Genome
An International Journal
for Genetics and Genomics

Research Article

Expression of defensin genes across house fly (*Musca domestica*) life history gives insight into immune system subfunctionalization

Danial Asgari[a], **Tanya Purvis**[b], **Victoria Pickens**[c], **Christopher Saski**[d], **Richard P. Meisel**[a], and **Dana Nayduch**[b]

[a]Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA; [b]Arthropod-Borne Animal Diseases Research Unit, United States Department of Agriculture, Agricultural Research Service, Manhattan, KS 66502, USA; [c]Department of Entomology, Kansas State University, Manhattan, KS 66506, USA; [d]Department of Plant and Environmental Sciences, Clemson University, Clemson, SC 29634, USA

Corresponding author: **Dana Nayduch** (email: Dana.Nayduch@usda.gov)

Abstract

Animals encounter diverse microbial communities throughout their lifetime, which exert varying selection pressures. Antimicrobial peptides (AMPs), which lyse or inhibit microbial growth, are a first line of defense against some of these microbes. Here we examine how developmental variation in microbial exposure has affected the evolution of expression and amino acid sequences of Defensins (an ancient class of AMPs) in the house fly (*Musca domestica*). The house fly is a well-suited model for this work because it trophically associates with varying microbial communities throughout its life history and its genome contains expanded families of AMPs, including Defensins. We identified two subsets of house fly Defensins: one expressed in larvae or pupae, and the other expressed in adults. The amino acid sequences of these two Defensin subsets form distinct monophyletic clades, and they are located in separate gene clusters in the genome. The adult-expressed Defensins evolve faster than larval/pupal Defensins, consistent with different selection pressures across developmental stages. Our results therefore suggest that varied microbial communities encountered across life history can shape the evolutionary trajectories of immune genes.

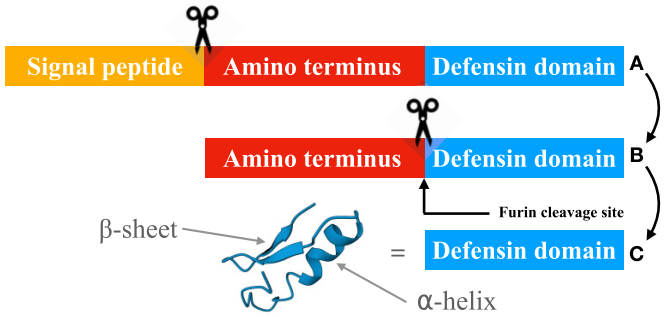
Key words: antimicrobial peptide, phylogeny, gene duplication, sequence evolution

Introduction

Antimicrobial peptides (AMPs) are crucial components of the innate immune response because of their activity against pathogens (Sarkar et al. 2021). AMPs inhibit microbial proliferation (Benfield and Henriques 2020), leading to evolutionary arms races between host immune systems and pathogens trying to avoid host defenses (Van Valen 1977; Tassanakajon et al. 2015). Possible effects of this selection pressure are high rates of AMP gene loss and duplication. However, specific factors responsible for the selective retention of AMP gene duplications are unresolved. One hypothesis is that maintaining multiple gene copies ensures a robust production of AMPs, effectively eliminating microbial pathogens (Lazzaro 2008; Lazzaro et al. 2020). This model implies that gene dosage, rather than specific amino acid sequences, is under selection. An alternative hypothesis suggests that context-dependent selection drives the retention of AMP gene duplications, as paralogs serve distinct functions across different contexts (Unckless et al. 2016). Under this model, gene duplications can be selectively retained via subfunctionalization, involving the accumulation of complementary loss of function mutations across gene copies (Hughes 1994; Force et al. 1999). Subfunctionalization leads to the partitioning of the ancestral gene function across paralogs—often manifesting in temporospatial partitioning of gene expression (e.g., developmental stages or tissues). An additional effect of subfunctionalization is the alleviation of pleiotropic constraints, enabling the adaptive specialization of paralogs for specific tissues or developmental stages (Force et al. 2005; Des Marais and Rausher 2008).

As a model system to explore the selection pressures shaping the evolution of AMP genes, we evaluated whether context-dependent effects contribute to the retention of AMP gene duplications in the house fly (*Musca domestica* L.). House flies live in septic environments and have an expanded immune gene repertoire containing many paralogous genes for AMPs and other effectors (West 1951; Sackton et al. 2017). The duplication of genes encoding AMPs and other effector proteins is hypothesized to play a pivotal role in the successful colonization and utilization of microbe-rich habitats, whose community composition can vary across life history and environmental niches (Nayduch and Burrus 2017). For example,

Fig. 1. Post-translational modifications of the canonical Defensin. Arthropod Defensin is comprised of a signal peptide (orange box), an amino terminus (red box), and a Defensin domain (blue box) (A). Prior to the secretion, the signal peptide and amino terminus are cleaved in a canonical Defensin. This produces the propeptide (B) and mature peptide (C). The scissors show the cleavage sites. The 3D structure of insect Defensins includes an α-helix and two antiparallel β-sheets (C). Here, the 3D structure of *Drosophila melanogaster* Defensin was modeled using Phyre2 (Kelley et al. 2015), and Mol* viewer was used to visualize the 3D structure of the Defensin (Sehnal et al. 2021).



larval house flies are immersed in a diverse microbial community of bacteria, fungi, and protists that they not only use for nutrition but also must consistently deter from breaching the larval cuticle. During pupation and metamorphosis, the fly is particularly vulnerable to bacteria retained from the third instar that can be released into the body during histolysis and histogenesis (Zurek and Nayduch 2016; de Jonge et al. 2020). In addition, unlike larvae, adult house flies exhibit no apparent nutritional requirement for microbes, but instead have ephemeral associations with microbes during feeding and breeding activities (Nayduch et al. 2023; Nayduch and Burrus 2017). These differences in trophic associations with microbes across life history may impose stage-specific selection pressures on the house fly immune system.

To unravel the intricacies of AMP gene paralog utilization by house flies, we focused on the expression of house fly Defensins, an ancient class of AMPs present in most multicellular organisms (Machado and Ottolini 2015) and the predominant invertebrate AMP (Rodríguez de la Vega and Possani 2005). The mature Defensin peptide consists of an alpha helix and two antiparallel beta sheets (Fig. 1), and the three-dimensional structure is stabilized by three disulfide bridges formed between six conserved cysteine residues (White et al. 1995). In insects, canonical Defensins are synthesized as pre-propeptides, featuring a signal peptide, a prosequence at the amino terminus, and a C-terminal Defensin domain (Fig. 1). Prior to secretion, the signal peptide is cleaved followed by cleavage of the prosequence furin site, resulting in the production of the mature Defensin domain (Dimarcq et al. 1990). RNA-seq was used to quantify the expression of house fly genes across eight life stages (from egg to adult), and we focused our comparisons on the 12 house fly Defensin genes along with the single *Defensin* gene in *Drosophila melanogaster*. Our study provided a framework to understand the interplay between subfunctionalization and amino acid sequence evolution of AMP genes.

Methods and materials

House fly life stage collections

House fly pupae were obtained from a colony of *M. domestica* (L.) established in 2011 and maintained at the USDA-ARS ABADRU, Manhattan, KS insectary. All cages and pans were maintained in a growth chamber (Percival, Perry, IA, USA) at 28 °C, 70% relative humidity, and a photoperiod of 13:11 (light:dark) h. Approximately 200 house fly pupae were added to a 20.5 cm^3 metal wire-screened cage for emergence and fed water and 10% sucrose solution ad libitum. Five days post-emergence, a protein source (1 tablespoon commercial calf feed) was fed to the adults to encourage egg production. A 90 mm Petri dish containing moistened wheat bran was added 4 days later and left 9 h for egg laying and collection.

For Day 0, eggs ($n = 10$/pool) were collected randomly from various egg batches found in wheat bran media. To collect first instar (L1) larvae, eggs ($n = 300$) were placed into individual Petri dishes filled with larval media (see composition below) to facilitate searching and collection of the larvae, which were identified and collected ($n = 10$/pool) after 24 h. The remaining hatched larvae and media were added to a larval pan containing larva media (~500 g, comprised of 0.5 L wheat bran, 0.3 L water, and ½ cup commercial calf feed). L2 larvae ($n = 10$/pool) and L3 larvae ($n = 2$/pool) were collected from each pan on Day 2 and Day 3, respectively. Once pupation was observed on Day 5, pans were inspected every 2 h to collect early (0–2 h after initiating pupation), middle (4–6 h after initiating pupation), and late (~48 h after initiating pupation) pupae ($n = 2$/pool). The remaining pupae were maintained in Petri dishes and checked daily (every 12 h) until male and female flies emerged (around Day 10–11), at which point they were collected ($n = 2$/pool, each). At time of collection, all samples were hand homogenized in 500 µL of Trizol reagent and stored except for L3, pupal, and adult samples, which were homogenized then were brought to a final volume of 1 mL before storing at −80 °C.

RNA extraction and sequencing

RNA was extracted from 9 different sample types from four independent replicates, and individuals from each sample type were pooled independently per replicate for RNA extraction as follows: eggs ($n = 10$), L1 ($n = 10$), L2 ($n = 10$), L3 ($n = 2$), pupal stages ($n = 2$ for each of three stages: early, middle, and late), adult male or adult female ($n = 2$, each). The sample size (n) is larger for early developmental stages ($n = 10$) compared to later stages ($n = 2$) because more individuals are needed to extract the same quantity of RNA at earlier developmental stages. However, across all developmental stages, there are four replicates; therefore, the statistical power for RNA-seq analysis is not affected by differences in the number of individuals per replicate. Pooled samples were homogenized with a pestle and stored at −80 °C until extraction. RNA was

extracted using the Zymo Direct-zol RNA Miniprep Plus kit (Zymo Research Corporate Headquarters, Irvine, CA, USA), with modification of the manufacturer's protocol for sample preparation as follows. Sample homogenates were thawed and mixed before centrifuging for 2 min at max speed (~10 000 RPM) to spin down insect tissues. The supernatant was then transferred to a clean nuclease-free microcentrifuge tube before adding 0.2× volume of Bromo-3-chloro-propane. After mixing by hand inversion, the solution was incubated at room temperature for 5 min, inverting the tube periodically. The sample was then centrifuged at 4 °C and max (>13 000) RPM for 15 min. The upper phase was carefully transferred to a clean nuclease-free microcentrifuge tube and processed following the manufacturer's RNA purification steps and washes, including DNase I treatment. RNA was quantified by Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA) and Qubit (Life Technologies, Thermo Fisher Scientific, Waltham, MA, USA) and pooled in equimolar ratios for library preparation. Library preparation was conducted with the NEBNext Ultra II RNA Library Prep Kit for Illumina following the manufacturer's recommended procedures. Paired-end reads for each sample (2 × 150 bp) were collected on an Illumina NovaSeq 6000 S4 flow cell by Novogene Co, Ltd. (Tianjin, China).

RNA-seq analysis

Gene expression was quantified with kallisto v 0.44.0 (Bray et al. 2016) by performing pseudoalignment of RNA-seq reads to annotation release 102 of the house fly reference transcriptome from genome assembly v2.0.2 (Scott et al. 2014). To quantify the expression of all known Defensins in the house fly, we supplemented the annotation with the four additional Defensin genes that were not included in the original annotation (Asgari et al. 2022). The updated transcriptome can be found here: https://github.com/danialasg74/House-fly-Transcriptome. Read mapping was performed separately for each of the four replicates of the nine sample types. Transcripts were assigned to their corresponding genes. Because genes with alternative splicing may have multiple transcripts, we measured the expression of a gene by summing over the pseudocount values of all of its transcripts. Transcript per million (TPM) values were summed from kallisto for genes with multiple transcripts to obtain gene-level TPM values (Tables S1 and S2).

Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) was used to identify modules of genes that are co-expressed across the life stages of house flies (Langfelder and Horvath 2008). WGCNA was run on the 9 RNA-seq sample types, including seven life stages from egg to late pupa, adult males, and adult females. Each sample includes four replicates, and replicate structure was included in the WGCNA input. Count data was used as the input to the WGCNA package to create a signed correlation matrix. The signed network allows us to identify connected genes whose expression is positively correlated. The correlation matrix was raised to the power of 20 to create a scale-free network. A merging threshold of 0.25, with a minimum module size of 30, was used to group genes with similar patterns of expression into clusters. Correlation of the module eigengenes was calculated for samples via Pearson's correlation and modules of co-expressed genes across life stages were identified.

RNA-seq data for developmental stages of *D. melanogaster*

Defensin gene expression across developmental stages of *D. melanogaster* was determined by using available RNA-seq data from the Model Organism ENCyclopedia Of DNA Elements (modENCODE) Project (Graveley et al. 2011). This dataset contains gene expression as RPKM (reads per kilobase of exon per million reads mapped) values for embryos (every 2 h from 0 to 24 h after oviposition), first and second instar larvae, different stages of third instar larvae (12 h post-molt, dark gut, light gut, and clear gut), prepupae, pupae samples for 12 and 24 h after pupa formation (APF), pharate adults (2, 3, and 4 APF), and adult male and female flies (1, 5, and 30 days old). We obtained RPKM values for 30 samples from FlyBase (Gelbart and Emmert 2013) with IDs FBlc0000086 to FBlc0000115 (Table S3).

Phylogenetic and transcription factor binding site analysis of house fly Defensins

Our analysis focused on the 12 house fly Defensin genes, which are all found in tandem on the scaffold (NW_004754939) at a single locus. We performed transcription factor binding site enrichment analysis across the Defensin locus following previous methods (Asgari et al. 2022).

The protein sequences of 12 house fly Defensins were used to construct phylogenetic trees, with *D. melanogaster Defensin* (*FBgn0010385*) used as the outgroup. The entire protein sequence—consisting of a signal peptide, amino terminus, and Defensin domain—was used to construct the alignment. Signal peptides and amino termini of Defensins were identified using the ProP method (Duckert et al. 2004). Alignments were performed using MUSCLE (Edgar 2004), implemented in MegaX with the default parameters (Kumar et al. 2018). Next, a phylogenetic tree was constructed using the Maximum Likelihood method with 500 bootstrap replications, setting *D. melanogaster* Defensin as the outgroup.

To measure the rate of evolution for house fly Defensin genes, we calculated the ratio of non-synonymous substitutions to synonymous substitutions ($\omega = \frac{dN}{dS}$) across the phylogeny using PAML (Yang 2007). Codon alignments were performed using the protein alignment and corresponding DNA sequences with PAL2NAL (Suyama et al. 2006). The codon alignment and the tree topology of the 12 house fly Defensins (excluding *D. melanogaster* Defensin) were used as inputs to PAML. We compared a branch-specific model in which a subset of the tree has a different ω value to a null model with a single ω for the entire Defensin tree using a likelihood ratio test (Yang 1998). We also compared a branch-site model (Model C) to a null model (M2a_rel) with a likelihood ratio test and used the results of the branch-site model to calculate the proportion of sites that have different ω values between

Defensin clades (Weadick and Chang 2012). Models were fitted both to the entire coding region of Defensin genes and to the specific domains (signal peptide, amino terminus region, and Defensin domain).

Comparative analysis of larval and adult Defensins

Separate phylogenetic trees were constructed for Defensin domains of adult and larval/pupal Defensins and the number of radical and conservative amino acid changes were compared between the two trees using BLOSUM62 and BLOSUM80 matrices (Henikoff and Henikoff 1992). The BLOSUM matrices contained log odd ratios $\left(2\log_2\left(\frac{\text{observed}}{\text{expected}}\right)\right)$ that estimated the likelihood of amino acid substitutions. Therefore, radical changes have negative likelihood values (expected > observed), and conservative changes have positive likelihood values (expected < observed). Radical changes were defined as those with likelihood values less than 0. To test for the robustness of our results, we also used a stricter definition of radical change as those with likelihood values less than −1. Synonymous substitutions at the codon level were defined as conservative changes, which increased the statistical power to detect the enrichment by increasing sample sizes. A Bayesian approach was used to calculate the posterior probability of finding radical and conservative changes in the adult or larval/pupal Defensin tree.

To assess physicochemical properties of predicted mature Defensin peptides, primary peptide sequences were trimmed to remove both signal peptide and propeptide amino terminus (if furin cleavage site was present) using ProP 1.0 (Duckert et al. 2004). Physicochemical profiles of mature Defensins were determined with NovoPro peptide property calculator (https://www.novoprolabs.com/).

Results

Network analysis reveals modules of co-expressed Defensins

RNA-seq was used to measure gene expression across eight developmental stages of the house fly, and modules of co-expressed genes across stages were identified. Nine different sample types were included (eggs, three larval stages, three pupal stages, adult males, and adult females). We grouped genes into 24 co-expression modules based on positive correlations across sample types. The modules of co-expressed genes and the expression of genes within each module are plotted in Figs. S1 and S2. Genes in a module were either all upregulated or downregulated within the same developmental stages.

The 12 house fly Defensin genes were assigned to six different co-expression modules, which are referred to by the colors Orange, Red, Brown, Blue, Green, and Gray (Fig. 2A). The six Defensin genes that belong to the Blue module (*LOC105261733*, *LOC105261775*, *LOC105261620*, *LOC101888225*, *NA1*, and *NA4*) were exclusively expressed in adult flies. The Defensin gene in the Green module (*NA2*) was expressed only in adult female flies. One Defensin (*NA1*) in the Gray module had undetectable expression across all life stages. However, using a previously published dataset (Sackton et al. 2017), we found that *NA1* is expressed in adult female house flies four days post eclosion, regardless of whether the flies are infected with bacteria (Fig. S3). Therefore, the eight Defensin genes in the Blue, Green, and Gray modules were expressed exclusively in adults, hereafter referred to as "adult Defensins". In contrast, the remaining four Defensin genes were predominantly expressed in larval (*LOC101888043* and *LOC101887540* are in the Orange module) or pupal (*LOC101887872* and *LOC101887709* are in the Red and Brown modules, respectively) house flies (Fig. 2A). We refer to these four genes as "larval/pupal Defensins".

We compared the expression of house fly Defensins across development with the expression of the single *D. melanogaster Defensin* (Fig. 3). Due to differences in the method used for quantification of RNA in our experiment (TPM) and the method used to quantify *D. melanogaster Defensin* (RPKM), we did not directly compare these values. Instead, we compared the trend in the expression of Defensins across the two species. *D. melanogaster Defensin* is not expressed in the embryo, similar to house fly Defensins. The expression of *D. melanogaster Defensin* gradually increases in larvae and pupae, drops in pharate adults, and then rises to its peak in 30-day-old adults. In *D. melanogaster*, adult female flies express *Defensin* at a higher level than male flies. Therefore, each house fly Defensin gene captures a subset of the full life history expression of *D. melanogaster Defensin*.

Two distinct phylogenetic clades of house fly Defensins correlate with genomic location

A phylogenetic tree was constructed using the protein sequences of the 12 house fly Defensins, with *D. melanogaster* Defensin as the outgroup (Fig. 2A). There are two separate clades of house fly Defensins. One clade has long branches leading to the eight adult Defensins. The other clade has short branches leading to all four larval/pupal Defensins.

House fly Defensin genes were organized into two separate clusters on a single genomic scaffold (Asgari et al. 2022). All eight adult Defensins are grouped together in one cluster (right side of Fig. 2C), and all four larval/pupal Defensins are in the other cluster (left side of Fig. 2C). Thus, the location of house fly Defensin genes on the chromosome is concordant with the expression patterns of Defensins, which are associated with the phylogenetic clustering of Defensin genes (Fig. 2). Notably, Imd-responsive motifs are enriched upstream of the larval/pupal Defensin genes but not upstream of the adult Defensin genes (all four larval/pupal Defensins have Imd-responsive motifs; Fig. S4). Imd-responsive elements are typically associated with gene expression induced by infection (Busse et al. 2007), suggesting that larval/pupal Defensin genes are more likely to be induced by infection (Asgari et al. 2022). We tested this hypothesis by analyzing a previously published RNA-seq data of house flies injected with a mixture of *Serratia marcescens* and *Enterococcus faecalis* (Fig. S5) (Sackton et al. 2017). Out of the 12 Defensins in the house fly, only the four larval/pupal Defensins were induced upon infection (henceforth larval/pupal/inducible), consistent with

Fig. 2. Identification of two classes of Defensins in the house fly based on expression pattern (A), phylogeny (B), and the position of genes in the genome (C). (A) The expression levels of the 12 house fly Defensin genes are plotted for house fly eggs (E), three instar stages of larvae (L1, L2, L3), early pupa (EP), mid pupae (MP), late pupae (LP), teneral females (TF), and teneral males (TM) as transcript per million (TPM). Colors correspond to gene co-expression modules. (B) The phylogeny of full-length house fly Defensin proteins, based on maximum likelihood, is shown. Gene names with the same color belong to the same co-expression module. *Drosophila melanogaster* Defensin (Def; FlyBase: *FBgn0010385*) was used as an outgroup. (C) The arrangement of house fly Defensin genes on the genomic scaffold (NW_004754939) is shown. The position of Defensins is based on the newly published house fly genome (Meisel et al. 2023). The solid horizontal lines show gaps in the assembly. The orientation of genes is shown by an arrow above the gene (positive strand) or below the gene (negative strand). The last four digits of each gene name (e.g., 7540 corresponds to *LOC101887540*) are used in this figure except for newly annotated Defensins (*NA1*, *NA2*, *NA3*, and *NA4*) (Asgari et al. 2022). Colors are congruent with those used for panels A and B, based on the co-expression analysis.

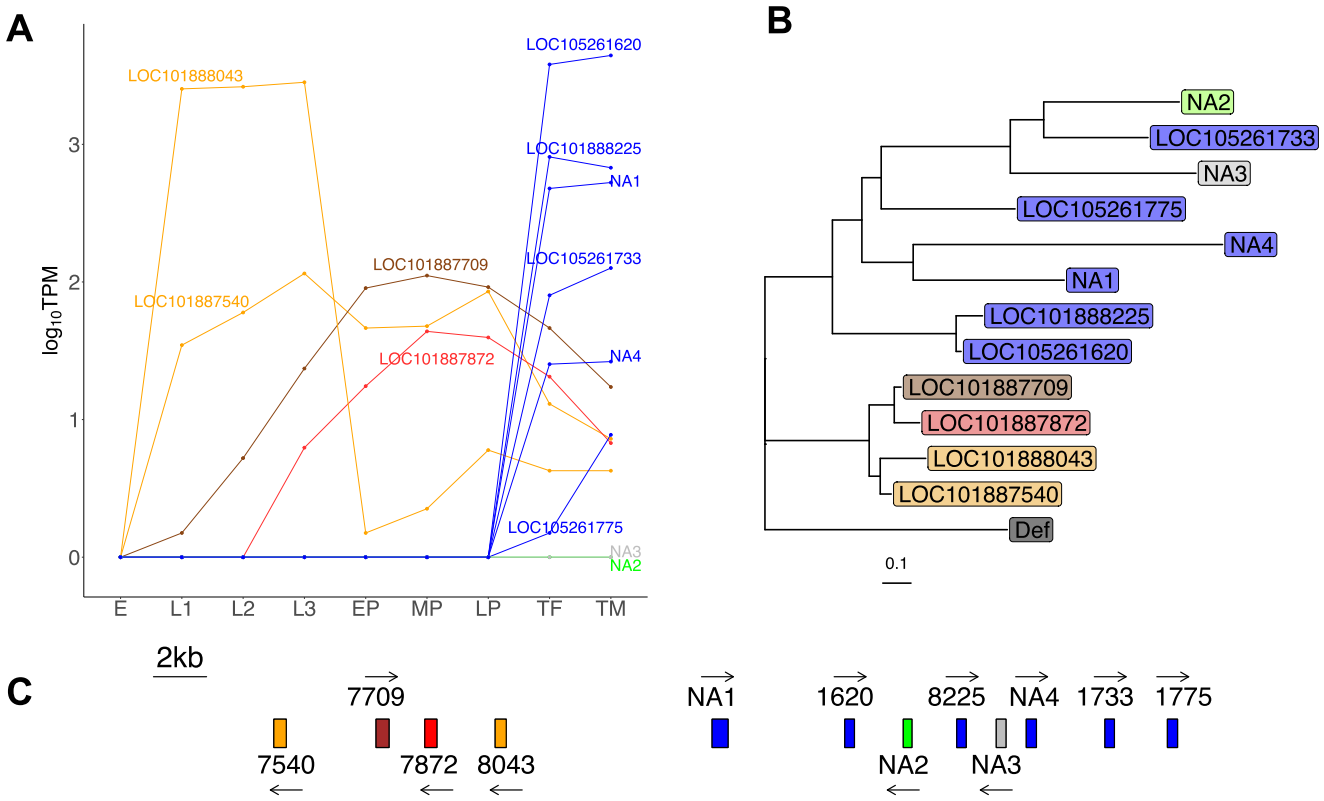
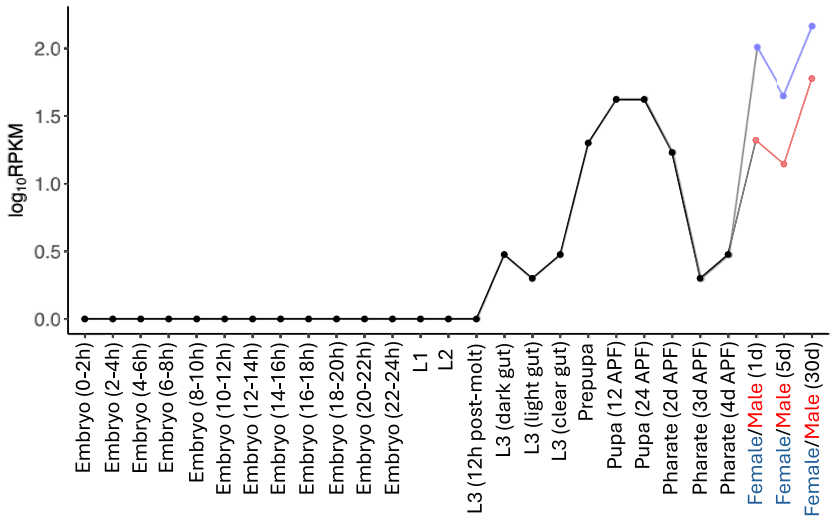


Fig. 3. Expression of the Defensin gene (*FBgn0010385*) across developmental stages of *Drosophila melanogaster*. The expression level of *Defensin* (*Y*-axis) is plotted for different developmental time points (*X*-axis).



the presence of imd-responsive motifs upstream of these genes.

Adult Defensins evolve faster than other Defensins

We next tested if adult Defensins evolve significantly faster than larval/pupal/inducible Defensins. This was motivated by the observation of longer branches for adult Defensins on the gene tree, compared to larval/pupal/inducible Defensins (Fig. 2B). First, an evolutionary model was fit to the phylogeny of house fly Defensins, which had separate ω (d_N/d_S) values for the adult clade (ω_1) and the larval/pupal clade (ω_2). The model with separate ω values was a better fit than a null model with a single ω for the entire tree (Table 1). The adult Defensin clade had a larger ω value than the larval/pupal clade ($\omega_1 > \omega_2$), which indicates faster evolution of adult Defensins. Because some amino acids were more conserved than others (e.g., cysteine residues in Defensins), a branch-site model (Model C) was fit to calculate the proportion of sites with different ω values in the two clades ($0 < \omega_1 \neq \omega_2 > 0$). The branch-site model was a better fit than a null model (M2a_rel) in which $\omega_1 = \omega_2 > 0$ (Table 1). The branch-site model estimates that ~77% of sites have different ω values between the two clades, and the adult Defensins have a faster rate of amino acid evolution than larval/pupal/inducible defensins at those differentially evolving sites ($\omega_1 > \omega_2$).

Defensins are produced as pre-propeptides consisting of a signal peptide, an amino terminus region, and a conserved Defensin domain. In a canonical Defensin, a conserved furin recognition site (RXKR) separates the amino terminus from the conserved Defensin domain (Thomas 2002). Larval/pupal/inducible Defensins have Lys–Arg residues at the furin recognition site (Fig. 4), suggesting that the amino terminus is cleaved from the conserved Defensin domain prior to secretion. In contrast, the furin recognition site is absent from adult house fly Defensins (Fig. 4). Based on this observation, we hypothesized that the amino terminus is retained in the secreted adult Defensins. Thus, the amino terminus region might evolve under different selection pressures in adult Defensins (where it may be retained) compared to larval/pupal/inducible Defensins (where it is likely cleaved prior to secretion).

To test the hypothesis of differential selection across domains between adult Defensins and larval/pupal/inducible defensins, we fit evolutionary models separately to each of the three regions of the Defensin genes (Table 2). There was no significant difference between a single ω model and a branch specific model for the signal peptide region, suggesting that the signal peptide evolves under similar selection pressures in both adult and larval/pupal/inducible Defensins. For the amino terminus region, there is a marginally significant difference between the ω values of the two clades (Table 2). The ω value was larger for the clade with adult Defensins than the larval/pupal/inducible Defensins ($\omega_1 > \omega_2$). Using a branch-site model, all sites in the amino terminus region had different ω values in the two clades, and the adult Defensins evolved faster than larval/pupal/inducible Defensins ($\omega_1 > \omega_2$). This is consistent with our hypothesis that selection acts differently on amino terminus regions of the two Defensin clades. For the conserved Defensin domain, ω values between the two clades were significantly different (Table 2), and the adult Defensins once again evolved faster than larval/pupal/inducible Defensins ($\omega_1 > \omega_2$). The sites with different ω values constitute ~69% of all sites in the Defensin domain, and those sites evolved faster in adult Defensins ($\omega_1 > \omega_2$). Altogether, our results suggest that both the amino terminus and the conserved Defensin domain evolved faster in adult Defensins compared to larval/pupal/inducible Defensins. All results from PAML analyses were robust to the number of bootstrap replicates used to construct the Defensin phylogeny. The analysis presented above used a phylogeny constructed with 500 bootstrap replicates, and we obtained qualitatively identical results using a phylogeny with 3000 bootstrap replicates (Tables S4 and S5).

To understand how the rapid evolution of adult Defensins shaped their structure and activity, we compared the number of radical and conservative amino acid substitutions between the two clades of Defensins. Radical substitutions are less likely to occur than conservative ones because radical substitutions change the volume and/or polarity of amino acids, thus affecting the structure and function of the protein (Zhang 2000). In the amino terminus region, a larger number of radical substitutions in adult Defensins was observed compared to larval Defensins; however, the signal for enrichment was not statistically significant (Fig. S6). In the Defensin domain, radical substitutions were more likely for adult Defensins compared to larval/pupal/inducible Defensins (Fig. 5). This pattern was observed regardless of the method that was used to define a radical amino acid substitution (Fig. S7). A larger number of radical substitutions in the adult Defensins is consistent with the hypothesis that the structure and/or activity of adult Defensins have been altered by their rapid evolution.

Larval/pupal/inducible mature defensins were predicted to be more cationic at neutral pH than adult defensins, potentially indicating activity in the hemolymph. Both groups were predicted to be cationic at low pH 4, indicating they would be active in the anterior midgut (Table 3). This shift in physicochemical properties of each group appears to be correlated with retention of the amino terminus domain amino acids in the adult group, that is otherwise cleaved from the larval/pupal/inducible Defensins.

Discussion

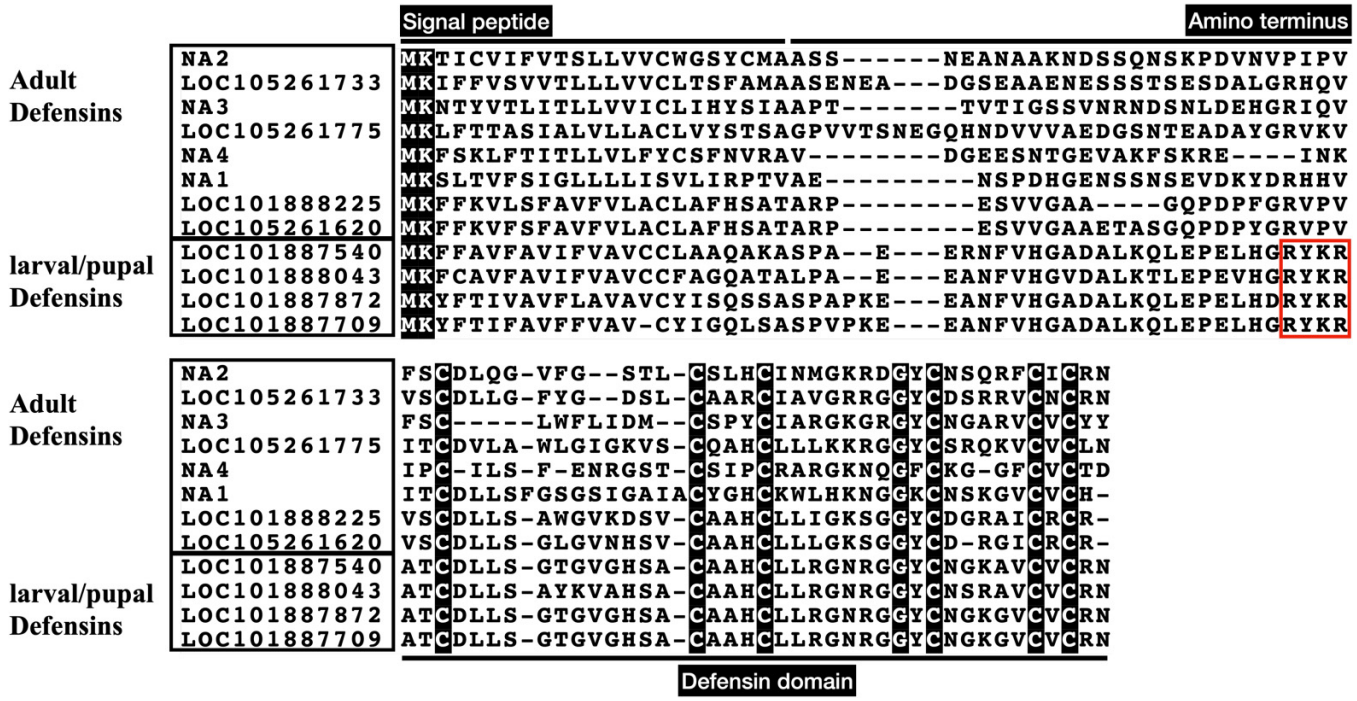
We identified two distinct classes of house fly Defensin genes based on gene expression, phylogeny, and genomic position. One class includes four Defensin genes that are predominantly expressed in larval and pupal stages, are inducible by bacterial infection in adults, have a conserved furin cleavage site for separation of the amino terminus from the Defensin domain, and have amino acid sequences that are evolving at a slower rate than the second class of Defensins. The second class includes eight genes that are exclusively expressed in adult house flies, do not have a furin cleavage site (thereby presumably retaining the amino

Table 1. Comparison of phylogenetic models of ω.

Model[a]	ω	Ln (likelihood)	χ^2 value	p value[b]
Single ω (null)	0.299	−2050.466816	8.741	0.003
Branch model (alternative)*	$\omega_1 = 0.385$ $\omega_2 = 0.165$	−2046.096189		
M2a_rel (null)	0.337	−2008.728248	8.98615	0.003
Model C (alternative)*	$\omega_1 = 0.464$ $\omega_2 = 0.165$	−2004.235173		

[a]The model with the better fit is shown with an asterisk.
[b]Degrees of freedom for all model comparisons are equal to 1.

Fig. 4. Alignment of protein sequences of 12 house fly Defensins. Four Defensins are primarily expressed in larvae and pupae (larval/pupal) and eight Defensins have an adult exclusive expression (adult Defensins). The red box shows the location of the furin cleavage site between the amino terminus and the conserved Defensin domain. The conserved cleavage site (RXKR) is missing from adult exclusive Defensin sequences. Conserved residues across 12 Defensins are colored in black, including six cysteines present in all insect Defensins. The figure was generated with BOXSHADE.



terminus domain in the mature peptide), and have faster evolving amino termini and Defensin domains. In contrast, the single *D. melanogaster Defensin* is expressed across most developmental stages, suggesting that the house fly Defensin paralogs represent subfunctionalized duplicates of an ancestral, broadly-expressed Defensin gene. Further, the different rates of protein evolution across house fly Defensins suggest that the subfunctionalized expression profiles result in different selection pressures on each Defensin's amino acid sequences as a consequence of context-specific gene expression.

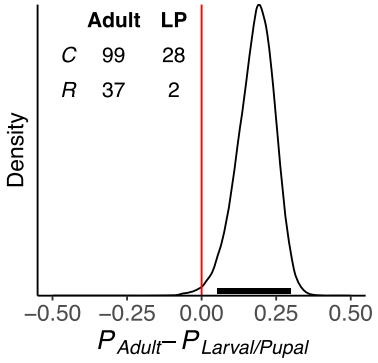
Our results are consistent with prior evidence for subfunctionalization of house fly Defensin duplications. Four of the eight adult Defensins that we identified in this study (*LOC105261620*, *LOC101888225*, *LOC105261733*, and *NA1*) were previously shown to be highly constitutively expressed in either gut or carcass of adult flies, whether or not they were infected with bacteria (Asgari et al. 2022). Our prior results suggested that the transcription factor Myc was responsible for the constitutive expression of those four Defensins (Meisel et al. 2023). Of note, Imd-responsive motifs are generally absent upstream of the adult Defensins (Asgari et al. 2022), yet are enriched upstream of the four larval/pupal/inducible Defensin genes (Fig. S4). The Imd pathway is one of the two primary signaling cascades responsible for the induction of AMP gene expression in flies (Myllymäki et al. 2014). In accordance, using a previously published dataset, we found larval/pupal/inducible Defensins to be induced upon infection (Asgari et al. 2022). Subfunctionalization thus appears to result in the partitioning of AMP gene expression across many separate contexts, including tissues, developmental stages, and infection status.

Table 2. Comparison of phylogenetic models of ω for different domains of Defensin.

Defensin protein Segment	Model	ω^*	Ln(likelihood)	χ^2value	p value†
Signal peptide	Single ω (null)	0.380	−684.215863	0.5083	0.476
	Branch model (alternative)	$\omega_1 = 0.434$ $\omega_2 = 0.305$	−683.961690		
Amino terminus	Single ω (null)	0.219	−489.300543	2.792214	0.095
	Branch model (alternative)	$\omega_1 = 0.296$ $\omega_2 = 0.069$	−487.904436		
	M2a_rel (null)	0.148	−489.300543	2.792214	0.095
	Model C (alternative)	$\omega_1 = 0.296$ $\omega_2 = 0.069$	−487.904436		
Defensin domain	Single ω (null)	0.132	−838.307184	5.762094	0.016
	Branch model (alternative)	$\omega_1 = 0.197$ $\omega_2 = 0.045$	−835.426137		
	M2a_rel (null)	0.181	−803.081693	8.792538	0.003
	Model C (alternative)	$\omega_1 = 0.267$ $\omega_2 = 0.038$	−798.685424		

*Adult clade (ω_1) and the larval/pupal/inducible clade (ω_2).
†Degrees of freedom for all model comparisons are equal to 1.

Fig. 5. The probability of finding radical amino acid substitutions between the two classes of Defensins (i.e., adult Defensins and larval/pupal/inducible Defensins). The *x*-axis shows the posterior probability of finding a radical change in the adult Defensins minus the probability of finding a radical change in the larval/pupal/inducible Defensins. The *y*-axis shows the density of the distribution. The high density interval (HDI) for the distribution is shown as a horizontal line. The HDI excludes zero; thus, radical changes are enriched in adult Defensin genes ($P_{adult} > P_{larval/pupa}$). A vertical red line shows the zero value. The number of radical (R) and conservative (C) changes for adult and larval/pupal (LP) Defensins are reported in the inset table. This analysis was performed using BLOSUM62 and a threshold of −1 for the likelihood that defines radical changes.



Our data further suggest that subfunctionalization of gene expression might cause differential selection pressures on the corresponding proteins encoded by the two classes of Defensin genes. An accelerated rate of evolution was determined for the Defensin domains of adult Defensins, relative to larval/pupal/inducible Defensins. We also found a larger number of radical amino acid substitutions in the Defensin domain of adult Defensins compared to larval/pupal/inducible Defensins. In a canonical Defensin, only the Defensin domain interacts with the pathogen, while the signal peptide and amino terminus are cleaved prior to export from the cell (Dimarcq et al. 1998). In house flies, adult Defensins have lost the furin cleavage site responsible for removal of the amino terminus, which could lead to dramatic changes in the structure and/or activity of adult Defensins, particularly in the amino terminus region. Furthermore, the difference in evolutionary rates of Defensin domains between adult and larval/pupal/inducible Defensins may be influenced by the diversity and composition of microbial communities encountered across house fly development

Table 3. Predicted physicochemical properties of mature house fly Defensins.

Paralog	pI	Charge at pH4	Charge at pH7
LOC101887540	8.31	5.4	2.9
LOC101887709	8.31	5.4	2.9
LOC101887872	8.31	5.4	2.9
LOC101888043	8.58	6.4	3.9
NA1	6.64	8.6	−0.7
LOC105261620	7.64	5.5	0.9
NA2	7.64	4.3	0.8
LOC101888225	8.02	5.3	1.8
NA3	8.01	4.9	1.8
NA4	8.02	6.1	1.6
LOC105261733	4.51	3.0	−4.2
LOC105261775	7.64	6.6	0.9

that would impose unique selection pressures on Defensin evolution (Nayduch and Burrus 2017).

The evolution of house fly Defensins is consistent with the duplication-degeneration-complementation (DDC) model of subfunctionalization (Force et al. 1999). The DDC model predicts that following gene duplication, null mutations in the regulatory regions can result in the partitioning of gene expression across contexts. Our results lend evidence for partitioning of gene expression when house fly and *D. melanogaster* expression are compared. Differential expression across paralogs could further impose unique selection pressures on paralogs expressed in each context, which could explain the differences in evolutionary rates observed between adult and larval/pupal/inducible Defensins. Another consequence of subfunctionalization is that it can remove adaptive conflicts that arise when genes must function in different contexts (Hughes 1994). In the case of Defensins, these conflicts could arise if the selection pressures on the amino acid sequence differ across development imparted by the diverse and disparate microbial communities to which larvae, pupae, and adults are exposed. Duplication and subfunctionalization of house fly Defensins could therefore have allowed for adaptive specialization of paralogs within each developmental context.

The accelerated evolution of the Defensin domain in adult Defensins is further evidence for adaptive specialization. One interpretation of the higher ω for the adult Defensins is that they have experienced an excess of amino acid substitutions that were fixed by positive selection. Alternatively, elevated ω can also be caused by relaxed purifying selection. Because the PAML analysis we implemented ignores segregating polymorphism in the population when calculating ω (Goldman and Yang 1994), we are unable to differentiate between positive selection and relaxed purifying selection (Nielsen 2005). Therefore, in the absence of polymorphism data for house flies, we cannot determine whether the accelerated evolution of adult Defensins is due to relaxed purifying selection or positive selection.

Surprisingly, we also found evidence for differential selection in the amino terminus region between adult and larval/pupal/inducible Defensins. The amino terminus is typically cleaved from the mature peptide in canonical Defensin, but the furin recognition site is absent from adult house fly Defensins. This raises the possibility that the amino terminus domain is retained in secreted adult Defensins, but not in larval/pupal/inducible Defensins (where the furin cleavage site is conserved). The faster evolution of the amino termini in adult Defensins could therefore be explained by positive selection imposed by interactions with the microbial cell envelope or perhaps even other AMPs or effectors. Notably, the retention of the amino terminus in the adult Defensins changes their physicochemical properties, indicating subfunctionalization for use primarily in the adult gut. Tissue-specific gene expression analyses have confirmed expression of 4 of 8 adult defensins in the gut of both males and females (data not shown). Alternatively, the amino terminus of adult Defensins may be evolving under relaxed purifying selection, as described above. Further functional analyses, such as spectrum of activity or protein–protein interactions with other house fly effectors, could lend insight into the function, if any, of the retained amino terminus in adult house fly Defensins.

Our results also address the hypothesis that AMP paralogs are selectively retained because insects with a microbe-rich lifestyle require a higher dosage of AMP expression that could be obtained through concurrent expression of multiple gene copies. House flies encounter high doses of pathogenic bacteria and other microbes across all life history stages (West 1951). Thus, although it has been speculated that the expanded repertoire of house fly AMPs is related to the need to produce high doses of AMPs in microbe-rich environments (Sackton et al. 2017), partitioning of the expression of distinct classes of Defensin genes across house fly developmental stages instead does not adequately support this theory. In addition, the gene dosage hypothesis predicts a high sequence identity across duplicated Defensin genes because the amino acid sequences should be under similar selection pressures and have similar functions. In contrast to that prediction, we found that adult Defensins are evolving rapidly, indicating potential diversification of Defensins that may have different spectra of activity from each other in this subgrouping. It is worth noting, however, that the subfunctionalization and gene dosage hypotheses are not mutually exclusive. For example, the conserved sequences of the four larval/pupal/inducible Defensins could be interpreted as evidence that a high dose of a specific type of Defensin peptide is required in some contexts.

In summary, our study identified two classes of Defensin genes in the house fly, those with larval/pupal/inducible expression and those expressed only in adults. These two groups are subject to unique selection pressures. When compared to the expression of *D. melanogaster Defensin*, our findings suggest subfunctionalization of house fly Defensins structure and expression that caters to the specific needs associated with house fly microbial encounters across life history. The unique selection pressures on adult Defensins may have resulted in the loss of a furin cleavage site, resulting in retention of their amino termini in the secreted mature peptide. The biological significance of the retention of the amino terminal region in adult Defensins remains to be understood, but it appears to affect the selection pressures on that region and may be related to microbial encounters adults face during their trophic activities. This study provides valuable insight regarding subfunctionalization of AMP genes and their importance in house fly defense against life history specific encounters with bacteria. This knowledge could identify vulnerabilities to target for novel control of house flies.

Acknowledgement

This work was completed in part using the Carya Cluster in the Research Computing Data Core at the University of Houston.

Article information

History dates

Received: 7 February 2024
Accepted: 12 April 2024

Accepted manuscript online: 9 May 2024
Version of record online: 10 June 2024

Author information

Author ORCIDs
Danial Asgari https://orcid.org/0000-0002-2987-0291
Tanya Purvis https://orcid.org/0000-0002-1976-9755
Victoria Pickens https://orcid.org/0000-0002-9179-9860
Christopher Saski https://orcid.org/0000-0002-2780-4274
Richard P. Meisel https://orcid.org/0000-0002-7362-9307
Dana Nayduch https://orcid.org/0000-0002-9258-6066

Author contributions
Conceptualization: DA, CS, RPM, DN
Data curation: DN
Formal analysis: DA
Funding acquisition: DN
Investigation: DA, TP, VP, RPM, DN
Methodology: DA, TP, VP, CS, DN
Project administration: RPM, DN
Resources: CS, DN
Supervision: RPM, DN
Visualization: DA
Writing – original draft: DA, RPM, DN
Writing – review & editing: DA, TP, VP, CS, RPM, DN

Competing interests
The authors declare there are no competing interests.

Funding information
This work was supported by a cooperative agreement (58-3020-8-016) between the United States Department of Agriculture and the University of Houston as well as funding from USDA-ARS NP104 project number 3020-32000-018-00D and High Priority Area Research Seed Grant from the University of Houston. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The conclusions in this report are those of the authors and do not necessarily represent the views of the USDA. USDA is an equal opportunity provider and employer.

Supplementary material

Supplementary data are available with the article at https://doi.org/10.1139/gen-2024-0016.

References

Asgari, D., Saski, C.A., Meisel, R.P., and Nayduch, D. 2022. Constitutively-expressed and induced immune effectors in the house fly (*Musca domestica*) and the transcription factors that may regulate them. Insect Mol. Biol. **31**(6): 782–797. doi:10.1111/imb.12804. PMID: 35875866.

Benfield, A.H., and Henriques, S.T. 2020. Mode-of-action of antimicrobial peptides: membrane disruption vs. intracellular mechanisms. Front. Med. Technol. **2**. doi:10.3389/fmedt.2020.610997.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. 2016. Near-optimal probabilistic RNA-seq quantification. Nature Biotechnol. **34**(5): 525–527. doi:10.1038/nbt.3519. PMID: 27043002.

Busse, M.S., Arnold, C.P., Towb, P., Katrivesis, J., and Wasserman, S.A. 2007. A kappaB sequence code for pathway-specific innate immune responses. EMBO J. **26**(16): 3826–3835. doi:10.1038/sj.emboj.7601798. PMID: 17660749.

De Jonge, N., Michaelsen, T.Y., Ejbye-Ernst, R., Jensen, A., Nielsen, M.E., Bahrndorff, S., and Nielsen, J.L. 2020. Housefly (*Musca domestica* L.) associated microbiota across different life stages. Sci. Rep. **10**(1): 7842. doi:10.1038/s41598-020-64704-y. PMID: 32398740.

Des Marais, D.L., and Rausher, M.D. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature, **454**(7205): 762–765. doi:10.1038/nature07092. PMID: 18594508.

Dimarcq, J.-L., Bulet, P., Hetru, C., and Hoffmann, J. 1998. Cysteine–rich antimicrobial peptides in invertebrates. Pept. Sci. doi:10.1002/(SICI)1097-0282(1998)47:6%3c465::AID-BIP5%3e3.0.CO;2-.

Dimarcq, J.L., Zachary, D., Hoffmann, J.A., Hoffmann, D., and Reichhart, J.M. 1990. Insect immunity: expression of the two major inducible antibacterial peptides, defensin and diptericin, in *Phormia terranovae*. EMBO J. **9**(8): 2507–2515. doi:10.1002/j.1460-2075.1990.tb07430.x. PMID: 2369900.

Duckert, P., Brunak, S., and Blom, N. 2004. Prediction of proprotein convertase cleavage sites. Protein Eng., Des. Sel. **17**(1): 107–112. doi:10.1093/protein/gzh013.

Edgar, R.C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinf. **5**: 113. doi:10.1186/1471-2105-5-113. PMID: 15318951.

Force, A., Cresko, W.A., Pickett, F.B., Proulx, S.R., Amemiya, C., and Lynch, M. 2005. The origin of subfunctions and modular gene regulation. Genetics, **170**(1): 433–446. doi:10.1534/genetics.104.027607. PMID: 15781713.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Yi-L, and Postlethwait, J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics, **151**(4): 1531–1545. doi:10.1093/genetics/151.4.1531. PMID: 10101175.

Gelbart, W.M., and Emmert, D.B. 2013. Flybase high throughput expression pattern data. FlyBase Analysis, https://flybase.org/reports/FBrf0221009.htm [last accessed 30 May 2024].

Goldman, N., and Yang, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**(5): 725–736. PMID: 7968486.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, Li, et al. 2011. The developmental transcriptome of *Drosophila melanogaster*. Nature, **471**(7339): 473–479. doi:10.1038/nature09715. PMID: 21179090.

Henikoff, S., and Henikoff, J.G. 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U.S.A. **89**(22): 10915–10919. doi:10.1073/pnas.89.22.10915. PMID: 1438297.

Hughes, A.L. 1994. The evolution of functionally novel proteins after gene duplication. Proc. R. Soc. B, **256**(1346): 119–124.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. **10**(6): 845–858. doi:10.1038/nprot.2015.053. PMID: 25950237.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. **35**(6): 1547–1549. doi:10.1093/molbev/msy096. PMID: 29722887.

Langfelder, P., and Horvath, S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. **9**: 559. doi:10.1186/1471-2105-9-559. PMID: 19114008.

Lazzaro, B.P. 2008. Natural selection on the *Drosophila* antimicrobial immune system. Curr. Opin. Microbiol. **11**(3): 284–289. doi:10.1016/j.mib.2008.05.001. PMID: 18555739.

Lazzaro, B.P., Zasloff, M., and Rolff, J. 2020. Antimicrobial peptides: application informed by evolution. Science, **368**(6490):. doi:10.1126/science.aau5480.

Machado, L.R., and Ottolini, B. 2015. An evolutionary history of defensins: a role for copy number variation in maximizing host innate

and adaptive immune responses. Front. Immunol. **6**: 115. doi:10.3389/fimmu.2015.00115. PMID: 25852686.

Meisel, R.P., Freeman, J.C., Asgari, D., Llaca, V., Fengler, K.A., Mann, D., et al. 2023. New insights into immune genes and other expanded gene families of the house fly, *Musca domestica*, from an improved whole genome sequence. Arch. Insect Biochem. Physiol. e22049. doi:10.1002/arch.22049. PMID: 37608635.

Myllymäki, H., Valanne, S., and Rämet, M. 2014. The *drosophila* imd signaling pathway. J. Immunol. **192**(8): 3455–3462. doi:10.4049/jimmunol.1303309.

Nayduch, D., and Burrus, R.G. 2017. Flourishing in filth: house fly–microbe interactions across life history. Ann. Entomol. Soc. Am. **110**(1): 6–18. doi:10.1093/aesa/saw083.

Nayduch, D., Neupane, S., Pickens, V., Purvis, T., and Olds, C. 2023. House Flies Are Underappreciated Yet Important Reservoirs and Vectors of Microbial Threats to Animal and Human Health. Microorganisms **11**: 583. doi:10.3390/microorganisms11030583.

Nielsen, R. 2005. Molecular signatures of natural selection. Ann. Rev. Genet. **39**: 197–218. doi:10.1146/annurev.genet.39.073003.112420. PMID: 16285858.

Rodriguezdelavega, R., and Possani, L. 2005. On the evolution of invertebrate defensins. Trend Genet.doi:10.1016/j.tig.2005.03.009.

Sackton, T.B., Lazzaro, B.P., and Clark, A.G. 2017. Rapid expansion of immune-related gene families in the house fly, *musca domestica*. Mol. Biol. Evol. **34**(4): 857–872. PMID: 28087775.

Sarkar, T., Chetia, M., and Chatterjee, S. 2021. Antimicrobial peptides and proteins: from Nature's reservoir to the laboratory and beyond. Front. Chem. **9**: 691532. doi:10.3389/fchem.2021.691532. PMID: 34222199.

Scott, J.G., Warren, W.C., Beukeboom, L.W., Bopp, D., Clark, A.G., Giers, S.D., et al. 2014. Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. Genome Biol. **15**(10): 466. doi:10.1186/s13059-014-0466-3. PMID: 25315136.

Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., et al. 2021. Mol* viewer: modern web app for 3D visualization and analysis of large biomolecular structures. Nucleic Acids Res. **49**(W1): W431–W437. doi:10.1093/nar/gkab314. PMID: 33956157.

Suyama, M., Torrents, D., and Bork, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. **34**(Web Server issue): W609–W612. doi:10.1093/nar/gkl315. PMID: 16845082.

Tassanakajon, A., Somboonwiwat, K., and Amparyup, P. 2015. Sequence diversity and evolution of antimicrobial peptides in invertebrates. Dev.& Comp. Immunol. **48**(2): 324–341.

Thomas, G. 2002. Furin at the cutting edge: from protein traffic to embryogenesis and disease. Nat. Rev. Mol. Cell Biol. **3**(10): 753–766. doi:10.1038/nrm934.

Unckless, R.L., Howick, V.M., and Lazzaro, B.P. 2016. Convergent balancing selection on an antimicrobial peptide in *Drosophila*. Curr. Biol. **26**(2): 257–262. doi:10.1016/j.cub.2015.11.063.

Van Valen, L. 1977. The Red Queen. Am. Nat. **111**(980):. The University of Chicago Press: 809–810. doi:10.1086/283213.

Weadick, C.J., and Chang, B.S.W. 2012. An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. Mol. Biol. Evol. **29**(5): 1297–1300. doi:10.1093/molbev/msr311.

West„ L.S. 1951 The housefly: its natural history, medical importance, and control. J. Am. Med. Assoc. **147**(6): 614. doi:10.1001/jama.1951.03670230080042.

White, S.H., Wimley, W.C., and Selsted, M.E. 1995. Structure, function, and membrane integration of defensins. Curr. Opin. Struct. Biol. **5**(4): 521–527. doi:10.1016/0959-440X(95)80038-7.

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**(5): 568–573. doi:10.1093/oxfordjournals.molbev.a025957.

Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. **24**(8): 1586–1591. doi:10.1093/molbev/msm088.

Zhang, J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J. Mol. Evol. **50**(1): 56–68. doi:10.1007/s002399910007.

Zurek, K., and Nayduch, D. 2016. Bacterial associations across house fly life history: evidence for transstadial carriage from managed manure. J. Insect Sci. **16**(1):. doi:10.1093/jisesa/iev156.