bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The mode of expression divergence in *Drosophila* fat body is infection-specific

Bryan A. Ramirez-Corona, Stephanie Fruth, Oluchi Ofoegbu, Zeba Wunderlich*

Department of Developmental and Cell Biology, University of California, Irvine, CA

*Corresponding Author:
4107 Natural Sciences 2
University of California, Irvine
Irvine, CA 92697
949-824-5959
zeba@uci.edu

Running title: Expression divergence in the *Drosophila* fat body

Keywords: innate immune response; expression divergence; *Drosophila melanogaster*; fat body

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Abstract

Transcription is controlled by the interactions of *cis*-acting DNA elements with diffusible *trans*-acting factors. Changes in *cis* or *trans* factors can drive expression divergence within and between species, and the relative prevalence of each can reveal the evolutionary history and pressures that drive expression variation. Previous work delineating the mode of expression divergence in animals has largely used whole body expression measurements in a single condition. Since *cis*-acting elements often drive expression in a subset of cell types or conditions, these measurements may not capture the complete contribution of *cis*-acting changes. Here, we quantify the mode of expression divergence in the *Drosophila* fat body, the primary immune organ, in several conditions. We performed allele-specific expression analysis using two geographically distinct lines of *D. melanogaster* and their F1 hybrids. We measured expression in the absence of infection and in separate infections with Gram-negative *S. marcescens* or Gram-positive *E. faecalis* bacteria, which trigger the two primary signaling pathways in the *Drosophila* innate immune response. The mode of expression divergence strongly depends on the condition, with *trans*-acting effects dominating in response to Gram-positive infection and *cis*-acting effects dominating in Gram-negative and pre-infection conditions. Expression divergence in several receptor proteins may underlie the infection-specific *trans* effects. Before infection, when the fat body has a metabolic role, there are many compensatory effects, changes in *cis* and *trans* that counteract each other to maintain expression levels. This work demonstrates that within a single tissue, the mode of expression divergence varies between conditions and suggests that these differences reflect the diverse evolutionary histories of host-pathogen interactions.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

Differences in gene expression are believed to be major drivers of phenotypic divergence in closely related species (King and Wilson 1975). These differences can arise through sequence changes in *cis*-regulatory elements, such as enhancers, or in the coding regions of *trans*-acting factors, such as transcription factors. Evolutionary processes rely on both changes in *cis* and changes in *trans*, and the prevalence and relative contributions of *cis* and *trans* changes are actively being explored in various model systems (Signor and Nuzhdin 2018). For example, within individual *Drosophila melanogaster* lines or between *Drosophila* species, the contributions of *cis*-acting changes generally increase with phylogenetic distance, and the precise balance of *cis* versus *trans* effects depend on the phylogenetic relationships and demographics of the genotypes being compared (Wittkopp et al., 2004, Wittkopp et al., 2008, McManus et al., 2010, Coolon et al., 2014, Osada et al., 2017). These studies have elucidated mode and tempo of the evolutionary processes driving gene expression divergence; however, most of these studies use whole body measurements, thus averaging signal across multiple tissue and cell types. Therefore, these studies cannot examine the prevalence of *cis* and *trans* changes in specific biological processes, which may be subject to different types of selection pressure. In addition, given that many *cis*-regulatory elements act in a tissue-specific manner, studies that measure *cis* and *trans* effects with tissue-specific resolution may reveal effects that are undetectable in heterogenous samples.

Drosophila have an innate, but not adaptative, immune response, and this response is a powerful system for measuring the contributions of *cis* and *trans* changes for several reasons. First, the immune response is inducible, with active and inactive states. This allows for the clear delineation of the transcriptional response of the immune system from that of other processes. Second, the fat body within the immune system is an optimal tissue for study. Though other tissues participate in the immune system, the fat body is a primary driver of the humoral response (Buchon et al., 2014) and is relatively easy to isolate. Lastly, there is ample variation in the resistance, survival, and transcriptional response to infection between individual *D. melanogaster* lines (Lazzaro et al., 2004, Lazzaro et al., 2006, Sackton et al., 2010, Hotson and Schneider 2015), suggesting there are many sequence changes driving these differences.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

To quantify changes in *cis* and *trans* that drive transcriptional divergence in the immune response, we used allele-specific expression analysis (ASE) of RNA-seq data (Wittkopp et al. 2004, Signor and Nuzhdin 2018, Frochaux et al 2020). In this approach, we compare a gene's expression levels in two parental lines to the expression levels of each parental allele in the resulting F1 hybrids. Differences in expression due to changes in *cis*, for example a sequence change in a promoter or enhancer, will only affect the expression of the corresponding parental allele. Thus, changes in *cis* are independent of cellular environment and will be observed as allelic imbalance between the parents that is then maintained in the hybrids. Differences in *trans*, for example a coding sequence change in a transcription factor, will affect the expression of both alleles in the F1 hybrids and thus will be observed as differential expression in the parental lines that is not maintained in the F1 hybrids. Combining ASE with RNA-seq allows us to determine the prevalence of *cis* and *trans* changes genome-wide.

When comparing the innate immune response of different *D. melanogaster* lines, it is not clear whether *cis* or *trans* changes will dominate. Changes in *cis* generally affect a single gene's expression and thus may easily tolerated, as they only introduce small amounts of phenotypic variation into a system. Changes in *trans* can affect the expression of many genes at once and thus efficiently introduce a large amount of phenotypic variation but may be harder for the organism to tolerate. However, the specific biology of the innate immune response may temper this expectation. Antimicrobial peptides (AMPs) are among the most highly up-regulated genes in response to infection; however, changes to their expression, and even deletion of individual AMP genes, often have little to no measurable effect of infection survival (Hanson et al., 2019). This suggests that to get an appreciable phenotypic effect, synchronous changes in gene expression are required, which can result from a single change in *trans*-acting factors. In addition, previous work has suggested that within *Drosophila melanogaster* lines, *trans* changes are typically more prevalent (Wittkopp et al., 2008, Coolon et al., 2014).

To measure the contributions of *cis* and *trans*-acting changes in the *Drosophila* innate immune response, we measured fat body gene expression in two sequenced inbred *Drosophila* lines and their F1 hybrids in control and infection conditions. We separately infected the animals with either Gram-positive

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Enterococcus faecalis or the Gram-negative *Serratia marcescens* to trigger the two primary immune signaling pathways in the fly. Using ASE analysis, we quantified the contribution of *cis* and *trans* effects in the control and in each infection condition and found that *cis* effects dominated the expression divergence among control and *Enterococcus faecalis*-infected samples, while *trans* effects were dominant in the *Serratia marcescens*-infected samples. Further analysis suggested that expression differences in several receptor proteins may be driving the observed *trans*-acting changes. In sum, our work suggests that the relative importance of *cis* and *trans* acting changes may be highly dependent on the dominant biological process, even within a single tissue.

Results

Two geographically-distinct lines show ample genotype-specific immune response

To measure the relative contributions of *cis*- and *trans*-acting effects in the innate immune response, we needed two inbred, sequenced strains of *D. melanogaster* with abundant genetic variation and phenotypic differences in the immune response. The founder lines of the *Drosophila* Synthetic Population Resource are inbred, sequenced, and genetically diverse, making them ideal candidates (King et al. 2012). To maximize the likelihood of finding both genetic and phenotypic variation in these lines, we selected two lines from different continents, the A4 line, also known as KSA2, collected from the Koriba Dam in South Africa, and the B6 line, collected from Ica, Peru. Using the available SNP data, we found 462,548 SNPs between A4 and B6, with about half of them falling into exonic regions, indicating that 279,656/30,107,083 = 0.9% of exonic bases varied between the genotypes. The extensive variation in the coding regions allowed us to map, on average, 11.2% (±1.3%) of RNA-seq reads in an allele-specific manner.

To assess the divergence in the A4 and B6 immune responses, we measured gene expression pre- and post-infection in the abdominal fat body, the primary site of immune response. To do so, we performed RNA-seq on the dissected fat bodies of 4-day old males from both lines that had been infected with either Gram-positive *Enterococcus faecalis* (*Efae*) or Gram-negative *Serratia marcescens* (*Smar*). We selected these bacteria because in *D. melanogaster*, Gram-positive infections generally stimulate the Toll pathway, and Gram-negative infections generally stimulate the IMD pathway, though there is additional nuance due to

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

signaling crosstalk and the contributions of other signaling pathways (Buchon et al., 2014; Busse et al., 2007; Lemaitre and Hoffmann 2007; Tanji et al., 2010; Troha et al., 2018). We measured expression pre-infection and 3 hours post-infection, to capture the early transcriptional response prior to the complicating effects of feedback.

In response to *Efae* infection, we found sizable genotype-specific effects in the immune response. To detect these effects, we performed two types of differential gene expression analysis: we first compared control and infected samples to find *Efae*-responsive genes, and then within this group, we looked for genes differentially expressed between the A4 and B6 genotypes. We found 1165 differentially expressed genes between the control and infected samples regardless of genotype (Figure 1A). We categorized these *Efae*-responsive genes into three groups based on their differential expression between genotypes. Group 1 genes are differentially expressed only in the control samples, Group 2 genes are differentially expressed only in the infected samples, and Group 3 genes are differentially expressed in both control and infected samples. Genes not categorized into any of these groups were designated as not showing genotype-specific expression. Of the 500 *Efae*-responsive genes showing genotype effects, 87% (433 genes) are in Group 2, while only 10 genes are in Group 1 and 57 genes in Group 3 (Figure 1B). This indicates that many *Efae*-responsive genes show genotype-specific expression, and these differences are typically only revealed in response to infection.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

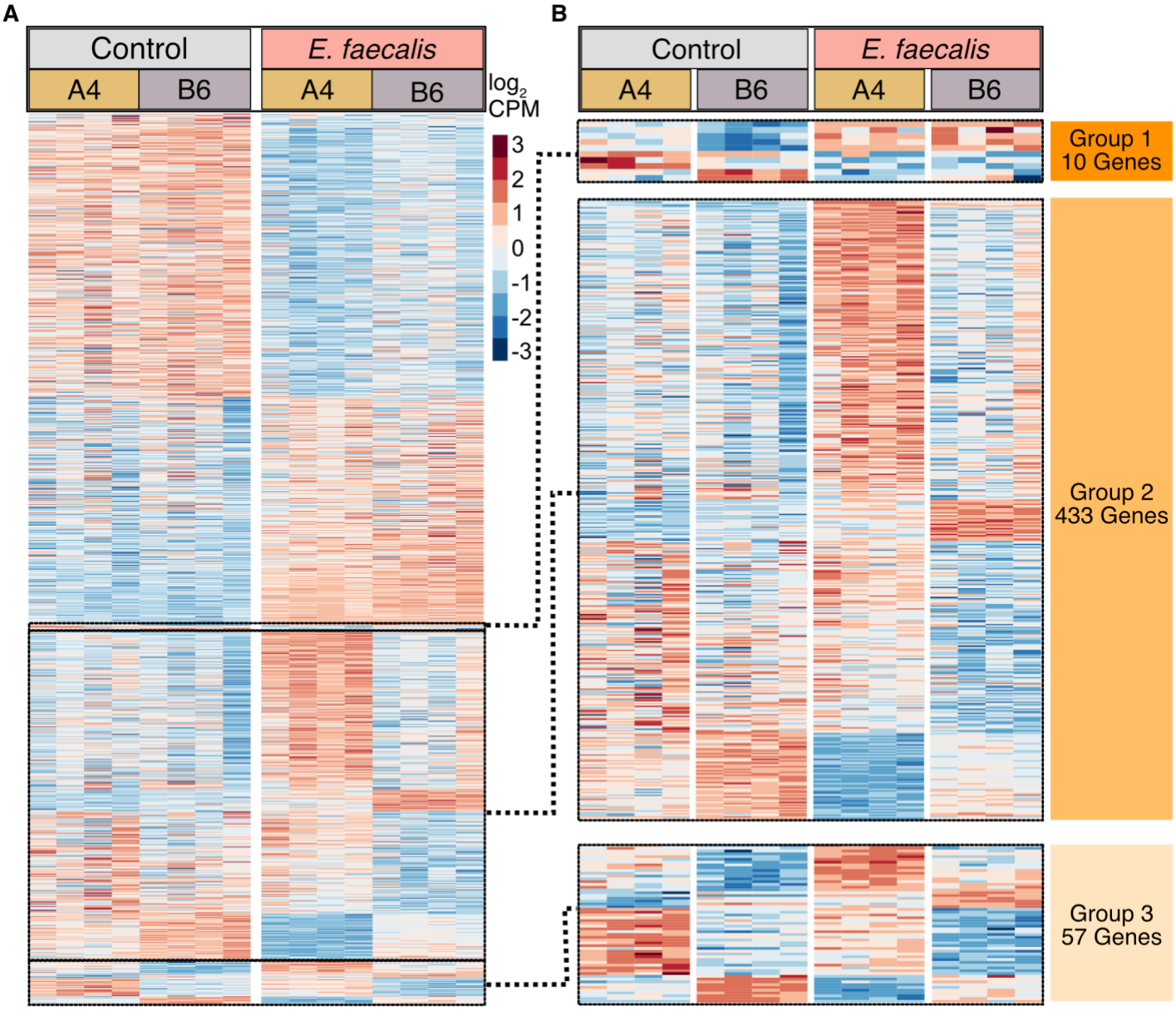


Figure 1. The A4 and B6 *D. melanogaster* lines have variation in their response to Gram-positive *E. faecalis* infection.

A) We measured expression in the fat bodies of the A4 and B6 lines infected with Gram-positive *Enterococcus faecalis* 3 hours post-infection. Of 11038 genes detected, we found 1165 differentially expressed in response to infection, relative to control samples, of which 201 were previously published *Drosophila* immune genes. Mean centered $\log_2$ CPM values are displayed. B) We categorized the 1165 *Efae*- responsive genes into three groups, based on their differential expression between the two fly genotypes: genes showing genotype-specific expression only in the control condition (Group 1), genes showing genotype-specific expression only in the infected condition (Group 2) and genes showing genotype-specific expression in both control and infected conditions (Group 3). The majority of genes fell into the Group 2 classification, indicating a large amount of genotype-specific expression variation is revealed upon infection with *Efae*. The fewest genes were classified into Group 1, suggesting that in the resting state there are few genotype-specific differences in the *Efae*-responsive gene set.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

In response to the *Smar* infection, we found 1203 differentially expressed genes between the control and infected samples (Figure 2A). To look for genotype-specific expression, we categorized the 1203 *Smar*-responsive genes into the three previously mentioned groups. For this infection, we found roughly equal numbers of genes in each group, with 89, 91, and 84 genes in Groups 1-3, respectively (Figure 2B). This indicates that a higher fraction of *Smar*-responsive genes show genotype effects prior to infection than *Efae*-responsive genes ($p = 1.1e-11$, Chi-square test, Bonferroni corrected), while a higher fraction of *Efae*-responsive genes show genotype effects after infection ($p = 9.5e-67$, Chi-square test, Bonferroni corrected)

To assess whether there is also phenotypic divergence on the organismal level, we performed the *Efae* and *Smar* infections and measured survival and bacterial load for 7 days post-infection. In response to *Efae* infection, we found differences in the ability to survive infection between genotypes, with B6 showing greater ability to survive infection over A4 (Supplemental Figure S1A). In response to *Smar*, we found that while there were no significant differences in survival, bacterial load was lower in A4 than in B6 (Supplemental Figure S1B, S1C). Together, these data demonstrate that there are differences between the two lines in their ability to resist or survive infection, and that these differences are pathogen specific.

To compare our tissue-specific measurements to previous work, we intersected our *Efae*- and *Smar*-responsive genes to an existing list of immune-responsive genes. This list is an expanded version of the *Drosophila* immune responsive genes set (DIRGS) and constitutes the summation of more than two decades of work in *Drosophila* (De Gregorio et al., 2001; Lemaitre and Hoffman 2007; Troha et al., 2018). Of 538 genes on this expanded list, we found more than half of these (297 genes) were identified as immune-responsive in our data (Figure 2C). Troha and colleagues identified a subset of immune-responsive genes as the core of the immune response, i.e. the set of genes that is differentially expressed regardless of the type of bacterial infection (Troha et al., 2018). We found that of these 252 core genes, approximately 40% were found to be both *Smar*- and *Efae*-responsive in our data. CrebA, which was identified as a core gene essential for both Toll and IMD-drive immune response, is one of the genes found in this overlap. Therefore, despite differences in the genetic background, tissue (previous studies were typically done with whole body

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

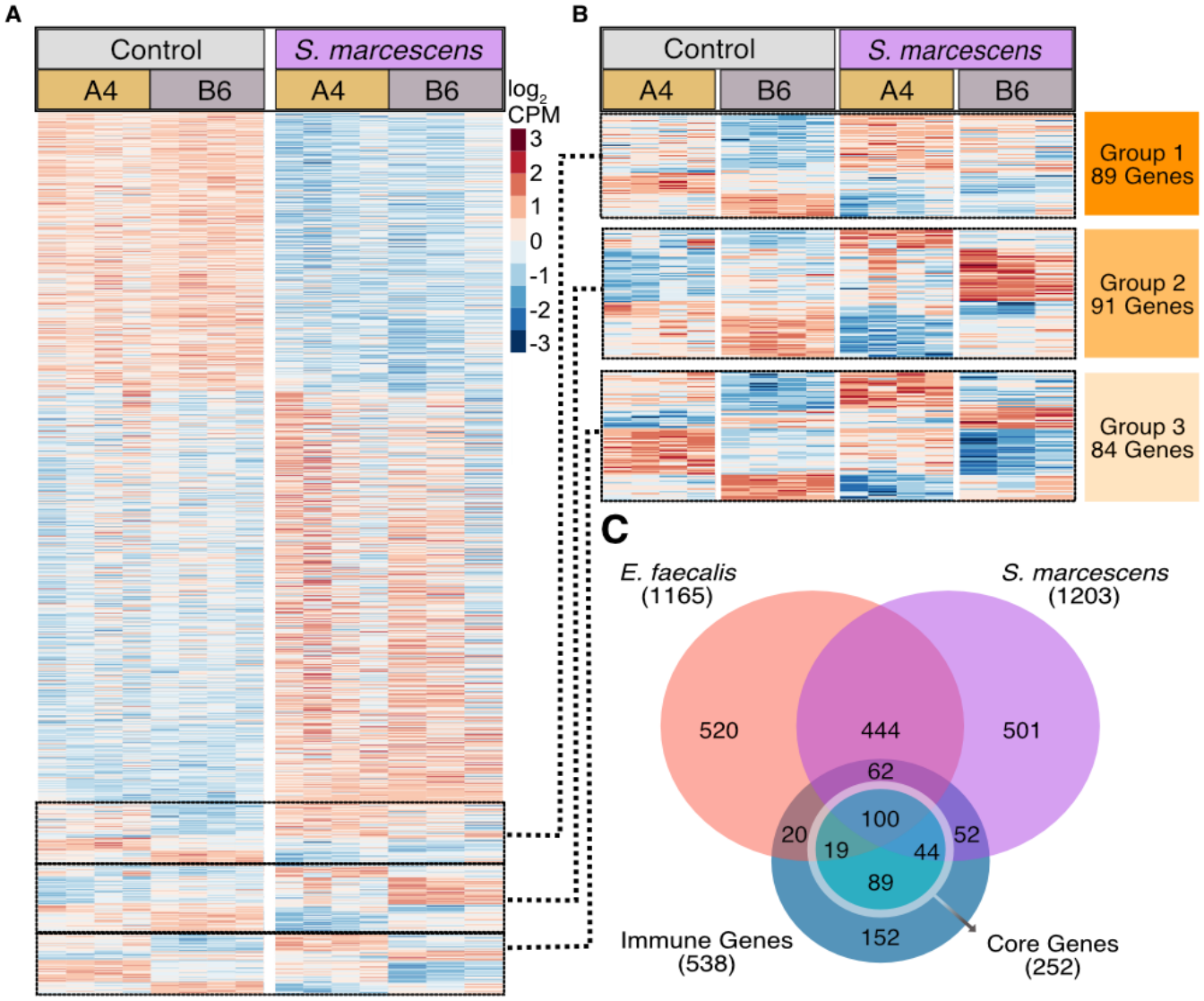


Figure 2. The A4 and B6 *D. melanogaster* lines vary in their response to Gram-negative *S. marcescens* infection.

A) We measured expression in the fat bodies of the A4 and B6 lines infected with Gram-negative *Serratia marcescens*. Of 11071 genes detected, we found 1205 differentially expressed genes between the control and infected samples; 258 were previously described *Drosophila* immune genes. Mean centered $\log_2$ CPM values are displayed. B) We categorized the 1205 genes into the three previously described groups (Figure 1). Among *Smar*-responsive genes, roughly equal numbers show expression differences between the genotypes before (Group 1), after (Group 2), and both before and after infection (Group 3). C) We intersected the genes we identified as differentially expressed in response to infection and a list of previously published immune responsive genes. This list of immune genes is an expanded version of the *Drosophila* Immune Responsive Genes Set (DIRGS). More than half of the verified immune genes were identified as differentially expressed in the abdominal fat body, with half of these immune genes being shared between conditions. Among these previously identified immune genes, core genes are differentially expressed across all infections. We detected roughly 40% of the core set as differentially expressed in both our infection conditions, despite differences in the genetic background, tissue type, and time point used in our study versus previous work.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

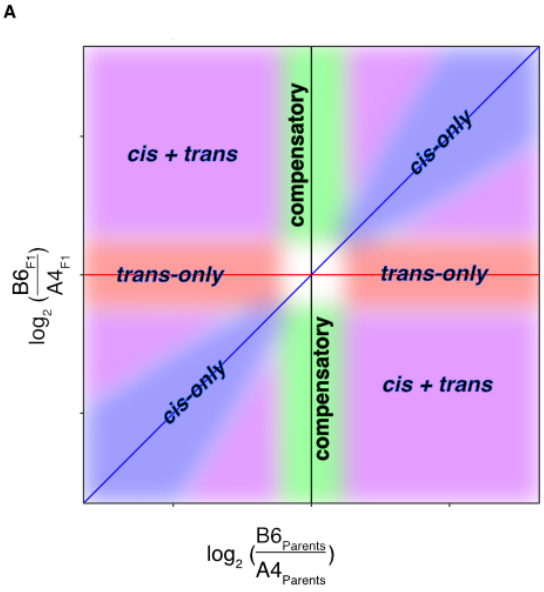
sampling), and time points, our findings show concordance with previous studies of gene expression in response to infection. More importantly, we show that the A4 and B6 lines have phenotypic divergence both in expression of immune-responsive genes and the ability to fight infection, making them suitable for subsequent F1 hybrid experiments.

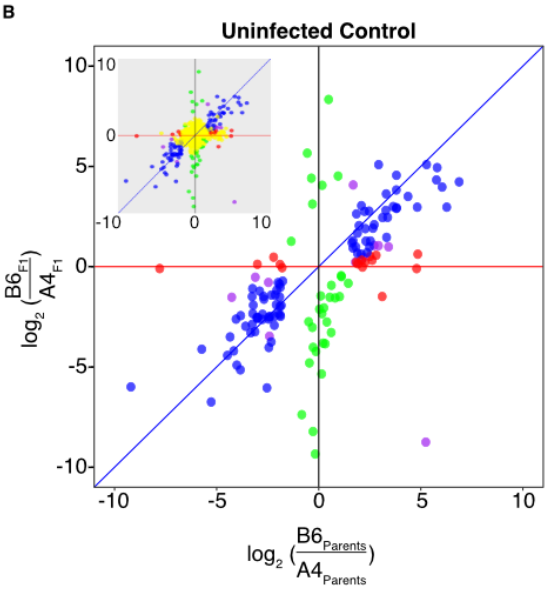
Cis-acting effects dominate expression variation in the uninfected fat body

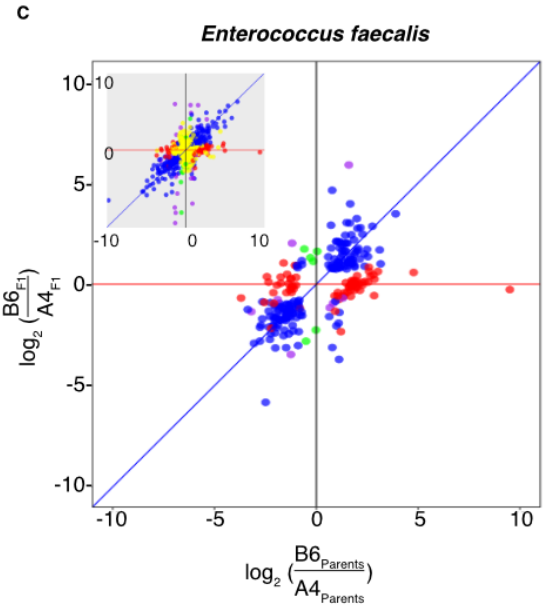
To effectively quantify *cis* and *trans* effects, we needed to verify our ability to accurately analyze the allelic expression in F1 hybrids. To do so, we used the RNA-seq data from the A4 and B6 parental lines mentioned above as well as data from the F1 hybrids (A4♂x B6♀) and reciprocal crosses (B6♂x A4♀), in the control, *Efae*-infected, and *Smar*-infected conditions. Since we are using males, if our allele-specific expression analysis is correct, none of X-chromosome reads should map to the paternal genotype. Using the published A4 and B6 genomes and the Allele-Specific Alignment Pipeline (ASAP) (Krueger, https://www.bioinformatics.babraham.ac.uk/projects/ASAP/), we quantified the fraction of X-chromosome reads that incorrectly map to the paternal genotype. On average, samples had 0.5% mis-assigned reads (standard deviation = 3%), with the highest fraction being 1.2% (Supplemental Table S1). The consistent, low level of mis-assigned reads verifies our ability to accurately quantify allelic expression.

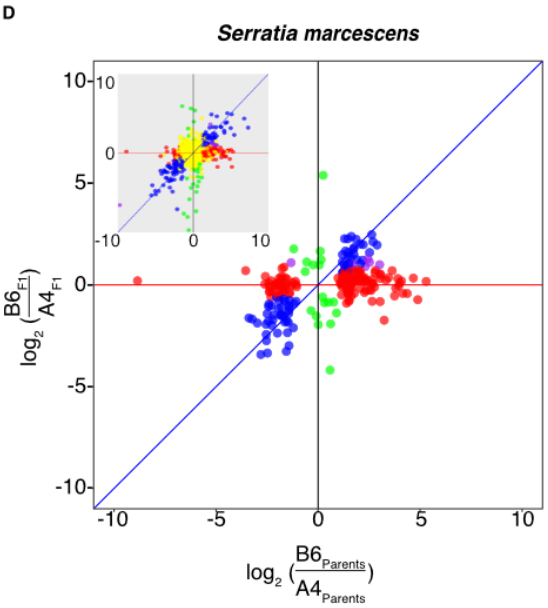
We next sought to quantify *cis* and *trans* effects in the control samples. We used the complete set of parental RNA-seq reads and the subset of the F1 hybrid reads that could be assigned to a specific allele. Using three separate generalized linear models, we tested for differential expression in the parents, allelic imbalance in the F1 hybrids, and *trans* effects between parents and F1 hybrids (see Methods) (Davidson and Balakrishnan, 2016; Osada et al., 2017; Takada et al., 2017). We then categorized each gene into one of six categories (Figure 3A). Genes showing no differential expression in the parents or F1 hybrids have no evidence of *cis* or *trans* effects and are called ***conserved***. Genes showing differential expression in both the parents and F1 hybrids and no *trans* signal are categorized as ***cis-only***. Genes showing differential

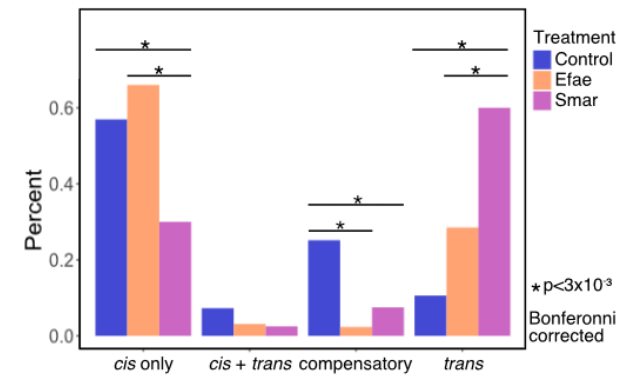
bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.











F

Category	Control	Efae	Smar
Conserved	3808	2586	4107
cis-only	86	169	72
cis + trans	11	2	6
compensatory	38	6	18
trans-only	16	73	144

210

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 3. The relative contributions of *cis* and *trans* effects to expression divergence are condition specific.
A) Here we show a schematic of the expected locations of four of the six categories of causes of expression divergence, conserved (yellow) and undetermined genes are excluded from downstream analysis. B) In the uninfected control condition, of 4960 genes that could be detected, 151 genes showed *cis* or *trans* signal. 86 genes show *cis*-only effects, 16 genes show *trans*-only effects, 11 genes show *cis* and *trans* effects, and 38 show compensatory effects. C) In response to *Efae* infection, expression divergence is driven predominantly by changes in *cis*. There are 256 genes that show *cis* or *trans* signal and do not overlap with the 151 genes showing signal in the control condition. Of 256 genes, 169 genes show *cis-only* signal, 86 genes show *trans-only* signal, 8 genes show a combination of *cis* and *trans* effects, and 6 genes show compensatory effects. D) In response to *Smar* infection, expression divergence is dominated by changes in *trans*. There are 240 genes that show *cis* or *trans* signal not found in the uninfected control. In these 240 genes, 72 genes showed *cis*-only signal, 144 genes showed *trans*-only signal, 6 genes showed both *cis* and *trans* effects, and 18 genes show compensatory effects. E) We compared the fraction of genes categorized into each divergence class in the three conditions and found that the modes of expression divergence were condition-specific.

expression in the parents and not the F1 hybrids are categorized as ***trans-only***. Some genes show evidence of both *cis* and *trans* effects and are categorized as either ***compensatory*** (if the changes have opposing effects on expression) or ***cis + trans*** (if the changes are coherent). Genes that do not fall into any of these categories have an ambiguous pattern of divergence and are called ***undetermined.***

Of the 4960 genes that were both expressed in the pre-infection fat body and could be detected in an allele-specific manner, 77% were categorized as conserved (3808 genes; Figure 3B). We found 151 genes showing unambiguous *cis* or *trans* effects. In these 151 genes, *cis* effects dominated the signal: 90% of genes (135 genes) showed *cis* signal (including *cis-only*, *cis* + *trans* and compensatory genes), and 57% (86 genes) showed *cis-only* effects. 43% of genes (65 genes) showed *trans* signal and only 10% of genes (16 genes) showed *trans-only* effects. One-quarter of genes (38 genes) were categorized as compensatory, even when using non-overlapping samples to detect *cis* and *trans* effects, which avoids the artificial inflation of compensatory signal (Fraser et al., 2019; Zhang and Emerson, 2019). Additionally, to ensure that any differences in the quality of our in-house A4 and B6 transcriptomes do not affect our conclusions, we quantified *cis* and *trans* effects using sets of high confidence genes at multiple levels of stringency and found that this had negligible effects on the detected signal (Methods; Supplemental Figures S2; Supplemental Table S2). From these data, we can conclude that in the unstimulated state, most genes have conserved

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

243 expression levels in the fat body, and among those genes that diverge, *cis* effects dominate, with a sizable
244 number of genes showing compensatory *cis* and *trans* changes.
245
246 **More *cis* than *trans* effects are found in *Efae*-infected fat body expression**
247 We quantified *cis* and *trans* effects in *Efae*-infected samples following the same methodology. We found
248 roughly 52% of genes (2586 genes) showed no evidence of *cis* or *trans* effects and 381 genes showed
249 unambiguous *cis* or *trans* effects (Figure 3C). To identify genes whose expression divergence is specific to
250 the immune response, we eliminated genes that show *cis* or *trans* signal in the control sample. After this
251 filtering, roughly 88% of the genes showing cis or trans effects (336 genes) remained; 66% of these genes
252 (169 genes) show *cis-only* signal, and 28% (73 genes) show *trans-only* signal. Only 8 genes (3%) show a
253 combination of *cis* and *trans* effects, with only 6 genes showing compensatory effects. Of the genes that
254 show *cis-only* signal, roughly even numbers of genes show higher expression in each genotype, consistent
255 with the idea that *cis*-acting changes affect a single gene at a time. In contrast, of the genes showing *trans-*
256 *only* signal, nearly twice as many were expressed more highly in the B6 genotype (47 genes) than in the A4
257 genotype (26 genes) (p = 0.00962, Chi-square test), suggesting that one or a few changes in upstream
258 regulatory factors are responsible for this observation. Since we do not observe this trend towards higher B6
259 expression in the control samples and have removed genes that showed any evidence of mapping bias
260 between the two genotypes (Methods), we are confident this trend is not an artifact and reflects true
261 biological differences in the immune response. In sum, we find both *cis* and *trans* effects drive *Efae-*
262 responsive expression divergence, with *cis* effects dominating and few genes showing compensatory
263 changes.
264
265 ***Trans* effects dominate expression variation in the *Smar*-infected fat body**
266 Lastly, we quantified *cis* and *trans* effects in response to *Smar* infection. We found roughly 82% of genes
267 (4107 genes) are conserved, and 357 genes showed unambiguous *cis* or *trans* signal (Figure 3D). We again
268 filtered out genes that show *cis* or *trans* effects in the control samples and were left with 240 genes that have

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

immune-specific signal. Of these, 30% (72 genes) showed *cis-only* signal, and roughly equal numbers of *cis-only* genes showed higher expression in each genotype. 10% (24 genes) showed both *cis* and *trans* effects, and within these genes, 18 genes had compensatory signal. Surprisingly, 60% of genes (144 genes) showed *trans-only* signal. Within *trans-only* genes, we found that 71% showed greater expression in B6. Though the *Efae* analysis found far fewer genes affected in *trans*, the proportion of *trans-only* genes showing higher B6 expression was similar. This suggests that some of the upstream differences giving rise to the *trans* effects may be shared between the two infection conditions. In summary, in response to *Smar* infection, *trans* effects drive the majority of expression divergence between the two genotypes and few genes show compensatory effects.

Comparisons of *cis* and *trans* signals in different conditions reveal both infection-specific and shared divergence

To systematically assess how *cis* and *trans* effects contribute to expression variance in the same tissue under different conditions, we compared the proportion of genes falling into the different divergence categories. The control and *Efae*-infected samples had a greater proportion of *cis-only* genes than the *Smar* samples (control vs. *Smar* p = 1.97e-06, *Efae* vs. *Smar* p = 1.97e-14, Chi-square test, Bonferroni-corrected). However, all three groups differ in the proportion of *trans-only* genes, with *Smar*-infected samples showing more than twice the proportion of genes with *trans*-only signal, followed by *Efae,* and then the control samples (control vs. *Efae* p = 3.7e-04, control vs. *Smar* p < 2.2e-16, *Efae* vs. *Smar* p = 2.79e-11, Chi-square test, Bonferroni-corrected). We also found that the uninfected fat body showed significantly more compensatory signal than either infected sample (control vs. *Efae* p = 2.35e-11, control vs. *Smar* p = 2.26e-05, Chi-square test, Bonferroni-corrected). Taken together, this suggests that before infection, when the fat body is carrying out its metabolic functions, there is less pressure for expression divergence, as supported by the large number of compensatory changes and conserved genes. In genes that show expression divergence in the control condition, *cis*-acting changes, which have local, non-pleiotropic effects, dominate. In response to infection, there is ample expression divergence, which is driven by both *cis* and *trans* effects.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The extent to which each type of effect contributes is dependent on the particular pathogen, suggesting that the relative importance of local and pleiotropic changes is specific to different infection pressures.

Though we generally expect the two infections to regulate gene expression via distinct signaling pathways, we also anticipated some genes would be regulated in both infections, either due to crosstalk between the IMD and Toll pathways (Busse et al., 2007; Tanji et al., 2010) or via more general infection and wound responses. We found 75 genes with unambiguous *cis* and/or *trans* signal in response to both *Efae* and *Smar* infection (Supplemental Data). Of these genes, 61 showed concordant classification and the remaining 14 genes did not. Thereofre, in the majority of genes shared between these two infections, the same genetic differences are likely driving the expression divergence in both infection conditions. Additionally, in rare cases, a single gene can experience either *cis* or *trans* effects depending on the infection context.

Differential expression of detection genes is a likely source for genotype expression bias in observed *trans* effects

Since we observed that genes with *trans-only* effects tended to be more highly expressed in B6 than in A4 in both infection conditions, we hypothesized that changes in a handful of upstream immune factors are responsible for this phenomenon. The changes in upstream regulators can either be infection-specific or shared, and genes affected by shared regulators would show *trans* signal in both infections. Out of 217 genes showing *trans-only* signal in either infection, only 13 genes were shared, 4 with higher expression in A4, and 9 with higher expression in B6. The small number of genes that show *trans* effects in both infection conditions indicates that the bulk of *trans*-acting changes are likely infection-specific and not driven by a shared infection or wound healing response.

To find likely sources of infection-specific *trans* effects, we hypothesized that immune detection genes, signaling genes, or transcription factors differentially expressed between genotypes in the control condition would be likely candidates, since these genes have the ability to affect the expression of many

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

downstream targets. Further, we posited that these genotype-specific differences had to be present in the control to have the effects at the 3-hour post-infection timepoint. Of the approximately 300 genes that are differentially expressed between genotypes in the control samples, we found 22 genes that are prime candidates (Table 1). Fourteen are previously identified immune-responsive genes, and eight are transcription factors not yet implicated in immune response, but that may be peripherally involved.

Five peptidoglycan recognition proteins (PGRP) genes are potential mediators of the large number of *trans* effects observed in the *Smar* infection. Four of these PGRPs (PGRP-SC1a, PGRP-SC1b, PGRP-SC2, PGRP-LB) are negative regulators of the IMD response, and the last gene, PGRP-SD is positive Toll and IMD regulator (Bischoff et al., 2006; Zaidman-Rémy et al., 2006; Iatsenko et al., 2016; Charroux et al., 2018; Lu et al., 2020). Three of the negative regulators, PGRP-SC1a, PGRP-SC1b, PGRP-SC2, are more highly expressed in A4. Given that these are negative regulators of the IMD pathway, this finding is congruent with the observation that genes showing *trans-only* signal tend to show greater expression in B6. PGRP-SD is more highly expressed in B6, and, given its role as a positive regulator of the IMD response, it is also consistent with the trend of higher expression of genes showing *trans-only* signal in the B6 line. The last negative regulator of IMD response, PGRP-LB, was also found to have higher expression in B6. Since there are more negative regulators showing higher expression in A4 than B6, it is possible the balance of negative to positive regulators can account for the expression trend observed in *Smar trans-only* genes. It is also possible that the greater expression of PGRP-SD is enough to account for the differences observed.

Though there were fewer *trans* effects in the *Efae*-infected samples than in the *Smar*-infected samples, the pattern wherein nearly twice as many genes showed greater expression in B6 than A4 was maintained. Of the 22 candidate genes, we found two Toll-specific genes that may be responsible for the observed signal: *Spatzle-Processing Enzyme* (*SPE*) and *spatzle* (*spz*), which are more highly expressed in

Table 1: Transcription factors and immune genes identified as potential sources of trans effects in infection.

List of genes potentially involved in observed *trans* effects for *Efae* and *Smar* infection. Candidate genes were identified by finding genes that had genotype-specific expression differences in the uninfected control conditions and were classified as either a transcription factor, immune signaling gene, or immune detection gene.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

FB Gene ID	Gene name	Type	Log_2 Fold Change (B6/A4)	More Highly Expressed in:	A4 Average CPM	B6 Average CPM	Immune involvement
FBgn0029822	CG12236	TF	-3.13	A4	28	2.7	Unclear
FBgn0039075	CG4393	Signaling	2.06	B6	8.8	39	Unclear
FBgn0038978	CG7045	TF	3.06	B6	7	57	Unclear
FBgn0001981	esg	TF	-3.27	A4	11	1	Unclear
FBgn0039932	fuss	TF	2.50	B6	1.1	6.3	Unclear
FBgn0250732	gfzf	TF	8.24	B6	0	2	Unclear
FBgn0000448	Hr46	TF	-4.12	A4	10	0.7	Unclear
FBgn0016675	Lectin-galC1	Detection	2.72	B6	79	570	Binding and agglutation
FBgn0035993	Nf-YA	TF	-10.16	A4	10	0	Unclear
FBgn0028542	NimB4	Detection	-1.08	A4	40	22	Unclear
FBgn0259896	NimC1	Detection	-3.06	A4	97	27	Unclear
FBgn0003130	Poxn	TF	-4.38	A4	1.2	0.07	Unclear
FBgn0014033	Sr-CI	Detection	-2.39	A4	84	38.6	Unclear
FBgn0004606	zfh1	Signaling / TF	1.77	B6	50	17	Unclear

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

FBgn0031973	Spn28Dc	Signaling	2.56	B6	9.4	2.4	Negative regulator of melanization
FBgn0037906	PGRP-LB	Detection	4.536	B6	111.9	75.2	Negative regulator of IMD pathway
FBgn0043576	PGRP-SC1a	Detection	-5.57	A4	4.3	6.8	Negative regulator of IMD pathway
FBgn0033327	PGRP-SC1b	Detection	-5.24	A4	3.9	0.2	Negative regulator of IMD pathway
FBgn0043575	PGRP-SC2	Detection	-4.02	A4	15	1.3	Negative regulator of IMD pathway
FBgn0035806	PGRP-SD	Detection	4.25	B6	97.6	19.2	Positive regulator of IMD pathway
FBgn0039102	SPE	Signaling	2.41	B6	491.9	255.2	Positive regulator of Toll pathway
FBgn0003495	spz	Signaling	0.68	B6	72.8	45.5	Positive regulator of Toll pathway

B6. *Spatzle* is the Toll receptor ligand, and *SPE* is required to generate the active form of *spz*, so differential expression of these genes can drive a large number of downstream changes. In addition, PGRP-SD can act as a positive regulator of both the Toll and IMD responses and is also found to have higher expression in the B6 line.

In addition to differences in expression levels between genotypes, function-altering differences in the coding sequences of key immune genes may also be the source of the observed *trans*-acting changes. To analyze the pattern of coding sequence variants, we used the Ensembl Variant Effect Predictor (VEP) to identify the proportions of synonymous to nonsynonymous coding changes between A4 and B6 in several gene sets (McLaren et al., 2016). We considered all genes expressed in the fat body above a threshold (CPM>1), and then sorted them into two groups: genes that are differentially expressed in response to either or both infections (*DE infection*) and those that are not (*fat body detected*). We also generated a gene set that is the intersection between *DE infection* and our list of curated immune-responsive genes (*DE immune*;

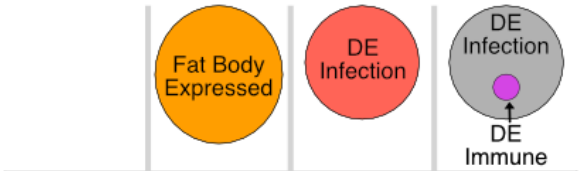
bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 4A). We posited that, given the large number of *trans* effects in response to infection, immune-related genes may have a greater number of nonsynonymous to synonymous changes, compared to the *fat body detected* gene set. We found that *DE immune* genes have a significantly higher fraction of nonsynonymous sequence changes (24%) compared to the *fat body detected* genes (21%) (p = 0.007, Chi-square test, Bonferroni-corrected), suggesting that some of these changes may be under selection and possibly the source of our *trans*-acting signal (Figure 4B-C). *DE infected* genes showed lower fractions of nonsynonymous changes (20%) compared to the *fat body detected* genes (p = 0.04, Chi-square test, Bonferroni-corrected). We may not see an elevated fraction of nonsynonymous changes in the *DE infected* gene set because this includes both immune-related genes, but also genes with unrelated functions whose expression may be regulated due to the metabolic constraints on the fat body tissue.

This analysis focuses on overall patterns of coding sequence changes in these large gene sets. We recognize that even an individual coding sequence change may drive many downstream expression differences and analyzed these changes, but no obvious candidates emerged (Supplemental Figures S3, Supplemental Table S3 and S4). Predicting the effect of these mutations on individual protein function, however, remains a challenge.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A



Category	Expressed	DE Infection	DE Immune
Synonymous SNPs	74144	14194	1957
Synonymous %	0.789	0.797	0.764
Non-synonymous SNPs	19819	3610	604
Non-synonymous %	0.211	0.203	0.236

B

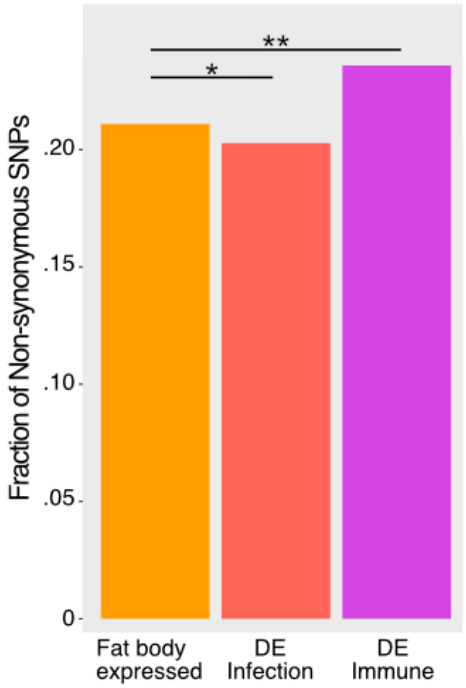


Figure 4: There is a greater proportion of non-synonymous SNPs in previously identified immune-responsive genes.
A) To look for the prevalence of non-synonymous SNPs, we defined three gene sets. Among genes detected in the fat body samples, we separated genes into those that were differentially expressed in response to either infection (DE infection), and those that were not (fat body expressed). Among the DE infection genes, we further refined the gene list to include only previously identified immune genes (DE immune). The numbers indicate the total number of SNPs found in each gene set. B) DE immune genes have a higher proportion of non-synonymous SNPs than the fat body expressed genes, which suggests they may be subject to different selection pressures. *p*-values are Bonferroni-corrected Chi-square test with the proportion of non-synonymous SNPS relative to the fat body expressed gene set.

Discussion

Here, we quantified the mode and extent of expression divergence in the *Drosophila* abdominal fat body, both in an uninfected control condition, where it carries out a variety of metabolic roles, and in response to two types of infection. We found that two geographically isolated lines of *D. melanogaster* are phenotypically distinct in their immune responses, differing both on the organismal and transcriptional levels. By comparing gene expression in the fat body between these lines and their F1 hybrids, we quantified the contributions of *cis* and *trans* effects to expression divergence in the uninfected control, *Efae*-infected and *Smar*-infected conditions. Both the control and *Efae* infection conditions were dominated by *cis* effects, while the *Smar*

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

infection condition had an abundance of *trans* effects. Notably, the uninfected control also showed a greater proportion of compensatory effects, suggesting that there is stabilizing selection to maintain fat body expression levels of certain genes in the absence of an infection. Among the genes showing changes in *trans*, we found that expression of the B6 allele is typically higher in both types of infections. This suggests that changes to a small number of upstream immune response regulators may be responsible for this bias, and we identified expression divergence in a group of receptor programs that may drive these *trans* effects. Overall, we find that the mode of evolution in expression divergence can vary between conditions in a single tissue and likely represents condition-specific selection pressures.

Our unique approach to measuring the mode of expression divergence gave rise to several novel observations about the relative contributions of *cis* and *trans* effects on expression variation. While there have been a number of studies aimed at disentangling the contribution of *cis* and *trans* changes to gene expression in *Drosophila,* few have sought to answer this question using a single organ or with different physiological stimuli (Wittkopp et al., 2004, Wittkopp et al.,2008, McManus et al., 2010, Coolon et al., 2014, Osada et al., 2017). Our approach allows us to examine evolutionary changes in response to perturbation while minimizing the confounding effects of multiple tissue types. There are three studies most closely related to ours. A previous study by Juneja, et al. (2016) found, among geographically distinct flies, a large number of *cis*-acting changes that cause whole body expression divergence in response to an infection with mixture of bacteria. This is concordant with our finding of a large number of *cis*-acting changes in both infection conditions, but this study did not quantify *trans*-acting changes or distinguish between Toll- and IMD-specific responses. By measuring expression in the heads and abdomens of multiple *D. melanogaster* lines, another group reported the predominance of changes in *cis* over those in *trans* but did not measure these differences in different physiological states or attempt to dissect individual tissues in the head or abdomen (Osada et al., 2017). Most recently, Frochaux et al., (2020) sought to uncover the underlying genetics of *P. entomophila* resistance in the gut and identified a novel driver of this phenotype but limited their analysis to locally acting eQTLs. Here, we sought to directly assess the contribution of *cis* and *trans*

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

sequence changes in a single tissue in the context of multiple treatment conditions, giving a uniquely high-resolution view of the evolutionary sequence changes underlying expression divergence.

With our approach we were able to uncover two notable trends. First, we found that compensatory mutations were more frequent in the control samples. We observe that while overall compensatory effects are less common than *cis-only* and -*trans-only* effects, compensatory effects in the uninfected samples are more pervasive than in either of the infected conditions. Previous studies in several organisms had suggested that compensatory effects were very prevalent (McManus et al., 2010, Gonclaves et al., 2012, Schaefke et al., 2013, Coolon et al., 2014). This was perplexing because it would seem to suggest that, even in between species, selective forces to alter expression and to stabilize it were at odds and did not explain how biological systems are able to evolve divergent expression. However, certain choices in experimental design can inflate estimates of compensatory effects (Zhang and Emerson 2019; Fraser et al., 2019). To avoid this artifact, we use non-overlapping F1-hybrid samples and therefore have generated more accurate estimates of compensatory effects across multiple conditions. Additionally, a large proportion of studies addressing *cis* and *trans* effects in animals do so in “control” conditions, which may not reveal the full extent of selection forces that act on gene expression (Gonclaves et al., 2012, Osada et al., 2017, Davidson and Balakrishnan 2016, Signor and Nuzhdin 2018). In our system, we find evidence that the genes involved in the maintenance of basic metabolic functions of the uninfected fat body are under different selective pressures than those involved in immune response. Unlike the immune-responsive genes, which must contend with a continuously evolving pathogen landscape, the genes carrying out metabolic functions may be subject to stabilizing selection, given relatively unchanging nutritional availability. In future studies, it will be interesting to further probe which systems and conditions show enrichment for these different patterns of expression divergence.

Secondly, we observe that the relative contribution of *cis*- and *trans*-acting changes are perturbation-specific. Most notably, in response to *Efae* infection, *cis* effects dominate expression changes, while in the *Smar* infection, *trans* changes are predominant. The prevalence of either *cis* or *trans* effects can be reasonably justified in our system, but we did not anticipate that the proportion of these effects would be

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

446 infection specific. Because changes in *trans* factors have pleiotropic effects, it has been suggested that
447 changes to these factors are under more selective constraint than *cis*-acting elements, and, thus, *cis* effects
448 can more readily introduce small-scale variation into a system (Schaefke et al ., 2013). In some cases,
449 however, arriving at a more fit phenotype may require the coordinated alteration of expression of many
450 genes, which may be more readily achieved by changes to *trans*-acting factors. In other cases, a coordinated
451 change in all genes affected by perturbing a *trans*-acting factor may not yield a phenotype with a net
452 beneficial effect, and it may be more likely that changes to individual genes offer paths to higher fitness. In
453 our *D. melanogaster* lines, *S. marcescens* is more virulent than *E. faecalis* – a higher dose of *E. faecalis* is
454 needed to achieve similar levels of mortality to that of *S. marcescens* (Supplemental Figure S1). It is possible
455 that adaptation to highly virulent pathogens or rapidly evolving pathogens requires large-scale, synchronous
456 changes to expression, whereas adaptation to less virulent pathogens is possible with smaller, localized
457 mutations. Consequently, we suggest that the differences in the abundance of *cis* or *trans* effects may reflect
458 the individual details different host-pathogen interactions and how that influences the genetic architecture of
459 adaption. Experiments with a wider range of pathogens will further illuminate the relationship between the
460 mode of expression divergence and the host-pathogen relationship.

461 In summary, we find that the mode of expression divergence, as represented by the proportion of *cis*
462 and *trans* effects in a system, is condition-specific in the *Drosophila melanogaster* abdominal fat body. This
463 specificity is likely a result of the distinct selective pressures that different host-pathogen interactions exert on
464 the *D. melanogaster* immune system. In the course of our study, we found several candidate genes that may
465 be the sources of the observed *trans* effects, which are most prominent in *Smar* infection. To verify these
466 effects, we aim to over-express these candidate genes in multiple genetic backgrounds in future
467 experiments, which are becoming more feasible with the development of new genetic tools. In the longer
468 term, we can combine the data sets presented here with other types of functional genomics experiments to
469 identify immune-responsive *cis*-acting elements and the sequences changes that drive *cis*-acting divergence.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Taken together, these studies will provide a more comprehensive view of how regulation of expression in this rapidly changing system is wired and evolves.

Methods

Animal genotypes, infection, and survival analysis

The A4 and B6 *D. melanogaster* lines, SNP tables and genomic reads were received from the *Drosophila* Synthetic Population Resource (King et al., 2012). Flies were reared at 25˚C on standard cornmeal/yeast media (recipe available upon request). For all RNA-seq experiments, four-day-old males were infected with approximately 15 nL of A_{600} = 0.5 OD solution of either *Enterococcus faecalis* or *Serratia marcescens* via microinjection, yielding an infection of ~10,000 CFUs/fly (Khalil et al., 2015). Uninfected controls were placed on a carbon dioxide pad for 6 minutes to mimic the effects of anesthesia used for microinjection. Both bacteria were grown in liquid culture on a shaker at 37˚C overnight and then diluted 1:1000 in fresh media in the morning. These cultures were grown until exponential phase (*S. marcescens* in Luria-Bertani broth for 4 hours, *E. faecalis* in brain-heart infusion media for 5 hours). Bacteria were then pelleted down and resuspended in PBS for OD measurement and injection. All injections took place between 3:00 and 5:00 pm to account for the impact of circadian rhythm on immune response (Scheiermann et al., 2013). For the survival analysis, we used lower doses of infection to more effectively ascertain differences in survival. We infected A4 and B6 genotypes with either 5,000 CFUs of *E. faecalis* or 1,000 CFUs of *S. marcescens*, and the survival status of the flies was recorded once per day following infection (see Supplemental Figure S1 for details). To measure bacterial load, a sample of living flies were collected once per day, and we measured load with dilution plating as in (Khalil et al., 2015). Kaplan-Meier estimates of survival were calculated using the `survival` package in R (Therneau et al., 2020), and log-rank tests and plotting were performed using the `survminer` package (Kassambara and Kosinski 2019).

To determine the number of unique SNPs between A4 and B6, we downloaded published SNP tables available through the DSPR website (King et al., 2012). We selected for SNPs not shared between lines and

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

that also showed a reference allele frequency of < 0.05 (implying an alternative allele frequency > 0.95). We then calculated total SNP differences for exonic and non-exonic regions using exon coordinates from Flybase (dm6/iso-1: FB2019_01) (Thurmond et al., 2019).

Preparation and sequencing of RNA-seq libraries

Three hours after infection, abdominal filets with the attached fat bodies were prepared as in (Krupp and Levine et al., 2010). Three fat bodies per sample were suspended in Trizol (Life Technologies) and stored at -80˚C for later extraction. RNA was extracted from samples using Zymo Research Direct-zol RNA Extraction Kits. Library construction was completed using a modified version of the Smart-Seq2 protocol outlined in (Serra et al., 2018). Samples were then sequenced on Illumina Next-seq Platform with NextSeq 500/550 High Output Kit v2.5 to generate 43bp paired end reads. Data was imported to the UCI High Performance Computational Cluster for trimming and mapping of sequenced reads.

Differential expression analysis

Reads were trimmed and filtered using Trimmomatic 0.35 (Bolger et al., 2014). Count and TPM data for each sample was then calculated using Salmon 0.12.0 aligner (Patro et al., 2017) using the dm6/iso-1 transcriptome. Count matrices of gene-level data were then constructed in R using the `Tximport` 1.12.3 package (Soneson et al., 2015). To find genes differentially expressed in response to each infection, compared to control, we used the `EdgeR` 3.26.5 package (Robinson et al., 2010, McCarthy et al., 2012), excluding lowly expressed genes (CPM<1) with false discovery rate corrected *p*-values (Benjamini and Yekutieli et al., 2001). Genes with an FDR < 0.05 were considered differentially expressed. To determine genotype-specific effects, among the *Efae*- or *Smar*-responsive genes, we used EdgeR to find genes differentially expressed between A4 and B6 in either the control conditions or the treated condition.

Generation of A4 and B6 transcriptome annotations

To map RNA-seq reads in an allele-specific manner, we created two reference transcriptomes by lifting over Iso-1 genome annotations to sequenced A4 and B6 genomes. Using tools from UCSC liftOver suite, custom chain files were created by mapping homologous sequences to the A4 or B6 genome (Salinas et al., 2016). To assess the quality of our annotations and remove genes with poor annotations, genomic sequencing

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

reads from the DSPR website where downloaded and aligned to our transcriptome files (Thurmond et al., 2019). We hypothesized that well-annotated genes would show similar coverage of genomic reads in both the A4 and B6 transcriptomes. We then filtered genes using two methods for outlier calling: a Poisson distribution-based method and a negative binomial generalized linear model (GLM) method, similar to that used for differential gene expression in RNA-seq experiments. For the Poisson method, we fit a Poisson distribution to gene count data for the A4 and B6 transcriptomes separately, using the `fitdistributionplus 1.0-14` package in R and called outlier genes using three thresholds of increasing stringency p = 0.001, 0.01 and 0.025. For the GLM-based approach, we looked for gene counts that were significantly different between the A4 and B6 transcriptomes and filtered genes using FDR thresholds of 0.01, 0.05 and 0.09. As our threshold for significance became more stringent, we filtered out an increasing number of genes (Supplemental Figure S2). Genes found not to be outliers in either the Poisson or GLM method were then combined into gene sets based on the stringency of filtering. These gene sets were then used to quantify *cis* and *trans* effects for all three conditions. We found that the stringency of filtering did not significantly impact the total number or proportions of *cis* and *trans* effects between conditions. For the allele-specific expression analysis presented in Figure 3, we used a set of genes filtered using a combination of both methods at medium stringency.

Allele-specific expression analysis

RNA reads were assigned parental alleles using Allele Specific Alignment Pipeline (Krueger, https://www.bioinformatics.babraham.ac.uk/projects/ASAP/) using the A4 and B6 genomes and allowing for no mismatches. Non-uniquely assignable reads were discarded. Count and TPM data were then generated by aligning allelic reads to the corresponding transcriptome. Count matrices of gene-level data were then constructed in R using the `Tximport` package (Soneson et al., 2015). Accuracy of allele calling was assessed by the proportion of X chromosome reads that aligned to the maternal genotype versus the paternal genotype. Given that all the flies are male, any reads aligning to the paternal X chromosome can definitively be classified as mis-assigned.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

546 To characterize expression divergence into *cis* and *trans* categories, differential expression was
547 determined with unparsed parental reads and allele-specific reads from the F1 hybrids, using EdgeR and
548 three distinct GLM structures. Genes not found in both A4 and B6 transcriptomes, lowly expressed genes
549 (CPM<1) and X chromosome genes were excluded from the analysis. For each condition, we first tested for
550 differential gene expression between parental samples (DE parents; Murad et al., 2019). Next we tested for
551 allelic imbalance, taking into account parent of origin and maternal genotype effects as outlined in (Osada et
552 al., 2017; Takada et al., 2017). For this test we used half of the F1 hybrid samples. Finally, we tested for
553 *trans* effects using parental samples and the remaining F1 hybrid samples (J. Coolon pers. comm.). In all
554 three tests, we assigned significance after adjusted *p*-values for multiple comparisons using the False
555 Discovery Rate method (Benjamini and Yekutieli, 2001). Using the results from each test, we categorized
556 each gene into one of five classes using the following logic, which is based on previous studies (Emerson
557 and Li 2010, McManus et al., 2010):
558

Category	DE parents	F1 allelic imbalance	*Trans* test
cis only	True	True	False
trans only	True	False	True
cis + *trans*	True	True	True
Compensatory	False	True	True
Conserved	False	False	False

559 Any genes that did not fit into the patterns found above were categorized as “undetermined” and were
560 excluded from further analysis. A complete list of genes and their categories for each condition is available in
561 the Supplemental Data.
562 *Identification of sources of trans effects*
563 To investigate potential sources of observed *trans* effects, we looked for genes differentially expressed in our
564 uninfected samples. This included genes that are differentially expressed between A4 and B6 only uninfected
565 samples and genes differentially expressed in response to infection and effects in the control (Groups 1 and

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

3 from Figures 1-2). These genes were then intersected with a list of known *Drosophila* transcription factors as well as known immune genes (De Gregorio et al., 2001; Lemaitre and Hoffman 2007; Celniker et al., 2013, Troha et al., 2018). Only genes that were transcription factors, immune detection genes, or immune signaling genes were considered to be candidates.

Analysis of SNPs

To better understand the effects of sequence changes on coding regions between our lines we used the Ensemble Variant Effect Predictor Tool (VEP) to predict the effects of SNPs on the resulting amino acid sequence (McLauren et al., 2016). We first created two mutually exclusive lists of genes. The first list consists of genes found to be expressed in the unstimulated fat body above a CPM of 1 and excluding genes found to be differentially expressed in response to infection, this list was referred to as *fat body expressed* genes. The second list consists of only genes differentially expressed in response to infection with either *Efae* or *Smar,* referred to as *DE infection.* We then subsetted the *DE infection* genes to make a list of genes that are differentially expressed genes that are also previously verified immune response genes, which we called *DE immune*. For each list of genes, we pulled out SNPs falling into coding regions and ran these through VEP. The proportion of synonymous to non-synonymous SNPs was then compared between conditions.

Description of statistical tests

p-values for all single and multiple proportion comparisons were calculated using R's prop.test function which performs a Chi-square test with Yate's continuity correction. For data where more than one test was performed, *p*-values were Bonferroni corrected by multiplying the p-value by the number of tests performed.

Data Access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE155033.

Acknowledgements

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We would like to thank J.J. Emerson and Xinwin Zhang for their thoughtful comments and suggestions on this work, Ali Mortazavi and Lorrayne Serra for their insight and access to sequencing equipment, Tom Schilling for access to microinjection equipment, Joseph Coolon, Carl de Boer and Rabi Murad for sharing their scripts as well as general guidance with computational protocols, and Anthony Long and Mahul Chakraborty for their insight helpful comments on the genomes utilized. This work was funded in part by the National Science Foundation, Award 1953324 to Z.W. B.R. was supported by an NSF Bridge to the Doctorate Fellowship. S.F. was supported by a UCI UROP award. O.O. was supported by NIH Grant R25 GM055246, T34 GM136498, and an UCI UROP award.

Disclosure Declarations

The authors have no conflicts of interest to declare.

References

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, *29*(4), 1165–1188. https://doi.org/10.1214/aos/1013699998

Bischoff, V., Vignal, C., Duvic, B., Boneca, I. G., Hoffmann, J. A., & Royet, J. (2006). Downregulation of the Drosophila immune response by peptidoglycan- recognition proteins SC1 and SC2. *PLoS Pathogens*. https://doi.org/10.1371/journal.ppat.0020014

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Buchon, N., Silverman, N., & Cherry, S. (2014). Immunity in Drosophila melanogaster — from microbial recognition to whole- organism physiology. *Nature Publishing Group*, *14*(12), 796–810. https://doi.org/10.1038/nri3763

Busse, M. S., Arnold, C. P., Towb, P., Katrivesis, J., & Wasserman, A. (2007). A κB sequence code for pathway-specific innate immune responses, *26*(16), 3826–3835. https://doi.org/10.1038/sj.emboj.7601798

Charroux, B., Capo, F., Kurz, C. L., Peslier, S., Chaduli, D., Viallat-lieutaud, A., & Royet, J. (2018). Cytosolic and Secreted Peptidoglycan-Degrading Enzymes in Drosophila Respectively Control Local and Systemic Immune Responses to Microbiota. *Cell Host and Microbe*, *23*(2), 215-228.e4. https://doi.org/10.1016/j.chom.2017.12.007

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R., & Wittkopp, P. J. (2014). Tempo and mode of regulatory evolution in Drosophila. *Genome Research*, *24*(5), 797–808. https://doi.org/10.1101/gr.163014.113

Davidson, J. H., & Balakrishnan, C. N. (2016). Gene Regulatory Evolution During Speciation in a Songbird. *G3: Genes|Genomes|Genetics*, *6*(5), 1357–1364. https://doi.org/10.1534/g3.116.027946

Emerson, J. J., & Li, W.-H. (2010). The genetic basis of evolutionary change in gene expression levels. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1552), 2581–2590. https://doi.org/10.1098/rstb.2010.0005

Frochaux, M. V., Bou Sleiman, M., Gardeux, V., Dainese, R., Hollis, B., Litovchenko, M., … Deplancke, B. (2020). cis-regulatory variation modulates susceptibility to enteric infection in the Drosophila genetic reference panel. *Genome Biology*, *21*(1), 6. https://doi.org/10.1186/s13059-019-1912-z

Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., … Marioni, J. C. (2012). Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Research*, *22*(12), 2376–2384. https://doi.org/10.1101/gr.142281.112

Hammonds, A. S., Bristow, C. A., Fisher, W. W., Weiszmann, R., Wu, S., Hartenstein, V., … Celniker, S. E. (2013). Spatial expression of transcription factors in Drosophila embryonic organ development, 1–15.

Hanson, M. A., Dostálová, A., Ceroni, C., Poidevin, M., Kondo, S., & Lemaître, B. (2019). Synergy and remarkable specificity of antimicrobial peptides in vivo using a systematic knockout approach. *ELife*, *8*, 1–24. https://doi.org/10.7554/eLife.48778

Hotson, A. G., & Schneider, D. S. (2015). Drosophila melanogaster natural variation affects growth dynamics of infecting Listeria monocytogenes. *G3: Genes, Genomes, Genetics*, *5*(12), 2593–2600. https://doi.org/10.1534/g3.115.022558

Iatsenko, I., Kondo, S., Mengin-Lecreulx, D., & Lemaitre, B. (2016). PGRP-SD, an Extracellular Pattern-Recognition Receptor, Enhances Peptidoglycan-Mediated Activation of the Drosophila Imd Pathway. *Immunity*, *45*(5), 1013–1023. https://doi.org/10.1016/j.immuni.2016.10.029

Juneja, P., Quinn, A., & Jiggins, F. M. (2016). Latitudinal clines in gene expression and cis -regulatory element variation in Drosophila melanogaster. *BMC Genomics*, 1–11. https://doi.org/10.1186/s12864-016-3333-7

Kassambara, A., & Kosinski, M. (2019). survminer: Drawing Survival Curves using “ggplot2.” Retrieved from https://cran.r-project.org/package=survminer

Khalil, S., Jacobson, E., Chambers, M. C., & Lazzaro, B. P. (2015). Systemic bacterial infection and immune defense phenotypes in Drosophila melanogaster. *Journal of Visualized Experiments*, *2015*(99), 1–9. https://doi.org/10.3791/52613

King, E. G., Macdonald, S. J., & Long, A. D. (2012). Properties and power of the Drosophila synthetic population resource for the routine dissection of complex traits. *Genetics*. https://doi.org/10.1534/genetics.112.138537

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

King, M. C., & Wilson, A. C. (1975). Evolution ay Two Levels in humans and Chimpanzees. *Science*, *188*(4184), 107–118. Retrieved from http://www.jstor.org/stable/1739875

Krueger, F. (n.d.). ASAP-Allele-sp[ecific alignment pipeline. Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/ASAP/

Lazzaro, B. P., Sceurman, B. K., Clark, A. G., & Sciences, B. (2004). Genetic Basis of Natural Variation in D . melanogaster. *Science*, *303*(March), 1873–1877.

Lazzaro, B. P., Sackton, T. B., & Clark, A. G. (2006). Genetic Variation in Drosophila melanogaster Resistance to Infection : *Genetics*, *1554*(November), 1539–1554. https://doi.org/10.1534/genetics.105.054593

Lemaitre, B., & Hoffmann, J. (2007). The Host Defense of Drosophila melanogaster. https://doi.org/10.1146/annurev.immunol.25.022106.141615

Lu, Y., Su, F., Li, Q., Zhang, J., Li, Y., Tang, T., … Yu, X. Q. (2020). Pattern recognition receptors in Drosophila immune responses. *Developmental and Comparative Immunology*. https://doi.org/10.1016/j.dci.2019.103468

McCarthy, D. J., Chen, Y., & Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, *40*(10), 4288–4297. https://doi.org/10.1093/nar/gks042

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., … Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1), 1–14. https://doi.org/10.1186/s13059-016-0974-4

McManus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R., & Wittkopp, P. J. (2010). Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Research*, *20*(6), 816–825. https://doi.org/10.1101/gr.102491.109

Murad, R., Macias-Muñoz, A., Wong, A., Ma, X., & Mortazavi, A. (2019). Integrative analysis of *Hydra* head regeneration reveals activation of distal enhancer-like elements. *BioRxiv*, 544049. https://doi.org/10.1101/544049

Osada, N., Miyagi, R., & Takahashi, A. (2017). Cis - and trans -regulatory effects on gene expression in a natural population of Drosophila melanogaster. *Genetics*, *XXX*(June), 1–12. https://doi.org/10.1534/genetics.XXX.XXXXXX

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Pearson, W. R. (2013). Selecting the Right EMR Vendor Selecting the Right EMR Vendor. *Curr Protoc Bioinformatics*, (43), 3.5.1–3.5.9. https://doi.org/10.1002/0471250953.bi0305s43.Selecting

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Sackton, T. B., Lazzaro, B. P., & Clark, A. G. (2010). Genotype and Gene Expression Associations with Immune Function in Drosophila. *PLoS Genetics*, *6*(1). https://doi.org/10.1371/journal.pgen.1000797

Salinas, F., De Boer, C. G., Abarca, V., García, V., Cuevas, M., Araos, S., … Cubillos, F. A. (2016). Natural variation in non-coding regions underlying phenotypic diversity in budding yeast. *Scientific Reports*, *6*, 1–13. https://doi.org/10.1038/srep21849

Schaefke, B., Emerson, J. J., Wang, T. Y., Lu, M. Y. J., Hsieh, L. C., & Li, W. H. (2013). Inheritance of gene expression level and selective constraints on trans-and cis-regulatory changes in yeast. *Molecular Biology and Evolution*, *30*(9), 2121–2133. https://doi.org/10.1093/molbev/mst114

Scheiermann, C., Kunisaki, Y., & Frenette, P. S. (2013). Circadian control of the immune system. *Nature Reviews Immunology*, *13*(3), 190–198. https://doi.org/10.1038/nri3386

Serra, L., Chang, D., Macchietto, M., Williams, K., Murad, R., Lu, D., … Mortazavi, A. (2018). Adapting the Smart-seq2 Protocol for Robust Single Worm RNA-seq. *Bio-Protocol*, *8*(4), 1–16. https://doi.org/10.21769/bioprotoc.2729

Signor, S. A., & Nuzhdin, S. V. (2018). The Evolution of Gene Expression in cis and trans. *Trends in Genetics*, *34*(7), 532–544. https://doi.org/10.1016/j.tig.2018.03.007

Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, *4*(1521). https://doi.org/10.12688/f1000research.7563.1

Takada, Y., Miyagi, R., Takahashi, A., Endo, T., & Osada, N. (2017). A Generalized Linear Model for Decomposing Cis -regulatory, Parent-of-Origin, and Maternal Effects on Allele-Specific Gene Expression. *G3 (Bethesda, Md.) Genes|Genomes|Genetics Genes|Genomes|Genetics*, *7*(7), 2227–2234. https://doi.org/10.1534/g3.117.042895

Tanji, T., Yun, E. Y., & Ip, Y. T. (2010). Heterodimers of NF-κB transcription factors DIF and Relish regulate antimicrobial peptide genes in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1009473107

Therneau, T. M. (2020). A Package for Survival Analysis in R. Retrieved from https://cran.r-project.org/package=survival

Therneau, T. M. (2000). *Modeling Survival Data: Extending the {C}ox Model*. New York: Springer. Retrieved from https://cran.r-project.org/package=survival

Thurmond, J., Goodman, J. L., Strelets, V. B., Attrill, H., Gramates, L. S., Marygold, S. J., … Baker, P. (2019). FlyBase 2.0: The next generation. *Nucleic Acids Research*, *47*(D1), D759–D765. https://doi.org/10.1093/nar/gky1003

Troha, K., Im, J. H., Revah, J., Lazzaro, B. P., & Buchon, N. (2018). *Comparative transcriptomics reveals CrebA as a novel regulator of infection tolerance in D. melanogaster*. *PLoS Pathogens* (Vol. 14). https://doi.org/10.1371/journal.ppat.1006847

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature Publishing Group*, *430*(1), 85–88.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.30.229641; this version posted July 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

735 Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2008). Regulatory changes underlying expression
736 differences within and between Drosophila species. *Nature Genetics*, *40*(3), 346–350.
737 https://doi.org/10.1038/ng.77

738 Zaidman-Rémy, A., Hervé, M., Poidevin, M., Pili-Floury, S., Kim, M. S., Blanot, D., … Lemaitre, B.
739 (2006). The Drosophila Amidase PGRP-LB Modulates the Immune Response to Bacterial
740 Infection. *Immunity*, *24*(4), 463–473. https://doi.org/10.1016/j.immuni.2006.02.012

741 Zhang, X., Emerson, J. J. (2019). Inferring Compensatory Evolution of cis- and trans-Regulatory
742 Variation. *Trends in Genetics*, *35*(1), 1–3. https://doi.org/10.1016/j.tig.2018.11.003