28(1):63–70, 2011.

[66] Y. Benjamini, Y.; Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, 57(1):289–300, 1995.

[67] M. Lagueux, E. Perrodou, E. A. Levashina, M. Capovilla, and J. A. Hoffmann. Constitutive expression of a complement-like protein in toll and JAK gain-of-function mutants of *Drosophila*. *Proc Natl Acad Sci U S A*, 97(21):11427–32, 2000.

[68] K. Hart and M. Wilcox. A *Drosophila* gene encoding an epithelial membrane protein with homology to CD36/LIMP II. *J Mol Biol*, 234(1): 249–53, 1993.

[69] H. Agaisse, L. S. Burrack, J. A. Philips, E. J. Rubin, N. Perrimon, and D. E. Higgins. Genome-wide RNAi screen for host factors required for intracellular bacterial infection. *Science*, 309(5738):1248–51, 2005.

[70] J. A. Philips, E. J. Rubin, and N. Perrimon. *Drosophila* RNAi screen reveals CD36 family member required for mycobacterial infection. *Science*, 309 (5738):1251–3, 2005.

[71] D. J. Obbard, J. J. Welch, K. W. Kim, and F. M. Jiggins. Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet*, 5(10): e1000698, 2009.

[72] N. Fujita, T. Itoh, H. Omori, M. Fukuda, T. Noda, and T. Yoshimori. The Atg16L complex specifies the site of LC3 lipidation for membrane biogenesis in autophagy. *Mol Biol Cell*, 19(5):2092–100, 2008.

[73] O. Florey and M. Overholtzer. Autophagy proteins in macroendocytic engulfment. *Trends Cell Biol*, 22(7):374–80, 2012.

[74] A. Peltan, L. Briggs, G. Matthews, S. T. Sweeney, and D. F. Smith. Identification of *Drosophila* gene products required for phagocytosis of *Leishmania donovani*. *PLoS One*, 7(12):e51831, 2012.

[75] U. Nair, A. Jotwani, J. Geng, N. Gammoh, D. Richerson, W. L. Yen, J. Griffith, S. Nag, K. Wang, T. Moss, M. Baba, J. A. McNew, X. Jiang, F. Reggiori, T. J. Melia, and D. J. Klionsky. SNARE proteins are required for macroautophagy. *Cell*, 146(2):290–302, 2011.

[76] M. J. Williams, I. Ando, and D. Hultmark. *Drosophila melanogaster* Rac2 is necessary for a proper cellular immune response. *Genes Cells*, 10(8): 813–23, 2005.

[77] T. E. Rusten, T. Vaccari, K. Lindmo, L. M. Rodahl, I. P. Nezis, C. Sem-Jacobsen, F. Wendler, J. P. Vincent, A. Brech, D. Bilder, and H. Stenmark. ESCRTs and Fab1 regulate distinct steps of autophagy. *Curr Biol*, 17(20):1817–25, 2007.

[78] A. R. Winslow, C. W. Chen, S. Corrochano, A. Acevedo-Arozena, D. E. Gordon, A. A. Peden, M. Lichtenberg, F. M. Menzies, B. Ravikumar, S. Imarisio, S. Brown, C. J. O'Kane, and D. C. Rubinsztein. alpha-Synuclein impairs macroautophagy: implications for Parkinson's disease. *J Cell Biol*, 190(6):1023–37, 2010.

[79] N. Dong, Y. Zhu, Q. Lu, L. Hu, Y. Zheng, and F. Shao. Structurally distinct bacterial TBC-like GAPs link Arf GTPase to Rab1 inactivation to counteract host defenses. *Cell*, 150(5):1029–41, 2012.

[80] Y. Ichimura, T. Kirisako, T. Takao, Y. Satomi, Y. Shimonishi, N. Ishihara, N. Mizushima, I. Tanida, E. Kominami, M. Ohsumi, T. Noda, and Y. Ohsumi. A ubiquitin-like system mediates protein lipidation. *Nature*, 408(6811):488–92, 2000.

[81] T. Kirisako, Y. Ichimura, H. Okada, Y. Kabeya, N. Mizushima, T. Yoshimori, M. Ohsumi, T. Takao, T. Noda, and Y. Ohsumi. The reversible modification regulates the membrane-binding state of Apg8/Aut7 essential for autophagy and the cytoplasm to vacuole targeting pathway. *J Cell Biol*, 151(2):263–76, 2000.

[82] S. Takats, P. Nagy, A. Varga, K. Pircs, M. Karpati, K. Varga, A. L. Kovacs, K. Hegedus, and G. Juhasz. Autophagosomal Syntaxin17-dependent lysosomal degradation maintains neuronal function in *Drosophila*. *J Cell Biol*, 201(4):531–9, 2013.

[83] H. Augustin, Y. Grosjean, K. Chen, Q. Sheng, and D. E. Featherstone. Nonvesicular release of glutamate by glial xCT transporters suppresses glutamate receptor clustering *in vivo*. *J Neurosci*, 27(1):111–23, 2007.

[84] E. A. Gonzalez, A. Garg, J. Tang, A. E. Nazario-Toole, and L. P. Wu. A glutamate-dependent redox system in blood cells is integral for phagocytosis in *Drosophila melanogaster*. *Curr Biol*, 23(22):2319–2324, 2013.

[85] S. Nonaka, K. Nagaosa, T. Mori, A. Shiratsuchi, and Y. Nakanishi. Integrin alphaPS3/betanu-mediated phagocytosis of apoptotic cells and bacteria in *Drosophila*. *J Biol Chem*, 288(15):10374–80, 2013.

[86] P. Juneja and B. P. Lazzaro. Haplotype structure and expression divergence at the *Drosophila* cellular immune gene *eater*. *Mol Biol Evol*, 27(10):2284–99, 2010.

[87] S. Blandin and E. A. Levashina. Thioester-containing proteins and insect immunity. *Mol Immunol*, 40(12):903–8, 2004.

[88] A. Dostalova, S. Rommelaere, M. Poidevin, and B. Lemaitre. Thioester-containing proteins regulate the Toll pathway and play a role in *Drosophila* defence against microbial pathogens and parasitoid wasps. *BMC Biol*, 15(1):79, 2017.

[89] S. C. Piyankarage, H. Augustin, Y. Grosjean, D. E. Featherstone, and S. A. Shippy. Hemolymph amino acid analysis of individual *Drosophila* larvae. *Anal Chem*, 80(4):1201–7, 2008.

[90] M. Kimura. The role of compensatory neutral mutations in molecular evolution. *J Genet*, 64(7), 1985.

[91] P. Capy and P. Gibert. *Drosophila melanogaster*, *Drosophila simulans*: so similar yet so different. *Genetica*, 120(1-3):5–16, 2004.

[92] H. Akashi. Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. *Genetics*, 139(2):1067–76, 1995.

[93] T. Starr, R. Child, T. D. Wehrly, B. Hansen, S. Hwang, C. Lopez-Otin, H. W. Virgin, and J. Celli. Selective subversion of autophagy complexes facilitates completion of the *Brucella* intracellular cycle. *Cell Host Microbe*, 11(1):33–45, 2012.

[94] J. M. Smith and J. Haigh. The hitch-hiking effect of a favourable gene. *Genet Res*, 23(1):23–35, 1974.

[95] A. L. Hughes and M. Nei. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, 335(6186):167–70, 1988.

[96] J. Hermisson and P. S. Pennings. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4): 2335–52, 2005.

[97] N. R. Garud, P. W. Messer, E. O. Buzbas, and D. A. Petrov. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLoS Genet*, 11(2):e1005004, 2015.

[98] N. R. Garud and D. A. Petrov. Elevated linkage disequilibrium and signatures of soft sweeps are common in *Drosophila melanogaster*. *Genetics*, 203(2):863–80, 2016.

[99] P. W. Messer and D. A. Petrov. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol Evol*, 28(11):659–69, 2013.

[100] E. A. Stahl, G. Dwyer, R. Mauricio, M. Kreitman, and J. Bergelson. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature*, 400(6745):667–71, 1999.

[101] D. Tian, M. B. Traw, J. Q. Chen, M. Kreitman, and J. Bergelson. Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature*, 423 (6935):74–7, 2003.

[102] D. Tian, H. Araki, E. Stahl, J. Bergelson, and M. Kreitman. Signature of balancing selection in *Arabidopsis*. *Proc Natl Acad Sci U S A*, 99(17): 11525–30, 2002.

[103] A. M. Andres, M. J. Hubisz, A. Indap, D. G. Torgerson, J. D. Degenhardt, A. R. Boyko, R. N. Gutenkunst, T. J. White, E. D. Green, C. D. Bustamante, A. G. Clark, and R. Nielsen. Targets of balancing selection in the human genome. *Mol Biol Evol*, 26(12):2755–64, 2009.

[104] E. M. Leffler, Z. Gao, S. Pfeifer, L. Segurel, A. Auton, O. Venn, R. Bowden, R. Bontrop, J. D. Wall, G. Sella, P. Donnelly, G. McVean, and M. Przeworski. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science*, 339(6127):1578–82, 2013.

[105] C. H. Langley, K. Stevens, C. Cardeno, Y. C. Lee, D. R. Schrider, J. E. Pool, S. A. Langley, C. Suarez, R. B. Corbett-Detig, B. Kolaczkowski, S. Fang, P. M. Nista, A. K. Holloway, A. D. Kern, C. N. Dewey, Y. S. Song, M. W. Hahn, and D. J. Begun. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics*, 192(2):533–98, 2012.

[106] R. L. Unckless and B. P. Lazzaro. The potential for adaptive maintenance of diversity in insect antimicrobial peptides. *Philos Trans R Soc Lond B Biol Sci*, 371(1695), 2016.

[107] C. Kocks, J. H. Cho, N. Nehme, J. Ulvila, A. M. Pearson, M. Meister, C. Strom, S. L. Conto, C. Hetru, L. M. Stuart, T. Stehle, J. A. Hoffmann, J. M. Reichhart, D. Ferrandon, M. Ramet, and R. A. Ezekowitz. Eater, a transmembrane protein mediating phagocytosis of bacterial pathogens in *Drosophila*. *Cell*, 123(2):335–46, 2005.

[108] A. Guillou, K. Troha, H. Wang, N. C. Franc, and N. Buchon. The *Drosophila* CD36 homologue croquemort is required to maintain immune and gut homeostasis during development and aging. *PLoS Pathog*, 12(10): e1005961, 2016.

[109] H. G. Boman, I. Nilsson, and B. Rasmuson. Inducible antibacterial defence system in *Drosophila*. *Nature*, 237(5352):232–5, 1972.

[110] S. Tauszig, E. Jouanguy, J. A. Hoffmann, and J. L. Imler. Toll-related receptors and the control of antimicrobial peptide expression in *Drosophila*. *Proc Natl Acad Sci U S A*, 97(19):10520–5, 2000.

[111] N. Buchon, N. Silverman, and S. Cherry. Immunity in *Drosophila melanogaster* - from microbial recognition to whole-organism physiology. *Nat Rev Immunol*, 14(12):796–810, 2014.

[112] E. De Gregorio, P. T. Spellman, P. Tzou, G. M. Rubin, and B. Lemaitre. The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. *EMBO J*, 21(11):2568–79, 2002.

[113] T. Kaneko, T. Yano, K. Aggarwal, J. H. Lim, K. Ueda, Y. Oshima, C. Peach, D. Erturk-Hasdemir, W. E. Goldman, B. H. Oh, S. Kurata, and N. Silverman. PGRP-LC and PGRP-LE have essential yet distinct functions in the *Drosophila* immune response to monomeric dap-type peptidoglycan. *Nat Immunol*, 7(7):715–23, 2006.

[114] P. Matzinger. Tolerance, danger, and the extended family. *Annu Rev Immunol*, 12:991–1045, 1994.

[115] R. L. Schmidt, T. R. Trejo, T. B. Plummer, J. L. Platt, and A. H. Tang. Infection-induced proteolysis of PGRP-LC controls the IMD activation and melanization cascades in *Drosophila*. *FASEB J*, 22(3):918–29, 2008.

[116] L. El Chamy, V. Leclerc, I. Caldelari, and J. M. Reichhart. Sensing of 'danger signals' and pathogen-associated molecular patterns defines binary signaling pathways 'upstream' of Toll. *Nat Immunol*, 9(10):1165–70, 2008.

[117] M. S. Dionne, L. N. Pham, M. Shirasu-Hiza, and D. S. Schneider. *Akt* and *FOXO* dysregulation contribute to infection-induced wasting in *Drosophila*. *Curr Biol*, 16(20):1977–85, 2006.

[118] N. Buchon, N. A. Broderick, T. Kuraishi, and B. Lemaitre. *Drosophila* EGFR pathway coordinates stem cell proliferation and gut remodeling following infection. *BMC Biol*, 8:152, 2010.

[119] R. I. Clark, S. W. Tan, C. B. Pean, U. Roostalu, V. Vivancos, K. Bronda, M. Pilatova, J. Fu, D. W. Walker, R. Berdeaux, F. Geissmann, and M. S. Dionne. MEF2 is an in vivo immune-metabolic switch. *Cell*, 155(2):435–47, 2013.

[120] A. Casadevall and L. A. Pirofski. Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. *Infect Immun*, 67(8): 3703–13, 1999.

[121] A. M. Bolger, M. Lohse, and B. Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–20, 2014.

[122] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, 2013.

[123] S. Anders, P. T. Pyl, and W. Huber. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–9, 2015.

[124] M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–40, 2010.

[125] D. W. Huang, B. T. Sherman, and R. A. Lempicki. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1):44–57, 2009.

[126] H. Mi, X. Huang, A. Muruganujan, H. Tang, C. Mills, D. Kang, and P. D. Thomas. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*, 45(D1):D183–D189, 2017.

[127] C. Herrmann, B. Van de Sande, D. Potier, and S. Aerts. i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. *Nucleic Acids Res*, 40(15):e114, 2012.

[128] H. Imrichova, G. Hulselmans, Z. K. Atak, D. Potier, and S. Aerts. i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res*, 43(W1):W57–64, 2015.

[129] M. Hedengren, B. Asling, M. S. Dushay, I. Ando, S. Ekengren, M. Wihlborg, and D. Hultmark. *Relish*, a central factor in the control of humoral but not cellular immunity in *Drosophila*. *Mol Cell*, 4(5):827–37, 1999.

[130] T. Michel, J. M. Reichhart, J. A. Hoffmann, and J. Royet. *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. *Nature*, 414(6865):756–9, 2001.

[131] R. N. Germain. The art of the probable: system control in the adaptive immune system. *Science*, 293(5528):240–5, 2001.

[132] M. R. Galac and B. P. Lazzaro. Comparative pathology of bacteria in the genus *Providencia* to a natural host, *Drosophila melanogaster*. *Microbes Infect*, 13(7):673–83, 2011.

[133] K. C. Johansson, C. Metzendorf, and K. Soderhall. Microarray analysis of immune challenged *Drosophila* hemocytes. *Exp Cell Res*, 305(1):145–55, 2005.

[134] R. I. Clark, K. J. Woodcock, F. Geissmann, C. Trouillet, and M. S. Dionne. Multiple TGF-beta superfamily signals modulate the adult *Drosophila* immune response. *Curr Biol*, 21(19):1672–7, 2011.

[135] A. M. Keestra-Gounder, M. X. Byndloss, N. Seyffert, B. M. Young, A. Chavez-Arroyo, A. Y. Tsai, S. A. Cevallos, M. G. Winter, O. H. Pham, C. R. Tiffany, M. F. de Jong, T. Kerrinnes, R. Ravindran, P. A. Luciw, S. J.

McSorley, A. J. Baumler, and R. M. Tsolis. NOD1 and NOD2 signalling links ER stress with inflammation. *Nature*, 532(7599):394–7, 2016.

[136] X. Meng, B. S. Khanuja, and Y. T. Ip. Toll receptor-mediated *Drosophila* immune response requires Dif, an NF-kappaB factor. *Genes Dev*, 13(7): 792–7, 1999.

[137] S. Barbosa, G. Fasanella, S. Carreira, M. Llarena, R. Fox, C. Barreca, D. Andrew, and P. O'Hare. An orchestrated program regulating secretory pathway genes and cargos by the transmembrane transcription factor CREB-H. *Traffic*, 14(4):382–98, 2013.

[138] R. M. Fox, C. D. Hanlon, and D. J. Andrew. The CrebA/Creb3-like transcription factors are major and direct regulators of secretory capacity. *J Cell Biol*, 191(3):479–92, 2010.

[139] K. Cartharius, K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M. Frisch, M. Bayerlein, and T. Werner. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*, 21(13):2933–42, 2005.

[140] P. Georgel, S. Naitza, C. Kappler, D. Ferrandon, D. Zachary, C. Swimmer, C. Kopczynski, G. Duyk, J. M. Reichhart, and J. A. Hoffmann. *Drosophila* immune deficiency (IMD) is a death domain protein that activates antibacterial defense and can promote apoptosis. *Dev Cell*, 1(4):503–14, 2001.

[141] S. Pili-Floury, F. Leulier, K. Takahashi, K. Saigo, E. Samain, R. Ueda, and B. Lemaitre. *In vivo* RNA interference analysis reveals an unexpected

role for GNBP1 in the defense against Gram-positive bacterial infection in *Drosophila* adults. *J Biol Chem*, 279(13):12848–53, 2004.

[142] R. E. Rose, N. M. Gallaher, D. J. Andrew, R. H. Goodman, and S. M. Smolik. The CRE-binding protein dCREB-A is required for *Drosophila* embryonic development. *Genetics*, 146(2):595–606, 1997.

[143] J. S. Ayres and D. S. Schneider. Tolerance of infections. *Annu Rev Immunol*, 30:271–94, 2012.

[144] D. Duneau, J. B. Ferdy, J. Revah, H. Kondolf, G. A. Ortiz, B. P. Lazzaro, and N. Buchon. Stochastic variation in the initial phase of bacterial infection predicts the probability of survival in *D. melanogaster*. *Elife*, 6, 2017.

[145] K. A. Moore and J. Hollien. The unfolded protein response in secretory cell function. *Annu Rev Genet*, 46:165–83, 2012.

[146] H. D. Ryoo. *Drosophila* as a model for unfolded protein response research. *BMB Rep*, 48(8):445–53, 2015.

[147] H. Yoshida, T. Matsui, A. Yamamoto, T. Okada, and K. Mori. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell*, 107(7):881–91, 2001.

[148] A. van Schadewijk, E. F. van't Wout, J. Stolk, and P. S. Hiemstra. A quantitative method for detection of spliced X-box binding protein-1 (XBP1) mRNA as a measure of endoplasmic reticulum (ER) stress. *Cell Stress Chaperones*, 17(2):275–9, 2012.

[149] H. Yoshida, T. Matsui, N. Hosokawa, R. J. Kaufman, K. Nagata, and K. Mori. A time-dependent phase shift in the mammalian unfolded protein response. *Dev Cell*, 4(2):265–71, 2003.

[150] H. D. Ryoo, J. Li, and M. J. Kang. *Drosophila* XBP1 expression reporter marks cells under endoplasmic reticulum stress and with high protein secretory load. *PLoS One*, 8(9):e75774, 2013.

[151] Y. Demay, J. Perochon, S. Szuplewski, B. Mignotte, and S. Gaumer. The PERK pathway independently triggers apoptosis and a Rac1/Slpr/JNK/Dilp8 signaling favoring tissue homeostasis in a chronic ER stress *Drosophila* model. *Cell Death Dis*, 5:e1452, 2014.

[152] C. Y. Chow, F. W. Avila, A. G. Clark, and M. F. Wolfner. Induction of excessive endoplasmic reticulum stress in the *Drosophila* male accessory gland results in infertility. *PLoS One*, 10(3):e0119386, 2015.

[153] H. Xu, W. Xu, H. Xi, W. Ma, Z. He, and M. Ma. The ER luminal binding protein (BiP) alleviates Cd(2+)-induced programmed cell death through endoplasmic reticulum stress-cell death signaling pathway in tobacco cells. *J Plant Physiol*, 170(16):1434–41, 2013.

[154] E. Szegezdi, S. E. Logue, A. M. Gorman, and A. Samali. Mediators of endoplasmic reticulum stress-induced apoptosis. *EMBO Rep*, 7(9):880–5, 2006.

[155] B. A. Hay, T. Wolff, and G. M. Rubin. Expression of baculovirus P35 prevents cell death in *Drosophila*. *Development*, 120(8):2121–9, 1994.

[156] B. Lemaitre, J. M. Reichhart, and J. A. Hoffmann. *Drosophila* host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. *Proc Natl Acad Sci U S A*, 94(26): 14614–9, 1997.

[157] Y. Engstrom. Induction and regulation of antimicrobial peptides in *Drosophila*. *Dev Comp Immunol*, 23(4-5):345–58, 1999.

[158] M. Hedengren-Olcott, M. C. Olcott, D. T. Mooney, S. Ekengren, B. L. Geller, and B. J. Taylor. Differential activation of the NF-kappaB-like factors Relish and Dif in *Drosophila melanogaster* by fungi and Gram-positive bacteria. *J Biol Chem*, 279(20):21121–7, 2004.

[159] C. R. Stenbak, J. H. Ryu, F. Leulier, S. Pili-Floury, C. Parquet, M. Herve, C. Chaput, I. G. Boneca, W. J. Lee, B. Lemaitre, and D. Mengin-Lecreulx. Peptidoglycan molecular requirements allowing detection by the *Drosophila* immune deficiency pathway. *J Immunol*, 173 (12):7339–48, 2004.

[160] V. M. Howick and B. P. Lazzaro. Genotype and diet shape resistance and tolerance across distinct phases of bacterial infection. *BMC Evol Biol*, 14 (1):56, 2014.

[161] T. Lefevre, A. J. Williams, and J. C. de Roode. Genetic variation in resistance, but not tolerance, to a protozoan parasite in the monarch butterfly. *Proc Biol Sci*, 278(1706):751–9, 2011.

[162] D. Wong, D. Bazopoulou, N. Pujol, N. Tavernarakis, and J. J. Ewbank. Genome-wide investigation reveals pathogen-specific and shared signatures in the response of *Caenorhabditis elegans* to infection. *Genome Biol*, 8(9):R194, 2007.

[163] V. Doublet, Y. Poeschl, A. Gogol-Doring, C. Alaux, D. Annoscia, C. Aurori, S. M. Barribeau, O. C. Bedoya-Reina, M. J. Brown, J. C. Bull, M. L. Flenniken, D. A. Galbraith, E. Genersch, S. Gisder, I. Grosse, H. L. Holt,

D. Hultmark, H. M. Lattorff, Y. Le Conte, F. Manfredini, D. P. McMahon, R. F. Moritz, F. Nazzi, E. L. Nino, K. Nowick, R. P. van Rij, R. J. Paxton, and C. M. Grozinger. Unity in defence: honeybee workers exhibit conserved molecular responses to diverse pathogens. *BMC Genomics*, 18(1):207, 2017.

[164] O. W. Stockhammer, A. Zakrzewska, Z. Hegedus, H. P. Spaink, and A. H. Meijer. Transcriptome profiling and functional analyses of the zebrafish embryonic innate immune response to *Salmonella* infection. *J Immunol*, 182 (9):5641–53, 2009.

[165] N. Srinivasan, O. Gordon, S. Ahrens, A. Franz, S. Deddouche, P. Chakravarty, D. Phillips, A. A. Yunus, M. K. Rosen, R. S. Valente, L. Teixeira, B. Thompson, M. S. Dionne, W. Wood, and C. Reis e Sousa. Actin is an evolutionarily-conserved damage-associated molecular pattern that signals tissue injury in *Drosophila melanogaster*. *Elife*, 5, 2016.

[166] A. T. Tate and A. L. Graham. Dissecting the contributions of time and microbe density to variation in immune gene expression. *Proc Biol Sci*, 284(1859), 2017.

[167] K. Zhang, X. Shen, J. Wu, K. Sakaki, T. Saunders, D. T. Rutkowski, S. H. Back, and R. J. Kaufman. Endoplasmic reticulum stress activates cleavage of CREBH to induce a systemic inflammatory response. *Cell*, 124(3): 587–99, 2006.

[168] A. J. Shaywitz and M. E. Greenberg. CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. *Annu Rev Biochem*, 68:821–61, 1999.

[169] C. E. Richardson, T. Kooistra, and D. H. Kim. An essential role for XBP-1

in host protection against immune activation in *C. elegans*. *Nature*, 463 (7284):1092–5, 2010.

[170] O. Tirosh, Y. Cohen, A. Shitrit, O. Shani, V. T. Le-Trilling, M. Trilling, G. Friedlander, M. Tanenbaum, and N. Stern-Ginossar. The transcription and translation landscapes during human Cytomegalovirus infection reveal novel host-pathogen interactions. *PLoS Pathog*, 11(11):e1005288, 2015.

[171] L. Raberg. How to live with the enemy: understanding tolerance to parasites. *PLoS Biol*, 12(11):e1001989, 2014.

[172] R. Medzhitov, D. S. Schneider, and M. P. Soares. Disease tolerance as a defense strategy. *Science*, 335(6071):936–41, 2012.

[173] K. Troha, J. H. Im, J. Revah, B. P. Lazzaro, and N. Buchon. Comparative transcriptomics reveals *CrebA* as a novel regulator of infection tolerance in *D. melanogaster*. *PLoS Pathog*, 14(2):e1006847, 2018.

[174] S. Oliver. Guilt-by-association goes global. *Nature*, 403(6770):601–3, 2000.

[175] F. Gao, B. C. Foat, and H. J. Bussemaker. Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. *BMC Bioinformatics*, 5:31, 2004.

[176] A. L. Barabasi and Z. N. Oltvai. Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 5(2):101–13, 2004.

[177] S. van Dam, U. Vosa, A. van der Graaf, L. Franke, and J. P. de Magalhaes. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform*, 2017.

[178] R. Albert, H. Jeong, and A. L. Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–82, 2000.

[179] S. F. Levy and M. L. Siegal. Network hubs buffer environmental variation in *Saccharomyces cerevisiae*. *PLoS Biol*, 6(11):e264, 2008.

[180] E. Zotenko, J. Mestre, D. P. O'Leary, and T. M. Przytycka. Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLoS Comput Biol*, 4(8):e1000140, 2008.

[181] D. Coman, P. Rutimann, and W. Gruissem. A flexible protocol for targeted gene co-expression network analysis. *Methods Mol Biol*, 1153:285–99, 2014.

[182] C. M. Gachon, M. Langlois-Meurinne, Y. Henry, and P. Saindrenan. Transcriptional co-regulation of secondary metabolism enzymes in *Arabidopsis*: functional and evolutionary implications. *Plant Mol Biol*, 58 (2):229–45, 2005.

[183] J. T. Rosenkrantz, H. Aarts, T. Abee, M. D. Rolfe, G. M. Knudsen, M. B. Nielsen, L. E. Thomsen, M. H. Zwietering, J. E. Olsen, and C. Pin. Non-essential genes form the hubs of genome scale protein function and environmental gene expression networks in *Salmonella enterica* serovar typhimurium. *BMC Microbiol*, 13:294, 2013.

[184] W. C. Chou, A. L. Cheng, M. Brotto, and C. Y. Chuang. Visual gene-network analysis reveals the cancer gene co-expression in human endometrial cancer. *BMC Genomics*, 15:300, 2014.

[185] S. Rodius, G. Androsova, L. Gotz, R. Liechti, I. Crespo, S. Merz, P. V. Nazarov, N. de Klein, C. Jeanty, J. M. Gonzalez-Rosa, A. Muller,

F. Bernardin, S. P. Niclou, L. Vallar, N. Mercader, M. Ibberson, I. Xenarios, and F. Azuaje. Analysis of the dynamic co-expression network of heart regeneration in the zebrafish. *Sci Rep*, 6:26822, 2016.

[186] B. Zhang and S. Horvath. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*, 4:Article17, 2005.

[187] P. Langfelder and S. Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9:559, 2008.

[188] P. Langfelder, B. Zhang, and S. Horvath. Defining clusters from a hierarchical cluster tree: the *Dynamic* Tree Cut package for R. *Bioinformatics*, 24(5):719–20, 2008.

[189] G.; Nepusz T. Csardi. The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695, 2006.

[190] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11):2498–504, 2003.

[191] Y. Assenov, F. Ramirez, S. E. Schelhorn, T. Lengauer, and M. Albrecht. Computing topological parameters of biological networks. *Bioinformatics*, 24(2):282–4, 2008.

[192] J. Dong and S. Horvath. Understanding network concepts in modules. *BMC Syst Biol*, 1:24, 2007.

[193] J. Yoon, A. Blumer, and K. Lee. An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. *Bioinformatics*, 22(24):3106–8, 2006.

[194] J. X. Hu, C. E. Thomas, and S. Brunak. Network biology concepts in complex disease comorbidities. *Nat Rev Genet*, 17(10):615–29, 2016.

[195] J. Wang, J. Zhong, G. Chen, M. Li, F. X. Wu, and Y. Pan. ClusterViz: A Cytoscape app for cluster analysis of biological network. *IEEE/ACM Trans Comput Biol Bioinform*, 12(4):815–22, 2015.

[196] J. Li, W. Ma, P. Zeng, J. Wang, B. Geng, J. Yang, and Q. Cui. LncTar: a tool for predicting the RNA targets of long noncoding RNAs. *Brief Bioinform*, 16(5):806–12, 2015.

[197] S. Hunter, R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A. F. Quinn, J. D. Selengut, C. J. Sigrist, M. Thimma, P. D. Thomas, F. Valentin, D. Wilson, C. H. Wu, and C. Yeats. InterPro: the integrative protein signature database. *Nucleic Acids Res*, 37 (Database issue):D211–5, 2009.

[198] R. C. McLeay and T. L. Bailey. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics*, 11:165, 2010.

[199] R. Ziegler and R. Van Antwerpen. Lipid uptake by insect oocytes. *Insect Biochem Mol Biol*, 36(4):264–72, 2006.

[200] Y. L. Huang, J. Morales-Rosado, J. Ray, T. G. Myers, T. Kho, M. Lu, and R. S. Munford. Toll-like receptor agonists promote prolonged triglyceride storage in macrophages. *J Biol Chem*, 289(5):3001–12, 2014.

[201] T. Saitoh, T. Satoh, N. Yamamoto, S. Uematsu, O. Takeuchi, T. Kawai, and S. Akira. Antiviral protein Viperin promotes Toll-like receptor 7- and Toll-like receptor 9-mediated type I interferon production in plasmacytoid dendritic cells. *Immunity*, 34(3):352–63, 2011.

[202] A. B. Barletta, L. R. Alves, M. C. Silva, S. Sim, G. Dimopoulos, S. Liechocki, C. M. Maya-Monteiro, and M. H. Sorgine. Emerging role of lipid droplets in *Aedes aegypti* immune response against bacteria and Dengue virus. *Sci Rep*, 6:19928, 2016.

[203] P. Anand, S. Cermelli, Z. Li, A. Kassan, M. Bosch, R. Sigua, L. Huang, A. J. Ouellette, A. Pol, M. A. Welte, and S. P. Gross. A novel role for lipid droplets in the organismal antibacterial response. *Elife*, 1:e00003, 2012.

[204] S. Gronke, M. Beller, S. Fellert, H. Ramakrishnan, H. Jackle, and R. P. Kuhnlein. Control of fat storage by a *Drosophila* PAT domain protein. *Curr Biol*, 13(7):603–6, 2003.

[205] R. Birner-Gruenberger, I. Bickmeyer, J. Lange, P. Hehlert, A. Hermetter, M. Kollroser, G. N. Rechberger, and R. P. Kuhnlein. Functional fat body proteomics and gene targeting reveal *in vivo* functions of *Drosophila melanogaster alpha-Esterase-7*. *Insect Biochem Mol Biol*, 42(3):220–9, 2012.

[206] H. A. Saka and R. Valdivia. Emerging roles for lipid droplets in immunity and host-pathogen interactions. *Annu Rev Cell Dev Biol*, 28:411–37, 2012.

[207] A. Martin-Montalvo, Y. Sun, A. Diaz-Ruiz, A. Ali, V. Gutierrez, H. H. Palacios, J. Curtis, E. Siendones, J. Ariza, G. A. Abulwerdi, X. Sun, A. X. Wang, K. J. Pearson, K. W. Fishbein, R. G. Spencer, M. Wang, X. Han, M. Scheibye-Knudsen, J. A. Baur, H. G. Shertzer, P. Navas, J. M. Villalba,

S. Zou, M. Bernier, and R. de Cabo. Cytochrome *b5* reductase and the control of lipid metabolism and healthspan. *NPJ Aging Mech Dis*, 2:16006, 2016.

[208] J. Mattila and V. Hietakangas. Regulation of carbohydrate energy metabolism in *Drosophila melanogaster*. *Genetics*, 207(4):1231–1253, 2017.

[209] M. C. Chambers, K. H. Song, and D. S. Schneider. *Listeria monocytogenes* infection causes metabolic shifts in *Drosophila melanogaster*. *PLoS One*, 7 (12):e50679, 2012.

[210] M. H. Lentjes, H. E. Niessen, Y. Akiyama, A. P. de Bruine, V. Melotte, and M. van Engeland. The emerging role of GATA transcription factors in development and disease. *Expert Rev Mol Med*, 18:e3, 2016.

[211] R. Ugrankar, E. Berglund, F. Akdemir, C. Tran, M. S. Kim, J. Noh, R. Schneider, B. Ebert, and J. M. Graff. *Drosophila* glucome screening identifies Ck1alpha as a regulator of mammalian glucose metabolism. *Nat Commun*, 6:7102, 2015.

[212] L. Kadalayil, U. M. Petersen, and Y. Engstrom. Adjacent GATA and kappa B-like motifs regulate the expression of a *Drosophila* immune gene. *Nucleic Acids Res*, 25(6):1233–9, 1997.

[213] B. Zhang, M. A. Cunningham, W. C. Nichols, J. A. Bernat, U. Seligsohn, S. W. Pipe, J. H. McVey, U. Schulte-Overberg, N. B. de Bosch, A. Ruiz-Saez, G. C. White, E. G. Tuddenham, R. J. Kaufman, and D. Ginsburg. Bleeding due to disruption of a cargo-specific ER-to-Golgi transport complex. *Nat Genet*, 34(2):220–5, 2003.

[214] S. Lall and N. H. Patel. Conservation and divergence in molecular mechanisms of axis formation. *Annu Rev Genet*, 35:407–37, 2001.

[215] L. R. Otake, P. Scamborova, C. Hashimoto, and J. A. Steitz. The divergent U12-type spliceosome is required for pre-mRNA splicing and is essential for development in *Drosophila*. *Mol Cell*, 9(2):439–46, 2002.

[216] K. A. Brogden. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol*, 3(3):238–50, 2005.

[217] Y. Long, X. Wang, D. T. Youmans, and T. R. Cech. How do lncRNAs regulate transcription? *Sci Adv*, 3(9):eaao2110, 2017.

[218] S. Carpenter, D. Aiello, M. K. Atianand, E. P. Ricci, P. Gandhi, L. L. Hall, M. Byron, B. Monks, M. Henry-Bezy, J. B. Lawrence, L. A. O'Neill, M. J. Moore, D. R. Caffrey, and K. A. Fitzgerald. A long noncoding RNA mediates both activation and repression of immune response genes. *Science*, 341(6147):789–92, 2013.

[219] Z. Li, T. C. Chao, K. Y. Chang, N. Lin, V. S. Patil, C. Shimizu, S. R. Head, J. C. Burns, and T. M. Rana. The long noncoding RNA *THRIL* regulates TNFalpha expression through its interaction with hnRNPL. *Proc Natl Acad Sci U S A*, 111(3):1002–7, 2014.

[220] J. W. Wilson, M. J. Schurr, C. L. LeBlanc, R. Ramamurthy, K. L. Buchanan, and C. A. Nickerson. Mechanisms of bacterial pathogenicity. *Postgrad Med J*, 78(918):216–24, 2002.

[221] L. Pasman. The complication of coinfection. *Yale J Biol Med*, 85(1):127–32, 2012.

[222] A. J. Westermann, S. A. Gorski, and J. Vogel. Dual RNA-seq of pathogen and host. *Nat Rev Microbiol*, 10(9):618–30, 2012.

[223] P. J. Teixeira, D. P. Thomazella, O. Reis, P. F. do Prado, M. C. do Rio, G. L. Fiorin, J. Jose, G. G. Costa, V. A. Negri, J. M. Mondego, P. Mieczkowski, and G. A. Pereira. High-resolution transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. *Plant Cell*, 26(11):4245–69, 2014.

[224] D. Camilios-Neto, P. Bonato, R. Wassem, M. Z. Tadra-Sfeir, L. C. Brusamarello-Santos, G. Valdameri, L. Donatti, H. Faoro, V. A. Weiss, L. S. Chubatsu, F. O. Pedrosa, and E. M. Souza. Dual RNA-seq transcriptional analysis of wheat roots colonized by *Azospirillum brasilense* reveals up-regulation of nutrient acquisition and cell cycle genes. *BMC Genomics*, 15:378, 2014.

[225] K. J. Pittman, M. T. Aliota, and L. J. Knoll. Dual transcriptional profiling of mice and *Toxoplasma gondii* during acute and chronic infection. *BMC Genomics*, 15:806, 2014.

[226] Y. J. Choi, M. T. Aliota, G. F. Mayhew, S. M. Erickson, and B. M. Christensen. Dual RNA-seq of parasite and host reveals gene expression dynamics during filarial worm-mosquito interactions. *PLoS Negl Trop Dis*, 8(5):e2905, 2014.

[227] N. Kumar, M. Lin, X. Zhao, S. Ott, I. Santana-Cruz, S. Daugherty, Y. Rikihisa, L. Sadzewicz, L. J. Tallon, C. M. Fraser, and J. C. Dunning Hotopp. Efficient enrichment of bacterial mRNA from host-bacteria total RNA samples. *Sci Rep*, 6:34850, 2016.

[228] R. L. Unckless, S. M. Rottschaefer, and B. P. Lazzaro. The complex contributions of genetics and nutrition to immunity in *Drosophila melanogaster*. *PLoS Genet*, 11(3):e1005030, 2015.

[229] V. M. Howick and B. P. Lazzaro. The genetic architecture of defence as resistance to and tolerance of bacterial infection in *Drosophila melanogaster*. *Mol Ecol*, 26(6):1533–1546, 2017.

[230] T. F. Mackay, S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, Y. Han, M. M. Magwire, J. M. Cridland, M. F. Richardson, R. R. Anholt, M. Barron, C. Bess, K. P. Blankenburg, M. A. Carbone, D. Castellano, L. Chaboub, L. Duncan, Z. Harris, M. Javaid, J. C. Jayaseelan, S. N. Jhangiani, K. W. Jordan, F. Lara, F. Lawrence, S. L. Lee, P. Librado, R. S. Linheiro, R. F. Lyman, A. J. Mackey, M. Munidasa, D. M. Muzny, L. Nazareth, I. Newsham, L. Perales, L. L. Pu, C. Qu, M. Ramia, J. G. Reid, S. M. Rollmann, J. Rozas, N. Saada, L. Turlapati, K. C. Worley, Y. Q. Wu, A. Yamamoto, Y. Zhu, C. M. Bergman, K. R. Thornton, D. Mittelman, and R. A. Gibbs. The *Drosophila melanogaster* Genetic Reference Panel. *Nature*, 482(7384):173–8, 2012.

[231] W. Huang, A. Massouras, Y. Inoue, J. Peiffer, M. Ramia, A. M. Tarone, L. Turlapati, T. Zichner, D. Zhu, R. F. Lyman, M. M. Magwire, K. Blankenburg, M. A. Carbone, K. Chang, L. L. Ellis, S. Fernandez, Y. Han, G. Highnam, C. E. Hjelmen, J. R. Jack, M. Javaid, J. Jayaseelan, D. Kalra, S. Lee, L. Lewis, M. Munidasa, F. Ongeri, S. Patel, L. Perales, A. Perez, L. Pu, S. M. Rollmann, R. Ruth, N. Saada, C. Warner, A. Williams, Y. Q. Wu, A. Yamamoto, Y. Zhang, Y. Zhu, R. R. Anholt, J. O. Korbel, D. Mittelman, D. M. Muzny, R. A. Gibbs, A. Barbadilla, J. S. Johnston, E. A.

Stone, S. Richards, B. Deplancke, and T. F. Mackay. Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res*, 24(7):1193–208, 2014.

[232] M. Kebschull and P. N. Papapanou. Exploring genome-wide expression profiles using machine learning techniques. *Methods Mol Biol*, 1537: 347–364, 2017.

[233] J. Christianini, N.; Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, Cambridge, 2000.