The effect of developmental pleiotropy on the evolution of immune genes in *Drosophila melanogaster*

Thi Minh Ngo and Ann T. Tate*

Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA

*Corresponding email: a.tate@vanderbilt.edu

Article Type: Article (Discoveries)

Running Title: Pleiotropy in immune system evolution

Keywords: molecular evolution, Toll pathway, insect immunity, evolutionary constraint, adaptive evolution

Abstract

The pressure to survive relentless pathogen exposure explains the frequent observation that immune genes are among the fastest-evolving ones in the genomes of many taxa, but an intriguing proportion of immune genes also appear to be under purifying selection. Though variance in evolutionary signatures of immune genes is often attributed to differences in gene-specific interactions with microbes, this explanation neglects the possibility that immune genes pleiotropically participate in other biological processes that could constrain adaptive selection. In this study, we analyzed available transcriptomic and manual annotation data from *Drosophila melanogaster* to uncover substantial pleiotropic overlap in the developmental and immunological functions of genes involved in immune signaling. As developmental pleiotropy could constrain both the deployment and evolution of a gene product for immunological purposes, we predicted that pleiotropic immune genes would show stronger signatures of purifying selection than non-pleiotropic immune genes. We further predicted that, within the pleiotropic gene class, genes expressed early in development or more broadly across developmental stages would be under stronger purifying selection than genes with stage-specific functions. Using population genomics data from *D. melanogaster* and related species, we show that pleiotropic immune genes do undergo slightly slower evolutionary rates than those having no known developmental functions, and that signatures of purifying selection are significantly stronger for broadly-expressed pleiotropic immune genes. This study underscores the need to investigate immune system evolution in the broader context of host life history and development, and raises new questions about the evolution and maintenance of pleiotropic genetic architecture.

Introduction

Over evolutionary time, organisms have developed defense mechanisms against microbial pathogens and parasites which counter-adapt, in turn, to maintain successful infection strategies. Host immune systems put selective pressure on microbes to evade host recognition, repel antimicrobial effectors, and even manipulate immune signaling components to dampen host defenses (Schmid-Hempel 2008; Heil 2016). Hosts that cannot circumvent these machinations could suffer massive fitness costs from infection. As a result, co-evolutionary dynamics represent a major driving force in molecular evolution (Paterson et al. 2010).

How should we expect selection to act on immune system genes? Host adaptation to microbial pressure should drive positive, directional selection or, in the face of coevolutionary negative frequency dependence, balancing selection that maintains polymorphism in populations (Casals et al. 2011; Sackton 2019). Studies in species as diverse as humans (Mukherjee et al. 2009; Casals et al. 2011)(Mukherjee et al. 2009), non-human mammals (Seabury et al. 2010; Areal et al. 2011) and insects (Sackton et al. 2007; Obbard et al. 2009; Rottschaefer et al. 2015) have found evidence for both positive and balancing selection in immune system recognition and effector genes (Unckless et al. 2016). For example, Obbard *et al*. found that *Drosophila melanogaster* immune genes, as a class, have higher rates of adaptive substitution than location-matched non-immune genes (Obbard et al. 2009). However, they found that these trends were driven by a few particularly rapidly evolving genes associated with a subset of immune signaling pathways, while also uncovering a surprising prevalence of purifying selection on immune genes in other pathways. If parasites frequently target or evade signaling components, why wouldn’t those targets show rapid adaptation?

The answer may depend on a crucial but underappreciated quality of immune systems. Genetic pleiotropy arises when a single gene product contributes to multiple discrete phenotypic traits, and many components of immune pathways appear to be pleiotropic. Since the discovery of the Toll pathway, for example, numerous studies (and indeed Nobel prizes) have marveled at its dual role in development and innate immune system signaling (Lemaitre et al. 1997; DiAngelo et al. 2009) and its pleiotropic status appears to be at least partially conserved from insects to mammals (Anthoney et al. 2018). More broadly, a recent study estimated that ~17% of human genes affect multiple discrete phenotypic traits, and functional enrichment analysis of this pleiotropic gene set revealed immune system functions to be among the most over-represented processes (Sivakumaran et al. 2011). When a pleiotropic mutation affects uncorrelated traits, opposing forces of selection on each trait can reduce the efficacy of selection and resist the fixation of adaptive substitutions (Fraïsse et al. 2019). Thus, the adaptive evolution of pleiotropic immune genes may be constrained by the deleterious effects of substitutions on other traits.

Pleiotropy between development and immunity is particularly intriguing because a developmental program must be carried out faithfully for an organism to progress through its life cycle, resulting in purifying selection on genes involved in embryonic and early life development. Indeed, developmental pleiotropy (defined by the number of genetic interactions (Stark et al. 2006)) has been shown in *D. melanogaster* to constrain positive selection in early-expressed genes due to a higher number of functional interactions in those genes that render mutations deleterious (Artieri et al. 2009). More broadly, evolutionary conservation is predicted to be highest when developmental architecture is at its most connected and least modular, as exemplified by the slime mold *Dictyostelium*, which undergoes its most complex (and evolutionarily constrained) developmental processes late in ontogeny (Tian et al. 2013). We

hypothesize that developmental pleiotropy could constrain immune gene evolution, particularly for genes involved in the most complex stages of development, leading to an under-representation of signatures of positive selection on immune genes relative to theoretical expectations.

Insects can serve as particularly valuable models for studying the evolutionary consequences of developmental and immunological pleiotropy due to their discrete life stages, a wealth of genomic resources, and availability of studies on immune gene function (Evans et al. 2013; Palmer and Jiggins 2015; Viljakainen 2015). The canonical components of an insect innate immune response include microbial recognition, signal transduction to initiate cellular and humoral responses, and production of effector molecules for pathogen clearance (Lemaitre and Hoffmann 2007). Many genes and signaling pathways previously identified as core participants in these processes are also broadly conserved among species (Waterhouse et al. 2007), including two of the best studied pathways, Toll and Imd. Although these two pathways have been shown to independently mediate defense against specific microorganisms, they can also act synergistically to activate downstream target response genes such as antimicrobial peptides, lysozymes and other pathogen-clearing effectors (Ferrandon et al. 2007; Tanji et al. 2007). Other crucial immune signaling pathways include the melanization pathway, JAK-STAT, and RNAi, which participate in injury and antiviral responses (Buchon et al. 2014). While the Toll pathway is the most recognized example of developmental and immunological pleiotropy in insect immune systems, previous work has highlighted potential pleiotropy within other pathways (Tate and Graham 2015). For example, the same components of the melanization pathway responsible for tanning the insect cuticle after each larval molt are also used for melanizing parasitoid eggs and neutralizing pathogenic fungi, leading to allocation issues when an insect needs to

accomplish both at once (McNeil et al. 2010; Parker et al. 2017). Thus, pleiotropy is likely to interfere with the deployment of immune responses if a host needs to use a gene product for both development and immunity in the same life stage. Even if these functions are segregated into different life stages, however, could pleiotropy still constrain immune system evolution?

We predict that immune genes that have a pleiotropic developmental function will be more likely to experience evolutionary constraint, as defined by higher rates of purifying selection and a lower frequency of positive selection, than immune genes that have no known developmental function. Further, we predict that pleiotropic genes with a more specialized role in early development or that are crucial to multiple developmental stages will be the most constrained, relative to genes involved in later and less conserved developmental processes. To investigate these predictions, we combine transcriptional and functional genomics data from fruit flies (*Drosophila* spp.) to characterize the overall and immune pathway-specific degree of pleiotropy among immune and developmental genes. We then analyze the rates of adaptive selection in immune genes over long evolutionary periods using genomics data from six species in the melanogaster group, as well as over the shorter term, using population genomics data from *D. melanogaster*. Empirical support for our predictions would raise the question of why evolution would maintain pleiotropy between development and immunity given the potential for conflict and constraint. On the other hand, if pleiotropic immune genes are not more constrained than non-pleiotropic ones, this study could inspire future investigations into compensatory evolution and the role of network architecture in minimizing evolutionary conflict.

Results

Extent of developmental pleiotropy in immune genes

To determine the prevalence of developmental pleiotropy among immune genes, we started by curating separate lists of immune and developmental genes. Previous studies have employed various methods to curate immune gene lists, ranging from using only Gene Ontology annotations (Fraïsse et al. 2019) to compiling experimentally confirmed and/or computationally predicted immune gene orthologs (Early et al. 2017). Taking these different approaches into account, we employed several sources to assemble a comprehensive suite of genes that participate in immunity (Table 1 and Methods). In total, we assembled a list of 808 immune genes, of which 551 genes have known canonical roles in immunity and 107 genes play a role in immune system development, as annotated by Gene Ontology and previous studies (Early et al. 2017). The degree of overlap between different immune gene list sources can be found in Supplemental Figure 1. The list of developmental genes contains 3346 genes, of which 262 genes are annotated specifically as “embryonic development” genes and 508 as “post-embryonic development.” Some embryonic development genes also participate in post-embryonic development (overlap visualized in Supplemental Figure 2).

Genes that appear in both the immune and developmental gene lists were labeled as “pleiotropic.” When considering immune genes as those identified by all methods including manually curated, GO annotated and differentially expressed genes, we found 354 immune genes (43.8%) to be pleiotropic (Table 1, line 1). When considering immune genes as those that directly contribute to an immune response while excluding genes participating in development of the immune system, 299 (39.7%) genes are considered pleiotropic (Table 1, line 2). Under the most conservative definition of development (only genes that directly participate in embryonic

development or 7.8% (262/3346) of all annotated developmental genes), 52 immune genes (6.9%) still meet the definition of pleiotropy (Table 1). The full list of immune, developmental and pleiotropic genes under different categorization methods is included in Supplemental Table 1. Note that although we used several methods to compile a list of pleiotropic genes, the conclusions generated throughout this study are robust to different categorical definitions of immunity, development, and pleiotropy (see expanded discussion in Methods). Therefore, from this point on, for simplicity, we refer to our immune gene group as those defined using the sources from Table 1, section 2, which comprises Immune Response GO-annotated genes, immune genes employed in previous large scale studies, and a core set of genes differentially expressed in ten bacterial infections (Troha et al. 2018).

Comparison of pleiotropic and non-pleiotropic immune gene characteristics

Immune genes can be categorized into different classes, such as recognition, signaling, and effector, depending on their canonical function in an immune response. We were curious whether certain classes of immune genes are more likely to have pleiotropic status than others. We divided immune genes into major categories, relying on both annotation from previous studies (Sackton et al. 2007; Early et al. 2017) and manual annotation based on gene description in FlyBase (Supplemental Table 2). According to this classification system, the number of genes confirmed to each category includes 33 recognition genes, 123 signaling genes and 27 effector genes (Supplemental Figure 3). As represented in Figure 1A, the signaling immune class contains the highest proportion of pleiotropic genes (66.67%, n = 123). Moreover, using the PANTHER pathway database, we found that pleiotropic genes are, on average, associated with more pathways than non-pleiotropic ones (Supplemental Table 3).

We also wanted to know whether our curated immune-developmental pleiotropic genes exhibit characteristics associated with alternative definitions of pleiotropy, such as a high number of associated protein-protein interactions and gene-gene interactions that reflect activity at the molecular level. When comparing pleiotropic and non-pleiotropic immune genes (Fig. 1B-C), we do find that pleiotropic genes have significantly more protein-protein interactions (Wilcoxon test, $p = 6.4e-06$) and more gene-gene interactions (Wilcoxon test, $p = 1.4e-09$). Moreover, pleiotropic genes are associated with more Biological Processes (Wilcoxon test, $p < 2e-16$) and Molecular Functions (Wilcoxon test, $p < 2e-16$) GO terms than non-pleiotropic genes (Supplemental Figure 4).

Expression specificity across stages and tissues between pleiotropic and non-pleiotropic genes

To investigate the hypothesis that broadly expressed pleiotropic genes are under stronger evolutionary constraint than specific ones, we determined gene expression specificity across stages and tissues for pleiotropic and non-pleiotropic immune genes using the τ specificity index ((Yanai et al. 2005), see methods). A large τ value indicates specific expression while a small value indicates broad expression across stages or tissues. While we could not confidently determine whether any given gene plays only a developmental or immunological role or both at any given stage, genes involved in development at multiple life stages may present a temporal as well as evolutionary constraint on the immunological function of that gene.

We found that, in uninfected insects, pleiotropic immune genes tend to be more broadly expressed across stages than both non-pleiotropic immune genes (Fig. 2A, Wilcoxon test, $p = 1.4e-06$) and non-pleiotropic developmental genes (Wilcoxon test, $p = 0.023$). We also found that the most specific pleiotropic genes, determined by the top 25th percentile in τ (expression specificity) value, mostly have maximal expression during the embryonic stage (43% among

pleiotropic genes vs 3.6% among non-pleiotropic immune genes) while the most specific non-pleiotropic immune genes tend to have maximum expression during the larval stage (Figure 2B, Supplemental Table 5). At the tissue level, pleiotropic genes are also expressed more broadly than non-pleiotropic immune genes, and this trend is consistent throughout all life stages (Figure 2C). We found no significant differences in the tissue expression specificity between developmental genes and pleiotropic genes except in the adult stage (Fig. 2C), where developmental genes showed more specific patterns of expression.

Signatures of molecular evolution in non-pleiotropic and pleiotropic immune genes

To address whether pleiotropic genes are more evolutionarily constrained than non-pleiotropic immune genes, we calculated the rate of molecular evolution for each gene. We found that pleiotropic immune genes have significantly lower dN/dS ratios (Fig. 3A, $\omega = 0.0597$ vs $\omega = 0.0789$, $p = 0.038$), suggesting more evolutionary constraint compared to non-pleiotropic immune genes. This observation holds true regardless of the phylogenetic group chosen for the dN/dS analysis (Supplemental Figure 5). Interestingly, when using the branch-site model of the dN/dS analysis to detect signatures of positive selection along the *Dmel* lineage specifically, 6.35% (total n = 299) of pleiotropic genes are found to be significantly positively selected, in comparison to 3.08% (total n = 454) of non-pleiotropic immune genes showing positive selection (Fisher's exact test, $p = 2.2e-16$) (Supplemental Figure 6). We also found that pleiotropic genes broadly expressed across life stages have a significantly lower rate of molecular evolution than specific pleiotropic genes (Figure 3C, Wilcoxon test, $p = 0.041$), suggesting that the ontogenetic duration of pleiotropy may be an important determinant of evolutionary constraint.

On the other hand, we detected no difference in the MK α statistic between non-pleiotropic and pleiotropic genes (Figure 3D). It is notable that many genes show extreme

negative values of MK α even when using the FWW correction method (Fay et al. 2001), indicating prevalence of slightly deleterious mutations segregating in the population or test sensitivity to population dynamics or the complex nature of these alleles. This pattern is observed when using population data from both the Raleigh and Zambia populations (Supplemental Figure 7).

When we subsetted the development gene list to include only embryonic morphogenesis development functions, subsequent dN/dS analysis also indicate a lower median ω value for pleiotropic genes (Figure 3B, $\omega = 0.0561$ vs $\omega = 0.0672$), but this difference is not statistically significant ($p = 0.24$). When comparing MK α values between the same groups, pleiotropic genes ($N = 35$, $\alpha = 0.7083$) actually display a higher average rate of adaptive substitution compared to non-pleiotropic immune genes ($N = 511$, $\alpha = 0.4062$, $p = 0.00017$).

The extent of pleiotropy and signatures of molecular evolution in multiple signaling pathways

Since signaling pathways commonly coordinate both immunological and developmental functions, we next wanted to examine major signaling pathways for patterns of pleiotropy and signatures of molecular evolution (Figure 4). We were particularly interested in investigating whether there are certain components of a signaling pathway that tend to be pleiotropic or show discernable patterns of ω values.

As illustrated in Figure 4, there is a wide variety of pleiotropic gene organization and rates of molecular evolution across pathways. For example, we find that most of the downstream components of the Toll signaling pathway are pleiotropic, in contrast to, for example, the TGF-β or the Hedgehog signaling pathway which only contain pleiotropic ligands. Furthermore, there are instances where a pleiotropic component interacts with a non-pleiotropic component in the pathway, such as Tak1 (pleiotropic) and Tab2 (immune) in the IMD pathway. Since this figure

only depicts canonical signal transduction in the context of immunity, this suggests that Tak1, the pleiotropic component, may interacts with other molecules in developmental contexts as well. Overall, each pathway displays a unique distribution of immune-development pleiotropy and it would be interesting to further study these components to understand how the pleiotropy pattern affect the availability for deployment of these pathways in immunity and development.

Discussion

Researchers have long recognized that some immune genes, such as those in the Toll pathway, play double-duty in development (Lemaitre et al. 1996). Pleiotropy seems like it would be a liability for a host, for multiple reasons – what if a gene product cannot be deployed to fight a parasite because it is already being fully allocated to development? Shouldn't purifying selection on developmental genes constrain the rate of adaptation against parasite pressure, putting the host at a disadvantage during coevolution with rapidly evolving parasites? In this study, we investigated the relationship between immunity-development pleiotropy and signatures of molecular evolution in *Drosophila melanogaster* immune genes. Our results give credence to the notion that pleiotropy between development and immunity is actually quite common (Tate and Graham 2015). While pleiotropy does not appear to constrain the adaptive evolution of immune genes as systematically as we expected, we did identify stronger signatures of purifying selection in pleiotropic immune genes that were broadly expressed across developmental stages, relative to those with more stage-specific patterns of expression.

We first explored the extent of developmental pleiotropy in characterized immune genes. We observed that about 40-44% of immune genes are pleiotropic with development, concurring with a phenotypic screening study in mammals that more generally classified approximately 65% of screened alleles as pleiotropic across a range of phenotypes (De Angelis et al. 2015). We want

to note that the actual percentage of immune genes that are pleiotropic with other traits might even be higher, since this study only focused on pleiotropy in developmental processes. Upon analyzing the different immune gene classes for their prevalence of pleiotropy, we found that immune signaling genes are most likely to participate in developmental functions. This is expected since a signaling pathway is capable of activating the transcription of multiple genes, as opposed to, for example, effector genes which likely only interact with microbial pathogens or have specific immune functions. Further, we found that genes annotated with both immune and developmental functions are also associated with common molecular parameters associated with pleiotropy (Alvarez-Ponce et al. 2017), as they have more protein-protein interactions, gene-gene interactions and are expressed in more stages. Although these interactions may not directly reflect immune or developmental activities, this suggests that the pleiotropic genes might participate in different processes by interacting with more molecular partners. The ubiquitous expression of pleiotropic genes across stages compared to non-pleiotropic genes suggests that one or both of the immune and developmental functions are required throughout ontogeny. In the future, it would be interesting to know the extent to which pleiotropic genes exhibit temporal segregation of developmental processes and immune roles in different life stages, as opposed to simultaneous participation in both functions in one or more stages.

Regardless of how we curate our immune gene lists, we observed that in general, pleiotropic genes have lower rates of protein coding sequence evolution. While this could be attributed to antagonistic pleiotropy between immune and developmental functions, it is interesting to note that the difference is not statistically significant when we consider only embryonic developmental genes. Further studies are needed to understand why immune genes with embryonic developmental pleiotropy are not more constrained than non-pleiotropic genes,

assuming that this result is not due solely to Type II error from low sample sizes. We also want to note that our dN/dS analysis relies on genes with 1:1 orthologs across all five species, excluding genes with copy number variations or missing orthologs in any species from the group. For future studies, it would be interesting to revisit these genes, since the effect of paralog functional diversification on pleiotropic constraint could be significant (Hahn 2009; Long et al. 2013).

We did not observe a difference in McDonald-Kreitman α values between the pleiotropic and non-pleiotropic gene groups. Several interpretations are possible, and we keep in mind both the technical and biological explanations for these observations. First, the McDonald-Kreitman test has several limitations that can result in conservative estimation of α (Uricchio et al. 2019), and it is unclear how exactly our gene lists are affected. Another apparent limitation is the fact that genes in the non-pleiotropic group can actually be pleiotropic in biological contexts other than development, further obscuring the true pleiotropic status of each gene. However, we found no relationship between the number of Biological Processes associated with an immune gene and its ω or α value (Supplemental Figure 10), suggesting that this is not likely to be a systematic explanation. It is also possible that compensatory mutations arise in response to deleterious sequence changes, potentially leading to the observation of more polymorphism and overestimation of α in pleiotropic genes. This concept has been explored Vedanayagam et al, where the authors found that genes with mutations affecting multiple phenotypes exhibited higher rates of protein evolution (Vedanayagam and Garrigan 2015). In the context of our study, we speculate that when mutations occur in pleiotropic proteins that have antagonistic effects on immunity and development, compensatory substitutions could arise to resolve this conflict. While we did not examine this explanation in the current study, a previous study suggested that

the presence of a non-synonymous mutation greatly increases the chance of finding other substitutions nearby, possibly reflecting the correlated evolution of codons within a protein module (Callahan et al. 2011).

Studies that have considered the more general relationship between signatures of molecular evolution and molecular pleiotropy have also reached contrasting conclusions. For example, in some cases, pleiotropy, as defined by connectivity in protein-protein or gene co-expression networks, is negatively correlated with molecular evolution rates (Alvarez-Ponce et al. 2017; Masalia et al. 2017), while others have detected very minimal or no correlations (Hahn et al. 2004; Fraïsse et al. 2019). The variance in these results could be attributed to differences in study organisms, different experimental contexts and the inherent differences in the various definitions of pleiotropy. Therefore, the exact effect of pleiotropy on molecular evolution is most likely to be context specific, emphasizing the importance of our study in focusing on very specific biological processes such as immunity and development. One caveat associated with our study is that our analyses do not consider domain- or site-specific evolution, and thus we cannot parse signatures of selection on regions within a pleiotropic gene that might provide specific immune or developmental functions or be closely associated with compensatory mutations that obscure finer-scale evolutionary processes. Although such analysis would require very specific knowledge of the effect of each mutation to the immune and development phenotypes, future analyses could focus on a subset of genes with well-defined protein domain structures and protein-protein interaction data to refine the functional and evolutionary significance of pleiotropic activity.

The prevalence of upstream and downstream pleiotropic components varies among pathways, resisting simple conclusions about the evolutionary maintenance of pleiotropy. While

most pleiotropic components of the Toll pathway are intracellular, the only pleiotropic molecules from the TGF-β pathway (for example) are the extracellular ligands Dawdle and Dpp. Second, we observed cases where pleiotropic proteins are interacting with non-pleiotropic ones, or where evolutionarily constrained proteins interact with others that exhibit relatively high rates of adaptation. This might reflect alternative interactions between the pleiotropic component and molecules not annotated as part of canonical immune or developmental signaling pathways. This analysis raises new questions for future investigation: how can a signaling pathway balance its role in multiple biological processes? What are the key players and their characteristics that affect how a pathway is used across several contexts or life stages?

Overall, our study serves as the first one to systematically quantify the degree of pleiotropy in a specific biological context and investigate correlations between pleiotropy and rates of molecular evolution in immune systems. These results lay the groundwork for future work to tease apart the mechanistic framework of these pleiotropic patterns to understand how genetic architecture shapes the mode and tempo of immune system evolution and their influence on immune phenotypes**.**

Methods

Immune and developmental gene list curation

We curated a comprehensive list of genes representing immunity by combining several resources, starting with manually curated list from previous immune studies, including Lemaitre et al 2007 (Lemaitre and Hoffmann 2007) and Early et al 2017 (Early et al. 2017), which include most experimentally validated “canonical” immune genes. Separately, we appended Gene Ontology (GO)-annotated genes under the term “immune system process” (GO:0002376) to the list. We further sub-divided genes under this GO term into either “Immune Response” or

"Immune Development" genes to differentiate between genes that play direct roles in mounting an immune response and genes contributing to the development and maturation of the immune system. Finally, we added to our list a core set of immune genes from Troha et al 2018, which comprises 252 genes that show differential expression across infection with ten different bacterial species of variable virulence (Troha et al. 2018).

For each immune gene, we also assigned an immune gene class – recognition, signaling, or effector - based on the gene's known function in the immune system. If a gene has not been assigned a class in previous studies, we manually assign it a class based on the gene description from FlyBase. For a detailed description of each gene class definition, see Supplemental Protocol.

Separately, we created a list of GO-annotated developmental genes by querying the term "Developmental Process" (GO:0032502), while separately annotating genes belonging to the child term "embryonic morphogenesis" (GO:0048698). All GO annotation queries were conducted through FlyBase (Thurmond et al. 2019). A full list of genes in each group is included in Supplemental Table 1 and visualization of the degree of overlap between different resources is in Supplemental Figure 1.

Pleiotropy categorization

Pleiotropy refers to the phenomenon where a single gene influences multiple traits. However, the definition of "trait" can be ambiguous across different biological contexts, and thus pleiotropy can manifest at different levels and be detected by various methods (Paaby and Rockman 2013; Tyler et al. 2016). At the molecular level, pleiotropy can refer to the multiple biochemical roles that a gene can have and is frequently measured as the number of physical interacting partners (Hahn et al. 2004). At the developmental or phenotypic level, pleiotropy can

involve genes affecting distinct phenotypes or biological processes, as measured by the number of stage or tissues in which such genes are expressed (Artieri et al. 2009). Lastly, under an evolutionary perspective, pleiotropy can refer to the separate components of fitness that a gene might modulate, a well-known example being the antagonistic pleiotropy model for the evolution of aging (Williams 1957). Though many interpretations of pleiotropy exist, in this study, we are specifically concerned about pleiotropic genes at the phenotypic level. In particular, we focused on genes annotated to play roles in both immune and developmental processes. As such, if a gene is annotated as functioning in both immunity and development from the lists curated from the method described above, it was considered pleiotropic. A full list of pleiotropic genes is included in Supplemental Table 1.

For comparison purposes, we also calculated molecular metrics of pleiotropy for each gene in the genome regardless of annotated function in immunity or development. These measurements include expression stage specificity (described below), number of associated Biological Processes GO terms, number of associated Molecular Functions GO terms, number of protein-protein interactions, and number of gene-gene interactions. All raw data files were obtained through the FlyBase ftp server, and the latest version of each file was downloaded (March 2020, Supplemental Protocol).

Categorization of stage and tissue specificity

Genes with functions limited to specific tissues or life stages (and particularly later life stages) may have less pervasive effects on organismal fitness (Cutter and Ward 2005; Artieri et al. 2009), possibly buffering evolutionary constraint from pleiotropy. To calculate expression specificity, we applied the following equation (Yanai et al. 2005) to expression level data of all

D. melanogaster genes in all stages (embryo, larva, pupa, adult) and tissues (Supplemental Methods):

$$\tau = \frac{\sum_{j=1}^{n} 1 - \log(Aj)/\log(Amax)}{n-1}$$

In this equation, n is the number of of stages or tissues. A_j is the expression level at stage/tissue j, A_{max} the maximum expression level of of stages/tissues. Lower tau (τ) values signify specific expression in a certain stage/tissue, while a higher one indicates broad expression across all stages/tissues. (Fraïsse et al. 2019).

Pathway annotation

We used the PANTHER database to annotate our gene lists to pathway, if available. In short, all genes are compiled into a list of IDs, which is then used as a query in PANTHER (http://pantherdb.org/). We then downloaded the annotations and computed the total number of unique pathways associated with each gene group (pleiotropic vs. non-pleiotropic).

Molecular evolution data and analysis

The patterns of molecular evolution in immune genes are often studied using comparative approaches, either between different species or between individuals within the same species (Barreiro and Quintana-Murci 2010; Viljakainen 2015). Specifically, the strength and mode of selection acting on protein-coding genes is often measured by methods such as the ratio of nonsynonymous and synonymous mutations (dN/dS, or ω) ratio and the McDonald-Kreitman α estimation (McDonald and Kreitman 1991; Yang and Swanson 2002). In short, high estimates of ω or α are indicative of positive selection, while low values suggest purifying selection.

We obtained a list of orthologs within the melanogaster group from FlyBase (Supplemental Protocol). Only genes that have 1:1 orthologs for each of the six species were considered for this analysis. The longest amino acid sequences from each species' orthologs were aligned using

MAFFT (Katoh and Standley 2013) and threaded back to coding sequences to obtain codon alignments using pal2nal (Suyama et al. 2006). We used the PAML package (v4.9) (Yang 2007) to calculate dN/dS under model 0 of the codeml program, in which dN/dS is assumed to have one value averaged across all five species for each gene, and this is the value used for all subsequent analysis. To test signatures of positive selection along the Dmel lineage specifically, we used the branch-site model from the same PAML package. We also used a precalculated set of dN/dS data from flyDIVAS (Stanley and Kulathinal 2016) for comparison purposes. This resource contains dN/dS values under different models of molecular evolution of many phylogenetic groups. Since the conclusions from this study are not affected by the dN/dS dataset used (Supplemental Figure 9), all values reported are from our own calculations to reflect a more up-to-date ortholog assignment from FlyBase. As dN/dS and MK calculations rely on 1:1 orthologs, we want to note that some genes are not included in the results from Figure 3; of the 299 pleiotropic genes defined in this study, 258 have 1:1 orthologs and were included in the analysis. Similarly, 263/454 non-pleiotropic immune genes and 2132/3047 developmental genes have 1:1 orthologs available for dN/dS calculations. A table of ortholog counts for all Dmel genes across all 12 Dmel species can be found in Supplemental Table 4.

The McDonald-Kreitman (MK) test values were obtained from Murga-Moreno 2019 (Murga-Moreno et al. 2019) using polymorphism data from the Raleigh and Zambia population of *Drosophila melanogaster* along with genomic sequences from *Drosophila simulans* as the outgroup. Detailed explanations of the methodology can be found in (Murga-Moreno et al. 2019). Briefly, the number of nonsynonymous and synonymous mutations between individuals in the population as well as between Dsim and Dmel are determined from polymorphism data obtained from PopFly (Lack et al. 2015). The alpha values are calculated using the standard MK

equation with the Fay-Wycoff-Wu correction method, which considers only polymorphic sites with a frequency above 0.05. This accounts for segregating nonsynonymous mutations that can drastically lead to underestimation of α (Fay et al. 2001).

Statistical Analysis

All statistical analyses were conducted in R (3.6.1). We used Shapiro tests to assess distribution normality in datasets. For comparison between multiple groups (Figure 1B,C, 2A,C, 3A,B,D,E), we conducted Kruskal-Wallis tests followed by pairwise Wilcoxon with Benjamini-Hochberg correction in cases where there are a significant difference between groups. Fisher's exact test was used when assessing independence between two nominal variables (Figure 1A, Supplemental Figure 7).

Acknowledgments

We thank Nora Schulz, Alissa Williams, Stephanie Birnbaum, and other members of the Tate lab for comments and discussion. This work was supported by the National Institute of General Medical Sciences at the National Institutes of Health (grant number R35GM138007 to A.T.T.).

References

Alvarez-Ponce D, Feyertag F, Chakraborty S. 2017. Position matters: Network centrality considerably impacts rates of protein evolution in the human protein-protein interaction network. *Genome Biol. Evol.* 9:1742–1756.

De Angelis MH, Nicholson G, Selloum M, White JK, Morgan H, Ramirez-Solis R, Sorg T, Wells S, Fuchs H, Fray M, et al. 2015. Analysis of mammalian gene function through broad-based phenotypic screens across a consortium of mouse clinics. *Nat. Genet.* 47:969–978.

Anthoney N, Foldi I, Hidalgo A. 2018. Toll and toll-like receptor signalling in development. *Dev.* 145:1–6.

Areal H, Abrantes J, Esteves PJ. 2011. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol. Biol.* 11:368.

Artieri CG, Haerty W, Singh RS. 2009. Ontogeny and phylogeny: Molecular signatures of selection, constraint, and temporal pleiotropy in the development of Drosophila. *BMC Biol.* 7:1–14.

Barreiro LB, Quintana-Murci L. 2010. From evolutionary genetics to human immunology: How selection shapes host defence genes. *Nat. Rev. Genet.* 11:17–30.

Buchon N, Silverman N, Cherry S. 2014. Immunity in Drosophila melanogaster-from microbial recognition to whole-organism physiology. *Nat. Rev. Immunol.* [Internet] 14:796–810. Available from: http://dx.doi.org/10.1038/nri3763

Callahan B, Neher RA, Bachtrog D, Andolfatto P, Shraiman BI. 2011. Correlated evolution of nearby residues in drosophilid proteins. *PLoS Genet.* 7.

Casals F, Sikora M, Laayouni H, Montanucci L, Muntasell A, Lazarus R, Calafell F, Awadalla P, Netea M, Bertranpetit J. 2011. Genetic adaptation of the antibacterial human innate immunity network. *BMC Evol. Biol.* 11.

Cutter AD, Ward S. 2005. Sexual and temporal dynamics of molecular evolution in C. elegans development. *Mol. Biol. Evol.* 22:178–188.

DiAngelo JR, Bland ML, Bambina S, Cherry S, Birnbaum MJ. 2009. The immune response attenuates growth and nutrient storage in Drosophila by reducing insulin signaling. *Proc. Natl. Acad. Sci. U. S. A.* 106:20853–20858.

Early AM, Arguello JR, Cardoso-Moreira M, Gottipati S, Grenier JK, Clark AG. 2017. Survey

of global genetic diversity within the Drosophila immune system. *Genetics* 205:353–366.

Evans JD, Brown SJ, Hackett KJJ, Robinson G, Richards S, Lawson D, Elsik C, Coddington J, Edwards O, Emrich S, et al. 2013. The i5K initiative: Advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J. Hered.* 104:595–600.

Fay JC, Wyckoff GJ, Wu CI. 2001. Positive and negative selection on the human genome. *Genetics* 158:1227–1234.

Ferrandon D, Imler JL, Hetru C, Hoffmann JA. 2007. The Drosophila systemic immune response: Sensing and signalling during bacterial and fungal infections. *Nat. Rev. Immunol.* 7:862–874.

Fraïsse C, Puixeu Sala G, Vicoso B. 2019. Pleiotropy Modulates the Efficacy of Selection in Drosophila melanogaster. *Mol. Biol. Evol.* 36:500–515.

Hahn MW. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J. Hered.* 100:605–617.

Hahn MW, Conant GC, Wagner A. 2004. Molecular Evolution in Large Genetic Networks: Does Connectivity Equal Constraint? *J. Mol. Evol.* 58:203–211.

Heil M. 2016. Host manipulation by parasites: Cases, patterns, and remaining doubts. *Front. Ecol. Evol.* 4.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.

Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Detig RB, Stevens KA, Langley CH, Pool JE. 2015. The drosophila genome nexus: A population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. *Genetics* 199:1229–1241.

Lemaitre B, Hoffmann J. 2007. The Host Defense of Drosophila melanogaster . *Annu. Rev. Immunol.* 25:697–743.

Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA. 1996. The dorsoventral regulatory gene cassette spatzle/Toll/Cactus controls the potent antifungal response in Drosophila adults. *Cell* 86:973–983.

Lemaitre B, Reichhart JM, Hoffmann JA. 1997. Drosophila host defense: Differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* 94:14614–14619.

Long M, Vankuren NW, Chen S, Vibranovski MD. 2013. New gene evolution: Little did we know. *Annu. Rev. Genet.* 47:307–333.

Masalia RR, Bewick AJ, Burke JM. 2017. Connectivity in gene coexpression networks negatively correlates with rates of molecular evolution in flowering plants. *PLoS One* 12:1–10.

McDonald J, Kreitman M. 1991. Adaptive protein evolution at Adh in Drosophila. *Nature* [Internet] 351:652–654. Available from: http://ib.berkeley.edu/labs/slatkin/popgenjclub/pdf/mcdonald-kreitman1991.pdf

McNeil J, Cox-Foster D, Slavicek J, Hoover K. 2010. Contributions of immune responses to developmental resistance in Lymantria dispar challenged with baculovirus. *J. Insect Physiol.* 56:1167–1177.

Mukherjee S, Sarkar-Roy N, Wagener DK, Majumder PP. 2009. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. *Proc. Natl. Acad. Sci. U. S. A.* 106:7073–7078.

Murga-Moreno J, Coronado-Zamora M, Hervas S, Casillas S, Barbadilla A. 2019. iMKT: the

integrative McDonald and Kreitman test. *Nucleic Acids Res.* 47:W283–W288.

Obbard DJ, Welch JJ, Kim KW, Jiggins FM. 2009. Quantifying adaptive evolution in the Drosophila immune system. *PLoS Genet.* 5.

Paaby AB, Rockman M V. 2013. The many faces of pleiotropy. *Trends Genet.* 29:66–73.

Palmer WJ, Jiggins FM. 2015. Comparative genomics reveals the origins and diversity of arthropod immune systems. *Mol. Biol. Evol.* 32:2111–2129.

Parker BJ, Barribeau SM, Laughton AM, Griffin LH, Gerardo NM. 2017. Life-history strategy determines constraints on immune function. *J. Anim. Ecol.* 86.

Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, Quail M, Smith F, Walker D, Libberton B, et al. 2010. Antagonistic coevolution accelerates molecular evolution. *Nature* 464:275–278.

Rottschaefer SM, Crawford JE, Riehle MM, Guelbeogo WM, Gneme A, Sagnon N, Vernick KD, Lazzaro BP. 2015. Population genetics of Anopheles coluzzii immune pathways and genes. *G3 Genes, Genomes, Genet.* 5:329–339.

Sackton TB. 2019. Comparative genomics and transcriptomics of host–pathogen interactions in insects: evolutionary insights and future directions. *Curr. Opin. Insect Sci.* [Internet] 31:106–113. Available from: https://doi.org/10.1016/j.cois.2018.12.007

Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. 2007. Dynamic evolution of the innate immune system in Drosophila. *Nat. Genet.* 39:1461–1468.

Schmid-Hempel P. 2008. Parasite immune evasion: a momentous molecular war. *Trends Ecol. Evol.* 23:318–326.

Seabury CM, Seabury PM, Decker JE, Schnabel RD, Taylor JF, Womack JE. 2010. Diversity and evolution of 11 innate immune genes in Bos taurus taurus and Bos taurus indicus cattle.

Proc. Natl. Acad. Sci. U. S. A. 107:151–156.

Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, Rudan I, McKeigue P, Wilson JF, Campbell H. 2011. Abundant pleiotropy in human complex diseases and traits. *Am. J. Hum. Genet.* [Internet] 89:607–618. Available from: http://dx.doi.org/10.1016/j.ajhg.2011.10.004

Stanley CE, Kulathinal RJ. 2016. flyDIVaS: A comparative genomics resource for Drosophila divergence and selection. *G3 Genes, Genomes, Genet.* 6:2355–2363.

Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34:535–539.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34:609–612.

Tanji T, Hu X, Weber ANR, Ip YT. 2007. Toll and IMD Pathways Synergistically Activate an Innate Immune Response in Drosophila melanogaster. *Mol. Cell. Biol.* 27:4578–4588.

Tate AT, Graham AL. 2015. Dynamic patterns of parasitism and immunity across host development influence optimal strategies of resource allocation. *Am. Nat.* 186:495–512.

Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V, et al. 2019. FlyBase 2.0: The next generation. *Nucleic Acids Res.* 47:D759–D765.

Tian X, Strassmann JE, Queller DC. 2013. Dictyostelium development shows a novel pattern of evolutionary conservation. *Mol. Biol. Evol.* 30:977–984.

Troha K, Im JH, Revah J, Lazzaro BP, Buchon N. 2018. Comparative transcriptomics reveals CrebA as a novel regulator of infection tolerance in D. melanogaster. Available from: http://dx.doi.org/10.1371/journal.ppat.1006847

Tyler AL, Crawford DC, Pendergrass SA. 2016. The detection and characterization of pleiotropy: Discovery, progress, and promise. *Brief. Bioinform.* 17:13–22.

Unckless RL, Howick VM, Lazzaro BP. 2016. Convergent Balancing Selection on an Antimicrobial Peptide in Drosophila. *Curr. Biol.* [Internet] 26:257–262. Available from: http://dx.doi.org/10.1016/j.cub.2015.11.063

Uricchio LH, Petrov DA, Enard D. 2019. Exploiting selection at linked sites to infer the rate and strength of adaptation. *Nat. Ecol. Evol.* [Internet] 3:977–984. Available from: http://dx.doi.org/10.1038/s41559-019-0890-6

Vedanayagam JP, Garrigan D. 2015. The effects of natural selection across molecular pathways in Drosophila melanogaster. *BMC Evol. Biol.* [Internet] 15:1–9. Available from: http://dx.doi.org/10.1186/s12862-015-0472-4

Viljakainen L. 2015. Evolutionary genetics of insect innate immunity. *Brief. Funct. Genomics* [Internet] 14:407–412. Available from: https://academic.oup.com/bfg/article-lookup/doi/10.1093/bfgp/elv002

Waterhouse RM, Kriventseva E V, Meister S, Xi Z, Alvarez S, Bartholomay LC, Barillas-mury C, Bian G, Christensen BM, Dong Y, et al. 2007. Evolutionary Dynamics of Immune-Related Genes and Pathways in Disease-Vector Mosquitoes. *Science (80-.).* 316:1738–1743.

Williams GC. 1957. Pleiotropy, Natural Selection, and the Evolution of Senescence. *Evolution (N. Y).* 11:398–411.

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* 21:650–

659.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–1591.

Yang Z, Swanson WJ. 2002. Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol. Biol. Evol.* 19:49–57.

Tables

Table 1: The extent of pleiotropy as defined with different annotation methods.

	Definition	Pleiotropic	Immune Non-pleiotropic	Dev Non-pleiotropic
1	Immune = all Immune GO + previous citations. + DE (808) Dev = all Dev GO	354 (43.8%)	454	2992
2	Immune = Immune Response GO + previous citations + DE (753) Dev = all Dev GO	299 (39.7%)	454	3047
3	Immune = Immune Response GO + previous citations+ DE (753) Dev = embryonic Dev GO	52 (6.9%)	701	210
4	Immune = Immune Response GO + previous citations (551) Dev = all Dev GO	276 (50.1%)	275	3070

Notes: GO: Gene ontology annotation terms. DE = differentially expressed via transcriptional analyses. Dev = developmental. Previous citations = genes or gene lists manually or computationally identified as having immune system functions in Drosophila

Figures

Figure 1. Overall characterization of pleiotropic and non-pleiotropic immune genes. Each immune gene was assigned a "gene class" (**A**) depending on their canonical function in an immune response. For each class, the percentage of pleiotropic (those with developmental roles) and non-pleiotropic genes was determined. The number of known protein-protein interactions (ppi; **B**) and number of known gene-gene interactions (ggi; **C**) were also calculated for genes annotated as immune non-pleiotropic (green), pleiotropic for development and immunity (pink), or developmental non-pleiotropic (blue), represented on a log-scale and statistically analyzed using Kruskal-Wallis tests followed by pairwise Wilcoxon tests (p values on figure).

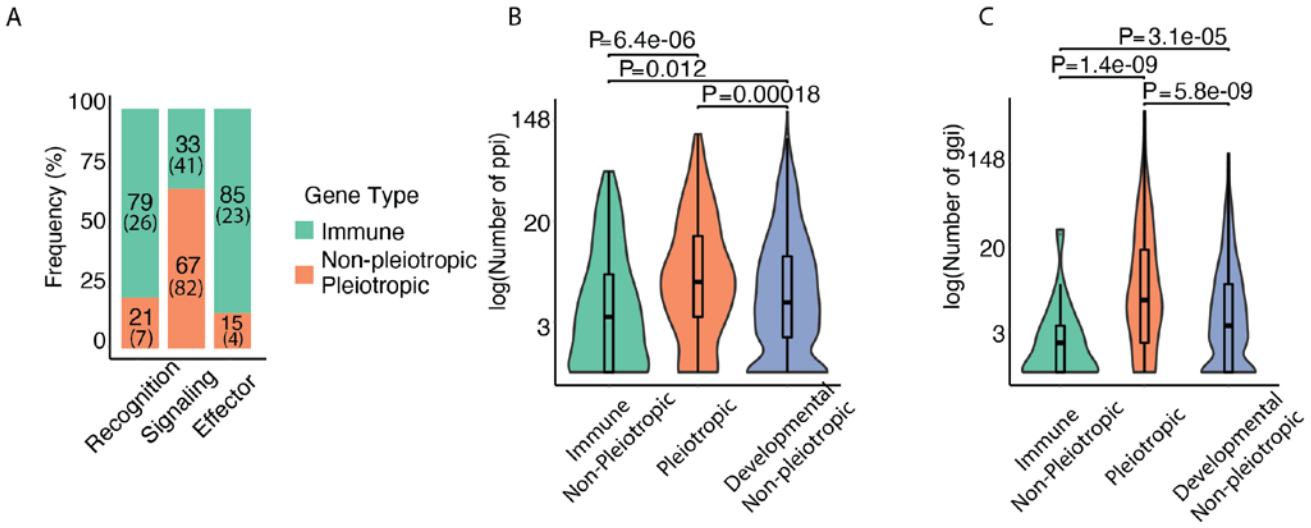

Figure 2. Comparison of gene expression stage specificity between pleiotropic and non-pleiotropic genes. The stage specificity tau value (**A**), which varies from 0 (broadly expressed across all stages) to 1 (expressed in only one stage) was calculated for each gene. For the non-pleiotropic and pleiotropic immune gene group (**B**), the genes within the top 25th percentile of τ value were characterized as "specific genes", and the stage with the highest expression for each gene was determined and tallied for the whole group. To compare tissue gene expression specificity between pleiotropic and non-pleiotropic genes within each life stage (**C**), the tau value (tissue specificity level) was calculated for each gene across tissues. Differences among groups were statistically analyzed using Kruskal-Wallis tests followed by pairwise Wilcoxon tests (p values on figure or *** indicates $p < 0.001$).

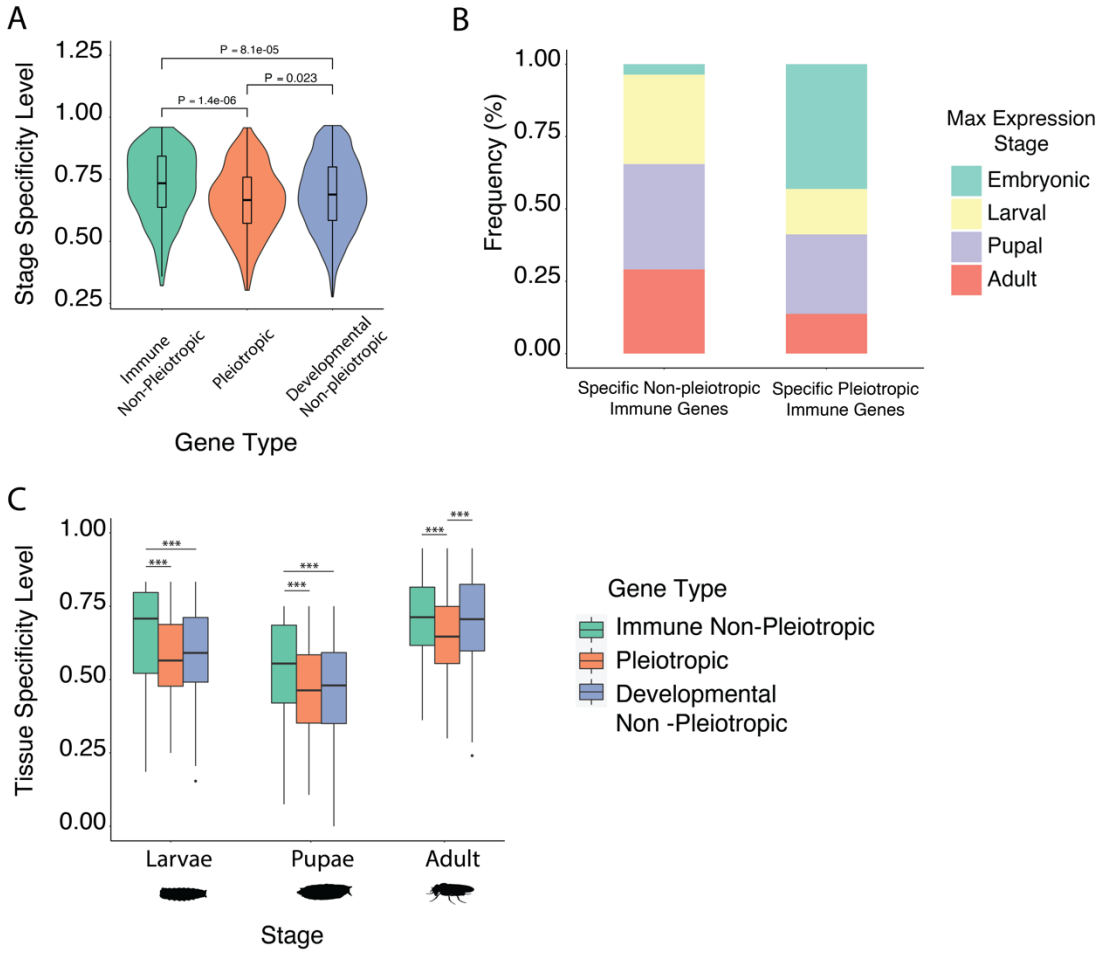

Figure 3. Comparision of dN/dS and McDonald-Kreitman alpha values between pleiotropic and immune- or developmental non-pleiotropic genes. dN/dS values were compared between non-pleiotropic and pleiotropic genes when considering all types of developmental genes (A) and only embryonic developmental genes (B). dN/dS values were also compared between non-specific and specific pleiotropic genes (C), where non-specific genes are defined as the genes within the lowest 25th percentile in stage-specific tau value from the pleiotropic group, while specific genes are the ones within the top 25th percentile. McDonald-Kreitman alpha values were compared between non-pleiotropic and pleiotropic genes when considering all types of developmental genes (D) and only embryonic developmental genes (E). Differences among groups were statistically analyzed using Kruskal-Wallis tests followed by pairwise Wilcoxon tests (p values on figure).

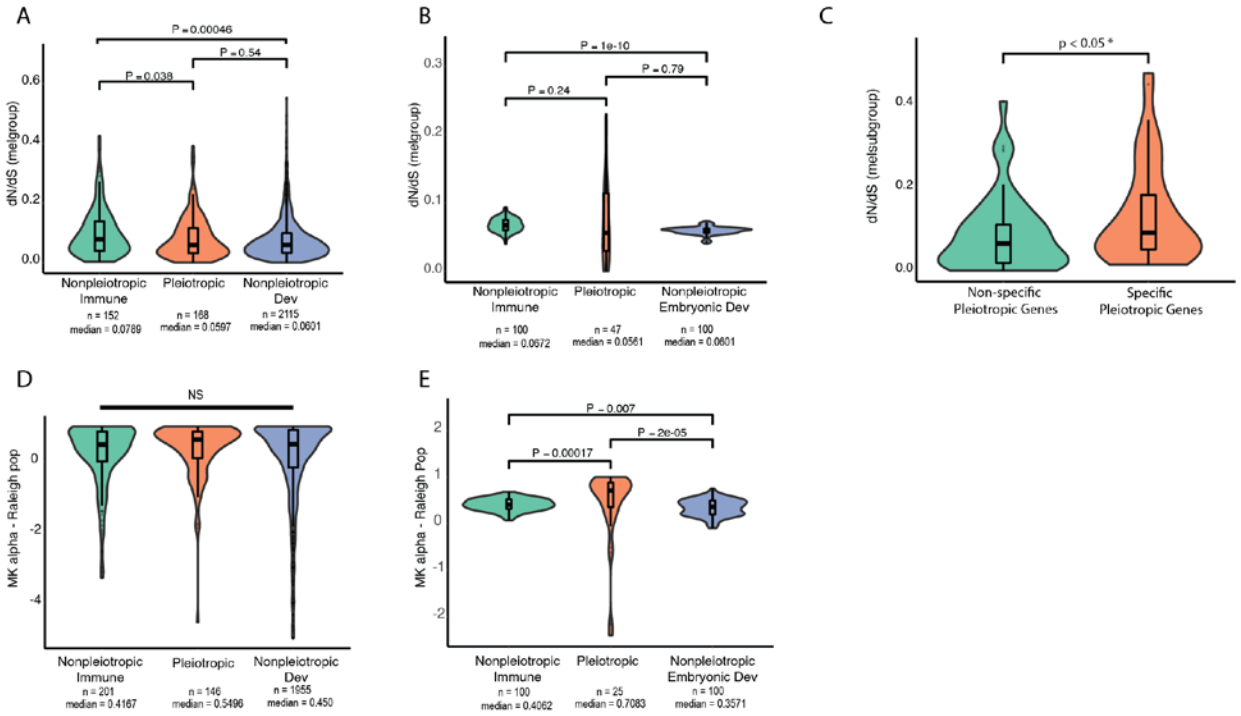
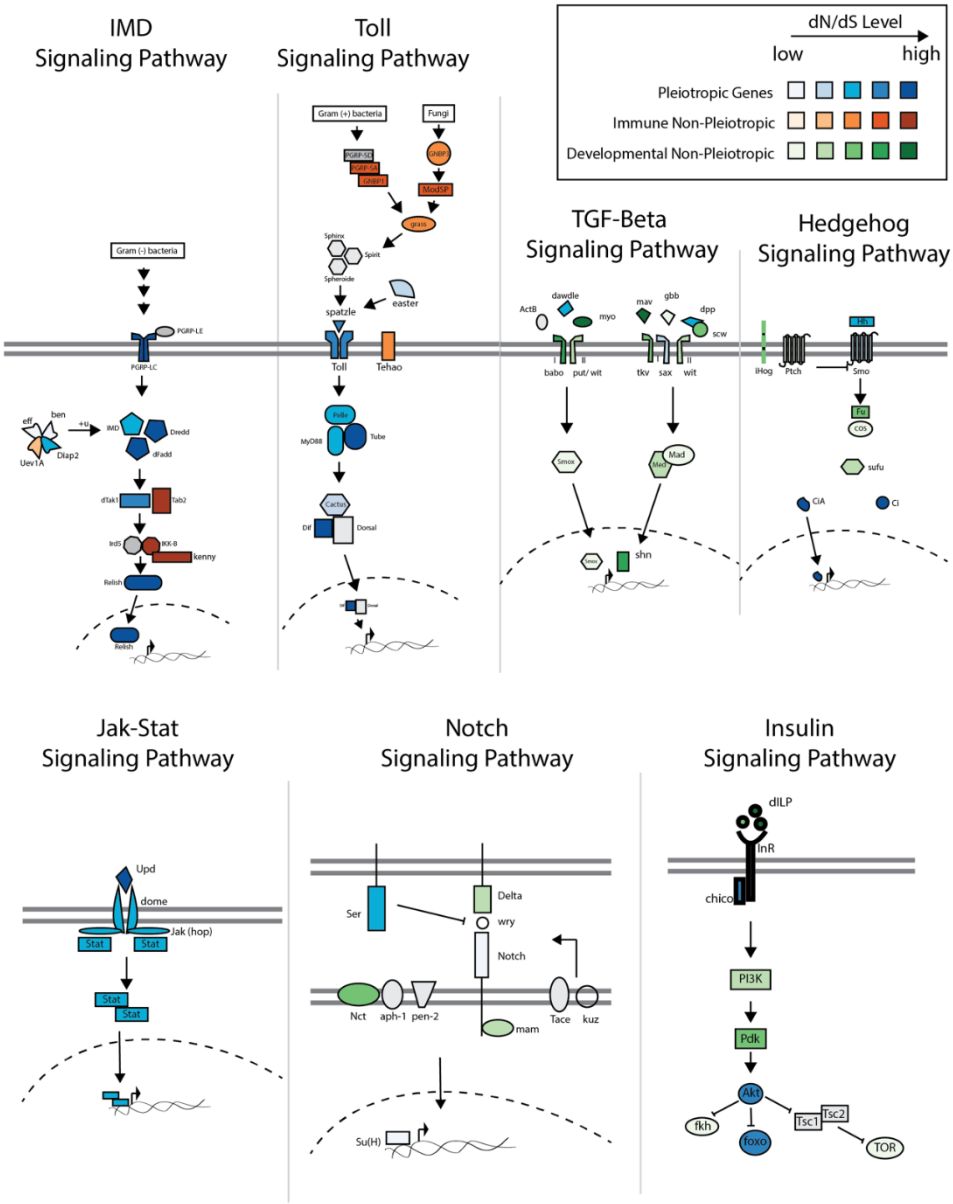
Figure 4. Examining the pleiotropy status and dN/dS levels for genes participating in major signaling pathways. The color indicates whether a gene is pleiotropic (blue), immune-exclusive (orange) or developmental-exclusive (green) as determined in methods. Each color is shaded according to dN/dS level of each gene, with the darker shade represent a higher ω value within the gene's respective pleiotropic or non-pleiotropic group. Pathway is depicted as shown from KEGG. Components for which no pleiotropy status available are shown in gray.