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

Erin S. Keebaugh* and Todd A. Schlenke

1Biology Department, Emory University

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

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 novel 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.

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,
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 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 Figidid wasp Leptopilina bouardi 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 macro-parasites. 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. bouardi 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 1-week-old female wasps were placed in each dish for a 2 h attack time. Two and nine hours post-attack, fly larvae were dissected and flash frozen for expression timepoint analyses. Due to the 2 h attack time and 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
Tools) dipped in sterile LB broth, Enterococcus faecalis gram(+) bacterial culture grown overnight and diluted to OD<sub>600</sub> = 1.0, or Escherichia coli gram(−) bacterial culture grown overnight and diluted to OD<sub>600</sub> = 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′-ACACTTCCAATAAAAACTCAA-TATGC-3′, 5′-CGGTGCTCAAGACGATGA-3′. Primers used for lectin-24A (which does not contain introns) are as follows: 5′-CGAGTGGGGTCCCTGGTAAAC-3′, 5′-GAAA CGCATGCCCTTGGTC-3′. Primers used for Drosomycin and Diptericin, antimicrobial peptides regulated by the Toll and Imd pathways, respectively, were modified from (Ayers and Schneider 2009) as follows: Drosomycin 5′-GTACTTGTTCGCCCTCCTCG-3′, 5′-CTTGCACACACGCA CGACAG-3′ and Diptericin 5′-ACCGCAGTACCCACCT-CAATC-3′, 5′-CCCAAGTGTCTCATCCAT-3′. Primers used for Drosophila lectin and/or Diptericin were found upregulated after wasp attack 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 dN/dS analyses for two reasons: 1) the possibility that sequence downstream and potentially upstream of the early termination codon 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
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. bouleri* wasps at the 2–5 h post-infection timepoint, compared with unattacked flies (fig. 1, supplementary material fig. S2a, 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 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, *P* < 0.05.

At the early timepoint, expression patterns for *Drosomycin* and *Diptericin* following sterile injury, gram(+) injury, and gram(−) injury were noticeably different than that of...
**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. S3a, 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. S3b, 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. S3c 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...
### Table 1. Population Genetic Statistics for *lectin-24A* and Flanking Loci

<table>
<thead>
<tr>
<th>Region</th>
<th><em>lectin-24A</em></th>
<th>Flanking Loci</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>ZnS</em></td>
<td><em>Hd</em></td>
</tr>
<tr>
<td></td>
<td><em>P</em>-value</td>
<td><em>P</em>-value</td>
</tr>
<tr>
<td>25 kb upstream</td>
<td>0.0109</td>
<td>0.0109</td>
</tr>
<tr>
<td>15 kb upstream</td>
<td>0.0109</td>
<td>0.0109</td>
</tr>
<tr>
<td>2 kb upstream</td>
<td>0.0109</td>
<td>0.0109</td>
</tr>
<tr>
<td>5 kb downstream</td>
<td>0.0109</td>
<td>0.0109</td>
</tr>
</tbody>
</table>

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. S3b, 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. S3a and S4a, 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. S4b, 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. S4a 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, caused by an insertion in *D. melanogaster* relative to *D. simulans* at the 3’ end of the coding sequence.

**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...
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 is 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

**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.
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

![Fig. 4. Hd of lectin-24A and 68 other genes from the California Drosophila simulans population (Schlenke and Begun 2003).](image-url)

![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.](image-url)
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.

Table 2. McDonald–Kreitman Analyses for lectin-24A.\textsuperscript{a}

<table>
<thead>
<tr>
<th></th>
<th>Synonymous</th>
<th>Nonsynonymous</th>
<th>\textit{P} Value\textsuperscript{c}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixations\textsuperscript{b}</td>
<td>14</td>
<td>51</td>
<td>–</td>
</tr>
<tr>
<td>All mel and all sim polymorphisms</td>
<td>30</td>
<td>39</td>
<td>0.0097*</td>
</tr>
<tr>
<td>All mel polymorphisms</td>
<td>4</td>
<td>10</td>
<td>0.7258</td>
</tr>
<tr>
<td>California mel polymorphisms</td>
<td>3</td>
<td>4</td>
<td>0.3443</td>
</tr>
<tr>
<td>Africa mel polymorphisms</td>
<td>2</td>
<td>7</td>
<td>1.0000</td>
</tr>
<tr>
<td>All sim polymorphisms</td>
<td>28</td>
<td>29</td>
<td>0.0021*</td>
</tr>
<tr>
<td>California sim polymorphisms</td>
<td>15</td>
<td>12</td>
<td>0.0026*</td>
</tr>
<tr>
<td>Africa sim polymorphisms</td>
<td>24</td>
<td>24</td>
<td>0.0023*</td>
</tr>
<tr>
<td>Fixations, non-lectin domain\textsuperscript{d}</td>
<td>9</td>
<td>29</td>
<td>–</td>
</tr>
<tr>
<td>All mel and all sim polymorphisms</td>
<td>19</td>
<td>24</td>
<td>0.0639</td>
</tr>
<tr>
<td>Fixations, lectin domain\textsuperscript{d}</td>
<td>5</td>
<td>22</td>
<td>–</td>
</tr>
<tr>
<td>All mel and all sim polymorphisms</td>
<td>11</td>
<td>15</td>
<td>0.0772</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Only full ORF strains included in analyses.

\textsuperscript{b} The number of fixed differences between the \textit{Drosophila melanogaster} (mel) and \textit{Drosophila simulans} (sim) population samples are compared with the number of polymorphisms from a variety of populations and species.

\textsuperscript{c} Significance determined by two-tailed Fisher’s exact test. * \textit{P} < 0.05.

\textsuperscript{d} The number of fixed differences between the \textit{D. melanogaster} (mel) and \textit{D. simulans} (sim) population samples in portions of lectin-24A are compared with the number of polymorphisms from the same portion.
measure the activation of the Imd pathway) (Lemaître 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 (Lemaître 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; Lemaître 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 (Lemaître 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., Zetervall 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 Ivantsch 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; Eskin 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
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 non-lectin 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
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 extra-cellularly 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


