View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by CiteSeerX

Adaptive Evolution of a Novel Drosophila Lectin Induced by Parasitic Wasp Attack

Erin S. Keebaugh*,[1] and Todd A. Schlenke[1]

[1]Biology Department, Emory University

***Corresponding author:** E-mail: ekeebau@emory.edu.

Associate editor: Patricia Beldade

Abstract

Drosophila melanogaster has long been used as a model for the molecular genetics of innate immunity. Such work has uncovered several immune receptors that recognize bacterial and fungal pathogens by binding unique components of their cell walls and membranes. Drosophila also act as hosts to metazoan pathogens such as parasitic wasps, which can infect a majority of individuals in natural populations, but many aspects of their immune responses against these more closely related pathogens are poorly understood. Here, we present data describing the transcriptional induction and molecular evolution of a candidate Drosophila anti-wasp immunity gene, *lectin-24A*. Lectin-24A has a secretion signal sequence and its lectin domain suggests a function in sugar group binding. Transcript levels of *lectin-24A* were induced significantly stronger and faster following wasp attack than following wounding or bacterial infection, demonstrating *lectin-24A* is not a general stress response or defense response gene but is instead part of a specific response against wasps. The major site of *lectin-24A* transcript production is the fat body, the main humoral immune tissue of flies. Interestingly, *lectin-24A* is a new gene of the *D. melanogaster*/*Drosophila simulans* clade, displaying very little homology to any other Drosophila lectins. Population genetic analyses of *lectin-24A* DNA sequence data from African and North American populations of *D. melanogaster* and *D. simulans* revealed gene length polymorphisms segregating at high frequencies as well as strong evidence of repeated and recent selective sweeps. Thus, *lectin-24A* is a rapidly evolving new gene that has seemingly developed functional importance for fly resistance against infection by parasitic wasps.

Key words: molecular evolution, new gene, immunity, lectin, Drosophila, parasitic wasp.

Research article

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016

Introduction

Over the past 15 years, *Drosophila melanogaster* has served as a valuable model system for the molecular genetics of innate immunity (Lemaitre and Hoffmann 2007). *D. melanogaster* is especially useful for understanding innate immune systems of other insects, such as insect vectors of human disease, agricultural pests, and crop pollinators (Schneider and Shahabuddin 2000; Evans et al. 2006). Innate immunity can be divided into two main components, the humoral response and the cellular response. The Drosophila humoral response has been intensely studied for its role in combating bacterial and fungal infections but may also be responsible for aspects of macroparasite killing. It is governed by the fat body, which controls release of immune active extracellular proteins such as antimicrobial peptides and complement-like proteins (e.g., thioester-containing proteins) into the hemolymph (Lemaitre and Hoffmann 2007). Two major humoral immune response pathways operating in the fat body are the NF-κB pathways Toll and Imd, to which the JAK/STAT and JNK pathways appear to play complementary roles (Boutros et al. 2002). There is some evidence of Toll pathway specificity for infection by gram(+) bacteria and fungi and Imd pathway specificity for infection by gram(−) bacteria, but this distinction is not absolute and cross talk between these and other pathways appears common (Lemaitre and Hoffmann 2007).

The Drosophila cellular response is mediated by the lymph gland (the hematopoietic organ) and the hemocytes, of which there are three types. The plasmatocytes represent ~95% of the standing hemocytes, act as sentinels of infection, and are responsible for phagocytosis. The crystal cells make up the remaining 5% of the standing hemocyte population. They are responsible for generating melanin and associated free radicals, which are important in coagulation, wound healing, and pathogen killing. The lamellocytes are large flattened hemocytes responsible for encapsulating macroparasites such as parasitic wasp eggs, and their production is induced in response to infection. Lamellocytes are derived from prohemocytes in the larval lymph gland but also may develop directly from circulating plasmatocytes (Rizki 1957; Honti et al. 2010). The Toll pathway plays a major role in hematopoiesis, whereas the JAK/STAT pathway appears to be important for the development of lamellocytes (Sorrentino et al. 2004).

We have decided to focus on the molecular biology and evolution of genes potentially involved in Drosophila's cellular immune response against parasitic wasps. Several wasp species from multiple Hymenopteran families attack Drosophila larvae and pupae in nature, including generalists and numerous specialists of particular Drosophila species and species groups. The larval parasites lay single eggs in their hosts that, if allowed to hatch, begin to consume internal fly tissues. Successful infections are always lethal,

© The Author 2011. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

with the young wasps eclosing from fly pupal cases. Wasps are one of the most prevalent parasites of Drosophila in nature, infecting upwards of 50% of individuals in some natural fly populations (Carton et al. 1986; Janssen et al. 1987; Fleury et al. 2004).

Wasp eggs elicit a strong cellular encapsulation response and can be killed by resistant flies. The current model for the steps involved in encapsulation is as follows (Carton and Nappi 1997): 1) Following receptor binding to the wasp egg, circulating hemocytes contact the wasp egg and lyse, releasing signaling factors. 2) This signal causes activation of nearby hemocytes and potentiates hematopoiesis in the lymph gland, leading to the production of lamellocytes. 3) The lamellocytes migrate toward and then attach and spread around the wasp egg. 4) Finally, the inner cells of the capsule surrounding the wasp egg lyse and release reactive oxygen species and an impermeable layer of melanin, resulting in death of the parasite. Encapsulation of wasp eggs is functionally similar to vertebrate granuloma formation (McKerrow et al. 1985), although little attempt has been made to establish mechanistic homology. Although many Drosophila genetic pathways including Toll and JAK/STAT have been shown to be involved in the encapsulation response (Sorrentino et al. 2004; Zettervall et al. 2004), the genetic bases for many aspects of the encapsulation response, for example, recognition, signaling between hemocytes and the lymph gland, and the encapsulation killing mechanism, remain relatively poorly characterized.

It remains an extremely interesting question as to what kind of innate immune receptors animals might use to detect other animals. It is relatively straightforward for animal hosts to recognize bacteria and fungi as pathogens because of the distinct cell wall and cell membrane epitopes they carry, but how does a fly recognize a parasite that is much more similar to itself, such as a parasitic wasp? To date, two whole-genome gene expression studies have been conducted on wasp-attacked flies to identify novel genes involved in Drosophila's immune response against the wasps (Wertheim et al. 2005; Schlenke et al. 2007). In both of these studies, one using the Figitid wasp *Leptopilina boulardi* and one using the Braconid *Asobara tabida*, a C-type lectin named *lectin-24A* (Theopold et al. 1999) was more than 7-fold upregulated following wasp attack. *lectin-24A* was also found upregulated in larvae from multiple mutant fly strains that produce melanotic aggregates of hemocytes (Bettencourt et al. 2004; Zettervall et al. 2004; Walker et al. 2011).

Lectins are sugar-binding proteins that can distinguish very specific sugar moieties and as such have long been considered ideal candidates for specific recognition receptors in host innate immune systems. Perhaps the best-characterized immune lectin is the mannose-binding lectin of the vertebrate complement cascade (Turner 1996), although many other lectins have known roles as opsonins and attack proteins in the immune systems of vertebrates and other organisms (as reviewed in Marques and Barracco 2000; Cambi et al. 2005; Willment and Brown 2008). Thus, it was seen as a surprise that no lectins were indentified in early microarray studies of Drosophila infected with bacteria and fungi. However, two different C-type lectins were shown to aid in the Drosophila encapsulation reaction against agarose beads in vitro (Ao et al. 2007), suggesting such proteins may act specifically in the cellular immune response against macroparasites. Together with the microarray and hemocyte aggregation mutant studies, these data suggest *lectin-24A* might play an important role in melanotic capsule formation and perhaps as a pattern recognition receptor for wasp eggs.

In this study, we test whether *lectin-24A* is a general stress response, wound response, or immune response gene or whether it plays a specific role in the response to attack by parasitic wasps. Furthermore, we characterize the tissue specificity of its expression following wasp attack, to better understand its potential mechanistic role in the anti-wasp immune response. Finally, immune genes are expected to evolve rapidly and adaptively over time in order to keep pace with constantly evolving pathogen-mediated selection pressures, and Drosophila immune genes are no exception (Schlenke and Begun 2003, 2005; Jiggins and Kim 2006; Sackton et al. 2007; Lazzaro 2008). We undertake population genetic and molecular evolution analyses of the *lectin-24A* locus to determine whether it also shows a history of rapid and adaptive evolution.

Materials and Methods

Gene Expression Analysis

All aspects of the fly and wasp rearing were conducted in a 24–25 °C incubator with a 12:12 light cycle. For gene expression analyses following wasp infection, we used *D. melanogaster* strain Oregon R and the relatively virulent *L. boulardi* strain Lb17 (Schlenke et al. 2007). Flies were allowed to lay eggs for 3 h, and batches of 60 larvae from these egg lays were later moved onto 35 mm petri dishes containing standard Drosophila medium. Seventy-two hours after the egg lay period, ten experienced female wasps were placed in each of the dishes for a 2 h attack time. Two and nine hours post-attack, fly larvae were dissected or flash frozen for expression timepoint analyses. Due to the 2 h attack time and a 1 h handling time, these larvae had developed between 2–5 and 9–12 h post-attack, respectively. Note that it is possible that some fly larvae may not be attacked by wasps in the given time, however, we expect the infection rate to be greater than 90% under these conditions given past results (Schlenke et al. 2007). Ten larvae per dish were used for whole-body expression analysis, and another ten larvae were dissected for individual tissue expression analyses. For the dissected larvae, the fat body, gut, and body wall (cuticle plus associated muscle) tissues were separated and were only used if a wasp egg was found during the dissection. Dissected tissues were immediately placed into Trizol (Invitrogen), whereas whole larvae were placed into 1.5 ml tubes and frozen in liquid nitrogen for future processing. The remaining 40 larvae per dish were used for hemocyte analyses by draining larval hemolymph onto a metal rod that was immediately submerged into Trizol.

For gene expression analyses following sterile and septic injuries, the same larval rearing conditions were used. Seventy-two hours post-egg lay, 20 Oregon R larvae were each pierced with a 0.1-mm-diameter stainless steel needle (Fine Science

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016

Tools) dipped in sterile LB broth, *Enterococcus faecalis* gram(+) bacterial culture grown overnight and diluted to OD_{600} = 1.0, or *Escherichia coli* gram(−) bacterial culture grown overnight and diluted to OD_{600} = 1.0. Following injury, larvae were placed on moist Kimwipes inside a 35 mm petri dish, then later transferred to plates containing standard Drosophila medium. At 2 and 9 h post-injury, ten of the larvae were flash frozen in 1.5 ml tubes in liquid nitrogen.

Total RNA extraction for all samples was done using Trizol following the Invitrogen recommended protocol. cDNA was synthesized using the Qiagen Quantitect Reverse Transcription Kit. Each cDNA sample was used as a template for semiquantitative (comparative Ct) real-time polymerase chain reaction (PCR) using Applied Biosystems Power SYBR Green Master Mix. Each sample was run in triplicate to account for within sample variance, and any significant outliers within a sample triplicate were discarded. *alpha-Tub84B* (which was not differentially regulated following wasp attack; Schlenke et al. 2007) was used as a reference gene to control for differences in total cDNA amounts across samples. Intron spanning primers used for *alpha-Tub84B* are as follows: 5′-ACACTTCCAATAAAACTCAATATGC-3′, 5′-CCGTGCTCCAAGCAGTAGA-3′. Primers used for *lectin-24A* (which does not contain introns) are as follows: 5′-CGAGTGGGGTCCTGGTGAAC-3′, 5′-GAAACGCATCGCTCTTGGTC-3′. Primers used for *Drosomycin* and *Diptericin*, antimicrobial peptides regulated by the Toll and Imd pathways, respectively, were modified from (Ayres and Schneider 2009) as follows: *Drosomycin* 5′-GTACTTGTTCGCCCTCTTCG-3′, 5′-CTTGCACACACGACGACAG-3′ and *Diptericin* 5′-ACCGCAGTACCCACTCAATC-3′, 5′-CCCAAGTGCTGTCCATATCC-3′. Melting curves for PCR products were checked to ensure that no off-target loci were amplified by any primer pair. All expression experiments were done in four biological replicates, and untreated control larvae or larval tissues were included for each replicate (except for the gram(+) treated samples which were compared with two untreated replicates).

Relative quantification (RQ, also known as delta delta CT) data was collected to represent the fold change of each gene following treatment relative to untreated control samples. Most gene expression data are presented as log2 transformation of RQ data (log2(RQ)), except in the case of tissue-specific expression of *lectin-24A*, in which the abundance of *lectin-24A* relative to the reference gene (values known as delta CT) is used for data presentation. Statistical analysis was performed on log2 transformation of relative abundance values (log2(delta CT)) when testing if a gene is differentially regulated following treatment or differentially regulated between different tissues, and on log2(RQ) values when testing if a gene is differentially regulated following one treatment relative to another (as suggested in Rieu and Powers 2009).

Molecular Evolution

California *D. melanogaster* and *Drosophila simulans* sequence data are from sets of eight highly inbred lines made from field-caught inseminated females collected in Winters, California. African *D. melanogaster* and *D. simulans* sequence data are from sets of ten and nine isofemale lines collected in Malawi and Zimbabwe, respectively. For the subset of African *D. melanogaster* strains found to be heterozygous at *lectin-24A*, these strains were crossed to *D. melanogaster* deficiency strain 5330 (Bloomington stock center, deficiency Df(2L)ed1) to generate individuals hemizygous for *lectin-24A* for use in sequencing.

PCR primers were designed to amplify an approximately 1,900-bp region that includes the full coding sequence of *lectin-24A* plus the presumed 5′ regulatory region (bp 3,716,293–3,718,252). For the California *D. simulans* population sample, we also designed PCR primers to amplify approximately 500- to 700-bp regions flanking *lectin-24A* at various distances. PCR products were sent to Beckman Coulter Genomics for purification and Sanger sequencing, using four internal primers for *lectin-24A* itself, and the PCR primers for flanking loci. Sequences for all primers used in the sequence analyses are provided (supplementary material fig. S1, Supplementary Material online). All sequences were deposited in Genbank (# JN410844-JN410943).

Sequence data were edited using Lasergene software and population genetic and molecular evolution analyses were run in DnaSP version 5.10.01 (Librado and Rozas 2009). For the divergence and Fay and Wu's H statistics, which require an outgroup sequence, we used the genome-sequenced *D. melanogaster* strain as an outgroup for the *D. simulans* sequences, and the *D. simulans* consensus genome sequence as an outgroup for the *D. melanogaster* sequences. Significance of some population genetic statistics for various population samples and loci was calculated by comparing the observed values with those obtained from 10,000 neutral coalescence simulations. Simulated data were generated in DnaSP by using the observed number of segregating sites from each sample and under the conservative assumption of no recombination. Fly strains found to have early stop codons relative to the *D. melanogaster* genome sequence were not included in McDonald–Kreitman or d*N*/d*S* analyses for two reasons: 1) the possibility that sequence downstream and potentially upstream of the early termination codons may be under relaxed functional constraint and 2) the large deletion responsible for one early termination codon causes a large portion of the *lectin-24A* coding sequence, including part of the lectin domain, to be lost from the DnaSP analyses. Furthermore, comparisons between *D. melanogaster* and *D. simulans* coding sequences used coordinates for the consensus *D. melanogaster* open reading frame (ORF) rather than the longer *D. simulans* consensus ORF.

Results

Expression Analysis

We measured expression levels of *lectin-24A* along with two known Drosophila immune genes, *Drosomycin* and *Diptericin*, which are antimicrobial peptides commonly used to gage activation of the two immunity signaling pathways Toll and Imd, respectively. In previous studies, *Drosomycin* and/or *Diptericin* were found upregulated after wasp attack at times ranging from 12 to 48 h post-infection (Coustau

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016

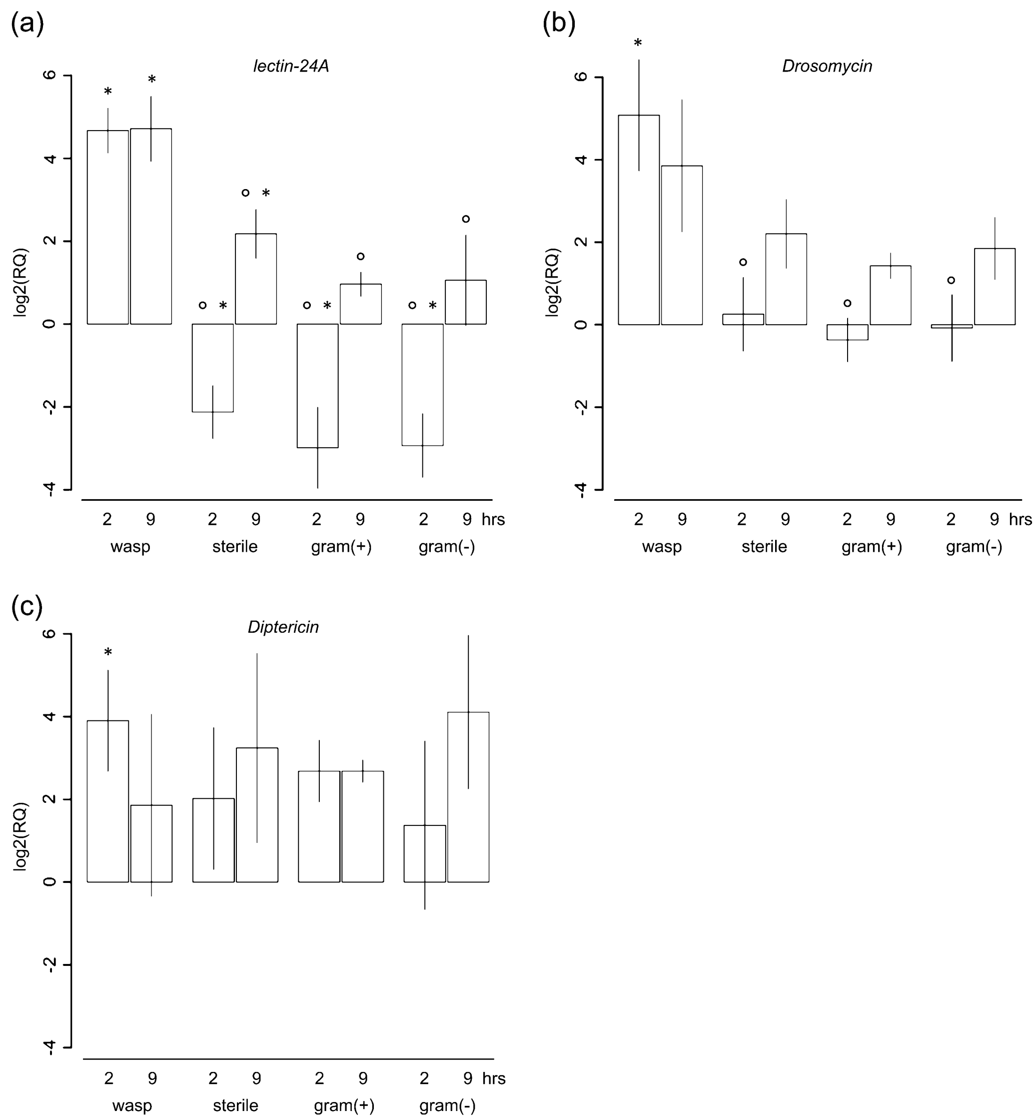

Fig. 1. Gene expression following immune challenge. log2(RQ) of (*a*) *lectin-24A*, (*b*) *Drosomycin*, and (*c*) *Diptericin* relative to untreated larvae 2–5 and 9–12 h after wasp attack, sterile injury, septic injury with the gram(+) bacteria *Enterococcus faecalis*, or septic injury with the gram(−) bacteria *Escherichia coli*. Error bars represent ± standard error of the mean. Significance values were judged by comparison of treated averages to untreated averages, $^{*}P < 0.05$. Significance values across treatments were judged by comparison of treated averages to wasp attack averages at the same timepoint, $^{\circ}P < 0.05$.

et al. 1996; Nicolas et al. 1996; Schlenke et al. 2007). We found that expression of all three genes significantly increased in whole *D. melanogaster* larvae attacked by *L. boulardi* wasps at the 2–5 h post-infection timepoint, compared with unattacked flies (fig. 1, supplementary material fig. S2*a*, Supplementary Material online). *lectin-24A* was upregulated 32-fold at this timepoint and *Drosomycin* and *Diptericin* were upregulated 81- and 38-fold, respectively, although the two antimicrobial peptide genes showed much greater variation in fold change than *lectin-24A*. Thus, wasp infection potentially activates both the Toll and the Imd pathways. At the 9–12 h post-infection timepoint, *lectin-24A* remained significantly upregulated by wasp attack, but upregulation of the two antimicrobial peptides dropped to lower nonsignificant levels.

Different regulatory trends are seen in response to piercing with a sterile needle (which presumably mimics the cuticular injury caused by wasp oviposition) or piercing with septic needles dipped in gram(+) and gram(−) bacterial cultures. *D. melanogaster* larvae significantly downregulate *lectin-24A* 3- to 5-fold at the early timepoint following sterile and septic injury with gram(+) and gram(−) bacteria. Pierced larvae then show modest nonsignificant upregulation in the 2- to 6-fold range at the later timepoint following gram(+) and gram(−) injury and significant upregulation following sterile injury, although these levels of upregulation are significantly lower than that reached by *lectin-24A* following wasp attack at the corresponding timepoint (fig. 1*a*, supplementary material fig. S2*a*, Supplementary Material online). Thus, the *lectin-24A* response to wasp infection is very different from that to sterile or septic injury.

At the early timepoint, expression patterns for *Drosomycin* and *Diptericin* following sterile injury, gram(+) injury, and gram(−) injury were noticeably different than that of

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016

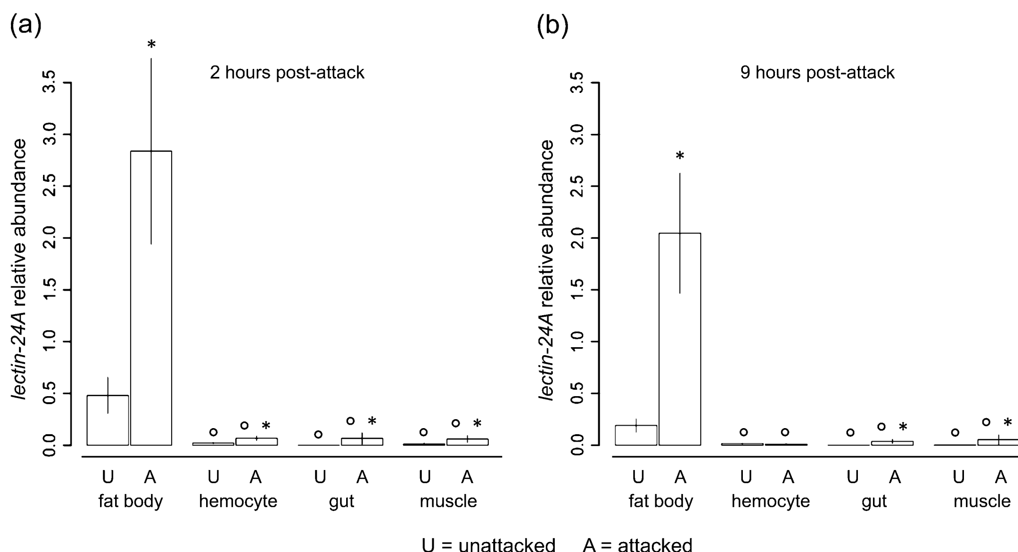

Fig. 2. Tissue-specific expression of *lectin-24A*. *lectin-24A* expression levels relative to *alphaTub84B* in fat bodies, hemocytes, guts, and body wall muscles in unattacked (U) and attacked (A) larval tissues (*a*) 2–5 h following wasp attack and (*b*) 9–12 h following wasp attack. Error bars represent ± standard error of the mean. Significance values were judged by comparison of wasp-attacked tissue averages to unattacked tissue averages, $^*P < 0.05$. Significance values across treatments were judged by comparison of *lectin-24A* abundance in fat body to *lectin-24A* abundance in other tissues at the same timepoint and under the same condition, $^{\circ}P < 0.05$.

lectin-24A, either showing no change in expression level (*Drosomycin*) or nonsignificant upregulation (*Diptericin*) (fig. 1*b* and *c*). Expression of *Drosomycin* and *Diptericin* at the later timepoint following sterile and septic injuries showed a trend of nonsignificant upregulation similar to that of *lectin-24A* following sterile and septic injuries. No significant differences in expression were observed between sterile injury, gram(+) bacterial infection, or gram(−) bacterial infection for any of the three genes, suggesting the fly larvae do not distinguish between the three treatments at these timepoints (supplementary material fig. S2*a*, Supplementary Material online). Altogether, these data show that *lectin-24A* is regulated in a different manner than genes that are known targets of the Toll and Imd pathways.

We next investigated tissue specificity of *lectin-24A* expression following wasp attack in two tissues important for hemolymph immunity (fat body, hemocytes) and two control tissues (gut, body wall). The constitutive expression level of *lectin-24A* was significantly greater in the fat body than the other three tissues at both timepoints (fig. 2, supplementary material fig. S2*a*, Supplementary Material online). Furthermore, *lectin-24A* expression was significantly upregulated approximately 9- and 16-fold in the fat body following wasp attack at 2–5 h and 9–12 h post-attack. Expression of *lectin-24A* in the hemocytes, gut, and body wall also significantly increased following wasp attack, excluding the 9–12 h timepoint in hemocytes (fig. 2, supplementary material fig. S2*b*, Supplementary Material online), but the overall levels of *lectin-24A* transcript (standardized by *alphaTub84B*) in these tissues still averaged approximately 40 times less than *lectin-24A* levels found in the fat body. These data indicate that the fat body, the most important humoral immunity organ, is the major site of both constitutive and wasp attack-induced *lectin-24A* production.

Species Range and Gene Structure

The coding region of *D. melanogaster lectin-24A* is 846 bp (282 aa) long, with the lectin domain located at amino acids 169–280. The gene has no other characterized domains and also contains no introns, similar to other Drosophila C-type lectins. We used basic local alignment search tool (BLAST) (specifically, tblastx) to search for orthologs of the *D. melanogaster lectin-24A* sequence in the nucleotide collection of Genbank. *lectin-24A* was present in only *D. melanogaster* and its *D. simulans* sister group (including *D. simulans* and *Drosophila sechellia*). Because the third member of the simulans group, *Drosophila mauritiana*, has not been genome sequenced, we tested and confirmed by PCR and sequencing that *D. mauritiana* also has a *lectin-24A* ortholog (supplementary material fig. S3*a*, Supplementary Material online). However, no *lectin-24A* ortholog was found in other genome-sequenced members of the melanogaster group (*Drosophila yakuba*, *Drosophila erecta*, *Drosophila ananassae*, supplementary material fig. S3*b*, Supplementary Material online), in any of the five other genome-sequenced Drosophila species, or in any other organism. BLAST also fails to identify close homologs to *lectin-24A* in the *D. melanogaster* genome. Although both the non-lectin and the lectin domains of *lectin-24A* BLAST to other *D. melanogaster* lectins (e.g., *lectin-24Db* and *lectin-28C*, respectively), the sequence homology in both cases is quite poor (supplementary material fig. S3*c* and *d*, Supplementary Material online).

In *D. melanogaster*, the gene *CG2818* is immediately upstream of *lectin-24A*, and the gene *Shaw* is immediately downstream, with *lectin-24A* in reverse orientation relative to the flanking genes. There is very little intergenic sequence between the transcript sequences of these three genes, as the 3′ transcript end of *lectin-24A* overlaps the

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016

Table 1. Population Genetic Statistics for *lectin-24A* and Flanking Loci.[a]

	S^b	π^b	Theta-W[b]	Tajima's D	Tajima's D P Value[c]	Fay and Wu's H	Fay and Wu's H P Value[c]	Hd	Hd P-value[c]	ZnS	ZnS P-value[c]	Divergence
25 kb upstream	14	0.0073	0.0077	−0.3093	0.3971	−3.0000	0.1101	0.7500	0.1235	0.4139	0.3818	0.0484
15 kb upstream	17	0.0117	0.0091	1.4729	0.9551	1.4286	0.5214	0.7860	0.0925	0.4641	0.2801	0.0482
5 kb upstream	21	0.0105	0.0105	0.0051	0.5250	−2.0714	0.2101	0.7860	0.0597	0.3279	0.5999	0.0340
2 kb upstream	5	0.0026	0.0040	−1.5952	0.0930	−4.0714	0.0184	0.2500	0.0201*	1.0000	0.0000*	0.0388
lectin-24A cal sim	56	0.0070	0.0107	−1.8973	0.0012*	−47.7857	0.0000*	0.2500	0.0000*	1.0000	0.0000*	0.0859
2 kb downstream	4	0.0020	0.0030	−1.5347	0.1188	−1.0714	0.1395	0.2500	0.0302*	1.0000	0.0000*	0.0594
5 kb downstream	31	0.0175	0.0187	−0.3451	0.3752	−2.2143	0.2153	0.7500	0.0268*	0.3388	0.5784	0.0361
15 kb downstream	17	0.0097	0.0084	0.8266	0.8302	2.2143	0.7645	0.7500	0.0895	0.4039	0.4027	0.0429
25 kb downstream	30	0.0160	0.0171	−0.3431	0.3872	−3.4286	0.1789	0.7500	0.0299*	0.3142	0.6612	0.0680
lectin-24A cal mel	30	0.0048	0.0059	−0.9812	0.1895	−2.5714	0.2128	0.8929	0.1541	0.2998	0.7065	0.0770
lectin-24A afr mel	28	0.0070	0.0058	0.9308	0.8580	−1.2444	0.2574	0.8222	0.0690	0.5054	0.1519	0.0755
lectin-24A afr sim	101	0.0217	0.0185	0.8807	0.8480	5.1944	0.4664*	0.9722	0.1779	0.2910	0.6853	0.0792

[a] Flanking loci sequenced from the California *Drosophila simulans* (cal sim) population.
[b] Three measures of heterozygosity are presented: *S* is the number of segregating sites in the sample, π is the average number of pairwise difference between strains per bpbase pair, and theta-W is Watterson's theta (Watterson 1975).
[c] *P* values determined from coalescent simulations, $^*P < 0.05$.

3′ transcript end of *CG2818* by 11 bp and the 5′ transcript start of *lectin-24A* is only 414 bp away from the 5′ transcript start of *Shaw*. Orthologs of *CG2818* and *Shaw* are found physically adjacent to one another but with little intervening sequence, across the melanogaster group of the genus Drosophila (supplementary material fig. S3*b*, Supplementary Material online), suggesting that *lectin-24A* arose from an insertion in the common ancestor of *D. melanogaster* and *D. simulans*.

We sequenced *lectin-24A* in California population samples of *D. melanogaster* and *D. simulans* and from more ancestral population samples from Africa. In these *D. melanogaster* strains, the consensus ORF length is 282 aa (as in the genome-sequenced strain), but in *D. simulans* the consensus ORF length is 291 aa (as in the genome sequences of *D. simulans* and *D. sechellia*). This is due to a difference in the position of the stop codon between these two species caused by an insertion in *D. melanogaster* relative to *D. simulans* at the 3′ end of the coding sequence.

Interestingly, ORF length variation also exists within the African population samples of both *D. melanogaster* and *D. simulans* and in the single *D. mauritiana* allele we sequenced, due to multiple independent mutations (supplementary material figs. S3*a* and S4*a*, Supplementary Material online). Six of ten *D. melanogaster* strains from Malawi had one of two different premature stop codons that fall within the lectin domain, resulting in truncation of *lectin-24A* and of the lectin domain itself. The first of these early stop codon variants, found in two strains, was generated by a point mutation resulting in a 29 aa truncation of the 3′ end of *lectin-24A* and the loss of 27 of the 112 amino acids from the lectin domain. The second early stop codon variant, found in four strains, was generated by an out-of-frame 169 bp deletion within the lectin domain, in combination with a short insertion, that formed a new stop codon that results in a 66 aa truncation of the 3′ end of *lectin-24A* and the loss of 64 of the 112 amino acids from the lectin domain. There appears to be an excess of shared nonsynonymous mutations upstream of the stop codons in the two *D. melanogaster* premature stop codon variants (supplementary material fig. S4*b*, Supplementary Material online), suggesting that the premature stop codons were independently selected for in this divergent haplotype background. Also, one of nine *D. simulans* strains from Zimbabwe had a premature stop codon located upstream of the lectin domain, resulting in a severe truncation of *lectin-24A* (supplementary material fig. S4*a* and *c*, Supplementary Material online). This early stop codon resulted from a 1 bp deletion and shortens the ORF to 75 aa. Finally, the *D. mauritiana* strain we sequenced had a premature stop codon compared with the consensus lengths of other species, truncating the ORF to 103 aa (supplementary material figs. S3*a* and S4*a*, Supplementary Material online).

Polymorphism Analysis

We tested for unusual haplotype structure at the *lectin-24A* locus of the four population samples by comparing observed haplotype diversity (Hd) (Nei 1987) with a distribution of Hd

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016

site		consensus	cal sim 1	cal sim 2	cal sim 3	cal sim 4	cal sim 5	cal sim 6	cal sim 7	cal sim 8
13	I	A	T	·	·	·	·	·	·	·
161	I	T	A	·	·	·	·	·	·	·
189	I	C	T	·	·	·	·	·	·	·
278	I	T	C	·	·	·	·	·	·	·
279	I	G	A	·	·	·	·	·	·	·
284	I	T	C	·	·	·	·	·	·	·
291	I	C	A	·	·	·	·	·	·	·
323	I	C	T	·	·	·	·	·	·	·
338	I	T	A	·	·	·	·	·	·	·
344	I	T	A	·	·	·	·	·	·	·
373-379	I	i7	d7	·	·	·	·	·	·	·
411	I	A	G	·	·	·	·	·	·	·
412	I	G	A	·	·	·	·	·	·	·
426	I	T	C	·	·	·	·	·	·	·
434	I	T	G	·	·	·	·	·	·	·
486	I	A	G	·	·	·	·	·	·	·
509	I	A	G	·	·	·	·	·	·	·
544	I	G	A	·	·	·	·	·	·	·
597-602	I	d6	i6	·	·	·	·	·	·	·
629	S	A	G	·	·	·	·	·	·	·
634	N	T	C	·	·	·	·	·	·	·
644	S	A	G	·	·	·	·	·	·	·
688	N	G	A	·	·	·	·	·	·	·
689	N	A	C	·	·	·	·	·	·	·
697	N	A	G	·	·	·	·	·	·	·
701	N	T	G	·	·	·	·	·	·	·
825	N	G	C	·	·	·	·	·	·	·
845	S	A	G	·	·	·	·	·	·	·
878	S	T	C	·	·	·	·	·	·	·
902	S	A	G	·	·	·	·	·	·	·
912	N	A	G	·	·	·	·	·	·	·
956	S	A	G	·	·	·	·	·	·	·
989	S	A	G	·	·	·	·	·	·	·
991	N	C	A	·	·	·	·	·	·	·
1028	S	A	G	·	·	·	·	·	·	·
1030	N	A	G	·	·	·	·	·	·	·
1061	S	C	T	·	·	·	·	·	·	·
1113	N	A	C	·	·	·	·	·	·	·
1136	S	A	T	·	·	·	·	·	·	·
1151	S	T	C	·	·	·	·	·	·	·
1223	S	G	A	·	·	·	·	·	·	·
1241	S	G	A	·	·	·	·	·	·	·
1262	S	A	C	·	·	·	·	·	·	·
1268	S	G	A	·	·	·	·	·	·	·
1430	N	C	A	·	·	·	·	·	·	·
1472	N	A	T	·	·	·	·	·	·	·
1499	S	T	C	·	·	·	·	·	·	·
1515	I	i1	d1	·	·	·	·	·	·	·
1521	I	i1	d1	·	·	·	·	·	·	·
1536	I	C	T	·	·	·	·	·	·	·
1544	I	G	A	·	·	·	·	·	·	·
1634	I	i1	d1	·	·	·	·	·	·	·
1672	I	G	T	·	·	·	·	·	·	·
1707	I	T	G	·	·	·	·	·	·	·
1780	I	T	C	·	·	·	·	·	·	·
1852	I	A	C	·	·	·	·	·	·	·
2012	I	C	T	·	·	·	·	·	·	·
2033	I	i1	d1	·	·	·	·	·	·	·
2034	I	T	C	·	·	·	·	·	·	·
2035	I	i1	d1	·	·	·	·	·	·	·
2037	I	T	A	·	·	·	·	·	·	·
2038	I	A	T	·	·	·	·	·	·	·
2040	I	T	C	·	·	·	·	·	·	·

Fig. 3. *lectin-24A* polymorphism table for the California *Drosophila simulans* population. Site number represents the position of a polymorphism. *N*, *S*, and *I* represent nonsynonymous substitutions, synonymous substitutions, or intergenic regions, respectively. Strains matching the consensus sequence at a polymorphic site contain a dot (·). i and d represent insertion and deletion, respectively, followed by the number of base pairs affected. Indel polymorphisms are displayed as one polymorphic site with the length and placement of the indel noted by the site range.

values generated by neutral coalescence simulation. Unlike the other samples, the California *D. simulans* population sample showed significantly low Hd, yielding only two haplotypes from the eight strains sequenced (table 1). One distinct haplotype was found in one of eight strains (cal sim 1), whereas the other haplotype was found in seven of eight strains (fig. 3). The cal sim 1 haplotype is very similar to those of some African *D. simulans* strains, whereas the other California alleles have a divergent haplotype that is quite distinct from any African strain (supplementary material fig. S4c, Supplementary Material online).

Low Hd at a locus can be explained by various demographic forces operating on a population or by the selective sweep of a beneficial allele. Demographic forces, however, are expected to affect the whole genome, whereas selection is usually locus specific. We compared Hd of *lectin-24A* in the California *D. simulans* population sample with the Hd values of 68 other genes located across the genome from the same eight California *D. simulans* strains (fig. 4). Immunity and non-immunity genes are indicated separately as it was previously found that immune genes have significantly lower Hd than non-immune genes (Schlenke and Begun 2003). We found that the Hd of *lectin-24A* is lower than 67 of the other 68 genes analyzed (second percentile) and that the only gene with similarly low Hd is the immune gene *Hemomucin*. Thus, low Hd observed at *lectin-24A* in the California *D. simulans* population is likely the result of a selective sweep.

Selection skews haplotype structure at a target locus but also at loci linked to the selected locus. Thus, determining the physical span of reduced Hd to the flanks of *lectin-24A* in the California *D. simulans* population sample can help to narrow the list of genes that were potentially selection targets. We sequenced genomic regions flanking *lectin-24A* by approximately 2, 5, 15 and 25 kb upstream and downstream and calculated Hd at those loci (fig. 5 and table 1). The region of reduced Hd appears centered on *lectin-24A* and is approximately 10 kb long, as Hd increases to approximately normal values further to either side. This 10-kb region contains two full and two partial genes other than *lectin-24A* (*cutlet*, *CG31955*, *CG2818*, and *Shaw*).

Three other partially independent population genetic descriptors also show a pattern of non-neutral polymorphism structure centered on the *lectin-24A* locus (table 1). Tajima's D, a measure of the allele frequency distribution (Tajima 1989), was significantly low at *lectin-24A* but not at flanking loci, indicating an excess of rare alleles. Fay and Wu's H, a measure of the frequency distribution of derived alleles (Fay and Wu 2000), was significantly low at *lectin-24A* and one flanking locus, indicating an excess of high frequency derived polymorphisms. ZnS is a measure of linkage disequilibrium, the degree to which alleles at different sites co-occur on haplotypes (Kelly 1997). ZnS was significantly high at *lectin-24A* and the two immediately flanking loci. Although the larger number of segregating sites at the *lectin-24A* locus disproportionately increases the power of significance tests at this locus, the absolute value of each statistic for the *lectin-24A* locus is greater than or equal to

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016

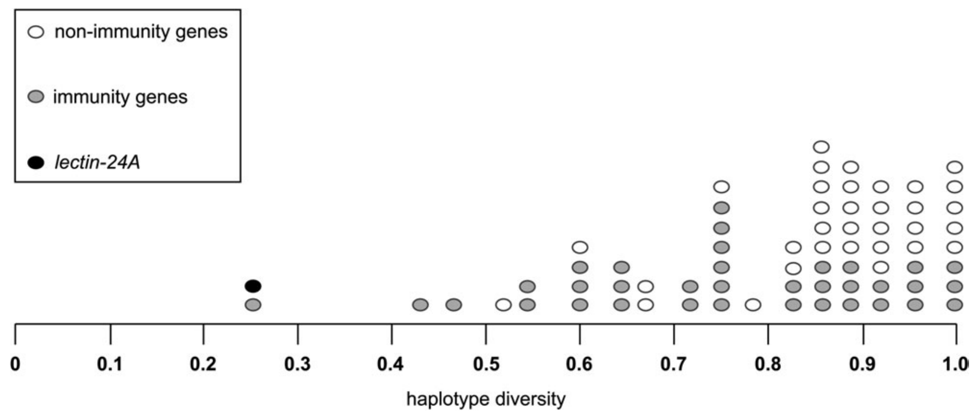

FIG. 4. Hd of *lectin-24A* and 68 other genes from the California *Drosophila simulans* population (Schlenke and Begun 2003).

the same value from every flanking locus. The direction of each of these skews is consistent with the effects of a strong recent selective sweep at the *lectin-24A* locus.

Divergence Analysis

We compared the ratio of nonsynonymous to synonymous substitutions per site (dN/dS) at *lectin-24A* to other genes in the genome using the *D. melanogaster* and *D. simulans* genome sequences. Because *lectin-24A* is only found in the *D. melanogaster* and *D. simulans* lineages, no outgroup sequence is available to polarize substitutions to one or the other of the *D. melanogaster* and *D. simulans* lineages. The dN/dS value of 0.878 is significantly high (98th percentile) compared with the distribution of dN/dS values from every other shared gene in this species pair, which averages at 0.151 (Begun et al. 2007) (fig. 6). Similar analysis on *lectin-24A* using our own *D. melanogaster* and *D. simulans* strains (excluding the early termination codon strains) yields a dN/dS of 0.806 over the full coding region, 0.691 for the non-lectin domain region specifically, and 1.018 for the lectin domain. High dN/dS values can be caused by the recurrent fixation of beneficial nonsynonymous mutations by selection but may also indicate relaxed functional constraint at a locus if dN/dS is less than or equal to 1.0.

One method for distinguishing adaptive evolution from relaxed functional constraint is the McDonald–Kreitman test, which compares the ratio of nonsynonymous to synonymous differences between species to that same ratio within species (McDonald and Kreitman 1991). For genes evolving neutrally under varying degrees of functional constraint, these ratios are expected to be equal. For a gene evolving adaptively, however, beneficial nonsynonymous mutations are expected to sweep to fixation very fast, contributing little to nonsynonymous polymorphism but accumulating as nonsynonymous substitutions. We performed multiple McDonald–Kreitman tests using different combinations of our *D. melanogaster* and *D. simulans* population samples (table 2). The *D. simulans* population samples consistently yielded highly significant results in the direction of excess nonsynonymous substitutions, whereas the *D. melanogaster* population samples trended in the same direction but did not reach statistical significance. For the analysis that includes polymorphism from all population samples, if we assume that it is only the nonsynonymous fixations causing the deviation from our expectation of equal nonsynonymous to synonymous ratios (Smith and Eyre-Walker 2002), we can infer that approximately 33 of the 51 nonsynonymous differences between

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016

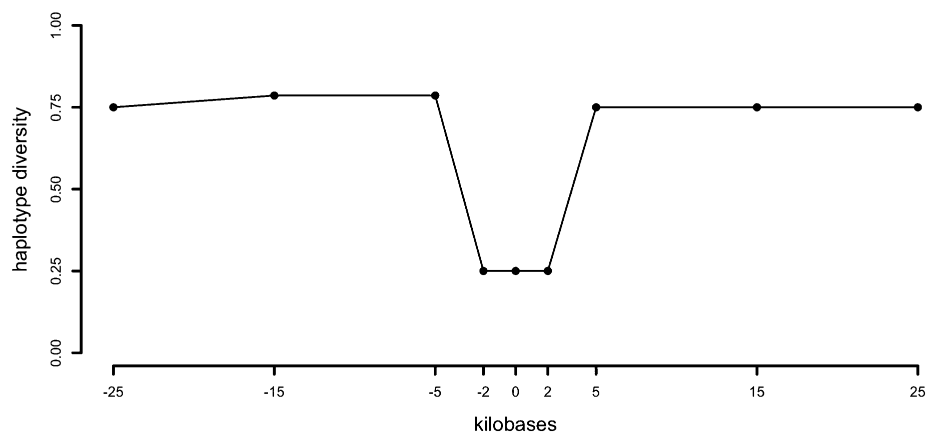

FIG. 5. Hd of *lectin-24A* and flanking loci from the California *Drosophila simulans* population. *lectin-24A* is located at 0 kb, negative values represent regions upstream of *lectin-24A*, and positive values represent regions downstream of *lectin-24A*.

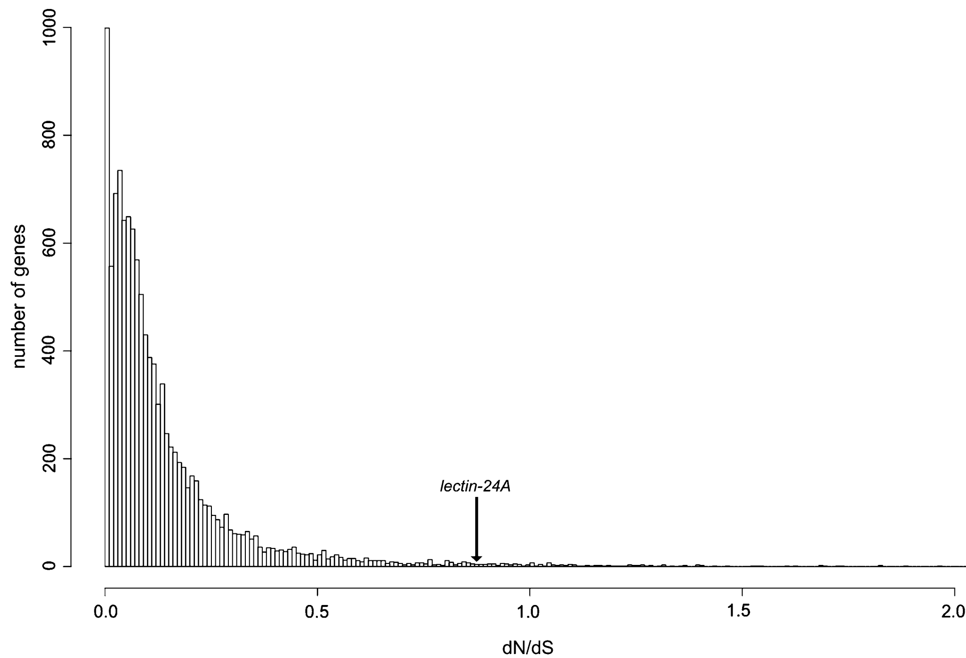

Fig. 6. dN/dS for every gene in the genome shared between *Drosophila melanogaster* and *Drosophila simulans* (Begun et al. 2007). An arrow marks the dN/dS value for *lectin-24A*. The 23 genes with dN/dS > 2 were excluded from the graph.

D. melanogaster and *D. simulans lectin-24A* sequences were fixed as the result of positive selection rather than genetic drift. The nonsynonymous fixations are distributed relatively equally between the non-lectin and lectin domains of *lectin-24A* (table 2).

Discussion

Unlike plant and vertebrate systems, most studies on Drosophila immunity have utilized pathogens that are not known to infect Drosophila in nature. While artificial infection of Drosophila with non-natural pathogens has been a powerful tool for uncovering basic aspects of the immune system, it is possible that essential parts of the immune system have been overlooked because they mediate specific responses against infection strategies of specialist parasites. In this paper, we have focused on a candidate Drosophila immune gene with potential specificity for infections by parasitic wasps, which are one of the most important groups of specialist Drosophila pathogens in nature.

Expression of *lectin-24A* was previously shown to significantly increase in Drosophila larvae after attack by parasitic wasps from two different families (Wertheim et al. 2005; Schlenke et al. 2007). Our first goal was to confirm *lectin-24A* induction following attack by the parasitic wasp *L. boulardi* using qRT-PCR. We indeed found a 32- to 42-fold increase in whole larvae *lectin-24A* transcript levels at both timepoints post-infection. The two antimicrobial peptide genes, *Drosomycin* (often used to measure the activation of the Toll pathway) and *Diptericin* (often used to

Table 2. McDonald–Kreitman Analyses for *lectin-24A*.[a]

	Synonymous	Nonsynonymous	P Value[c]
Fixations[b]	14	51	—
All mel and all sim polymorphisms	30	39	0.0097*
All mel polymorphisms	4	10	0.7258
California mel polymorphisms	3	4	0.3443
Africa mel polymorphisms	2	7	1.0000
All sim polymorphisms	28	29	0.0021*
California sim polymorphisms	15	12	0.0026*
Africa sim polymorphisms	24	24	0.0023*
Fixations, non-lectin domain[d]	9	29	—
All mel and all sim polymorphisms	19	24	0.0639
Fixations, lectin domain[d]	5	22	—
All mel and all sim polymorphisms	11	15	0.0772

[a] Only full ORF strains included in analyses.
[b] The number of fixed differences between the *Drosophila melanogaster* (mel) and *Drosophila simulans* (sim) population samples are compared with the number of polymorphisms from a variety of populations and species.
[c] Significance determined by two-tailed Fisher's exact test, *$P < 0.05$.
[d] The number of fixed differences between the *D. melanogaster* (mel) and *D. simulans* (sim) population samples in portions of *lectin-24A* are compared with the number of polymorphisms from the same portion.

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016

measure the activation of the Imd pathway) (Lemaitre and Hoffmann 2007), were also upregulated following wasp attack, although their expression levels began declining at the later timepoint. These data suggest wasp infection induces a general immune response shortly after infection that potentially includes activation of both the Toll and the Imd pathways.

It was previously shown that sterile and septic injuries of adult flies result in induction of both *Drosomycin* and *Diptericin* at early timepoints following treatment, regardless of the bacterial type used (Lemaitre et al. 1997). The response to septic injury begins to show specificity at later timepoints past 6–12 h post-treatment, that is, *Drosomycin* stays induced following gram(+) bacterial infection and *Diptericin* stays induced following gram(−) bacterial infection. A study using fly larvae also found a common induction of antimicrobial peptides at early timepoints following either sterile or septic injury (Bettencourt et al. 2004). Similarly, we found little difference in the upregulation of *Drosomycin* or *Diptericin* across sterile and septic injury treatments in fly larvae in our relatively early timepoint trials (fig. 1). Both genes were upregulated following injury, but *Drosomycin* was not upregulated until the 9 h timepoint, and both genes showed a large amount of variance in upregulation across replicates that caused nonsignificant results. In contrast, *lectin-24A* was significantly downregulated by sterile and septic injuries at the early timepoint, before being modestly upregulated 2- to 6-fold at the later timepoint, indicating *lectin-24A* is part of a different immune regulatory network than *Drosomycin* and *Diptericin*.

We found that *lectin-24A* transcript was made at significantly higher abundance in the fat body, the main humoral immunity secretory organ, than in other tissues. The Toll, Imd, JAK/STAT, and JNK pathways are known to influence fat body production of immune proteins (Boutros et al. 2002; Delaney et al. 2006; Lemaitre and Hoffmann 2007) and thus would seem to be good candidates for inducing *lectin-24A* expression. Given that *Drosomycin* and *Diptericin* expression levels can be used to measure the relative activation of the Toll and Imd pathways, respectively (Lemaitre and Hoffmann 2007), and that they show expression patterns different from *lectin-24A* following injury, we find it unlikely that Toll or Imd are the primary pathways responsible for *lectin-24A* induction. Interestingly, however, both *Drosomycin* and *Diptericin* were significantly upregulated following wasp attack. These two genes may be responding to the cuticle injuries made by wasp ovipositors, but their expression may also be enhanced by a wasp infection-specific activation of JAK/STAT, JNK, or other pathways that undergo cross talk with Toll and Imd (e.g., Zettervall et al. 2004).

Altogether, our *lectin-24A* expression analyses are consistent with numerous other transcriptomic and proteomic studies using assorted Drosophila life stages, tissues, and pathogens for infection. For example, *lectin-24A* was not found upregulated in microarray studies on adult *D. melanogaster* infected with bacterial, fungal, viral, and microsporidian pathogens (De Gregorio et al. 2001; Irving et al. 2001; Roxstrom-Lindquist et al. 2004; Dostert et al. 2005; Carpenter et al. 2009), in larvae infected with bacteria (Vodovar et al. 2005), or in Drosophila hemocyte-like S2 and mbn2 cells treated with lipopolysaccharide or bacteria (Boutros et al. 2002; Johansson et al. 2005). Nor were Lectin-24A protein levels increased in larval or adult flies infected with bacteria, fungi, or lipopolysaccharide (Levy et al. 2004; Vierstraete, Verleyen, Baggerman, et al. 2004; Vierstraete, Verleyen, Sas, et al. 2004) or in mbn2 cells treated with lipopolysaccharide (Loseva and Engstrom 2004). Thus, *lectin-24A* shows a distinct wasp attack-specific expression pattern and cannot be categorized as a general stress response, wound response, or immune response gene.

It is inferred that the Lectin-24A protein is secreted because it carries a secretion signal sequence. Given the ability of lectins to recognize specific cell surface sugar moieties, it is particularly interesting to consider whether Lectin-24A might act as the initial immune recognition protein for wasp eggs. Induction of *lectin-24A* in the fat body 2 and 9 h post-infection does not immediately suggest a primary recognition role, as some recognition of attack must have occurred in the hemocoel prior to the induction of *lectin-24A* expression. However, it is possible that constitutively produced Lectin-24A may be responsible for recognizing wasp eggs and initiating a response that includes a positive feedback loop of self-induction, for example, if more Lectin-24 protein aids in opsonizing the entire wasp egg surface. Furthermore, it is possible that flies might recognize and respond to some other aspect of the wasp attack, such as the wound caused by the wasp ovipositor or the wasp venom and its effects, before expressing molecules that can recognize wasp eggs. Alternatively, because *lectin-24A* expression is induced in response to two different wasps from different Hymenopteran families (Wertheim et al. 2005; Schlenke et al. 2007) and is also upregulated in mutant Drosophila strains that constitutively produce melanotic aggregates of hemocytes (Bettencourt et al. 2004; Zettervall et al. 2004; Walker et al. 2011), Lectin-24A may instead be a general melanotic encapsulation response gene, for example, acting to facilitate the hemocyte–hemocyte interactions necessary for capsule formation. Further study of Lectin-24A's molecular function will be required to tease apart any role Lectin-24A plays in the anti-wasp immune response, be it in recognition or some other function.

We cannot rule out the possibility that genes we find upregulated after wasp attack, including *lectin-24A*, are beneficial to the wasps and may even be purposefully induced by the wasps themselves. It has long been known that parasitic wasp venoms can manipulate many aspects of their hosts' physiology (Vinson and Iwantsch 1980), and the wasp strains used in previous microarray studies and in this study are highly successful at evading and/or suppressing the immune response of *D. melanogaster* (Rizki and Rizki 1990; Eslin et al. 1996; Labrosse et al. 2003). Evidence in support of this hypothesis are the number of naturally segregating early termination codons in *lectin-24A* that might deprive the wasps of whatever potential benefit they receive from the full-length protein, as well as the fact that a fly strain artificially selected for resistance against the

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016

wasp *A. tabida* had significantly reduced constitutive *lectin-24A* expression compared with a control unselected strain (Wertheim et al. 2011). However, we find it unlikely that wasps benefit from *lectin-24A* induction for the following three reasons: 1) given most Drosophila species do not require *lectin-24A*, it seems unlikely that the majority of *D. melanogaster* and *D. simulans* strains would continue to carry a gene that benefits one of their most common types of pathogens; 2) it seems unlikely that two wasps from different families (*A. tabida* and *L. boulardi*) could have evolved the same *lectin-24A* induction strategy, especially given that *Leptopilina heterotoma* (a close relative to *L. boulardi*) does not cause *lectin-24A* induction in infected hosts; and 3) given that *A. tabida* and *L. boulardi* have European and worldwide ranges, respectively, it is surprising that early termination codons are only segregating in African fly populations. Thus, we continue to favor the hypothesis that Lectin-24A is an anti-wasp immune protein.

Surprisingly, no obvious homolog of *lectin-24* was found outside of the *D. melanogaster* and *D. simulans* sister clade, despite the fact that the genes immediately flanking *lectin-24A* upstream and downstream are present in tandem across the melanogaster group of the genus Drosophila. In previous work, *lectin-24A* was predicted to have originated via DNA-based duplication and not by an RNA-based insertion, because there is no evidence of a poly(A) tail or direct repeats flanking *lectin-24A* (Chen et al. 2010). De novo evolution from standing DNA sequence is also an unlikely explanation because the DNA sequence that became *lectin-24A* seems to have been an insertion unique to the genome of the common ancestor of the *D. melanogaster* and *D. simulans* lineages.

It was suggested that the parental gene of *lectin-24A* was either *lectin-28C* (Zhou et al. 2008) or *lectin-24Db* (Chen et al. 2010), the two *D. melanogaster* lectins that produced the best BLAST hits to *lectin-24A's* lectin domain and nonlectin domain, respectively. However, because full-length *lectin-24A* does not BLAST with high confidence to any specific lectin in the *D. melanogaster* genome, it must have evolved very rapidly from its parental sequence(s). Furthermore, none of the 40 other *D. melanogaster* C-type lectin domain-containing genes (as annotated in FlyBase), nor any gene immediately flanking *lectin-24A*, were as strongly or consistently upregulated following *L. boulardi* attack or as strongly or consistently downregulated following attack by the highly immune suppressive wasp *L. heterotoma* (supplementary material fig. S5, Supplementary Material online) (Schlenke et al. 2007), suggesting *lectin-24A* regulatory elements have also rapidly evolved. Rapid evolution of newly duplicated genes is expected, as gene redundancy results in relaxed selection on the new gene and the potential for accumulation of otherwise deleterious nonsynonymous mutations (as reviewed in Long et al. 2003). Such alterations can cause pseudogenization, subfunctionalization (when a new gene specializes on a subset of the functions of its parental gene), or neofunctionalization (when a new gene develops a novel function) of a young gene. Given *lectin-24A*'s apparently unique role in melanotic encapsulation, it appears that neofunctionalization is contributing to the adaptive evolution of *lectin-24A*.

Some of the naturally segregating premature termination codon *lectin-24A* haplotypes may represent a more advanced state of neofunctionalization or possibly pseudogenization. It is highly unlikely that the early termination mutations are deleterious alleles because of the relatively high frequency of haplotypes that have them, the fact that four unique mutations in three species contribute to this pool and the fact that all such mutations are geographically localized to the African region (*D. mauritiana* is endemic to the Mauritius Islands). It is more likely that the truncated proteins perform some beneficial function or that a null allele of *lectin-24A* is harmless or even beneficial under certain conditions in African fly populations. Interestingly, the melanogaster subgroup of the genus Drosophila (which includes *D. melanogaster* and the *D. simulans* clade) originated in Africa (Lemeunier et al. 1986), and the diversity of Drosophila parasitic wasps that infect members of the subgroup appears to be highest there (Allemand et al. 2002).

A variety of evidence supports the idea that *lectin-24A* has evolved adaptively, especially in the *D. simulans* lineage. Haplotype structure in the California *D. simulans* population is highly unusual, with one diverged invariant haplotype present in seven of eight strains, and a second, quite distinct African-like haplotype present in one of eight strains. Hd is significantly low when compared with neutrally simulated data or to data from other genes from the same population sample, and extends only a very short distance around *lectin-24A*. A similar non-neutral pattern is observed for other kinds of population genetic descriptors, including Tajima's D, Fay and Wu's H, and linkage disequilibrium. These analyses suggest the common *lectin-24A* haplotype (or a haplotype from one of four closely linked genes) has been the target of a recent selective sweep, having increased in frequency in the population so rapidly and so recently that no recombinants or new mutations are observed. Furthermore, the dN/dS value for *lectin-24A* between *D. melanogaster* and *D. simulans* is in the top 1.46% of all genes in the genome, and McDonald–Kreitman analyses reveal a tremendous excess of nonsynonymous fixations within and outside the *lectin-24A* lectin domain. Altogether, *lectin-24A* polymorphism and divergence statistics suggest this recently acquired gene has evolved (and is evolving) novel function.

Previous work has shown that Drosophila immune genes as a class evolve more rapidly and adaptively than other genes in the genome (Schlenke and Begun 2003, 2005; Jiggins and Kim 2006; Sackton et al. 2007; Lazzaro 2008). Furthermore, a number of immune genes described in *D. melanogaster*, such as *Hemese* and the drosomycins, are relatively newly arisen, being limited to the melanogaster species group (Sackton et al. 2007). These data suggest fly hosts adapt to their pathogen environments using a combination of de novo gene origination and standing gene evolution, and *lectin-24A* appears to encompass both these methods of immune adaptation. If wasp venom proteins evolve to target and impair specific fly immune proteins and if Lectin-24A showed novel

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016

anti-wasp function that wasps were not yet able to counteract, *lectin-24A* origination and adaptation may have been (and may continue to be) part of a cyclic arms race between Drosophila and parasitic wasps. However, given our limited understanding of the biological function of Lectin-24A, coevolution with wasps is only one potential explanation for the adaptive evolution of *lectin-24A*.

In conclusion, *lectin-24A* is a new gene that is evolving rapidly and adaptively and that has a unique expression pattern of upregulation following wasp attack but down-regulation immediately following wounding or bacterial infection. These data, together with the facts that *lectin-24A* has a secretion signal sequence and a sugar-binding lectin domain, suggest it plays some role in recognition of extracellularly exposed sugars during the fly immune response against parasitic wasps, although at what stage of the response is unclear. It will be interesting to further dissect the regulatory network governing *lectin-24A* expression and to uncover the functional role of Lectin-24A in fly–wasp interactions in the future.

Supplementary Material

Supplementary figures S1–S5 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We would like to thank members of the Schlenke lab for helpful comments on the manuscript and funding from the National Institutes of Health grant AI081879 to T.A.S. Two anonymous reviewers also provided valuable comments that improved this manuscript.

References

Allemand R, Lemaitre C, Frey F, Bouletreau M, Vavre F, Nordlander G, van Alphen J, Carton Y. 2002. Phylogeny of six African Leptopilina species (Hymenoptera: Cynipoidea, Figitidae), parasitoids of Drosophila, with description of three new species. *Ann Soc Entomol Fr*. 38:319–332.

Ao J, Ling E, Yu XQ. 2007. Drosophila C-type lectins enhance cellular encapsulation. *Mol Immunol*. 44:2541–2548.

Ayres JS, Schneider DS. 2009. The role of anorexia in resistance and tolerance to infections in Drosophila. *PLoS Biol*. 7:e1000150.

Begun DJ, Holloway AK, Stevens K, et al. 2007. (13 co-authors). Population genomics: whole-genome analysis of polymorphism and divergence in Drosophila simulans. *PLoS Biol*. 5: e310.

Bettencourt R, Tanji T, Yagi Y, Ip YT. 2004. Toll and Toll-9 in Drosophila innate immune response. *J Endotoxin Res*. 10: 261–268.

Boutros M, Agaisse H, Perrimon N. 2002. Sequential activation of signaling pathways during innate immune responses in Drosophila. *Dev Cell*. 3:711–722.

Cambi A, Koopman M, Figdor CG. 2005. How C-type lectins detect pathogens. *Cell Microbiol*. 7:481–488.

Carpenter J, Hutter S, Baines JF, Roller J, Saminadin-Peter SS, Parsch J, Jiggins FM. 2009. The transcriptional response of Drosophila melanogaster to infection with the sigma virus (Rhabdoviridae). *PLoS One*. 4:e6838.

Carton Y, Bouletreau M, Van Alphen JJM, van Lenteren JC. 1986. The Drosophila parasitic wasps. In: Ashburner M, Carson HL, Thompson JN, editors. The genetics and biology of Drosophila. London: Academic Press. p 347–394.

Carton Y, Nappi AJ. 1997. Drosophila cellular immunity against parasitoids. *Parasitol Today*. 13:218–227.

Chen S, Zhang YE, Long M. 2010. New genes in Drosophila quickly become essential. *Science* 330:1682–1685.

Coustau C, Carkion Y, Nappi A, Shotkoski F, ffrench-Constant R. 1996. Differential induction of antibacterial transcripts in Drosophila susceptible and resistant to parasitism by Leptopilina boulardi. *Insect Mol Biol*. 5:167–172.

De Gregorio E, Spellman PT, Rubin GM, Lemaitre B. 2001. Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. *Proc Natl Acad Sci U S A*. 98:12590–12595.

Delaney JR, Stoven S, Uvell H, Anderson KV, Engstrom Y, Mlodzik M. 2006. Cooperative control of Drosophila immune responses by the JNK and NF-kappaB signaling pathways. *EMBO J*. 25:3068–3077.

Dostert C, Jouanguy E, Irving P, Troxler L, Galiana-Arnoux D, Hetru C, Hoffmann JA, Imler JL. 2005. The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of Drosophila. *Nat Immunol*. 6:946–953.

Eslin P, Giordanengo P, Fourdrain Y, Prevost G. 1996. Avoidance of encapsulation in the absence of VLP by a braconid parasitoid of Drosophila larvae: an ultrastructural study. *Can J Zool*. 74:2193–2198.

Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, Jiang H, Kanost M, Thompson GJ, Zou Z, Hultmark D. 2006. Immune pathways and defence mechanisms in honey bees Apis mellifera. *Insect Mol Biol*. 15:645–656.

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413.

Fleury F, Ris N, Allemand R, Fouillet P, Carton Y, Bouletreau M. 2004. Ecological and genetic interactions in Drosophila-parasitoids communities: a case study with D. melanogaster, D. simulans and their common Leptopilina parasitoids in south-eastern France. *Genetica* 120:181–194.

Honti V, Csordas G, Markus R, Kurucz E, Jankovics F, Ando I. 2010. Cell lineage tracing reveals the plasticity of the hemocyte lineages and of the hematopoietic compartments in Drosophila melanogaster. *Mol Immunol*. 47:1997–2004.

Irving P, Troxler L, Heuer TS, Belvin M, Kopczynski C, Reichhart JM, Hoffmann JA, Hetru C. 2001. A genome-wide analysis of immune responses in Drosophila. *Proc Natl Acad Sci U S A*. 98: 15119–15124.

Janssen A, Driessen G, De Haan M, Roodboi N. 1987. The impact of parasitoids on natural populations of temperate woodland Drosophila. *Neth J Zool*. 38:61–73.

Jiggins FM, Kim KW. 2006. Contrasting evolutionary patterns in Drosophila immune receptors. *J Mol Evol*. 63:769–780.

Johansson KC, Metzendorf C, Soderhall K. 2005. Microarray analysis of immune challenged Drosophila hemocytes. *Exp Cell Res*. 305:145–155.

Kelly JK. 1997. A test of neutrality based on interlocus associations. *Genetics* 146:1197–1206.

Labrosse C, Carton Y, Dubuffet A, Drezen JM, Poirie M. 2003. Active suppression of D. melanogaster immune response by long gland products of the parasitic wasp Leptopilina boulardi. *J Insect Physiol*. 49:513–522.

Lazzaro BP. 2008. Natural selection on the Drosophila antimicrobial immune system. *Curr Opin Microbiol*. 11:284–289.

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

Lemaitre B, Reichhart JM, Hoffmann JA. 1997. Drosophila host defense: differential induction of antimicrobial peptide genes

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016

after infection by various classes of microorganisms. *Proc Natl Acad Sci U S A*. 94:14614–14619.

Lemeunier F, David JR, Tsacas L, Ashburner M. 1986. The melanogaster species group. In: Ashburner M, Carson HL, Thompson JN, editors. The genetics and biology of Drosophila. London: Academic Press. p 147–256.

Levy F, Bulet P, Ehret-Sabatier L. 2004. Proteomic analysis of the systemic immune response of Drosophila. *Mol Cell Proteomics*. 3:156–166.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.

Long M, Betran E, Thornton K, Wang W. 2003. The origin of new genes: glimpses from the young and old. *Nat Rev Genet*. 4:865–875.

Loseva O, Engstrom Y. 2004. Analysis of signal-dependent changes in the proteome of Drosophila blood cells during an immune response. *Mol Cell Proteomics*. 3:796–808.

Marques MRF, Barracco MA. 2000. Lectins, as non-self-recognition factors, in crustaceans. *Aquaculture* 191:23–44.

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

McKerrow JH, Jeong KH, Beckstead JH. 1985. Enzyme histochemical comparison of Biomphalaria glabrata amebocytes with human granuloma macrophages. *J Leukoc Biol*. 37:341–347.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Nicolas E, Nappi AJ, Lemaitre B. 1996. Expression of antimicrobial peptide genes after infection by parasitoid wasps in Drosophila. *Dev Comp Immunol*. 20:175–181.

Rieu I, Powers SJ. 2009. Real-time quantitative RT-PCR: design, calculations, and statistics. *Plant Cell*. 21:1031–1033.

Rizki MTM. 1957. Alterations in the haemocyte population of Drosophila melanogaster. *J Morphol*. 100:437–458.

Rizki RM, Rizki TM. 1990. Parasitoid virus-like particles destroy Drosophila cellular immunity. *Proc Natl Acad Sci U S A*. 87:8388–8392.

Roxstrom-Lindquist K, Terenius O, Faye I. 2004. Parasite-specific immune response in adult Drosophila melanogaster: a genomic study. *EMBO Rep*. 5:207–212.

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.

Schlenke TA, Begun DJ. 2003. Natural selection drives Drosophila immune system evolution. *Genetics* 164:1471–1480.

Schlenke TA, Begun DJ. 2005. Linkage disequilibrium and recent selection at three immunity receptor loci in Drosophila simulans. *Genetics* 169:2013–2022.

Schlenke TA, Morales J, Govind S, Clark AG. 2007. Contrasting infection strategies in generalist and specialist wasp parasitoids of Drosophila melanogaster. *PLoS Pathog*. 3:1486–1501.

Schneider D, Shahabuddin M. 2000. Malaria parasite development in a Drosophila model. *Science* 288:2376–2379.

Smith NG, Eyre-Walker A. 2002. Adaptive protein evolution in Drosophila. *Nature* 415:1022–1024.

Sorrentino RP, Melk JP, Govind S. 2004. Genetic analysis of contributions of dorsal group and JAK-Stat92E pathway genes to larval hemocyte concentration and the egg encapsulation response in Drosophila. *Genetics* 166:1343–1356.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Theopold U, Rissler M, Fabbri M, Schmidt O, Natori S. 1999. Insect glycobiology: a lectin multigene family in Drosophila melanogaster. *Biochem Biophys Res Commun*. 261:923–927.

Turner MW. 1996. Mannose-binding lectin: the pluripotent molecule of the innate immune system. *Immunol Today*. 17:532–540.

Vierstraete E, Verleyen P, Baggerman G, D'Hertog W, Van den Bergh G, Arckens L, De Loof A, Schoofs L. 2004. A proteomic approach for the analysis of instantly released wound and immune proteins in Drosophila melanogaster hemolymph. *Proc Natl Acad Sci U S A*. 101:470–475.

Vierstraete E, Verleyen P, Sas F, Van den Bergh G, De Loof A, Arckens L, Schoofs L. 2004. The instantly released Drosophila immune proteome is infection-specific. *Biochem Biophys Res Commun*. 317:1052–1060.

Vinson S, Iwantsch G. 1980. Host regulation by insect parasitoids. *Q Rev Biol*. 55:143–165.

Vodovar N, Vinals M, Liehl P, Basset A, Degrouard J, Spellman P, Boccard F, Lemaitre B. 2005. Drosophila host defense after oral infection by an entomopathogenic Pseudomonas species. *Proc Natl Acad Sci U S A*. 102:11414–11419.

Walker J, Kwon SY, Badenhorst P, East P, McNeill H, Svejstrup JQ. 2011. Role of elongator subunit Elp3 in Drosophila melanogaster larval development and immunity. *Genetics* 187:1067–1075.

Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 7:256–276.

Wertheim B, Kraaijeveld AR, Hopkins MG, Walther Boer M, Godfray HC. 2011. Functional genomics of the evolution of increased resistance to parasitism in Drosophila. *Mol Ecol*.. 20:932–949.

Wertheim B, Kraaijeveld AR, Schuster E, Blanc E, Hopkins M, Pletcher SD, Strand MR, Partridge L, Godfray HC. 2005. Genome-wide gene expression in response to parasitoid attack in Drosophila. *Genome Biol*. 6:R94.

Willment JA, Brown GD. 2008. C-type lectin receptors in antifungal immunity. *Trends Microbiol*. 16:27–32.

Zettervall CJ, Anderl I, Williams MJ, Palmer R, Kurucz E, Ando I, Hultmark D. 2004. A directed screen for genes involved in Drosophila blood cell activation. *Proc Natl Acad Sci U S A*. 101:14192–14197.

Zhou Q, Zhang G, Zhang Y, Xu S, Zhao R, Zhan Z, Li X, Ding Y, Yang S, Wang W. 2008. On the origin of new genes in Drosophila. *Genome Res*. 18:1446–1455.

Downloaded from http://mbe.oxfordjournals.org/ at Pennsylvania State University on October 6, 2016