

RESEARCH ARTICLE

Open Access



# The genome of the stable fly, *Stomoxys calcitrans*, reveals potential mechanisms underlying reproduction, host interactions, and novel targets for pest control

Pia U. Olafson<sup>1\*</sup>, Serap Aksoy<sup>2</sup>, Geoffrey M. Attardo<sup>3</sup>, Greta Buckmeier<sup>1</sup>, Xiaoting Chen<sup>4</sup>, Craig J. Coates<sup>5</sup>, Megan Davis<sup>1</sup>, Justin Dykema<sup>6</sup>, Scott J. Emrich<sup>7</sup>, Markus Friedrich<sup>6</sup>, Christopher J. Holmes<sup>8</sup>, Panagiotis Ioannidis<sup>9</sup>, Evan N. Jansen<sup>8</sup>, Emily C. Jennings<sup>8</sup>, Daniel Lawson<sup>10</sup>, Ellen O. Martinson<sup>11</sup>, Gareth L. Maslen<sup>10</sup>, Richard P. Meisel<sup>12</sup>, Terence D. Murphy<sup>13</sup>, Dana Nayduch<sup>14</sup>, David R. Nelson<sup>15</sup>, Kennan J. Oyen<sup>8</sup>, Tyler J. Raszick<sup>5</sup>, José M. C. Ribeiro<sup>16</sup>, Hugh M. Robertson<sup>17</sup>, Andrew J. Rosendale<sup>18</sup>, Timothy B. Sackton<sup>19</sup>, Perot Saelao<sup>1</sup>, Sonja L. Swiger<sup>20</sup>, Sing-Hoi Sze<sup>21</sup>, Aaron M. Tarone<sup>5</sup>, David B. Taylor<sup>22</sup>, Wesley C. Warren<sup>23</sup>, Robert M. Waterhouse<sup>24</sup>, Matthew T. Weirauch<sup>25,26</sup>, John H. Werren<sup>27</sup>, Richard K. Wilson<sup>28,29</sup>, Evgeny M. Zdobnov<sup>9</sup> and Joshua B. Benoit<sup>8\*</sup>

## Abstract

**Background:** The stable fly, *Stomoxys calcitrans*, is a major blood-feeding pest of livestock that has near worldwide distribution, causing an annual cost of over \$2 billion for control and product loss in the USA alone. Control of these flies has been limited to increased sanitary management practices and insecticide application for suppressing larval stages. Few genetic and molecular resources are available to help in developing novel methods for controlling stable flies.

**Results:** This study examines stable fly biology by utilizing a combination of high-quality genome sequencing and RNA-Seq analyses targeting multiple developmental stages and tissues. In conjunction, 1600 genes were manually curated to characterize genetic features related to stable fly reproduction, vector host interactions, host-microbe dynamics, and putative targets for control. Most notable was characterization of genes associated with reproduction and identification of expanded gene families with functional associations to vision, chemosensation, immunity, and metabolic detoxification pathways.

(Continued on next page)

\* Correspondence: [pia.olafson@usda.gov](mailto:pia.olafson@usda.gov); [joshua.benoit@uc.edu](mailto:joshua.benoit@uc.edu)

<sup>1</sup>Livestock Arthropod Pests Research Unit, USDA-ARS, Kerrville, TX, USA

<sup>8</sup>Department of Biological Sciences, University of Cincinnati, Cincinnati, OH, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Conclusions:** The combined sequencing, assembly, and curation of the male stable fly genome followed by RNA-Seq and downstream analyses provide insights necessary to understand the biology of this important pest. These resources and new data will provide the groundwork for expanding the tools available to control stable fly infestations. The close relationship of *Stomoxys* to other blood-feeding (horn flies and *Glossina*) and non-blood-feeding flies (house flies, medflies, *Drosophila*) will facilitate understanding of the evolutionary processes associated with development of blood feeding among the Cyclorrhapha.

**Keywords:** Stable fly genome, Muscid genomics, Insect orthology, Chemoreceptor genes, Opsin gene duplication, Metabolic detoxification genes, Insect adaptation, Gene regulation, Insect immunity

## Background

Livestock ectoparasites are detrimental to cattle industries in the USA and worldwide, impacting both confined and range-land operations. Flies from the Muscidae family commonly occupy these settings, including the nonbiting house fly and face fly and the blood-feeding (hematophagous) stable fly and horn fly. These muscid flies exhibit different larval and adult biologies, including diversity in larval developmental substrates, adult nutrient sources, and feeding frequency [1, 2]. As such, control efforts against these flies are not one size fits all. The stable fly, *Stomoxys calcitrans* (L.), in particular, is a serious hematophagous pest with a cosmopolitan host range, feeding on bovids, equids, cervids, canines, and occasionally humans throughout much of the world [2]. The stable fly's painful bites disrupt livestock feeding behavior [3–6]; these bites can be numerous during heavy infestation, leading to reductions of productivity by over \$2 billion USD [7].

Stable fly larvae occupy and develop in almost any type of decomposing vegetative materials that are often contaminated with animal wastes [8]. In Australia, Brazil, and Costa Rica, dramatic increases in stable fly populations have coincided with the expansion of agricultural production where the vast accumulation of post-harvest byproducts are recognized as nutrient sources for development of immature stages [9–11]. The active microbial communities residing in these developmental substrates (e.g., spent hay, grass clippings, residues from commercial plant processing, manure) are required for larval development and likely provide essential nutrients [12]. Even though stable flies are consistently exposed to microbes during feeding and grooming activities, biological transmission (uptake, development, and subsequent transmission of a microbial agent by a vector) of pathogens has not been demonstrated for organisms other than the helminth *Habronema microstoma* [13, 14]. Stable flies have been implicated in mechanical (non-biologic) transmission of Equine infectious anemia, African swine fever, West Nile, and Rift Valley viruses, *Trypanosoma* spp., and *Besnoitia* spp. (reviewed by [13]). The apparent low vector competence of stable flies implicates the importance of immune system pathways not only in regulating larval survival in microbe-rich environments but also in the inability of pathogens to

survive and replicate in the adult midgut following ingestion [15–17].

Stable flies rely on chemosensory input to localize nutritional resources, such as volatiles emitted by cattle [18–23] and volatiles/tastants produced from plant products [24–26]. Stable fly mate location and recognition are largely dependent upon visual cues and contact pheromones [27, 28], and gravid females identify suitable oviposition sites through a combination of olfactory and contact chemostimuli along with physical cues [12, 19, 20, 23, 29]. Since stable flies infrequently associate with their hosts, feeding only 1 to 2 times per day, on-animal and pesticide applications are less effective control efforts than those that integrate sanitation practices with fly population suppression by way of traps [30]. Given the importance of chemosensory and vision pathways, repellents have been identified that target stable fly chemosensory inputs and current trap technologies exploit stable fly visual attraction [31–33]. However, despite these efforts, consistent control of stable fly populations remains challenging and development of novel control mechanisms is greatly needed.

Although both sexes feed on sugar, adults are reliant on a bloodmeal for yolk deposition and egg development, as well as seminal fluid production [26, 34]. Blood feeding evolved independently on at least five occasions within the Diptera, in the Culicimorpha, Psychodomorpha, Tabanomorpha, Muscoidea, and Hippoboscoidea [35]. The Muscinae appear to have a high propensity for developing blood feeding; which has occurred at least four times within this subfamily—once in each of the *domestica*-, *sorbens*- and *lusoria*- groups and again in the Stomoxini [36]. Unlike other groups of blood-feeding Diptera where non-blood feeding ancestors are distantly related and / or difficult to discern, stomoxynes are imbedded with the subfamily Muscinae of the Muscidae, featuring many non-blood feeding species. Contrasting blood-feeding culicimorphs and tabanimorphs, stable flies exhibit gonotrophic discordance [37, 38], requiring three to four blood meals for females to develop their first clutch of eggs and an additional two to three for each subsequent clutch of eggs. These unique aspects

of stable flies offer opportunities for comparative analysis of the genomic features underlying these key biological traits.

Even with the importance of the stable fly as a pest, little is known about the molecular mechanisms underlying the biology of *S. calcitrans*. To further our understanding of this critical livestock pest, we report a draft genome sequence of the stable fly. The quality of this genome is high and includes in silico annotation that was aided by extensive developmental and tissue-specific RNA-Seq data focusing on the feeding and reproduction of *S. calcitrans*. Manual curation and comparative analyses focused on aspects underlying the role of this fly as a pest related to host interactions, reproduction, control, and regulation of specific biological processes. Our study significantly advances the understanding of stable fly biology including the identification of unique molecular and physiological processes associated with this blood-feeding fly. These processes can serve as novel targets which will assist in both developing and improving control of this important livestock pest.

## Results and discussion

### Genome assembly and annotation supported by comparative and functional genomics

Whole genome shotgun sequencing of pooled, teneral adult males from an *S. calcitrans* line developed for this project resulted in the 66x coverage draft assembly of 971 MB of total sequence (see Additional file 1: Table S1). Scaffolds (12,042) and contigs (125,702) had N50 lengths of 504.7 and 11.3 kb, respectively. The sequence was ~20% smaller compared to the genome predicted by propidium iodide analyses (~1150 MB, [39]). This difference is likely the result of heterochromatin and other repetitive regions that were unassembled, as genome size is not significantly different between the sexes [39] and is comparable to differences documented for other insect genomes [40–42]. There were 16,102 predicted genes/pseudogenes that included 2003 non-protein coding genes, and a total of 22,450 mRNA transcripts were predicted with over 94% supported by RNA-Seq (see Additional file 2). The *S. calcitrans* mitochondrial genome has been previously described [43]. Its 18 kb genome encodes 13 protein-coding, 2 rRNA and 23 tRNA gene sequences. In addition, sequence analysis revealed an extra copy of the *trnI* gene, similar to a blow fly duplication, and partial sequences of the control region. Manual curation and analyses of the nuclear genome sequences allowed preliminary chromosome arm assignment and identification of repeat elements from the genome (see Additional file 1: Tables S2 and S3 [44–61];). A select set of protein coding genes ( $n = 1600$ ) were manually annotated to identify genes associated with functional classes including immunity, host sensing, reproduction, feeding, and metabolic detoxification.

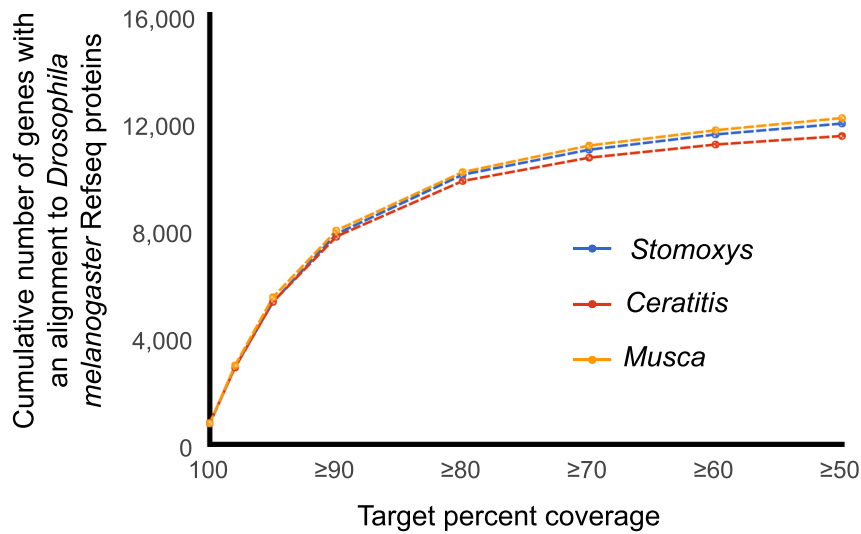
Completeness of the genome assembly was assessed via comparison of the *S. calcitrans* genome and predicted gene set against a database of fly derived benchmarking universal single-copy ortholog genes (BUSCOs [62]). Based on the near-universal single-copy orthologs from dipterans (OrthoDB v8, [63]), 95.1% were found in the assembly and 92.2% in the final predicted gene set (see Additional file 1: Figure S1 [63–67];). Comparison of the *Stomoxys* gene set against that of *Drosophila melanogaster* revealed approx. 10,000 *Stomoxys* genes with at least 75% target coverage, which is a similar number of genes with significant alignments relative to that of other published fly genomes (Fig. 1). Lastly, CEGMA genes and those associated with autophagy were all identified from the *S. calcitrans* genome (see Additional file 1: Table S4 and Additional file 3). These gene set comparisons provide an additional metric of genome completeness as these are highly conserved among flies [40, 68]. These metrics indicate that the genome is of sufficient quality for subsequent comparative analyses with other insects, specifically in relation to gene content. Comparison of protein orthologs revealed 122 *Stomoxys* species-specific protein families relative to other higher flies (Fig. 1). Based on gene ontology, there was enrichment for zinc finger transcription factors in relation to all genes ( $p = 0.02–0.007$ ; see Additional file 2), which has also been documented in other insect systems [69, 70].

Evaluation of gene expression by RNA-Seq produced a comprehensive catalog of sex-biased gene expression as well as genes enriched in different developmental stages and organs (see Additional file 4). Samples of RNA were sequenced from the following conditions: whole females (teneral and mated, reproductive), whole males (teneral), male reproductive tracts (mated), female reproductive tracts (mated), male heads (fed, mated), female heads (fed, mated), a third instar larva, and female/male salivary glands. Each condition consisted of a single replicate. These datasets will serve as a valuable resource for future studies. A stringent statistical significance cutoff (FDR,  $P < 0.01$ ) was used as only a single replicate was analyzed for each treatment [40, 69, 70]. A significant correlation (Pearson's  $R^2 = 0.8643$ ,  $P < 0.0001$ ; Figure S2) was observed between  $\log_2$  fold changes of reverse transcription quantitative real-time PCR (RT-qPCR) versus normalized expression values of 25 genes (see Additional file 1; Figure S2, Table S5), validating that we can glean biological relevance from the RNA-Seq datasets.

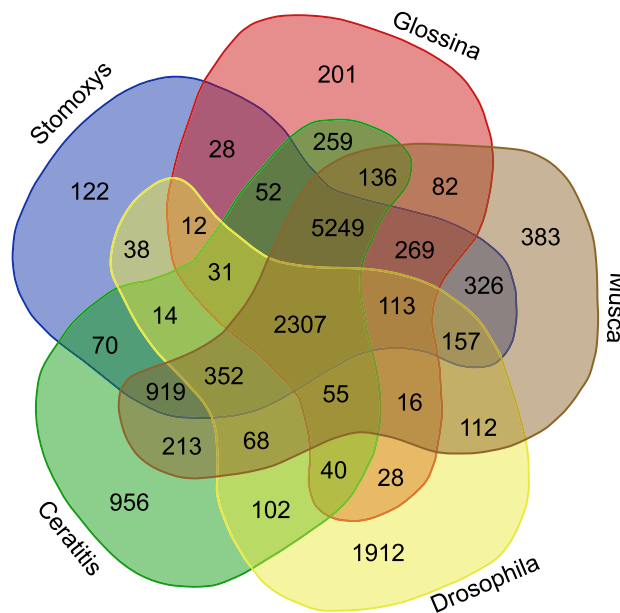
### Limited evidence for lateral gene transfer and no evidence of endosymbionts in the *Stomoxys* genome of a laboratory-colonized strain

Stable fly larvae are absolutely dependent on bacteria for survival and development [12, 29, 71], and larval physiology impacts the composition of this microbial

A.



B.



**Fig. 1** Quality assessment of the *Stomoxys calcitrans* genome. **a** Number of genes with alignment to *Drosophila melanogaster* genome. **b** Ortholog group comparison between *S. calcitrans* (*Stomoxys*), *Musca domestica* (*Musca*), *Ceratitis capitata* (*Ceratitis*), *Glossina morsitans* (*Glossina*), and *D. melanogaster* (*Drosophila*) based on comparison to the OrthoDB8 database

community irrespective of the substrate in which the larvae are developing [17]. As such, use of bacterial communities to deliver targeted control technologies, e.g., gene silencing constructs, is an approach being considered for use with stable flies [72], which led us to analyze culturable bacteria from surface-sterilized, adult *S. calcitrans* collected from Texas dairies. This revealed a variety of genera representing the adult gut bacterial community (see Additional file 5). Among those

cultured, the most abundant phyla represented were Proteobacteria and Firmicutes with only 3% of isolates classified as Actinobacteria and Bacteroidetes. Similar to mosquitoes, harbored bacterial communities were likely ingested while grooming or acquired during ingestion of nectar or water sources, as strict blood feeders usually have reduced gut microbiota [73–75]. As such, the most prevalent genus was *Aeromonas* sp., which are frequently found in aquatic and wet environments, such as

irrigation water, fecal matter, and moist organic waste, and have been cultured from other arthropods [76]. In addition, the adult *S. calcitrans* bacterial communities share similarity with those isolated from other flies associated with filth and decomposition (e.g., blow flies and other related species, [77, 78]), suggesting these bacteria may also be acquired by *S. calcitrans* adults when visiting oviposition sites.

Given this association with microbial communities, *Stomoxys* may contain laterally transferred genes for bacteria that impact on their biology, as in other reported insect genomes [79]. In addition to identifying potential symbiotic associations revealed in the genome assembly, it is also important to rule out bacterial contaminating scaffolds that could mistakenly be attributed as components of the *Stomoxys* genome. For these reasons, a DNA based pipeline was used to screen for potential contaminating bacteria scaffolds and possible lateral gene transfers (LGTs) from bacteria into the *Stomoxys* genome. The pipeline revealed only trace bacterial contaminants in the genome assembly (see Additional file 1: Table S6). This contrasts with some other arthropod genome assemblies, which contained complete or nearly complete bacterial genomes, such as the parasitoid *Trichogramma pretiosum* [80], amphipod *Hyaella azteca* [81], and the bedbug *Cimex lectularius* [40]. The results suggest that a maternally inherited and/or environmentally acquired microbiota does not occur in any abundance in teneral *Stomoxys* males. Whether circumstances are different in adult females would require further study.

The LGT analysis revealed presence of three candidates (see Additional file 1: *Stomoxys* lateral gene transfer, Table S7), all of which were derived from *Wolbachia*, a common endosymbiont found in arthropods [82] that infects 40–60% of insect species [83, 84]. *Wolbachia* are a common source of LGTs [85, 86], likely due to their association with the germline of their insect hosts. The three candidates were examined for sequencing read depth in the candidate LGT and flanking DNA to determine if there were large changes in read depth across the junction that would be indicative of miss-assembly to contaminating bacterial DNA (see Additional file 1: Figure S3). The results did not reveal large changes in read depth across the junctions. Further, there is no evidence that any of the three LGTs contain functional protein coding genes, and only one had detectable expression in our RNA-Seq datasets that occurred within the 3' UTR of XM\_013245585 (see Additional file 1: Table S7). This gene encodes a transcription factor containing a basic leucine zipper domain, but whether expression of the LGT is a consequence of its location in the UTR or is biologically significant is unknown. While *Wolbachia* was described from the closely related horn fly, *Haematobia irritans* [87], the *Stomoxys* strain used for the genome sequencing was not infected

with *Wolbachia*, no *Wolbachia* scaffolds were found in the assembly, and there are no reports of natural occurrence of *Wolbachia* in *Stomoxys* populations [88]. Presence of these LGTs, then, likely reflects LGT events from a past *Wolbachia* infection in the ancestors of *Stomoxys*. Stable flies will consume blood and nectar for nourishment [24, 26, 89], and this is different from the closely-related tsetse flies, which are obligate blood feeders. Due to this limited food source, tsetse flies (*Glossina morsitans*) harbor an obligate symbiont, *Wigglesworthia glossinidae*, that provides B vitamins that are present at low levels in blood [41, 90, 91]. Analysis of the assembled *S. calcitrans* genome, described above, did not reveal a distinct microbial symbiont.

### The *Stomoxys* immune system encodes gene family expansions that may reflect adaptation to larval development in microbe-rich substrates

The closely related house fly, *Musca domestica*, occupies microbe-rich environments that overlap with those of the stable fly, particularly in livestock production settings. Noted expansions in immune system-related gene families of the house fly are hypothesized to be a consequence of this ecology [92–94]. Analysis of the *S. calcitrans* genome revealed extensive conservation of immune system signaling pathways coupled with dramatic expansions of some gene families involved in both recognition and effector functions. The insect immune system—best characterized from work in the model organism *D. melanogaster*—includes both cellular defenses (e.g., macrophage-like cells that phagocytose pathogenic microorganisms) and a humoral defense system that results in the production of antimicrobial effector molecules [95]. The humoral immune system can be divided into recognition proteins, which detect pathogenic bacteria and fungi; signaling pathways, which are activated by recognition proteins and result in the translocation of transcription factors to the nucleus to induce gene expression; and effectors, which are (typically) secreted and ultimately act to clear infections.

Previous comparative work suggests that at least some parts of the immune system are deeply conserved across Dipterans and indeed most insects. Genes encoding immune signaling proteins, in particular, are generally preserved as single-copy orthologs across a wide range of insects [41, 93, 94, 96–101], with only rare exceptions [102]. Despite the strong conservation of the basic structure of the main signaling pathways in insect immunity, there is considerable evidence for variation in both the gene content and protein sequence of the upstream inputs (recognition proteins) and downstream outputs (effector proteins) of the immune system (e.g., [93, 98–100, 103, 104]).

Major components of the Toll, Imd, JAK/STAT, p38, and JNK pathways in the *S. calcitrans* genome (see Additional file 1: Table S8) were found to be largely conserved as single-copy orthologs [95]. A description and full lists of putative computationally annotated and manually curated immune-related genes in *S. calcitrans* is provided (see Additional file 1: Table S9, [15, 105–112]; Additional files 6 and 7). These findings are consistent with previous reports for many other Dipterans and supports the conclusion that the intracellular signaling mechanisms of innate immunity have been stable during the evolutionary history of Dipterans [97, 98, 100]. In contrast, the gene families encoding upstream recognition proteins and downstream effector proteins tend to be expanded in *S. calcitrans* and *M. domestica* relative to other Dipterans (Table 1).

Hidden Markov Model profiles and manual curation were used to analyze four canonical recognition families

with well-characterized immune roles and an additional eight families with less well-defined roles (Table 1). For three of the four canonical pattern recognition receptor families (NIM, PGRP, and TEP), and four of the other families (CTL, FREP, GALE, and SRCB), the *S. calcitrans* genome encodes either the most members or second-most members after *M. domestica* among the 5 Dipteran genomes screened (*S. calcitrans* plus *Aedes aegypti*, *D. melanogaster*, *M. domestica*, and *G. morsitans*). A similar pattern holds for downstream effector proteins: the *S. calcitrans* genome encodes either the most or second-most after *M. domestica* for attacins (ATT), defensins (DEF), cecropins (CEC), and lysozymes (LYS). In contrast, comparable numbers of non-canonical effector gene family members were identified across the Dipteran genomes screened. Not unexpectedly, three additional classes of antimicrobial peptides (AMP) were originally characterized in *D. melanogaster* but are missing from

**Table 1** Number of immune-related gene family members from sequenced Dipteran genomes, annotated by hidden Markov models

	<i>Scal</i>	<i>Mdom</i>	<i>Gmor</i>	<i>Aaeg</i>	<i>Dmel</i>
<b>Canonical pattern recognition</b>					
Nimrod (NIM)	25	23	10	8	17
Peptidoglycan recognition protein (PGRP)	17	17	4	10	13
beta-1,3-glucan-binding proteins (BGBP)	4	3	3	7	7
Thioester containing proteins (TEP)	16	22	4	8	6
<b>Other recognition</b>					
C-type lectin (CTL)	78	41	11	43	38
Fibrinogen-related proteins (FREP)	49	38	7	34	14
Galectins (GALE)	15	13	8	12	6
Immunoglobulin superfamily (IGSF)	1	1	1	0	1
MD2-like proteins (MD2L)	8	12	5	26	8
Scavenger receptor class A (SRCA)	3	3	2	2	3
Scavenger receptor class B (SRCB)	15	18	11	13	14
Scavenger receptor class C (SRCC)	7	8	4	5	9
<b>Canonical effectors</b>					
Attacin antimicrobial peptides (ATT)	11 (12)*	10	4	1	4
Defensin antimicrobial peptides (DEF)	5 (11)*	5	0	4	1
Diptericin antimicrobial peptides (DIPT)	3 (1)*	4	0	1	3
Cecropin antimicrobial peptides (CEC)	5 (10)*	12	2	9	5
Lysozymes (LYS)	23	32	4	7	13
<b>Non-canonical effectors</b>					
Thioredoxin peroxidases (TPX)	5	6	6	5	8
Prophenoloxidases (PPO)	19	23	4	25	10
Glutathione peroxidases (GPX)	1	1	0	3	2
Heme peroxidases (HPX)	12	12	8	19	10
Transferrins (TSF)	4	6	3	5	3

\*Numbers in parentheses are those numbers annotated after manual curation of the *S. calcitrans* genome. *Scal*: *Stomoxys calcitrans*, *Mdom*: *Musca domestica*, *Gmor*: *Glossina morsitans*, *Aaeg*: *Aedes aegypti*, *Dmel*: *Drosophila melanogaster*

the *M. domestica* genome ([94]; Metchnikowin, Drosocin, Drosomycin) and are also not detected in the *S. calcitrans* genome.

Several of the expanded AMP gene families were found tandemly arranged on individual scaffolds (see Additional file 1: Table S9). For example, the 11 *Stomoxys defensing* genes are located on a single scaffold (KQ079966) in two clusters, one of which includes five *defensins* present upstream of the other that includes six *defensins* (see Additional file 1: Figure S4). Within the downstream cluster, three genes were detected by RNA-Seq exclusively in larva and have 84–97% amino acid identity to each other, while the other three share 81–87% amino acid identity and were detected in heads of fed adults and male reproductive tracts but not in teneral adults (see Additional file 1: Table S10). The sequence identity and shared expression pattern support the likelihood that these sub-groups arose separately from gene duplication events. In contrast, the five *defensin* genes in the upstream cluster share 30–66% amino acid identity, and four were detected in teneral adults with variable detection in tissues of fed adults. This expression pattern is consistent with that reported for *Stomoxys midgut defensin 1* (*Smd1*) and *Smd2*, both of which are present in this upstream cluster [109].

In combination with the previously reported expansions of many effector and recognition immune components in the house fly [93, 94], our analysis of the *S. calcitrans* genome suggests that Muscidae likely have expanded the diversity of their immune repertoires, sometimes dramatically, despite differences in adult feeding ecology (blood feeder vs generalist). Muscid flies complete their life cycle in animal waste and other septic environments that are rich in communities of pathogenic and non-pathogenic bacteria. These communities and their metabolic products serve as the essential nutrient source for larval development, and larvae are repeatedly interacting with these bacteria. Further, the communities are in constant flux within the substrate over the course of this development. One hypothesis is that the shared diversification of immune receptors and effectors is driven by this larval ecology, while additional *M. domestica* specific expansions (e.g., in TEPs) are accounted for by the saprophytic adult feeding behavior of that species.

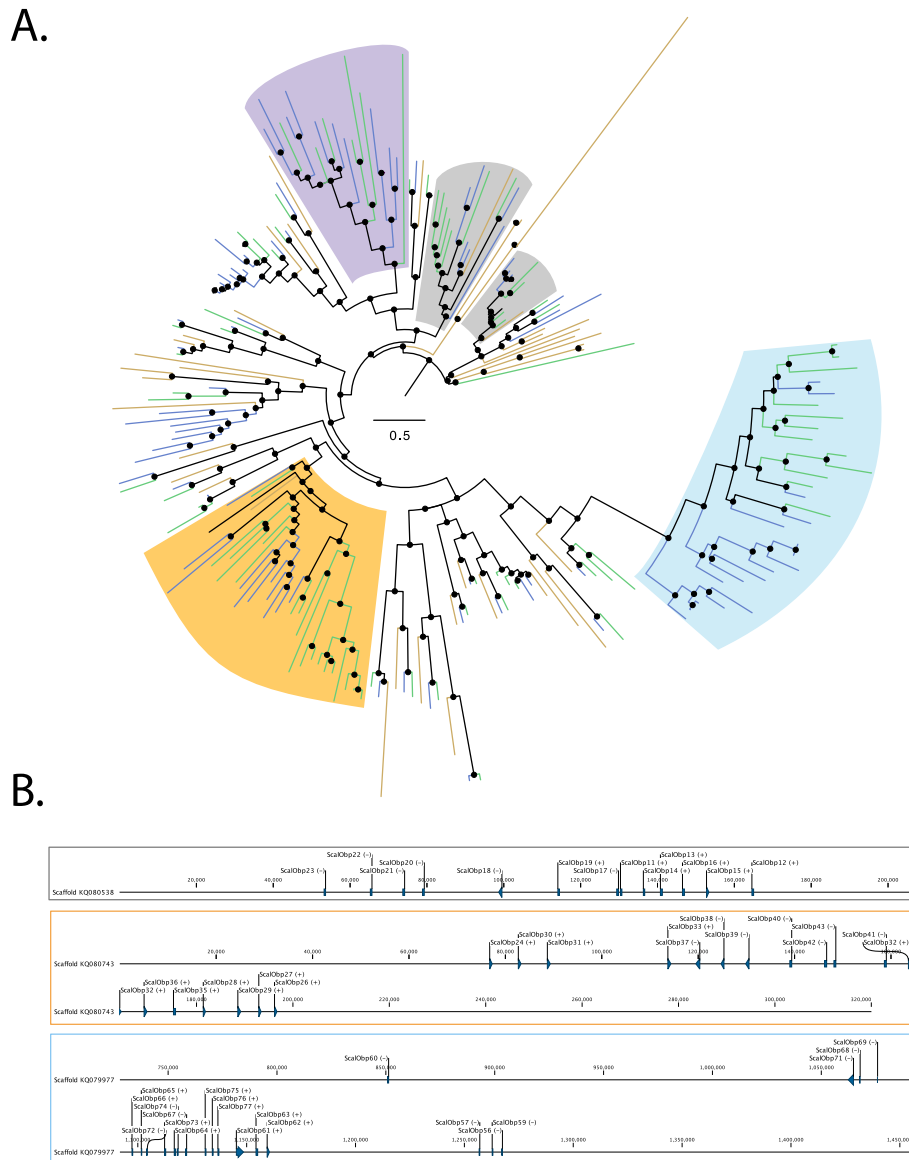
#### Gustatory and ionotropic receptor gene family expansions support importance of bitter taste perception in *Stomoxys*

Insect ecology and environment impact the size of chemosensory gene families with evidence for specialist insects having a smaller number of genes compared with generalists, with exceptions [113]. These gene families encode odorant binding proteins (OBP), carrier proteins for lipid molecules, as well as odorant (OR), gustatory (GR), and ionotropic (IR) receptors that display different

affinities for ligands that mediate the insect's response. *Glossina* are obligate blood-feeders and have a reduced number of chemoreceptors relative to more polyphagous insects that have been sequenced [41], while the *Musca* genome harbors an expanded number of chemoreceptors relative to *Drosophila* [94]. Analysis of the *S. calcitrans* genome revealed that it shares with *Musca* the presence of lineage-specific expansions and signatures of many pseudogenizations/deletions of chemosensory pathway genes relative to *Drosophila* (Figs. 2, 3 and 4, see Additional file 1: *Stomoxys* Chemosensory Gene Families [18–20, 23, 42, 94, 115–187] and Additional file 8), consistent with the birth-and-death model of evolution proposed for these gene families [188].

More than half of the 90 *S. calcitrans* OBP gene models are organized as tandem clusters across three scaffolds, consistent with OBP gene organization in other dipteran genomes [189] (Fig. 2, see Additional file 9 and Additional file 1: Figure S8). While this represents an increase in number of OBPs relative to *Drosophila*, it is comparable to the OBP gene family size in *M. domestica* [94], and it also corresponds with increases in gene family size for *S. calcitrans* chemoreceptor gene families described below. Expansions relative to and losses in *Drosophila* were evident in these clustered OBPs (Fig. 2). For example, an *S. calcitrans* cluster that resides on scaffold KQ079977 are in a lineage that includes the *Drosophila* Minus-C OBP subfamily and represent an apparent expansion of 17 *Stomoxys* and 16 *Musca* OBPs relative to *DmelObp99c*. Eighteen *Stomoxys* OBPs (*ScalObp24-43*) reside on scaffold KQ080743 and form a clade with nine *Musca* OBPs and *DmelObp56h*, depicting an expansion in this muscid lineage. Further, a separate clade comprised of seven *Stomoxys* OBPs (*ScalObp79-85*) and 11 *Musca* OBPs (*MdomObp77-87*) have no obvious *Drosophila* ortholog. Future studies to define functional roles for these OBPs are warranted, especially given the diverse tissues in which their expression was detected (see Additional file 1: Figure S8). Larval expression was detected by RNA-Seq for 35 *Obps*, while 63 and 67 *Obps* were detected in heads of mated adult males and females. However, 77 *Obps* were detected in the reproductive tract of mated adult, males and females (see Additional file 1: Figure S8), and this expression pattern supports non-chemosensory roles reported for OBPs in other dipteran species [190, 191]. For example, the transfer of OBPs from males to females in seminal fluid occurs in *Drosophila*, *Glossina*, and *Ae. aegypti* [192–194], and this may account for detection of *Obp* expression in tissues of *S. calcitrans* male and female reproductive tracts.

The OR and GR families make up the insect chemoreceptor superfamily [115, 117–119], with the OR family arising from a GR lineage at the base of the Insecta [113,



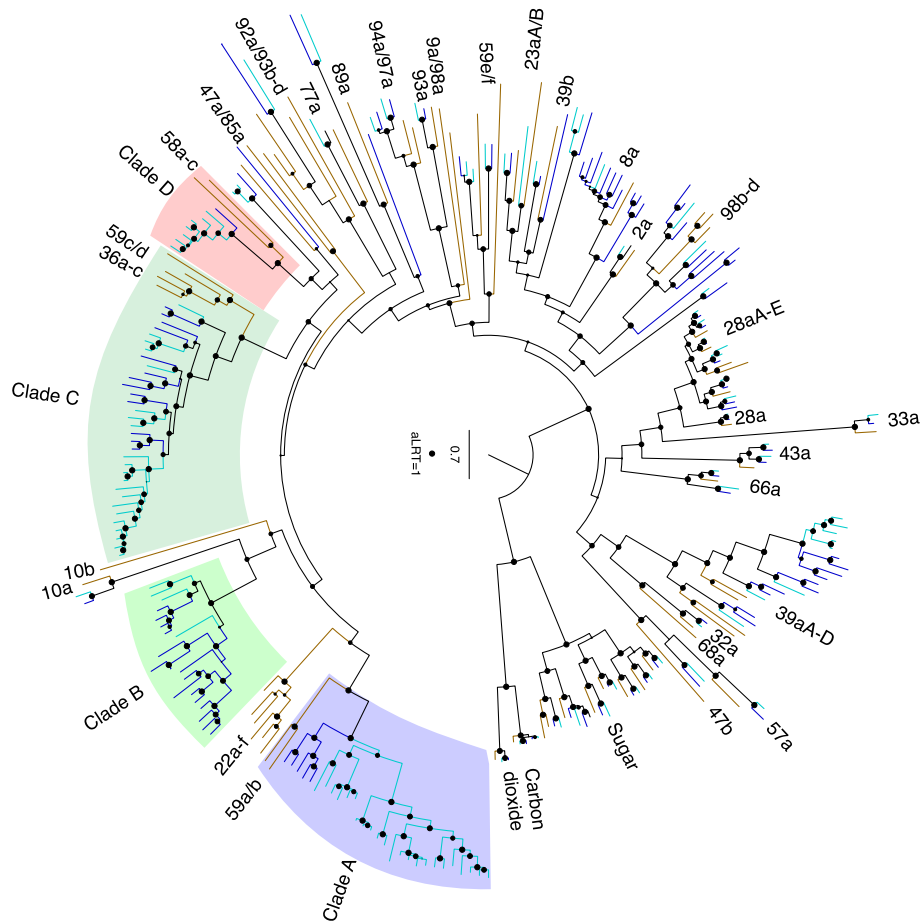
**Fig. 2** Phylogenetic tree of the *Stomoxys calcitrans* OBPs with those of *Drosophila melanogaster* and *Musca domestica* and locations of OBPs on *Stomoxys* scaffolds. **a** Maximum likelihood phylogeny was constructed using the web server version of IQ-TREE software ([114]; best-fit substitution model, branch support assessed with 1000 replicates of UFBoot bootstrap approximation). The full phylogenetic tree can be found in Additional file 1: Figure S8. The *S. calcitrans* and *M. domestica* lineages are in green and blue, respectively, while *D. melanogaster* lineages are in mustard. Clades that are expanded in the muscids relative to *Drosophila* or lost in *Drosophila* are shaded in orange and gray. Purple shading comprises a clade of muscid OBPs with no apparent ortholog in *Drosophila*, while blue shading comprises a clade that includes the *Drosophila* Minus-C OBPs subfamily and represents an apparent muscid expansion relative to *DmelObp99c*. **b** Three *Stomoxys* scaffolds on which 51 of 90 OBPs are organized. The location of each OBPs gene is indicated by a horizontal line. Transcriptional directions are indicated by (+) for same direction as the scaffold or (-) for the opposite direction. The color of the box reflects the shaded clades in **a**

195]. Olfactory sensory neurons that detect volatile compounds express ORs, while GRs are mostly involved in taste, but some have an olfactory role [116]. The GR family is generally divided into three major and divergent subfamilies of sugar or sweet receptors, the carbon dioxide receptors, and the bitter taste receptors. A lineage within the bitter taste receptor clade has evolved into an important receptor for fructose [133], and there

are others involved in courtship [123]. The ionotropic receptor family is a variant lineage of the ancient ionotropic glutamate receptor family [115, 138, 140, 169] and, like the GRs, they are involved in both olfaction and gustation, as well as sensing light, temperature, and humidity [169].

The carbon dioxide, sugar, and fructose receptors are relatively well conserved in *Stomoxys* and *Musca*, as is





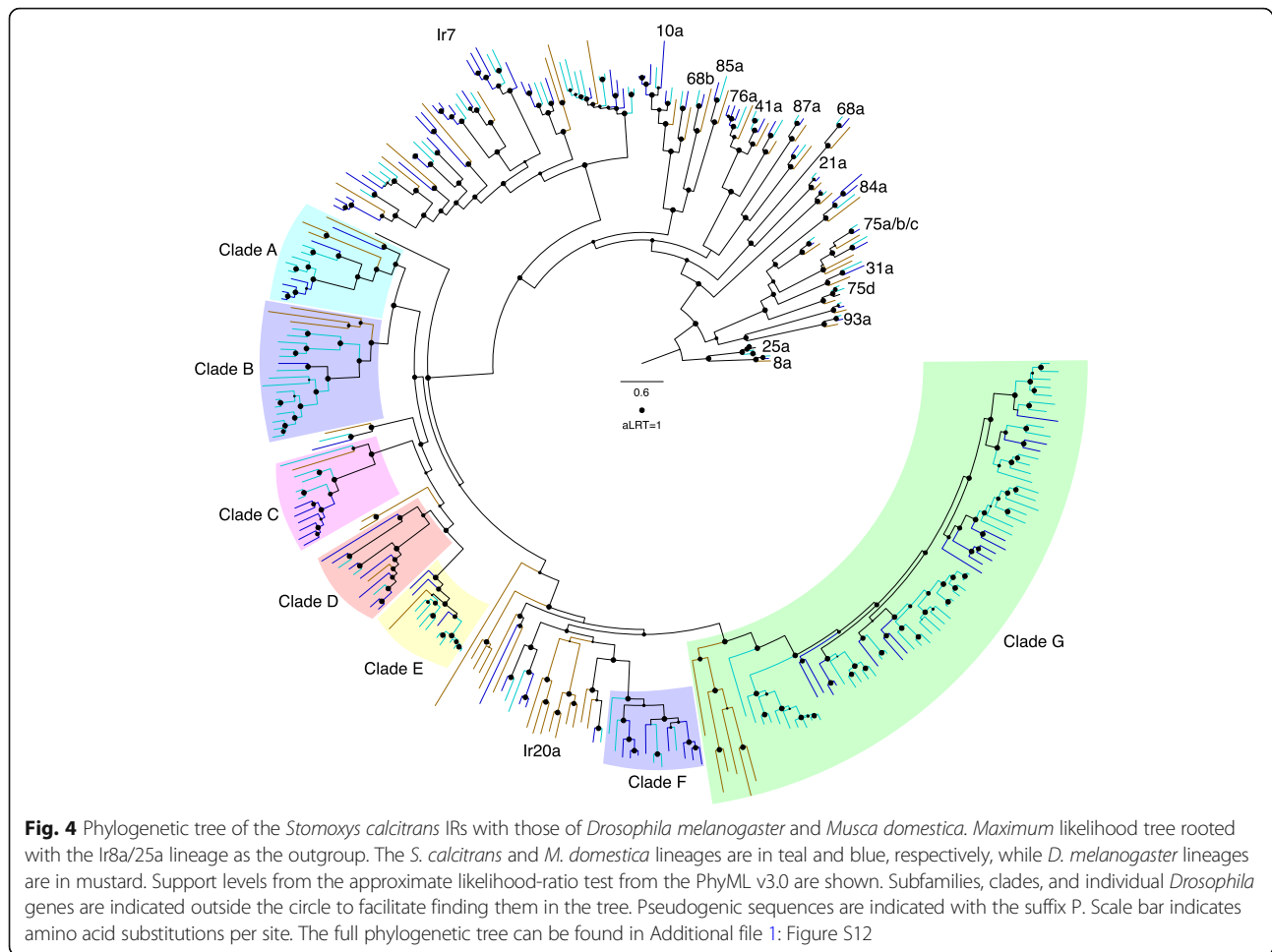
**Fig. 3** Phylogenetic tree of the *Stomoxys calcitrans* GRs with those of *Drosophila melanogaster* and *Musca domestica*. Maximum likelihood tree rooted by divergent carbon dioxide and sugar receptor subfamilies as the outgroup. The *S. calcitrans* and *M. domestica* lineages are highlighted in teal and blue, respectively, while *D. melanogaster* lineages are in mustard. Support levels from the approximate likelihood-ratio test (aLRT) from the PhyML v3.0 web server are shown. Subfamilies and individual or clustered *Drosophila* genes are indicated outside the circle to facilitate finding them in the tree. Four clades of candidate bitter receptors that are expanded in the muscids are highlighted. Scale bar indicates amino acid substitutions per site. The full phylogenetic tree can be found in Additional file 1: Figure S11

the case for many other insects. However, the bitter taste receptors reveal considerable gene family evolution both with respect to the available relatives of these muscid flies (*Drosophila* and the Mediterranean fruit fly, *Ceratitis capitata*) and between these two muscids. For example, a major expansion (*ScalGr29-57*) encoding 49 candidate bitter taste receptors occurs in *S. calcitrans*, comparable to a similarly complicated set in *M. domestica* (*MdomGr43-64*, encoding 35 proteins). Together, these form four major expanded clades in the muscids (Clades A–D, Fig. 3).

In *Drosophila* and most other insects examined to date, ionotropic receptors Ir8a and 25a, both of which function as co-receptors with other IRs [169], are highly conserved both in sequence and length and in being phylogenetically most closely related to the ionotropic glutamate receptors [138, 140]. While *Stomoxys* has the expected single conserved ortholog of Ir8a, surprisingly

it has four paralogs of Ir25a (*Scallr25a1-4*), the functions of which are enigmatic, as such duplications of Ir25a have seldom been observed in other insects (Fig. 4).

The *Ir7a-g* and *11a* genes in *Drosophila* are expressed in larval and adult gustatory organs [140], but ligands for these receptors are unknown. This subfamily is considerably expanded in both *S. calcitrans* and *M. domestica* and, given their complexities, they are not named for their *Drosophila* relatives. Rather, these are part of the numbered series from Ir101, in the case of *Stomoxys* to Ir121 and in *Musca* to Ir126 (Fig. 4). These IR gene family expansions strongly suggest an expanded gustatory capacity. A large clade of “divergent” IRs in *Drosophila* is involved in gustation and is known as the Ir20a subfamily of 33 proteins [165, 170, 171]. This clade of mostly intronless genes is considerably expanded in *M. domestica* to 53 members (*MdomIr127-179*), and even more so in *S. calcitrans* to 96 members (*Scallr122-217*),



labeled clades A-G in the tree (Fig. 4). Putative IR ligands in *Drosophila* are involved in sensing carbonation, amino acids, and specific carbohydrates [159, 162, 170].

Transcript detection by RNA-Seq provided some insight into those *Stomoxys* tissues expressing chemosensory receptor gene family members. Seventeen *Or* genes were enriched in the reproductive tract of mated males, and these may have a role in sperm activation, as proposed for *Anopheles gambiae* [196]. Interestingly, three *Ors* (*ScalOr22*, 54, and 55) were highly enriched in the reproductive tract of mated females relative to all other tissues examined (see Additional file 1: Figure S10 and Additional file 9), suggesting these *Ors* may have a role in female reproduction, possibly perceiving male derived chemicals transferred during copulation. Eleven *Ors* were detected in third instar larvae, and pilot evaluation by non-quantitative RT-PCR detected an additional nine *Ors* expressed in first and second instar but not in third instar larvae (see Additional file 9). This suggests that stable flies differentially utilize odorant receptors throughout immature development. Expression of all 20 of these *Ors* was not exclusive to the larval

stages, and the apparent absence of larval-specific receptors in the stable fly may be a result of exposure to related compounds during the immature and adult stages (e.g., host dung, detritus).

The expression of 35 *Gr* transcripts was detected by RNA-Seq in heads of mated females and males, and 27 *Gr*s were enriched in the reproductive tracts of mated males (see Additional file 1: Fig. S11 and Additional file 9). Evidence from *Drosophila* supports the expression of GRs in neurons that innervate testes and oviducts [146], suggesting that these *S. calcitrans* GRs may have a role in mediating reproductive system function. Twenty-three *Gr*s were detected in larvae, all of which clustered as candidate bitter taste receptors, suggesting they may mediate larval bitter sensing. Determination of the ligand specificities of these muscid receptors is required to fully understand the ecological significance of the differential expansions and contractions of their bitter taste abilities.

Twenty-three *Irs* were detected by RNA-Seq in heads of mated females and males of which two and seven were enriched in the female and male tissue, respectively. Interestingly, five *Irs* were detected in the female

reproductive tract while 56 *Irs* were detected in the male reproductive tract with 46 highly enriched in this male tissue (see Additional file 1: Figure S12 and Additional file 9), suggesting a potential critical role in fly reproduction or reproductive behaviors. Given the striking expansions and expression pattern of genes within this family, functional studies in *S. calcitrans* are warranted. Further characterization of these chemosensory gene families will facilitate the design of behavior-based control technologies that can be employed as part of integrated pest management strategies.

**Expansion of the long wavelength-sensitive Rh1 opsin subfamily in *Stomoxys* with evidence of tuning for diversified wavelength sensitivities via substitution**

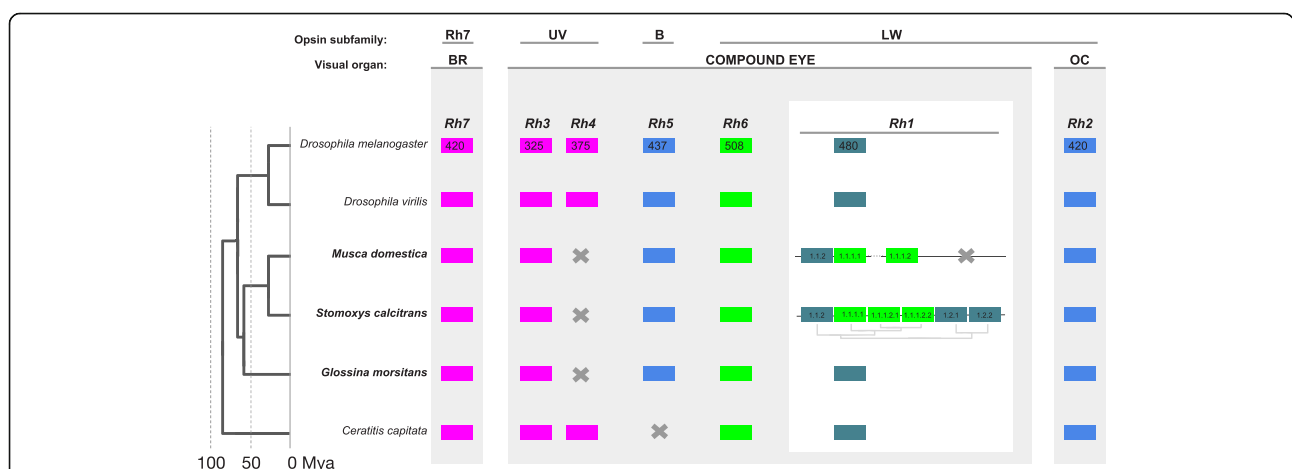
Visual cues are integral to stable fly mating and host orientation [35], and stable fly attraction to particular wavelengths of light and UV reflectance has been manipulated for the development of trapping technologies to suppress populations [197]. As is typical for the generally fast flying calyptrate flies, stable fly adults of both sexes are equipped with large laterally positioned compound eyes in the head and three ocelli positioned in the dorsal head cuticle [198, 199]. Both achromatic motion tracking and color-specific perception tasks begin with the harvest of photons by members of the opsin class of G-protein coupled transmembrane receptors, which differ in their wavelength absorption optima. The genomic survey in the stable fly revealed conservation of most opsin gene subfamilies observed in *Drosophila* (Fig. 5; see Additional file 1: Figure S13 and Table S11). This included the UV sensitive opsin paralog *Rh3*, the blue

sensitive opsin *Rh5*, several homologs of the long wavelength (LW) sensitive opsin *Rh1* and a 1:1 ortholog of LW opsin *Rh6*, all of which are expressed in subsets of photoreceptors in the compound eye retina [204]. In addition, we found 1:1 orthologs of the ocellus-specific opsin *Rh2* and of the recently characterized UV-sensitive, deep brain opsin *Rh7* [205] (Fig. 5; see Additional file 1: Figure S13 and Table S11). Overall, these findings are consistent with the electrophysiological and positive phototactic sensitivity of *S. calcitrans* to light in the UV, blue, and green range of visible light [33, 206].

*Drosophila* and other higher Diptera, including *C. capitata*, possess a second UV sensitive opsin gene *Rh4* (Fig. 5) [42, 207], which is not detected in the stable fly genome. The same result was previously obtained in the tsetse fly [41]. Global BLAST searches of calyptrate genomes were conducted (i.e. *G. morsitans*, *M. domestica*, the black blowfly (*Phormia regina*), and the flesh fly (*Sarcophaga bullata*)), and they failed to detect *Rh4* orthologs. Thus, it can be concluded that the *Rh4* opsin subfamily was lost during early calyptrate evolution.

**An *Rh1* opsin gene cluster in muscid Diptera**

A unique aspect of the *S. calcitrans* opsin gene repertoire is the existence of six homologs of the LW opsin *Rh1* (Fig. 5; see Additional file 1: Figure S14, Table S11). Most higher Diptera sampled so far, including related species like the tsetse fly [41] and the black blowfly [208], possess a singleton *Rh1* gene. Three *Rh1* homologs, however, were detected in the *M. domestica* draft genome [94]. Moreover, in both *S. calcitrans* and *M. domestica*, these *Rh1* homologs are closely linked and



**Fig. 5** Opsin gene family members detected in *Stomoxys calcitrans* and other higher Diptera. A phylogenetic tree of dipteran opsin gene relationships is presented, as is the genomic organization and evolution of the *S. calcitrans* Rh1 opsin subfamily [200]. Protein sequences of *S. calcitrans* Rh1 genes were aligned with MUSCLE [201], and ambiguous alignment regions were filtered with Gblocks [202] using least stringent settings. Maximum likelihood tree was estimated in MEGA version 6.0 [203] applying the Jones-Taylor-Thornton (JTT) model of amino acid sequence evolution and assuming Gamma Distributed substitution rates across sites with 3 categories. Rhodomer opsins depicted: Rh7, Rh7 gene; UV, UV-sensitive; B, Blue-sensitive; LW, long wavelength-sensitive expressed in BR, compound eye, and OC, ocelli

anchored as a cluster by homologous flanking genes (Fig. 5; see Additional file 1: Figure S14). This suggests that the Rh1 tandem gene clusters of the two species are homologous and date back to an ancestral cluster in the last common ancestor of muscid Calyptratae. Consistent with this, the *S. calcitrans* and *M. domestica* Rh1 homologs form a monophyletic unit in maximum likelihood trees estimated from amino acid or nucleotide sequence alignments of dipteran Rh1 homologs (see Additional file 1: Figure S14). Moreover, each of the three *Musca* Rh1 homologs grouped with strong support as 1:N orthologs with different members of the *Stomoxys* Rh1 gene cluster; however, two of the *Stomoxys* Rh1 genes (Rh1.2.1 and Rh1.2.2) lack *Musca* orthologs (see Additional file 1: Figure S14). Integrating the information on gene linkage, it is possible to conclude that the six Rh1 paralogs of *Stomoxys* originated by three early duplications before separating from the *Musca* lineage. While the latter subsequently lost one of the two earliest paralogs, the *Stomoxys* Rh1 cluster continued to expand by minimally one but possibly two subsequent tandem gene duplications (Fig. 5).

As expected, RNA-Seq results indicated that transcripts for all *Stomoxys* opsin genes, quantified by transcripts per million (TPM), were more abundant in the head relative to whole teneral adults of both sexes or male reproductive tracts (see Additional file 1: Table S11). The singleton opsin Rh1 from *Drosophila* is expressed in six outer photoreceptors (R1-6) that are found within each ommatidium. These receptors are specialized for motion detection. Opsins Rh5 and Rh6 are differentially expressed in a single color-vision specialized photoreceptor (R8) that are also within each ommatidium [204]. In the *Drosophila* modENCODE expression catalog, this is reflected as an up to 200 fold higher transcript abundance of Rh1 opsin compared to Rh5 or Rh6 in the adult heads of both sexes [209]. The *S. calcitrans* RNA-Seq data derived from head tissue provided evidence that a single member of the *S. calcitrans* Rh1 paralog cluster, named Rh1.1.1.1, likely maintains the ancestral function of Rh1 as the major motion detection specific opsin. This is based on comparison of its TPM values with that of the other Rh1 paralogs, as well as with those of the putative *S. calcitrans* opsins Rh5 and Rh6 (see Additional file 1: Table S11). The remaining *S. calcitrans* Rh1 cluster member genes showed relatively low to very low expression values (see Additional file 1: Table S11).

#### **An amino acid modification at a tuning site differentiates two muscid Rh1 paralog subclusters**

While exceptional for other higher Diptera, tandem duplicated LW opsin gene clusters have been found in mosquito and water strider species [210, 211]. In both

cases, evidence of functional paralog diversification has been detected in the form of amino acid changes that affect opsin wavelength sensitivity (i.e., at tuning sites). Integrating data from butterflies and *Drosophila*, the water strider study identified one high confidence tuning site that very likely affects the blue vs green range wavelength specificity in LW opsins. The site is residue 17 based on the numbering system developed for butterflies, which corresponds to residue 57 in *Drosophila* Rh1 [211–213]. Based on this criterion, the three oldest *S. calcitrans* Rh1 gene cluster paralogs preserve the blue-shifted wavelength specificity ( $\lambda_{\max}$  480 nm in *Drosophila*) of the Rh1 singleton homologs of other dipteran species due to conservation of the ancestral methionine state at tuning site 17 (see Additional file 1: Figure S15). In contrast, the three younger *S. calcitrans* Rh1 paralogs share a leucine residue at tuning site 17, which is extremely rare across insect LW opsins. In a survey of over 100 insect LW opsins, it was detected only in the two corresponding Rh1 orthologs from *M. domestica* in addition to one in the distantly related species of thrips (Thysanoptera) [211]. The physicochemical similarity of the leucine at tuning site 17 in the three youngest *S. calcitrans* Rh1 paralogs relative to the pervasively conserved isoleucine residue at tuning site 17 in the green-sensitive Rh6 opsins ( $\lambda_{\max}$  515 nm in *Drosophila*) represents compelling evidence that this shared derived replacement substitution defines a green-sensitive subcluster in the *S. calcitrans* Rh1 paralog group (see Additional file 1: Fig. S15).

#### **Unique duplications in the *Stomoxys* yolk protein gene family**

Cyclorrhaphan flies such as *Stomoxys* [214], *Drosophila* [215], *Musca* [216], *Calliphora* [217], and *Glossina* [218] utilize yolk proteins (YPs) as a primary source of nutrients for developing embryos. While many insects utilize YPs classified as vitellogenins, Cyclorrhaphan flies utilize an alternative class of proteins derived from lipase enzymes [219]. These proteins function as a source of amino acids and also as transporters of essential nutrients such as lipids and vitamins [220]. The number of yolk protein genes varies among higher fly species. The species-specific expansions/contractions observed within this class of genes may reflect reproductive demand within those species. Analysis of the *Stomoxys* genome identified eight putative *S. calcitrans* YP homologs relative to seven YP gene family members in *M. domestica* [94]; four of the seven *M. domestica* genes were annotated as part of this study.

Prediction of the evolutionary relationships between the predicted YPs from *S. calcitrans*, *M. domestica*, *G. morsitans*, and *D. melanogaster* by phylogenetic analysis (Fig. 6) suggests the yolk protein gene family expanded

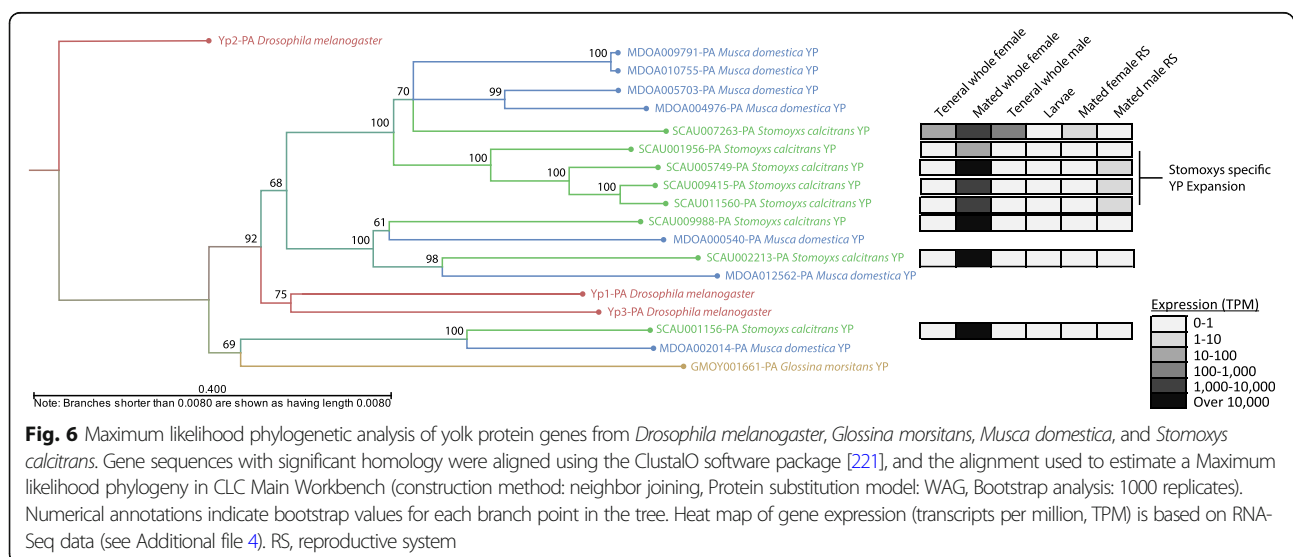
in *S. calcitrans* and *M. domestica* sometime after their divergence from *Drosophila*, which has three YP gene family members [222]. Of the *Stomoxys* and *Musca* specific YPs, three members are orthologous between the two species suggesting derivation from a common ancestor. However, the remaining yolk protein genes are paralogous and may originate from independent duplication events occurring after the divergence of the *Stomoxys* and *Musca* lineages. An alternative explanation is that these genes were originally orthologs, but have been altered via gene conversion resulting from homology based DNA repair mechanisms [223]. The lineage-specific expansions suggest that duplications in this gene family may confer a reproductive advantage by increasing reproductive capacity. In support of this role, all eight YP genes were detected by RNA-Seq in reproductively active *S. calcitrans* females, but not teneral females. Further, expression was not detected by RNA-Seq analysis of the female reproductive tract, which suggests these genes are expressed and translated in the fat body, secreted into the hemolymph and transported to the ovaries as observed in other higher flies (Fig. 6).

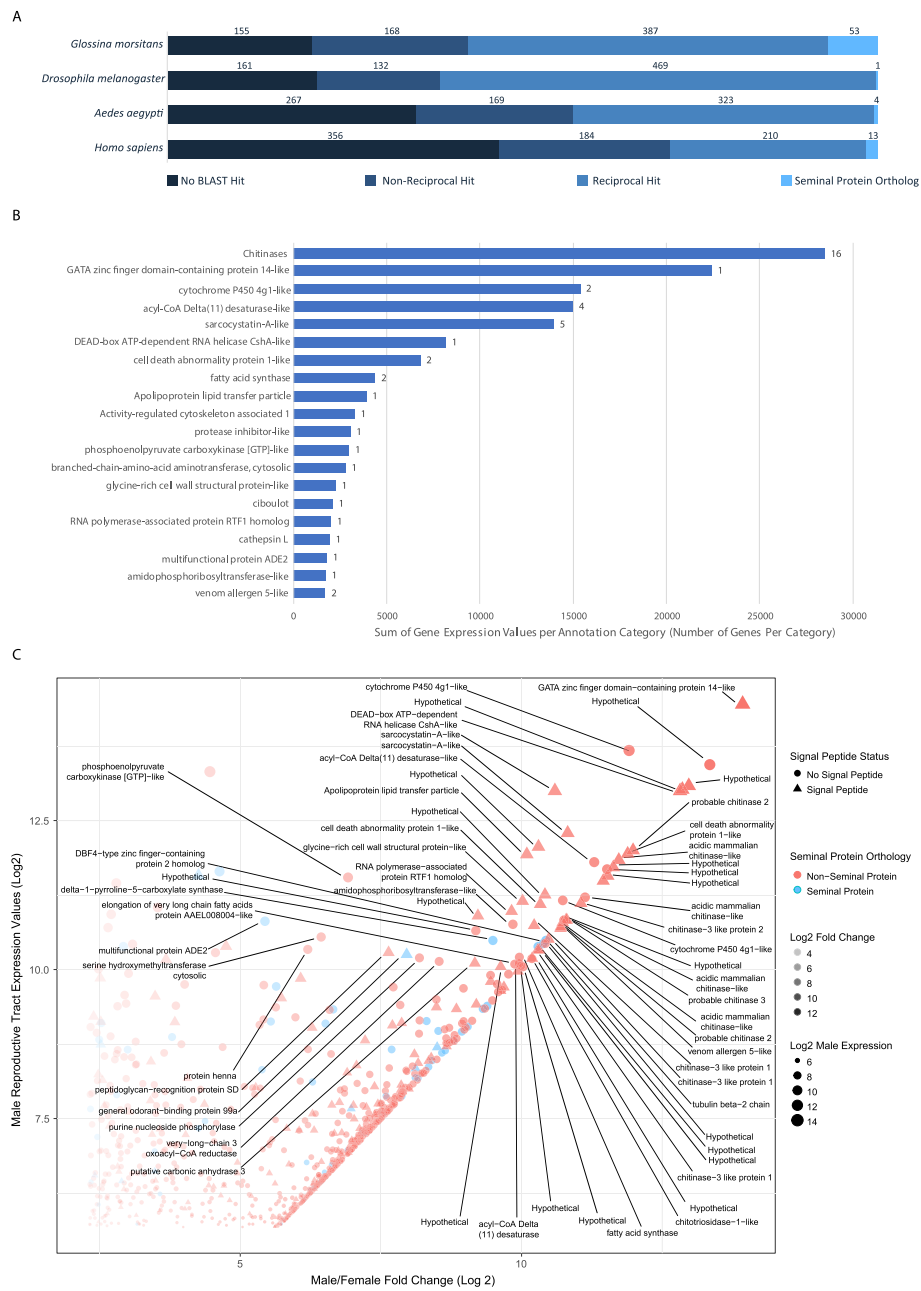
#### Evidence of *Stomoxys* male-biased reproductive tract genes with putative seminal fluid function

Insect seminal fluid proteins are produced by the male reproductive tract and are transferred to females during mating. The seminal fluid proteins induce a post-mating response in females that results in behavioral and physiological changes including refractoriness to additional matings. The *S. calcitrans* reproductive tract is comprised of testes, vas deferens, and an ejaculatory duct, the latter of which has a region that produces the seminal fluid and serves an accessory gland function [224]. While Meola [224] observed storage of

proteinaceous accessory gland material in the *S. calcitrans* ejaculatory duct, its composition has not been described. To identify male-biased reproductive tract genes encoding potential seminal proteins, RNA-Seq data derived from *S. calcitrans* male and female reproductive tract tissues were compared (Fig. 7). Genes in the male reproductive tract dataset were filtered to identify those expressed at least five-fold higher than in the female reproductive tract dataset. This analysis resulted in the classification of 763 genes with male reproductive tract-biased expression (see Additional file 10).

A reciprocal BLAST analysis identified orthologs of the male reproductive tract-biased genes in other species in which seminal proteins are characterized. These include *G. morsitans* [194], *D. melanogaster* [192, 225], *A. aegypti* [226], and *Homo sapiens* [227] (Fig. 7a). The overall number of identified orthologs corresponds roughly to the evolutionary distances between the species tested. However, these relationships did not hold for the number of gene orthologs associated with seminal function. Reciprocal analysis with *Drosophila* identified 469 1:1 orthologs of male reproductive tract-biased *S. calcitrans* genes, amounting to the largest number of orthologs identified between species included in this analysis. In contrast, of those 469 orthologs only one is associated with seminal fluid function in *D. melanogaster*. Comparison with *G. morsitans* identified the second highest number of orthologous proteins ( $n = 387$ ). Of those, 53 were associated with seminal function, suggesting a greater similarity in the constitution of seminal secretions between *Glossina* and *Stomoxys* consistent with their closer phylogenetic relationship compared to *Drosophila*. Of note, none of the *S. calcitrans* male reproductive tract-biased proteins were orthologous to seminal proteins across all four species. In *Drosophila*,





**Fig. 7** Analysis of male reproductive biased genes in *Stomoxys calcitrans*. **a** Results of reciprocal BLAST analysis of male reproductive tract biased genes. **b** Expression analysis of top 20 most abundant gene classes in the male reproductive tract RNA-Seq dataset versus the female reproductive tract, as annotated by BLAST best hits. Bar length represents combined expression values in TPM for the genes included in that category. Numbers associated with the bars represent the number of genes in that functional classification that had a male reproductive tract-biased expression. **c** Scatter plot of the 763 *Stomoxys* male reproductive biased genes. The plot shows on the x-axis - log<sub>2</sub> fold change expression in males relative to females and the y-axis represents the log<sub>2</sub> transformed expression value in TPM in the male reproductive tissue. Triangular points are genes predicted to contain signal peptides and blue points are genes with orthology to seminal proteins in other species. Genes with a log<sub>2</sub> expression value above 10 and log<sub>2</sub> Male/female fold change value above 5 are annotated with putative functional descriptions

male-biased genes evolve at a faster rate, especially those expressed in reproductive tissues. These genes tend to lack identifiable orthologs relative to genes expressed in an unbiased pattern [228, 229]. There is evidence for this in *M. domestica* as well [94], and this phenomena is

likely due to sexual selective pressure resulting in rapid evolution of male-biased genes [229]. However, one gene within this group encoding a catalase (XM\_013259723) is orthologous to seminal protein genes in *Aedes*, *Glossina*, and *H. sapiens*. In *S. calcitrans*, two catalase genes

were identified on separate scaffolds, one for which expression was detected in all life stages and tissues analyzed (XM\_013257324) and the other, described above, that was detected primarily in the male reproductive tract. As catalases function to reduce oxidative stress, this finding could reflect a conserved mechanism that protects the sperm from oxidative damage.

The 763 *S. calcitrans* male reproductive tract-biased genes were annotated by BLAST and gene ontology analysis and then categorized by best hit annotation. Of those genes, 216 lacked significant BLAST hits or were homologous to hypothetical proteins with no functional associations. Of the remaining genes that had significant hits to annotated proteins, certain categories were highly expressed in terms of both the number of genes and the relative level of expression within the male reproductive tract (Fig. 7b, c). These genes were also tested for the presence of a secretion signal peptide within the predicted open reading frame to differentiate between secreted and non-secreted proteins. Approximately 22% of the male reproductive tract-biased genes (165/763) were found to include a predicted signal peptide. Gene ontology analysis of this group's constituents revealed several significant functional enriched categories (see Additional file 10). The most highly significant groups included chitin binding, serine/cysteine-type endopeptidase inhibitors, metalloendopeptidases, antibacterial humoral response, and innate immune response among others. While the 1:1 orthologs to most of these genes do not act as seminal proteins in other organisms, they represent generally conserved functions in seminal proteins. Functions such as protease inhibition, immunity, protease activity and chitin binding have been characterized in the seminal secretions of other flies and insects [194, 226, 230–233].

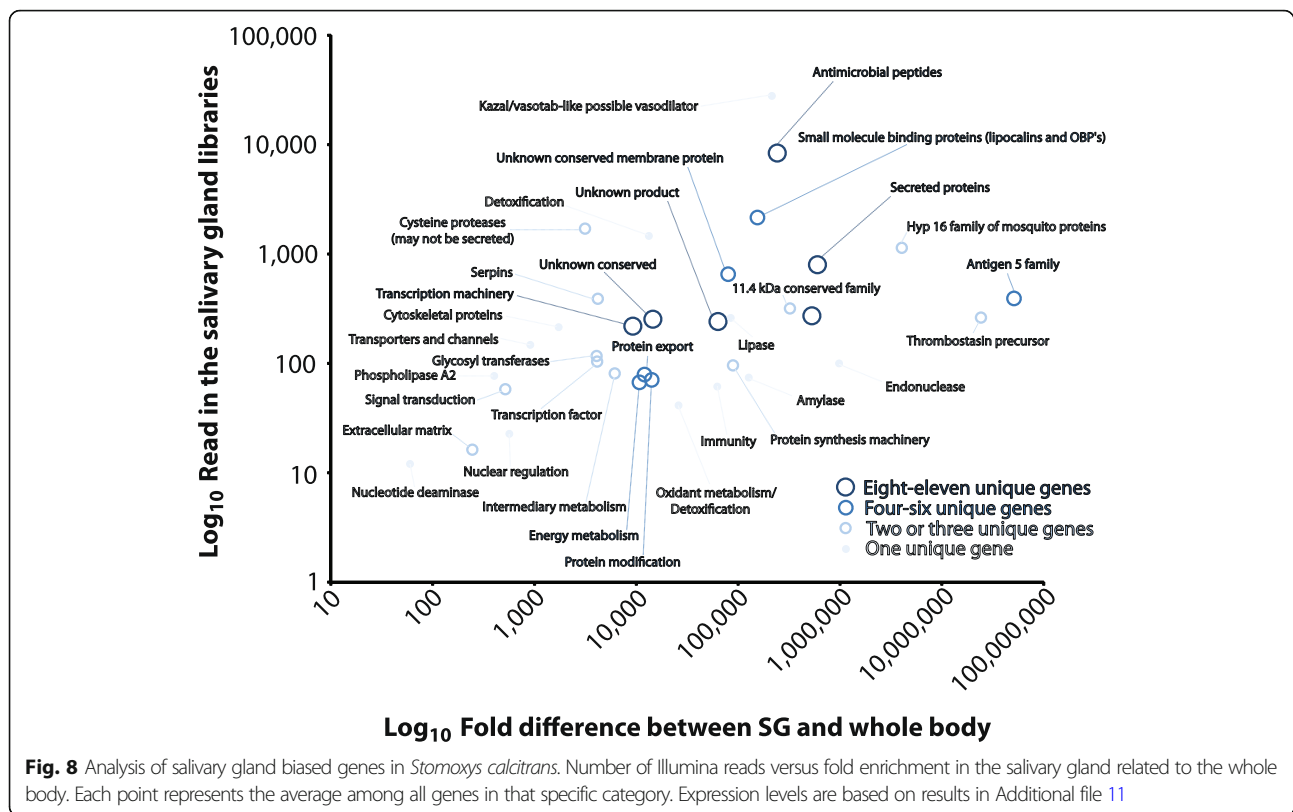
Chitinases represented the most highly expressed functional group within the transcriptome and comprise 16 chitinase-like genes, of which 12 are clustered on scaffolds KQ079939 (seven genes) and KQ080089 (five genes). Of the approximately 37 genes annotated as chitinase in the *S. calcitrans* genome, the RNA-Seq results suggest these 16 are biased towards the male reproductive tract tissue (see Additional file 10). Chitinases confer anti-fungal activity in honey bee seminal secretions, preventing the transfer of pathogenic spores during copulation [234]. Although it is unclear if these have the same role, such antimicrobial properties would be beneficial to *Stomoxys* given the high probability of exposure to fungi in the moist and microbe rich substrates in which females oviposit. The second most highly expressed category consists of a single gene, XM\_013245551, which is the most highly expressed gene in the male reproductive tract dataset. While it is annotated as a GATA zinc-finger domain containing protein, further analysis reveals little in the way of conserved domains to indicate its

function. Interestingly, “domesticated” transposable elements tend to have a number of zinc-finger domains [235] and further studies are needed to evaluate whether this transcript may actually be a highly expressed transposable element. This analysis provides some insight into genetic associations with male reproductive functions in *Stomoxys* and further highlights several interesting targets for functional analysis in the future.

#### **Immunomodulatory and anti-hemostatic products are prominent in the *Stomoxys* sialome**

Blood-feeding insects salivate while probing their host skin for a blood meal. Development of a sophisticated salivary potion that disarms their hosts' hemostasis is among the adaptations to blood feeding found in hematophagous animals [236, 237]. Blood clotting inhibitors, anti-platelet compounds, vasodilators, and immunomodulators are found in salivary gland homogenates or saliva of blood sucking arthropods [237]. To determine the genes associated with salivation, transcripts from male and female salivary glands (SG) were compared with those from teneral male and female whole bodies (WB). To be consistent with salivary gland transcriptome analyses completed in other blood-feeding arthropods, an  $X^2$  test was employed to identify those that were significantly over-expressed in salivary glands (Fig. 8), as in [238] (see Additional file 11). A subset of SG transcripts with 100-fold higher expression compared with teneral adults was analyzed in more detail (see Additional file 11). The SG 100-fold overexpressed set was comprised of 139 transcripts, 18 of which were found to be splice variants, or identical to other transcripts, as verified by their scaffold coordinates. The non-redundant set comprised of 121 transcripts was classified into three major groups: Putative Secreted, Putative Housekeeping, and Unknown; these groups were further classified into finer functional categories (see Additional file 11).

In congruence with *S. calcitrans* SG polypeptides previously sequenced from one-dimensional protein gel fragments [239], the antigen 5 family comprised 62% of the total reads mapped to the 121 SG-enriched transcripts (Fig. 8). Members of this family in *S. calcitrans* may function as inhibitors of the classical complement system [240]. Thrombostasin [241] members, which are precursors for anti-thrombin peptides previously identified in *S. calcitrans*, were represented by two transcripts. They accrued 29% of the reads and are strongly represented in the protein study [239]. The Hyp 16 family of peptides (unknown function, 4.9% of accrued reads) and one transcript encoding an endonuclease (1.2% of accrued reads) were also noted. Together, these groups of transcripts account for 97% of the reads that are over



expressed in the salivary glands relative to the whole body of *S. calcitrans*.

There was a wide variety of other transcripts represented in the last 3% of the reads, and serine proteases, nucleotide deaminase, amylase, phospholipase A2, and lipases were found enriched in the *S. calcitrans* salivary gland transcriptome. These enzymes are also found enriched in other sialomes and their functions have been reviewed [237, 242, 243]. Two of eight serine proteases were found 5–15 times overexpressed in female salivary glands when compared to male glands. These two products produce best matches to vitellin-degrading proteases from *M. domestica* (XM\_005191887.2) and may be indeed female enriched enzymes that were hitchhiked to the salivary set due to their similarities to overexpressed salivary enzymes. No other peptides were found above fivefold expressed in either salivary gland gender.

Several antimicrobial peptides appeared enriched in the *S. calcitrans* sialome, including lysozyme, attacins, defensins, dipterin, a GGY rich peptide, and sarcotoxin. Of these, only the GGY peptide and dipterin were identified in the previously reported Sanger-based *S. calcitrans* sialotranscriptome [239]. Given that a bloodmeal can be stored in the stable fly midgut for up to 48 h [244], these peptides may be a first line of defense to control microbial growth in the ingested meal. Regarding polypeptides with anti-proteolytic

activity, in addition to thrombostasin precursors discussed above, two transcripts encode serpins (however, with very low expression) and one encodes a Kazal domain-containing peptide (accruing 0.3% of the reads). While serpins may modulate clotting and inflammation-related proteases, the Kazal domain peptide may be related in function to vasotab, a vasodilatory peptide from a tabanid fly [245].

Finally, 24 transcripts accruing 0.19% of the reads could not be functionally classified and were thus assigned to the “unknown” class. These include membrane proteins XP\_013114823.1 and XP\_013117270.1 that are over one thousand-fold over expressed in the salivary glands relative to whole body, the former of which was identified in the Sanger sialotranscriptome. Given their abundance in salivary gland tissue, these are attractive targets for gene disruption experiments to elucidate the contribution of these proteins to the salivary function of *S. calcitrans*.

#### Expanded cytochrome P450 gene family suggests enhanced metabolic detoxification in *Stomoxys*

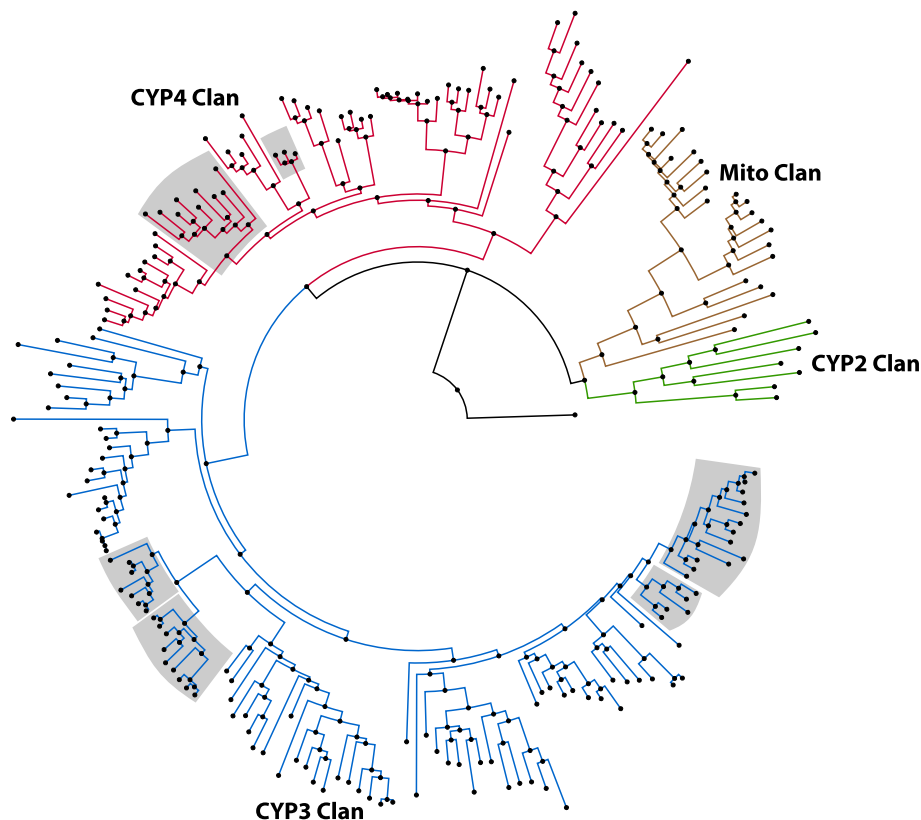
Various detoxification mechanisms have evolved in insects to enable their survival upon exposure to environmental toxins. Metabolic detoxification is mediated by members of the carboxylesterase and glutathione-S-transferase gene families (identified and described from



*S. calcitrans*; see Additional file 1: Gene families associated with metabolic detoxification, Figures S16 - S19 [246–263] and Additional file 12), as well as the cytochrome P450 (CYP) gene family. Arthropod CYPs have diverse roles in insect physiology, including ecdysteroid biosynthesis and xenobiotic detoxification [264, 265]. The CYP gene family size varies among insects, with dipterans having large arrays (i.e., 145 in *Musca*, 86 in *Drosophila*, and 77 in *Glossina*). The 214 *S. calcitrans* CYPs that were identified from genome analysis encode representatives from each of the CYP clans that are typically found in insects (i.e., mitochondrial, CYP2, CYP3, and CYP4), and they represent a substantial increase in number relative to other sequenced Dipteran genomes (Fig. 9; see Additional files 13 and 14). The *S. calcitrans* mitochondrial and CYP2 clans contain orthologs of the Halloween genes that mediate the pathway for ecdysteroid biosynthesis, namely *Cyp306A1*, *Cyp302A1*, *Cyp314A1*, and *Cyp315A1*, and the mitochondrial clan (21 genes) contains tandemly arranged genes along three scaffolds: nine CYP12A genes on KQ080363 and seven CYP12G

genes on KQ080439 and KQ082110. As in *M. domestica*, expansions in *S. calcitrans* were primarily observed in clans 3 and 4. The CYP4 clan (62 genes) was represented by the CYP4 (51 genes) family, while the CYP3 clan (107 genes) comprised the largest increase in number of CYPs in *Stomoxys*, predominated by the CYP6 (81 genes) and CYP9 (16 genes) families. Together, members of the CYP4 and CYP6 families represent 62% of the *S. calcitrans* CYPs, which is comparable to *M. domestica* [94].

At least three families of tandemly arranged CYP genes were present in the *S. calcitrans* genome (Fig. 9, shaded) and illustrate duplication and loss processes in this superfamily. The *S. calcitrans* and *M. domestica* CYP9F genes encode 15 and 4 proteins, respectively, and they are tandemly arranged on *Stomoxys* scaffold KQ080085 and *Musca* scaffold KB855374 with micro-synteny of flanking genes DNA ligase 3, fatty acyl co-A reductase, and a cluster of glutathione-S-transferase delta genes (see Additional file 1: Figure S20). A similar arrangement is found for the two *D. melanogaster* CYP9F proteins on chromosome 3R, and phylogenetic



**Fig. 9** Phylogenetic analysis of cytochrome P450 genes from *Stomoxys calcitrans*. Amino acid sequences from each family were aligned with the MUSCLE algorithm [201], and the alignments trimmed with the trimAl tool using the `-strictplus` option [266]. The trimmed alignment was used to construct a maximum likelihood phylogeny, rooted with *Mus musculus* CYP51 as the outgroup, with the web server version of IQ-TREE software (best-fit substitution model, branch support assessed with 1000 replicates of UFBoot bootstrap approximation; bootstrap percentages reported [114]). The CYP clades are identified by different colored lineages, and CYP gene clusters that are found in tandem within the genome are shaded in gray. P450 gene names were assigned based on comparative analyses (see Additional file 13), and the full phylogenetic tree can be found in Additional file 1: Figure S21

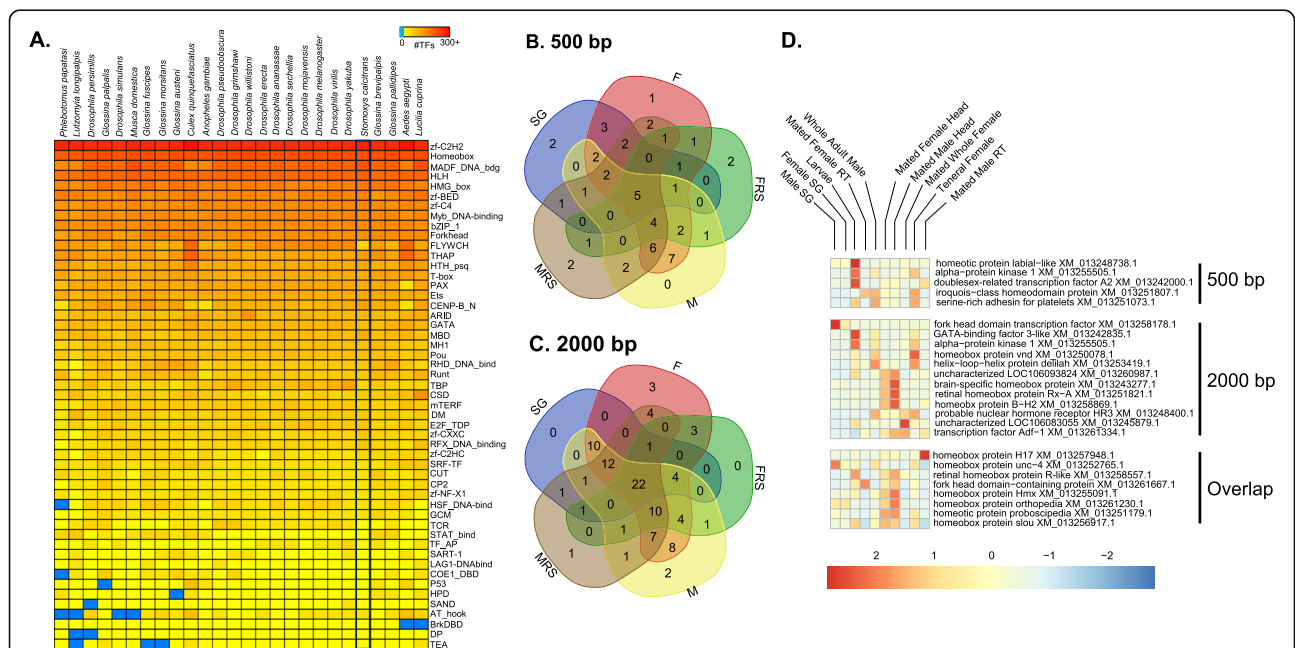
comparison suggests an expansion in *Stomoxys* and *Musca* that was lost in *Drosophila* (see Additional file 1: Figure S20-A). Genes encoding 24 CYP6A proteins in *S. calcitrans* are clustered on two scaffolds (KQ080770 and KQ080111) that are likely on one array in the genome, and a comparable cluster of genes encoding 19 CYP6A proteins on two *M. domestica* scaffolds (KB855857 and KB859418) were identified. Phylogenetic comparison with the 13 CYP6A *D. melanogaster* genes depicted expansion in *Stomoxys* and *Musca* relative to *Drosophila* CYP6A2, but a loss in *Drosophila* relative to 15 *Stomoxys* and 8 *Musca* CYP6As (see Additional file 1: Figure S20-B). While upregulation of genes in the CYP4, CYP6, and CYP9 families have been associated with resistance to spinosad and pyrethroid insecticides in *Musca*, *Anopheles*, and *Drosophila* [267–269], this has yet to be investigated in *Stomoxys*. Regardless, the large CYP gene family suggests *Stomoxys* may have an enhanced capacity for metabolic detoxification.

**Evidence for transcription factors with putative role in regulation of reproduction and salivation**

To determine transcription factors (TFs) that might control specific gene expression profiles in *S. calcitrans*, TF-encoding genes were first predicted by identifying putative DNA binding domains (DBDs), using a previously described approach [40, 70]. These analyses resulted in 837 predicted TFs, with the highest number coming from the

C<sub>2</sub>H<sub>2</sub> zinc finger and homeobox structural families (Fig. 10), consistent with previously analyzed insect genomes [40, 69, 70]. DNA binding motifs were subsequently predicted for as many of these putative TFs as possible using a previously developed method [270], and “inferred” motifs were identified for the *S. calcitrans* TFs. For example, the DBD of the uncharacterized XP\_013101333 protein is 92.3% identical to the DBD of the *D. melanogaster* gene *cropped* (FBgn0001994). Since the DNA binding motif of *cropped* has already been experimentally determined, and the cutoff for the bHLH family of TFs is 60% (see the “Methods” section), we can infer that XP\_013101333 will have the same binding motif as *cropped*. This procedure resulted in inferred motifs for 285 of the *S. calcitrans* TFs (34%).

TF binding site motif enrichment was then completed using promoter regions for groups of genes with similar gene expression patterns in our RNA-Seq datasets. Promoters were defined as either 500 or 2000 bp upstream of the predicted transcription start site for each gene and gene set/motif pairs were filtered as described in the “Methods” section. Expression of each TF was verified in specific tissues using our RNA-Seq datasets (see Additional file 15). Based on these criteria and comparative analyses between samples, seven and nine TFs, respectively, were enriched in SG tissues for the 2000 bp and 500 bp promoter regions (Fig. 10). Based on the 500 bp promoter regions, two specific TFs, *proboscipedia* (XM\_



**Fig. 10** Transcription factors associated with *Stomoxys calcitrans*. **a** Number of transcription factors identified in *S. calcitrans* compared to other flies. **b, c** Overlap between transcription factors with increased binding sites in differentially expressed genes that have noted expression in the same tissue (F, teneral female; FRS, female reproductive system; M, male; MRS, male reproductive system; SG, salivary glands). **d** Expression of specific TFs associated with female, male, and salivary glands among multiple tissues and developmental stages

013251179) and *orthopedia* (XM\_013261230), likely regulate SG-based transcript expression. These two TFs have been associated with head and salivary development in *Drosophila* [271, 272], and the increased binding sites and specific expression profile suggest a role in *S. calcitrans* saliva production.

Male- and female-enriched analysis based on stage and tissue specific RNA-Seq datasets identified TF targets in each of the 500 bp and 2000 bp promoter regions (Fig. 10). The four most likely TFs associated with female specific genes are XM\_013251807.1 (*iroquois-class homeodomain protein*) and XM\_013252765.1 (*Unc4 homeodomain protein*) based on the 500 bp promoter region and two other likely homeodomain proteins, XM\_013245879.1 (uncharacterized) and XM\_013261334.1 (uncharacterized), in the 2000 bp regulatory region. The latter have high TPM values in females and female reproductive tract tissues (Fig. 10). Two TFs within the 500 bp promoter region were for male enriched genes XM\_013258869.1 (*BarH-2*) and XM\_013251073.1 (uncharacterized), both of which are highly expressed in *S. calcitrans* teneral males, male heads, and/or the male reproductive tract tissue. When expanded to the 2000 bp promoter region, two additional putative TFs related to male enriched genes were XM\_013260987.1 (uncharacterized) and XM\_013257948.1 (*drop*), which are both highly expressed in male samples. Functional characterization of these TFs is warranted, as they could provide novel targets for the control of stable fly reproduction or the prevention of feeding, including development of genetically modified strains based on reproductive tract-specific regulatory elements.

## Conclusions

This study advances the knowledge of stable fly genomics and genetics to the breadth of other non-model, but extremely important, dipterans such as tsetse and house flies. Analysis of the genome sequence of stable fly males coupled with the life-stage and tissue-specific gene expression datasets revealed unique aspects of muscid fly biology that will guide future research within the livestock pest community. The distinct increase in the number of chemosensory receptors that are putatively involved in bitter taste perception is intriguing (Figs. 3 and 4), as it may support the stable fly's cosmopolitan distribution by enhancing avoidance behaviors and enabling a more discerning host and ovipositional substrate selection process [12, 19, 20]. This distribution is likely further supported by putative enhanced metabolic detoxification mediated by an increase in number of CYPs encoded by the genome (Fig. 9). Downstream functional analyses are critically essential to clarifying these roles. Importantly, this study provides the resources to support development of novel control approaches for this livestock pest. Identifying stable fly chemosensory factors that invoke aversion could

inform development of species-specific behavior modifying compounds and/or strategies [273], while the unique stable fly catalog of putative seminal proteins reported here (see Additional file 10) could be targeted for development of reproductive inhibitors. Conventional sterile insect technique methods are not sustainable for a stable fly control program, as stable flies are broadly dispersed in areas where they are problematic and both sexes of *Stomoxys* blood-feed. A continuous release of large numbers of irradiated, sterile males would be required to achieve successful population reduction [274–276]. Gene drive-based systems are attractive technologies, and availability of the genome enables identification of regulatory regions to develop such strains. Lastly, the recent analysis of sex chromosome evolution in stable flies and the recently sequenced horn fly underscores the unique opportunity that this genomic resource provides to enable future comparative genome analyses [49, 277].

## Methods

### Genome sequencing, assembly, and annotation

Total genomic DNA was isolated from pooled, teneral adult males ( $n = 120$ ) of a *Stomoxys calcitrans* line that resulted from single paired mating for 7 generations (Sc8C7A2A5H3J4). High quality/high molecular weight DNA was isolated from pooled flies using the Genomic-tip purification column and the associated buffer kit (QIAGEN, Valencia CA), and samples were processed according to the protocol for tissue-based DNA extraction. The pooled DNA isolates were utilized for sequencing on Illumina® HiSeq2000 instruments. The sequencing plan followed the recommendations provided in the ALLPAT HS-LG assembler [278]. Using this model, we targeted 45x sequence coverage each of fragments (overlapping paired reads ~ 180 bp length) and 3 kb paired end (PE) sequences as well as 5x coverage of 8 kb PE sequences. The first draft assembly scaffold gaps were closed where possible with mapping assembly input sequences (overlapping paired reads ~ 180 bp length) and local gap assembly [279]. Contaminating sequences and contigs 200 bp or less were removed. The genome assembly, *Stomoxys calcitrans*-1.0.1, was made publicly available in the Genbank sequence database, accession number GCF\_001015335.1. All sequence files associated with the project can be accessed on Genbank under BioProject PRJNA188117 ([280]; see Additional file 1: Table S1).

Automated gene annotation of the *Stomoxys* genome assembly (*Stomoxys calcitrans*-1.0.1) was completed using the NCBI Eukaryotic Genome Annotation Pipeline, version 6.4 with supporting RNA-Seq from existing transcripts in the expressed sequence tag [281] database and a de novo assembled transcriptome (Trinity-v2; transcriptome shotgun assembly (TSA) sequence database, Accession number GDIM00000000.1). The

resulting *Stomoxys* Annotation Report 100 can be accessed at [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Stomoxys\\_calcitrans/100/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Stomoxys_calcitrans/100/)

The Web Apollo tool [282], adopted by VectorBase [283], was used by our community to review and edit automated gene predictions. Edited gene models were incorporated into a community annotation patch build, ScalU1.6 (release date: November 5, 2020).

### Lateral gene transfer prediction

A DNA-based computational pipeline was used to identify “contaminating” bacterial scaffolds and bacterial to *Stomoxys* lateral gene transfer candidates in the *Stomoxys* genome assembly. The pipeline was originally developed by David Wheeler and John Werren [284] and has subsequently been modified. Details of the pipeline are provided in [81], and the procedure is summarized here. First, genome scaffolds were broken into 1000-bp units, which are screened against a bacterial genome database (see Additional file 16) containing over 2100 representative bacterial species. Scaffolds were then evaluated for proportion of bacterial matches along their length – those with greater than 20% bacterial matches were identified as likely bacterial scaffolds in the insect genome assembly and were manually examined for average coverage depth compared to all the scaffolds in the genome. That analysis supported twelve very small scaffolds (1028–3546 bp) as bacterial contamination (see Additional file 1: Table S6). The small size and low read depth compared to genome scaffold average supports the conclusion that these represent trace bacterial contamination (see Additional file 1: Tables S6 and S7).

To examine the assembly for candidate bacterial LGTs, analyses were conducted on scaffolds of greater than 100 kb in length, because confidently assigning a candidate LGT requires confirmation of eukaryotic flanking sequences to the candidate. Positive bacterial hits in each 1000-bp region with bit scores greater than 50 were compared to an “eukaryotic” database (see Additional file 16), which contains transcripts from the following eukaryotes: *Xenopus*, *Daphnia*, *Strongylocentrotus*, *Mus*, *Homo sapiens*, *Aplysia*, *Caenorhabditis*, *Hydra*, *Monosiga*, and *Acanthamoeba*. The purpose of this step is to exclude highly conserved sequences that are shared between bacterial and eukaryotes from further analysis. Focus was then placed on strong LGT candidates with bitscore > 75 in the bacterial match and zero bitscore in the eukaryotic match. Regions were combined when adjacent 1Kb fragments each had a zero “animal” match and a bacterial match to the same or similar bacterial source. The 1Kb fragments were also screened against transcripts from seven arthropod species to identify possible conserved arthropod genes (see Additional file 16). The initial candidate LGTs were then manually curated, through a series

of steps including (1) BLASTn to NCBI nr database, (2) BLASTx to NCBI protein database, (3) examination of flanking regions in the scaffold to confirm presence of flanking eukaryotic (insect) orthologous sequences and gene models, (4) removal of matches due to repetitive DNA, and (5) examination of RNA-Seq datasets for evidence of expression. In addition, LGT candidates were examined for sequencing read depth in the candidate LGT and flanking 1 kb up and downstream regions to determine if there were large changes in read depth across the junction. Sequencing reads were aligned to the reference genome using BWA v 0.7.17 [285]. Paired end alignment was called using the -mem function and aligned to the reference genome using default parameters. The coverage software program Mosdepth [286] was then used to calculate the average 50 bp read depth spanning the LGT and its junctions. Coverage ratio was calculated as the average read depth spanning a 50 bp window divided by the average 50 bp window read depth of the entire scaffold.

### RNA isolation and sequencing

A variety of developmental stages and dissected tissues were collected for total RNA isolation, and the accession numbers and metadata associated with these samples is summarized in see Additional file 1: Table S1. Specifically, RNA collected from whole females (teneral and mated, reproductive), whole males (teneral), male reproductive tracts, female reproductive tracts, male heads (fed, mated), female heads (fed, mated), and a single third instar larva were examined to assist in addressing core questions of this study and to assist with genome annotation. RNA was extracted with the use of TRIzol. DNA contamination was reduced via DNase treatment according to methods previously described [287, 288]. Poly(A)<sup>+</sup> RNA was isolated, then measured with the Agilent Bioanalyzer for quality and only those samples with a minimum RIN score of 7 were used to build non-normalized cDNA libraries using a modified version of the Nu-GEN Ovation<sup>®</sup> RNA-Seq System V2 (<http://www.nugeninc.com>). We sequenced each cDNA library (0.125 lane) on an Illumina HiSeq 2000 instrument (~ 36 Gb per lane) at 100 base pair length. RNA-Seq datasets used in gene prediction have been deposited to the NCBI Sequence Read Archive under the accession codes SRX995857–5860, SRX229930, SX229931, and SRX275910 (see Additional file 1: Table S1).

### RNA-Seq analyses

In conjunction with the genomic sequencing, RNA-Seq analyses were performed to examine specific transcript differences between different stages and tissues (see Additional file 4). RNA-Seq analyses were conducted based on methods in Benoit et al. [287] and updated according to Rosendale et al. [288, 289]. RNA-Seq datasets analyzed are summarized within Table S1. The main goals for analyzing these datasets were to (1) determine male, female, and

larva-enriched gene sets, (2) identify reproductive specific genes, and (3) establish chemosensory genes associated with sexes and specific tissue. Quality of each RNA-Seq dataset was assessed with FastQC prior to analyses, and RNA-Seq datasets were trimmed and low-quality reads were removed with trimmomatic [290]. Each dataset was mapped to the predicted gene set (NCBI Annotation Release 100) for *S. calcitrans* using CLC Genomics with each read requiring at least 95% similarity for over 60% of the read length with only two mismatches allowed. Splicing was allowed as long as 60% of each read mapped to the respective mRNA sequence. Read alignments were converted to per million mapped to allow comparison between RNA-Seq data sets with varying coverage (see Additional file 1: Table S1). Expression was based upon transcripts per million (TPM) and fold changes were determined as the TPM in one sample relative to the TPM of another dataset [288]. Significant enrichment was based on a Baggerly's test (compares proportional changes between groups) followed by Bonferroni correction at 0.01 (number of genes  $\times$   $\alpha$  value). This similar, stringent analysis was used on previous samples with only a single replicate and proved to be useful in the identification of genes of interest associated with specific developmental stages and sexes [41, 70]. RNA-Seq analyses are summarized in Additional file 4.

Transcripts in the official gene set were identified by BLASTx searching against an NCBI non-redundant proteins database for arthropods with an expectation value (*e* value) of at least 0.001. Gene ontology assessment for specific groups were conducted by the use of g:Profiler following conversion of *Stomoxys* gene IDs to *D. melanogaster* gene IDs [291]. The proteome of the stable fly was organized on a hyperlinked spreadsheet (see Additional file 2) with accompanying information (e.g., the presence or absence of signal peptide indicative of secretion, presence of transmembrane domains, and similarities to several databases). Expression values present within this sheet are based on the RNA-Seq data sets (see Additional file 1: Table S1) and analyzed according to previously described methods [238, 243].

#### Reverse transcription quantitative PCR (RT-qPCR) verification of RNA-Seq results

Since each RNA-Seq dataset is based on a single replicate, the results were validated by RT-qPCR (see Additional file 1: Figure S2, Table S5). RNA-Seq and RT-qPCR results showed Pearson correlation of 0.8643, indicating the RNA-Seq results are valid. Twenty-five transcripts were randomly selected to evaluate correlation between  $\log_2$  fold changes of RT-qPCR versus RNA-Seq. Total RNAs were isolated from tissues to represent those used for RNA-Seq {i.e. female or male heads (7d fed, mated), female or male reproductive systems (7d

fed, mated, dissected), adult female or male (whole, teneral), adult female or male (whole, 7d fed, mated), third instar larvae}. Samples were placed in TRIzol™, macerated, and stored at  $-80^\circ\text{C}$  until isolation using the Zymo Direct-zol™ method (Zymo Research, Irvine CA) with on-column DNase treatment (TURBO DNase, ThermoFisher Scientific, Waltham MA). cDNA templates were synthesized from 500 ng total RNA in a 20ul volume using SuperScript® III reverse transcriptase (ThermoScientific) primed with a dTV<sub>n</sub> oligonucleotide. Primers for RT-qPCR were designed with Beacon software to span exon-intron junctions (see Additional file 1: Table S5), and primer efficiencies were estimated using serially diluted cDNAs. Reactions were prepared in a 20ul volume consisting of 250 nM each of the forward and reverse primers, cDNA from 25 ng RNA, and the iTaq™ Universal SYBR® Green Supermix (Bio-Rad Laboratories, Hercules CA). Reactions were run in triplicate on a LightCycler® 96 System (Roche Life Sciences, Indianapolis IN). Data were analyzed using the  $2^{-\Delta\Delta CT}$  method incorporating primer efficiency data [292], and all values were normalized to the *S. calcitrans* ribosomal protein S3 reference gene *RpS3*. Pearson's correlation of  $\log_2$  fold change for RNA-Seq and RT-qPCR results was calculated using GraphPad Prism version 7.00 for MacOSX (GraphPad Software, La Jolla CA).

#### OrthoDB analysis

The OrthoDB hierarchical orthology delineation procedure was employed to predict orthologous groups (OGs) of genes across 87 arthropods for OrthoDB v8 [63]. Briefly, protein sequence alignments were assessed to identify all best reciprocal hits (BRHs) between genes from each pair of species, which are then clustered into OGs following a graph-based approach that starts with BRH triangulation. The annotated proteins from the genome of *Stomoxys calcitrans* were first filtered to select one protein-coding transcript per gene and then mapped to OrthoDB v8 at the Diptera (37 species), Endopterygota (72 species), Insecta (80 species), and Arthropoda (87 species) levels. OrthoDB orthology mapping uses the same BRH-based clustering procedure used to build the OGs but only allowing proteins from the mapped species to join existing OGs. Gene ontology assessment for specific groups were conducted by the use of g:Profiler following conversion of *Stomoxys* gene IDs to *D. melanogaster* gene IDs through BLASTx comparison [291].

#### Bacterial community analysis

To identify culturable bacterial communities harbored by adult stable flies, fly specimens were collected at each of four Texas dairies in April and June 2015 (Lingleville and Comanche, Texas), and total bacterial isolates from

the collections is reported. Twenty flies per site per date were collected by aerial sweep nets in the area surrounding each dairy's milking parlor. Within 4 h, whole flies were surface sterilized in 1% sodium hypochlorite for 15 min, followed by two washes in 70% ethanol and three rinses in sterile water. Individual flies were macerated in Butterfield's phosphate buffer, and the homogenate was diluted and plated on tryptic soy agar. Individual, morphologically distinct colonies were selected, suspended in Butterfield's phosphate buffer, and the DNA isolated by rapid boiling. These DNAs were used as template in 16S PCR amplification with a universal primer pair (16SEub\_61F: 5' – GCTTAACACATGCAAG – 3'; 16SEub\_1227R: 5' – CCATTGTAGCACGTGT – 3'). Individual amplicons were sequenced in both directions and the sequences assembled ( $n = 281$ ). Data were analyzed in mothur, v 1.38.1 [281], including UCHIME to detect and remove 19 chimeras, resulting in 262 isolates. Sequences were aligned using the silva.seed\_v132.align file, distance matrices calculated, and isolates classified at a 97% similarity cut-off, average neighbor, silva.seed\_v132.align/.tax. Phyla and genera for each isolate is reported (see Additional file 5).

#### Analysis of immune system gene families

Two computational approaches were combined with targeted manual annotation to produce an annotation of immune-related genes and gene families in the *S. calcitrans* genome. First, an initial list of *S. calcitrans* gene models was generated that likely have an immune function based on homology to annotated and functionally characterized *D. melanogaster* genes. Homologous genes were identified based on OrthoDB groups and include both orthologs and paralogs (e.g., every *S. calcitrans* gene in an OrthoDB group that includes a *D. melanogaster* immune-related gene is annotated as immune-related). To complement these homology-based annotations, all predicted *S. calcitrans* proteins were screened for similarity to HMM profiles of well-characterized Dipteran immune-related protein families (see Additional file 6) [93, 100], using hmmscan (from the HMMER software package). Finally, because antimicrobial peptides can be difficult for computational pipelines to properly annotate (due to their small size), we also include some manual annotation of *S. calcitrans* AMPs (see Additional file 6).

#### Chemosensory gene family identification and phylogenetic analysis

tBLASTn searches of the genome assembly were performed with gustatory receptors (GR) from *M. domestica*, *D. melanogaster*, and, where relevant the medfly, *C. capitata* [42], and with all newly identified *Stomoxys* GRs. Models were built primarily in the WebApollo server at

VectorBase, using a combination of existing automated models, RNA-Seq information from multiple lifestages, tissues, and sexes, and comparisons with the *Musca* and *Drosophila* Gr gene structures. A few models with regions missing in assembly gaps were repaired using raw genomic and/or RNA-Seq reads. Pseudogenes were translated as best possible to provide an encoded protein that could be aligned with the intact proteins for phylogenetic analysis. A 200 amino acid minimum was enforced for including pseudogenes in the analysis (roughly half the length of a typical GR), and there are several shorter fragments of genes that were not included in the analysis. All *Stomoxys* GRs were aligned in CLUSTALX v2.1 [293] using default settings with the GRs of *D. melanogaster* [118] and *M. domestica* [94]. Problematic gene models and pseudogenes were refined in light of these alignments. The final alignment was trimmed using TrimAl v1.4 [266] with the "strict" option. Phylogenetic construction was performed by maximum likelihood analysis using the PHYML v3.0 webserver with default settings, implementing an automatic model selection parameter [294]. The resultant tree was formatted and colored using FigTree v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>), while the final version with labels was made in Adobe Illustrator. Analysis of the Ionotropic Receptor (IR) family in *S. calcitrans* was generally similar to that for the GRs, except that pseudogenes were only included if they encoded at least 50% of the length of a related intact IR given that IRs vary considerably in length. The GR maximum likelihood tree was rooted by declaring the distantly-related and divergent carbon dioxide and sugar receptor subfamilies as the outgroup, while the IR tree was rooted by declaring the Ir8a/25a lineage as the outgroup.

Odorant receptor (OR), odorant binding protein (OBP), and chemosensory protein (CSP) sequences from *D. melanogaster* and *M. domestica* were used in tBLASTn searches to identify orthologs in the *Stomoxys calcitrans* 1.0.1 genome assembly. Models were built primarily in the WebApollo server at VectorBase, using a combination of existing automated models, RNA-Seq information from multiple lifestages, tissues, and sexes, and comparisons with the *Musca* and *Drosophila* gene models. Amino acid sequences from each family were aligned with the MUSCLE algorithm [201], and the alignments trimmed with the trimAl tool using the –strictplus option [266]. The trimmed alignment was used to construct a maximum likelihood phylogeny with the web server version of IQ-TREE software ([114]; best-fit substitution model, branch support assessed with 1000 replicates of UFBoot bootstrap approximation). The OR tree was rooted with the highly conserved ORCO. OBP domains of dimer OBPs were separated for phylogenetic analysis and labeled "a" and "b," and the Plus-C OBPs were not included in assembling alignments for tree construction.

### Opsin identification and analysis

For preparation of the global opsin gene tree, *S. calcitrans* sequences were collected by tBLASTn searches against the genome sequence draft version 1.0.1 (GCF\_001015335.1). A multiple sequence alignment was generated with MUSCLE [201] and variable sites were filtered using Gblocks with least stringent settings [65]. Bayesian tree analysis was performed out with MrBayes v3.2.6 [295] in the CIPRES Science Gateway V 3.3 environment [296], applying the GTR model of protein sequence evolution and correcting for across site substitution variation with a four rate category gamma distribution. A bootstrapped maximum likelihood tree was estimated in MEGA version 6.0 [203] applying the Jones-Taylor-Thornton (JTT) model of amino acid sequence evolution and assuming Gamma Distributed substitution rates across sites with three categories. For Bayesian analysis of the calyptrate expansion of Rh1 opsins, protein sequences were aligned with webPRANK [297]. Ambiguous alignment regions were filtered using TrimAl (v. 1.3) [266] as implemented on the Phylemon 2.0 server [298] applying User defined settings (Minimum percentage of positions to conserve: 10, Gap threshold: 0.9, Similarity threshold: 0.0, Window size: 1.0).

### Yolk protein gene identification and phylogenetic analysis

Annotated yolk protein gene sequences from *Drosophila melanogaster*, *Glossina morsitans*, and *Musca domestica* were used to identify yolk protein ortholog sequences from the *Musca domestica* and *Stomoxys calcitrans* predicted transcriptomes available at VectorBase ([283]; [www.vectorbase.org](http://www.vectorbase.org)). Gene sequences with significant homology were aligned using the ClustalO software package [221]. The alignment was used to generate a maximum likelihood phylogeny using the tree generation software included in the CLC Main Workbench (Qiagen, Redwood City CA) software package using the following settings (construction method: neighbor joining, Protein substitution model: WAG, Bootstrap analysis: 1000 replicates). Sequences with homology closer to lipase enzyme sequences (the ancestors to yolk protein genes) that form outgroups relative to annotated yolk proteins were removed from the alignment and the phylogeny was recalculated.

### Identification of male reproductive tract-biased genes and reciprocal orthology analysis

RNA-Seq libraries from male and female reproductive tissues of fed, mated *Stomoxys calcitrans* adults were used in this analysis. TPM values from the read mapping of these libraries against the putative *Stomoxys* transcriptome from the male and female reproductive tissues were compared to identify genes with a male/female

expression ratio of at least 5 and a minimum TPM value of 50 in the male reproductive tract. The sequences meeting these criteria were extracted from the *Stomoxys* transcriptome and then further filtered based on the presence of a secretory peptide. The narrowed list was categorized by gene ontology analysis using the R package topGO v 2.40 [299], providing statistical analysis of the categorization. Orthologous sequences for these genes were identified using a reciprocal hit analysis using the BLAST+ [300] software package against the predicted transcriptomes from *Glossina morsitans*, *Drosophila melanogaster*, *Aedes aegypti*, and *Homo sapiens*. Best hits from each of these transcriptomes from these species against male biased *Stomoxys* genes were extracted and used to BLAST back against the predicted *Stomoxys* transcriptome. BLAST output was parsed to detect *Stomoxys* genes with reciprocal hits (see Additional file 10).

### Salivary gland RNA-Seq analysis

Sixty salivary gland pairs were dissected from 7d fed, adult female and male stable flies. Tissues were placed in TRIzol™, macerated, and stored at -80 °C until isolation using the Zymo Direct-zol™ method (Zymo Research) with on-column DNase treatment (TURBO DNase, ThermoFisher Scientific). Resulting cDNA libraries were sequenced on an Illumina HiSeq 2000 instrument, and a total of 24.8 M and 47.6 M, 75 bp, single end reads were obtained. The salivary gland RNA-Seq dataset is based on a single replicate. To be consistent with previous studies on the sirolome of insect vectors, salivary gland RNA-Seq analyses were conducted using a published pipeline [301–303]. Briefly, the reads from four RNA-Seq libraries (male and female salivary glands, as well as teneral male and female whole bodies) were mapped to the *S. calcitrans* predicted gene set. To further be consistent with SG analyses completed in other blood-feeding arthropods, an  $\chi^2$  test was employed to identify those that were significantly over-expressed in SG relative to teneral whole bodies, as in [238] (see Additional file 11). Normalized fold-ratios of the sample reads were computed by adjusting the numerator by a factor based on the ratio of the total number of reads in each sample and adding one to the denominator to avoid division by zero. Expression results for the salivary glands showed high level of correlation between males and females (Pearson = 0.93). The complete dataset was organized in a hyperlinked spreadsheet as previously reported [238] and is provided in Additional file 2.

### Transcription factor analyses

To assess potential transcription factors regulating tissue and sex-specific expression, TFs were identified according to previously developed methods in other insect systems [40, 70]. DNA binding motifs were then predicted for as many of these putative TFs. In brief, the percent

of identical amino acids was calculated between each *S. calcitrans* TF and each eukaryotic TF with a known motif, with values exceeding a TF family-specific threshold resulting in “inferred” motifs for the *S. calcitrans* TFs. TF binding site motif enrichment was then completed using promoter regions for groups of genes with similar expression patterns. Promoters were defined as either 500 or 2000 bp upstream of the predicted transcription start site for each gene. The search was restricted to gene set/motif pairs with significant enrichment based on RNA-Seq analysis using Baggerly’s test followed by Bonferroni correction at 0.01, as described earlier. Gene set/motif pairs were further filtered to cases where (1) the given motif was present in at least 60% of the promoters of the gene set, (2) the given motif was present in less than 20% of all gene promoters, and (3) the difference between the presence of the motif in the gene set and promoters of all genes exceeded 40%. Expression of each TF was verified in specific tissues using our RNA-Seq datasets (see Additional file 15). Enriched TF binding motifs were identified in the 500 and 2000 bp regions upstream of the putative transcription start site using the HOMER tool [304] supplemented with the *Stomoxys* inferred binding motifs obtained from the CisBP database (build 0.90).

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-021-00975-9>.

**Additional file 1:** Main supplementary text file, including supplementary **Tables S1-S12** and supplementary **Figures S1-S21**; **Table S1.** RNA-Sequencing and Whole Genome Sequencing Accession Numbers and Statistics. **Table S2.** Summary of Dfam repeat elements with > 1000 copies including number of unique elements for each family of elements and total number of genomic copies for each family. **Table S3.** Detailed list of repeat elements with total genomic copy number > 1000. **Table S4.** Autophagy genes identified from the *Stomoxys* genome. **Table S5.** Validation of *Stomoxys* transcript expression by RT-qPCR. **Table S6.** Bacterial contaminating scaffolds located in the *Stomoxys* genome assembly. **Table S7.** Predicted lateral gene transfer events identified from the *Stomoxys* genome. **Table S8.** Components of the immune deficiency, Toll, and JAK/STAT pathways identified from the *Stomoxys* genome. **Table S9.** Manually annotated *Stomoxys* immune system gene family members. **Table S10.** RNA-Seq normalized expression values for transcripts annotated as antimicrobial peptides. **Table S11.** *Stomoxys* opsin gene compilation. **Table S12.** Aquaporin gene names and corresponding symbols, model numbers, scaffolds, and arthropod homologues. **Figure S1.** Phylogenetic placement and genomic comparisons for *Stomoxys calcitrans* and other fly species. **Figure S2.** Pearson’s correlation of RNA-Seq and RT-qPCR results. **Figure S3.** Average read depth across predicted lateral gene transfer candidates. **Figure S4.** *Stomoxys calcitrans* genomic scaffold housing 11 defensin gene models. **Figure S5.** Maximum likelihood phylogenetic tree of PGRP protein sequences from *S. calcitrans* (red), *M. domestica* (black), and *D. melanogaster* (blue). **Figure S6.** Alignment of *Stomoxys* PGRP-S sequences with characterized *D. melanogaster* PGRP-SC1 (C0HK98) and -SC2 (Q9VX2) and N-acetylmuramoyl-L-alanine amidase (P00806). **Figure S7.** Alignment of *Stomoxys* PGRP-L sequences with characterized *D. melanogaster* PGRP-L proteins and N-acetylmuramoyl-L-alanine amidase. **Figure S8.** *Stomoxys calcitrans* Odorant Binding Protein Gene Family. **Figure S9.** Duplicated OS-E-like and an OS-X ortholog in

*Stomoxys* and *Musca*. **Figure S10.** *Stomoxys calcitrans* Odorant Receptor Gene Family. **Figure S11.** *Stomoxys calcitrans* Gustatory Receptor Gene Family. **Figure S12.** *Stomoxys calcitrans* Ionotropic Receptor Gene Family. **Figure S13.** Maximum likelihood tree of dipteran opsin gene relationships. **Figure S14.** Phylogenetic analysis of the *Stomoxys* Rh1 gene cluster and Genomic organization and evolution of the *S. calcitrans* Rh1 opsin subfamily. **Figure S15.** Analysis of tuning site 17 variation in *Stomoxys* and *Musca* Rh1 paralogs. **Figure S16, S18, S19.** Phylogenetic relationship of catalytic carboxylesterases (S16), glutathione-S-transferases (S18), and Cys-Loop Ligand Gated Ion Channels (S19) from *Stomoxys* relative to *Musca* and *Drosophila*. **Figure S17.** Phylogenetic analysis of carboxylesterases with a role in neuronal development. **Figure S20.** Cytochrome P450 (CYP450) genes clustered on *Stomoxys* scaffolds and evidence for expansions in muscids relative to *Drosophila*. **Figure S21.** Phylogenetic analysis of cytochrome P450 genes from *Stomoxys calcitrans*.

**Additional file 2.** Functional annotation of transcripts assembled from the *Stomoxys* genome and output from gProfiler analysis.

**Additional file 3.** Output from analysis of core CEGMA proteins in the *Stomoxys* genome.

**Additional file 4.** Output from RNA-Seq Dataset Analyses.

**Additional file 5.** Output from cluster analysis of 16S bacterial isolate sequencing.

**Additional file 6.** *Stomoxys* immune-related proteins identified by similarity to HMM profiles of *Drosophila* immune protein families and by manual annotation.

**Additional file 7.** *Stomoxys* antimicrobial peptide amino acid sequences.

**Additional file 8.** *Stomoxys* chemosensory gene family member amino acid sequences.

**Additional file 9.** *Stomoxys* chemosensory gene family annotation and normalized RNA-Seq expression data.

**Additional file 10.** *Stomoxys* male reproductive tract-biased genes.

**Additional file 11.** *Stomoxys* salivary gland RNA-Seq analysis output.

**Additional file 12.** *Stomoxys* metabolic detoxification gene family member amino acid sequences.

**Additional file 13.** *Stomoxys* cytochrome P450 gene family identification.

**Additional file 14.** *Stomoxys* cytochrome P450 amino acid sequences.

**Additional file 15.** *Stomoxys* transcription factor expression analysis.

**Additional file 16.** Prokaryote and eukaryote databases used for the lateral gene transfer prediction pipeline.

## Acknowledgements

We thank The Genome Institute at Washington University for library construction and sequencing. JHW thanks S Cheng for assistance with the LGT pipeline.

## Footnote

This article reports the result of research only. 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 U.S. Department of Agriculture. The USDA is an equal opportunity provider and employer.

## Authors’ contributions

PUO and JBB organized and directed the sequencing, analysis, and manuscript development. Grant funding was obtained for this project as part of the *Glossina* genome project by SA. Sequencing and assembly of the genome was performed by WCW and RKW. Hosting of the genome for annotation was provided by GLM, SJE, and DL. Gene predictions were performed by TDM. Male and female reproductive analyses were conducted by GMA and JBB. Transcription factor analyses were conducted by XC, ENJ, MTW, and JBB. BUSCO and phylogenetic analyses were conducted by EOM, ECJ, and JBB. JMCR, JBB, and PUO collected samples and performed sialome analyses. Chemosensation associated genes were annotated and analyzed by HMR, MD, GB, and PUO. Cuticle proteins were analyzed



by AJR and JBB. Vision associated genes were analyzed by MF and JD. CJH and JBB assisted in the annotation of aquaporins. Annotation of CEGMA and autophagy genes was conducted by KJO and JBB. Immune genes were annotated and analyzed by TBS, DN, and PUO. DBT contributed to the writing of the introduction and interpretations of the results. Muller elements were assigned and analyzed by RPM. Bacterial community was assessed by SLS and PUO. Cytochrome P450s were analyzed by DRN. Carboxylesterases, GSTs, and CLGICs were analyzed by PUO. JHW and PS conducted analyses of lateral gene transfers. OrthoDB analyses were conducted by EZ, RMW, and PI. Transposable elements and genomic repeats were analyzed AMT, SHS, CJC, and TJR. PUO, JBB, GMA, MF, RPM, DN, TJR, JMCR, HMR, TBS, AMT, DBT, MTW, RMW, and JHW wrote and edited the manuscript. All authors read and approved the final manuscript.

### Funding

Sequencing was supported by the NIH-NHGRI grant 5U54HG00307907 to RKW. The lateral gene transfer work was supported by the US National Science Foundation (1257053 and 1950078) and the Nathaniel and Helen Wisch Chair to JHW. Funding for MTW was partially provided by CpG Award 53553 and a CCRF Endowed Scholar Award from Cincinnati Children's Hospital Medical Center. RMW was supported by Swiss National Science Foundation grant PP00P3\_170664. Funding for JBB was partially provided by bioinformatics support from the National Science Foundation (DEB-1654417), United States Department of Agriculture (2018-67013-28495), and the National Institutes of Health (1R01AI148551-01A1). Funding for PUO and PS was provided by the US Department of Agriculture | Agricultural Research Service (3094-32000-041-00D). The automated gene annotation work was supported by the Intramural Research Program of the National Library of Medicine, National Institutes of Health.

### Availability of data and materials

All whole genome sequence and RNA-Seq transcriptome data are publicly available at the NCBI BioProject database: PRJNA188117 [280], and the reference sequence (RefSeq) genome assembly is available at the NCBI BioProject database: PRJNA288986 [305]. The genome assembly can also be accessed at NCBI accession number GCA\_001015335.1. RNA-Seq datasets used in gene prediction have been deposited to the NCBI SRA site under the accession codes SRX995857-5860, SRX229930, SRX229931, and SRX275910. Annotation and gene model data are available at VectorBase [283], [www.vectorbase.org](http://www.vectorbase.org), *Stomoxys calcitrans* USDA, ScalU1.6.

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Livestock Arthropod Pests Research Unit, USDA-ARS, Kerrville, TX, USA. <sup>2</sup>Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA. <sup>3</sup>Department of Entomology and Nematology, University of California – Davis, Davis, CA, USA. <sup>4</sup>The Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. <sup>5</sup>Department of Entomology, Texas A & M University, College Station, TX, USA. <sup>6</sup>Department of Biological Sciences, Wayne State University, Detroit, MI, USA. <sup>7</sup>Department of Electrical Engineering & Computer Science, University of Tennessee, Knoxville, TN, USA. <sup>8</sup>Department of Biological Sciences, University of Cincinnati, Cincinnati, OH, USA. <sup>9</sup>Department of Genetic Medicine and Development, University of Geneva Medical School and Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland. <sup>10</sup>The European Molecular Biology Laboratory, The European Bioinformatics Institute, The Wellcome Genome Campus, Hinxton CB10 1SD, UK. <sup>11</sup>Department of Entomology, University of Georgia, Athens, GA, USA. <sup>12</sup>Department of Biology and Biochemistry, University of Houston, Houston, TX, USA. <sup>13</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA. <sup>14</sup>Arthropod-borne Animal Diseases Research Unit, USDA-ARS, Manhattan, KS, USA. <sup>15</sup>Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, Memphis, TN, USA. <sup>16</sup>Section

of Vector Biology, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, Rockville, MD, USA. <sup>17</sup>Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. <sup>18</sup>Department of Biology, Mount St. Joseph University, Cincinnati, OH, USA. <sup>19</sup>Informatics Group, Faculty of Arts and Sciences, Harvard University, Cambridge, MA, USA. <sup>20</sup>Department of Entomology, Texas A&M Agrilife Research and Extension Center, Stephenville, TX, USA. <sup>21</sup>Department of Computer Science & Engineering, Department of Biochemistry & Biophysics, Texas A & M University, College Station, TX, USA. <sup>22</sup>Agroecosystem Management Research Unit, USDA-ARS, Lincoln, NE, USA. <sup>23</sup>University of Missouri, Bond Life Sciences Center, Columbia, MO, USA. <sup>24</sup>Department of Ecology and Evolution, University of Lausanne, and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. <sup>25</sup>Center for Autoimmune Genomics and Etiology, Divisions of Biomedical Informatics and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. <sup>26</sup>Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA. <sup>27</sup>Department of Biology, University of Rochester, Rochester, NY, USA. <sup>28</sup>Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH, USA. <sup>29</sup>College of Medicine, Ohio State University, Columbus, OH, USA.

Received: 20 October 2020 Accepted: 3 February 2021

Published online: 10 March 2021

### References

- Krafsur ES, Moon RD. Bionomics of the face fly, *Musca autumnalis*. *Annu Rev Entomol.* 1997;42:503–23.
- Foil LD, Hogsette JA. Biology and control of tabanids, stable flies and horn flies. *Rev Sci Tech.* 1994;13:1125–58.
- Dougherty CT, Knapp FW, Burrus PB, Willis DC, Burg JG, Cornelius PL, Bradley NW. Stable flies (*Stomoxys calcitrans* L) and the behavior of grazing beef cattle. *Appl Anim Behav Sci.* 1993;35:215–33.
- Dougherty CT, Knapp FW, Burrus PB, Willis DC, Cornelius PL. Moderation of grazing behavior of beef cattle by stable flies (*Stomoxys calcitrans* L). *Appl Anim Behav Sci.* 1994;40:113–27.
- Dougherty CT, Knapp FW, Burrus PB, Willis DC, Cornelius PL. Behavior of grazing cattle exposed to small populations of stable flies (*Stomoxys calcitrans* L). *Appl Anim Behav Sci.* 1995;42:231–48.
- Mullens BA, Lii KS, Mao Y, Meyer JA, Peterson NG, Szijj CE. Behavioural responses of dairy cattle to the stable fly, *Stomoxys calcitrans*, in an open field environment. *Med Vet Entomol.* 2006;20:122–37.
- Taylor DB, Moon RD, Mark DR. Economic impact of stable flies (Diptera: Muscidae) on dairy and beef cattle production. *J Med Entomol.* 2012;49:198–209.
- Cook DF, Telfer DV, Lindsey JB, Deyl RA. Substrates across horticultural and livestock industries that support the development of stable fly, *Stomoxys calcitrans* (Diptera: Muscidae). *Austral Entomol.* 2018;57:344–8.
- Cook DF, Dadour IR, Voss SC. Management of stable fly and other nuisance flies breeding in rotting vegetable matter associated with horticultural crop production. *Int J Pest Manage.* 2011;57:315–20.
- Dominghetti TF, de Barros AT, Soares CO, Cancado PH. *Stomoxys calcitrans* (Diptera: Muscidae) outbreaks: current situation and future outlook with emphasis on Brazil. *Rev Bras Parasitol Vet.* 2015;24:387–95.
- Solorzano JA, Gilles J, Bravo O, Vargas C, Gomez-Bonilla Y, Bingham GV, Taylor DB. Biology and trapping of stable flies (Diptera: Muscidae) developing in pineapple residues (*Ananas comosus*) in Costa Rica. *J Insect Sci.* 2015;15:145.
- Romero A, Broce A, Zurek L. Role of bacteria in the oviposition behaviour and larval development of stable flies. *Med Vet Entomol.* 2006;20:115–21.
- Baldacchino F, Muenworn V, Desquesnes M, Desoli F, Charoenviriyaphap T, Duvallet G. Transmission of pathogens by *Stomoxys* flies (Diptera, Muscidae): a review. *Parasite.* 2013;20.
- Traversa D, Otranto D, Iorio R, Carluccio A, Contri A, Paoletti B, Bartolini R, Giangaspero A. Identification of the intermediate hosts of *Habronema microstoma* and *Habronema muscae* under field conditions. *Med Vet Entomol.* 2008;22:283–7.
- Boulanger N, Munks R, Hamilton JV, Vovelle F, Brun R, Lehane MJ, Bulet P. Epithelial innate immunity - a novel antimicrobial peptide with antiparasitic activity in the blood-sucking insect *Stomoxys calcitrans*. *J Biol Chem.* 2002; 277:49921–6.

16. Nayduch D, Burrus RG. Flourishing in filth: house fly-microbe interactions across life history. *Ann Entomol Soc Am.* 2017;110:6–18.
17. Scully E, Friesen K, Wienhold B, Durso LM. Microbial communities associated with stable fly (Diptera: Muscidae) larvae and their developmental substrates. *Ann Entomol Soc Am.* 2017;110:61–72.
18. Birkett MA, Agelopoulou N, Jensen KM, Jespersen JB, Pickett JA, Prijs HJ, Thomas G, Trapman JJ, Wadhams LJ, Woodcock CM. The role of volatile semiochemicals in mediating host location and selection by nuisance and disease-transmitting cattle flies. *Med Vet Entomol.* 2004;18:313–22.
19. Jeanbourquin P, Guerin PM. Chemostimuli implicated in selection of oviposition substrates by the stable fly *Stomoxys calcitrans*. *Med Vet Entomol.* 2007;21:209–16.
20. Jeanbourquin P, Guerin PM. Sensory and behavioural responses of the stable fly *Stomoxys calcitrans* to rumen volatiles. *Med Vet Entomol.* 2007;21:217–24.
21. Schofield S, Cork A, Brady J. Electroantennogram responses of the stable fly, *Stomoxys calcitrans*, to components of host odour. *Physiol Entomol.* 1995;20:273–80.
22. Tangtrakulwanich K, Albuquerque TA, Brewer GJ, Baxendale FP, Zurek L, Miller DN, Taylor DB, Friesen KA, Zhu JJ. Behavioural responses of stable flies to cattle manure slurry associated odourants. *Med Vet Entomol.* 2015;29:82–7.
23. Tangtrakulwanich K, Chen H, Baxendale F, Brewer G, Zhu JJ. Characterization of olfactory sensilla of *Stomoxys calcitrans* and electrophysiological responses to odorant compounds associated with hosts and oviposition media. *Med Vet Entomol.* 2011;25:327–36.
24. Jones CJ, Milne DE, Patterson RS, Schreiber ET, Milio JA. Nectar feeding by *Stomoxys calcitrans* (Diptera: Muscidae): effects on reproduction and survival. *Environ Entomol.* 1992;21:141–7.
25. Muller GC, Hogsette JA, Beier JC, Traore SF, Toure MB, Traore MM, Bah S, Doumbia S, Schlein Y. Attraction of *Stomoxys* sp. to various fruits and flowers in Mali. *Med Vet Entomol.* 2012;26:178–87.
26. Taylor DB, Berkebile DR. Sugar feeding in adult stable flies. *Environ Entomol.* 2008;37:625–9.
27. Gatehouse AG, Lewis CT. Host location behavior of *Stomoxys calcitrans*. *Entomologia Exp Appl.* 1973;16:275–90.
28. Uebel EC, Sonnet PE, Bierl BA, Miller RW. Sex pheromone of the stable fly: isolation and preliminary identification of compounds that induce mating strike behavior. *J Chem Ecol.* 1975;1:377–85.
29. Albuquerque TA, Zurek L. Temporal changes in the bacterial community of animal feces and their correlation with stable fly oviposition, larval development, and adult fitness. *Front Microbiol.* 2014;5:590.
30. Gerry AC, Peterson NG, Mullens BA. Predicting and controlling stable flies on California dairies. Oakland: University of California, Division of Agriculture and Natural Resources; 2007. Publication 8258
31. Beresford DV, Sutcliffe JF. Studies on the effectiveness of Coroplast sticky traps for sampling stable flies (Diptera: Muscidae), including a comparison to Alsynite. *J Econ Entomol.* 2006;99:1025–35.
32. Taylor DB, Berkebile D. Comparative efficiency of six stable fly (Diptera: Muscidae) traps. *J Econ Entomol.* 2006;99:1415–9.
33. Zhu JWJ, Zhang QH, Taylor DB, Friesen KA. Visual and olfactory enhancement of stable fly trapping. *Pest Manag Sci.* 2016;72:1765–71.
34. Anderson JR. Mating behavior of *Stomoxys calcitrans* (Diptera: Muscidae): effects of a blood meal on mating drive of males and its necessity as a prerequisite for proper insemination of females. *J Econ Entomol.* 1978;71:379–86.
35. Yuval B. Mating systems of blood-feeding flies. *Annu Rev Entomol.* 2006;51:413–40.
36. Patton WS. Studies on the higher Diptera of medical and veterinary importance. *Ann Trop Med Parasitol.* 1933;27:135–56.
37. Detinova TS. Age-grouping methods in Diptera of medical importance, with special reference to some vectors of malaria. Geneva: World Health Organization; 1962.
38. Thomas AW, Gooding RH. Digestive processes of hematophagous insects. VIII. Estimation of meal size and demonstration of trypsin in horse flies and deer flies (Diptera: Tabanidae). *J Med Entomol.* 1976;13:131–6.
39. Picard CJ, Johnston JS, Tarone AM. Genome sizes of forensically relevant Diptera. *J Med Entomol.* 2012;49:192–7.
40. Benoit JB, Adelman ZN, Reinhardt K, Dolan A, Poelchau M, Jennings EC, Szuter EM, Hagan RW, Gujar H, Shukla JN, et al. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. *Nat Commun.* 2016;7:10165.
41. International Glossina Genome Initiative. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. *Science.* 2014;344:380–6.
42. Papanicolaou A, Schetelig MF, Arensburger P, Atkinson PW, Benoit JB, Bourtzis K, Castanera P, Cavanaugh JP, Chao H, Childers C, et al. The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. *Genome Biol.* 2016;17:192.
43. Oliveira MT, Barau JG, Junqueira AC, Feijao PC, Rosa AC, Abreu CF, Azeredo-Espin AM, Lessinger AC. Structure and evolution of the mitochondrial genomes of *Haematobia irritans* and *Stomoxys calcitrans*: the Muscidae (Diptera: Calyptratae) perspective. *Mol Phylogenet Evol.* 2008;48:850–7.
44. Boyes JW, Paterson HE, Corey MJ. Somatic chromosomes of higher diptera .9. karyotypes of some muscid species. *Can J Zool.* 1964;42:1025.
45. Foster GG, Whitten MJ, Konovalov C, Arnold JTA, Maffi G. Autosomal genetic maps of the Australian sheep blowfly, *Lucilia cuprina*-dorsalis R-D (Diptera, Calliphoridae), and possible correlations with the linkage maps of *Musca domestica* L and *Drosophila melanogaster* (Mg). *Genet Res.* 1981;37:55–69.
46. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, Smit AF, Wheeler TJ. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* 2016;44:D81–9.
47. Joslyn DJ, Seawright JA, Willis NL. The karyotype of the stable fly, *Stomoxys calcitrans* (L.) (Diptera: Muscidae). *Caryologia.* 1979;32:349–54.
48. LaChance LE. Chromosome studies in three species of Diptera (Muscidae and Hypodermatidae). *Ann Entomol Soc Am.* 1964;57:69–73.
49. Meisel RP, Olafson PU, Adhikari K, Guerrero FD, Konganti K, Benoit JB. Sex chromosome evolution in muscid flies. *G3 (Bethesda).* 2020;10:1341–52.
50. Meisel RP, Scott JG, Clark AG, et al. *Genome Biol Evol.* 2015;7:2051–61.
51. Muller HJ. Bearings of the 'Drosophila' work on systematics. In: Huxley J, editor. *The new systematics.* Oxford: Clarendon Press; 1940. p. 186–268.
52. Seawright JA, Birky BK, Smittle BJ. Use of a genetic technique for separating the sexes of the stable fly (Diptera: Muscidae). *J Econ Entomol.* 1986;79:1413–7.
53. Taskin V, Kence M. The genetic basis of malathion resistance in housefly (*Musca domestica* L.) strains from Turkey. *Genetika.* 2004;40:1475–82.
54. Vicoso B, Bachtrög D. Reversal of an ancient sex chromosome to an autosome in *Drosophila*. *Nature.* 2013;499:332–5.
55. Wagoner DE. Linkage group-karyotype correlation in the house fly determined by cytological analysis of X-ray induced translocations. *Genetics.* 1967;57:729–39.
56. Wagoner DE. Linkage group-karyotype correlation in the house fly, *Musca domestica* L., confirmed by cytological analysis of x-ray induced Y-autosomal translocations. *Genetics.* 1969;62:115–21.
57. Walder JM, Seawright JA. Genetic method for the separation of males and females of the house fly, *Musca domestica* (Diptera: Muscidae). *J Econ Entomol.* 1985;78:1030–4.
58. Weller GL, Foster GG. Genetic maps of the sheep blowfly *Lucilia cuprina*: linkage-group correlations with other dipteran genera. *Genome.* 1993;36:495–506.
59. Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. *Bioinformatics.* 2013;29:2487–9.
60. Willis NL, Hillburn LR, Seawright JA. Black pupa, a recessive mutant on chromosome 3 of the stable fly, *Stomoxys calcitrans* (L.). *J Hered.* 1983;74:114–5.
61. Willis NL, Seawright JA, Nickel C, Joslyn DJ. Reciprocal translocations and partial correlation of chromosomes in the stable fly. *J Hered.* 1981;72:104–6.
62. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31:3210–2.
63. Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simao FA, Pozdnyakov IA, Ioannidis P, Zdobnov EM. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic Acids Res.* 2015;43:D250–6.
64. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30:3059–66.
65. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 2007;56:564–77.
66. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006;22:2688–90.

67. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 2015;16:157.
68. Parra G, Bradnam K, Ning Z, Keane T, Korff I. Assessing the gene space in draft genomes. *Nucleic Acids Res.* 2009;37:289–97.
69. Panfilio KA, Vargas Jentsch IM, Benoit JB, Erezylmaz D, Suzuki Y, Colella S, Robertson HM, Poelchau MF, Waterhouse RM, Ioannidis P, et al. Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *Genome Biol.* 2019;20:64.
70. Schoville SD, Chen YH, Andersson MN, Benoit JB, Bhandari A, Bowsher JH, Brevik K, Cappelle K, Chen MJM, Childers AK, et al. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Sci Rep.* 2018;8:1931.
71. Lysyk TJ, Kalischuk-Tymensen L, Selinger LB, Lancaster RC, Wever L, Cheng KJ. Rearing stable fly larvae (Diptera : Muscidae) on an egg yolk medium. *J Med Entomol.* 1999;36:382–8.
72. Whitten M, Dyson P. Gene silencing in non-model insects: overcoming hurdles using symbiotic bacteria for trauma-free sustainable delivery of RNA interference: sustained RNA interference in insects mediated by symbiotic bacteria: applications as a genetic tool and as a biocide. *Bioessays.* 2017;39:1600247.
73. Wang Y, Gilbreath TM 3rd, Kukutla P, Yan G, Xu J. Dynamic gut microbiome across life history of the malaria mosquito *Anopheles gambiae* in Kenya. *PLoS One.* 2011;6:e24767.
74. Coon KL, Vogel KJ, Brown MR, Strand MR. Mosquitoes rely on their gut microbiota for development. *Mol Ecol.* 2014;23:2727–39.
75. Guégan M, Zouache K, Démichel C, Minard G, Potier P, Mavingui P, Moro CV. The mosquito holobiont: fresh insight into mosquito-microbiota interactions. *Microbiome.* 2018;6:49.
76. Janda JM, Abbott SL. The genus *Aeromonas*: taxonomy, pathogenicity, and infection. *Clin Microbiol Rev.* 2010;23:35–73.
77. Tomberlin JK, Crippen TL, Tarone AM, Chaudhury MFB, Singh B, Cammack JA, Meisel RP. A review of bacterial interactions with blow flies (Diptera: Calliphoridae) of medical, veterinary, and forensic importance. *Ann Entomol Soc Am.* 2017;110:19–36.
78. Weatherbee CR, Pechal JL, Benbow ME. The dynamic maggot mass microbiome. *Ann Entomol Soc Am.* 2017;110:45–53.
79. Dunning Hotopp JC, Clark ME, Oliveira DC, Foster JM, Fischer P, Munoz Torres MC, Giebel JD, Kumar N, Ishmael N, Wang S, et al. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science.* 2007;317:1753–6.
80. Lindsey AR, Werren JH, Richards S, Stouthamer R. Comparative genomics of a parthenogenesis-inducing *Wolbachia* symbiont. *G3 (Bethesda).* 2016;6:2113–23.
81. Poynton HC, Hasenbein S, Benoit JB, Sepulveda MS, Poelchau MF, Hughes DST, Murali SC, Chen S, Glastad KM, Goodisman MAD, et al. The toxicogenome of *Hyalella azteca*: a model for sediment ecotoxicology and evolutionary toxicology. *Environ Sci Technol.* 2018;52:6009–22.
82. Werren JH, Baldo L, Clark ME. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol.* 2008;6:741–51.
83. Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH. How many species are infected with *Wolbachia*? – a statistical analysis of current data. *FEMS Microbiol Lett.* 2008;281:215–20.
84. Zug R, Hammerstein P. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS One.* 2012;7:e38544.
85. Ahmed MZ, Breinholt JW, Kawahara AY. Evidence for common horizontal transmission of *Wolbachia* among butterflies and moths. *BMC Evol Biol.* 2016;16:118.
86. Dunning Hotopp JC, Slatko BE, Foster JM. Targeted enrichment and sequencing of recent endosymbiont-host lateral gene transfers. *Sci Rep.* 2017;7:857.
87. Madhav M, Parry R, Morgan JAT, James P, Asgari S. *Wolbachia* endosymbiont of the horn fly (*Haematobia irritans irritans*): a supergroup A strain with multiple horizontally acquired cytoplasmic incompatibility genes. *Appl Environ Microbiol.* 2020;86:e02589-19.
88. Floate KD, Kyei-Poku GK, Coghlin P. Overview and relevance of *Wolbachia* bacteria in biocontrol research. *Biocontrol Sci Tech.* 2006;16:767–88.
89. Tseng JM, Jones CJ, Hogsette JA. Nectar feeding and the stable fly, *Stomoxys calcitrans* (Diptera: Muscidae). *J Fla Anti-Mosq Assoc.* 1983;54:40–1.
90. Michalkova V, Benoit JB, Weiss BL, Attardo GM, Aksoy S. Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies. *Appl Environ Microbiol.* 2014;80:5844–53.
91. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat Genet.* 2002;32:402–7.
92. Sackton TB. Comparative genomics and transcriptomics of host-pathogen interactions in insects: evolutionary insights and future directions. *Curr Opin Insect Sci.* 2019;31:106–13.
93. Sackton TB, Lazzaro BP, Clark AG. Rapid expansion of immune-related gene families in the house fly, *Musca domestica*. *Mol Biol Evol.* 2017;34:857–72.
94. Scott JG, Warren WC, Beukeboom LW, Bopp D, Clark AG, Giers SD, Hediger M, Jones AK, Kasai S, Leichter CA, et al. Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. *Genome Biol.* 2014;15.
95. Buchon N, Silverman N, Cherry S. Immunity in *Drosophila melanogaster* – from microbial recognition to whole-organism physiology. *Nat Rev Immunol.* 2014;14:796–810.
96. Barribeau SM, Sadd BM, du Plessis L, Brown MJF, Buechel SD, Cappelle K, Carolan JC, Christiaens O, Colgan TJ, Erler S, et al. A depauperate immune repertoire precedes evolution of sociality in bees. *Genome Biol.* 2015;16:83.
97. Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, Jiang H, Kanost M, Thompson GJ, Zou Z, Hultmark D. Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect Mol Biol.* 2006;15:645–56.
98. Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. Dynamic evolution of the innate immune system in *Drosophila*. *Nat Genet.* 2007;39:1461–8.
99. Sackton TB, Werren JH, Clark AG. Characterizing the infection-induced transcriptome of *Nasonia vitripennis* reveals a preponderance of taxonomically-restricted immune genes. *PLoS One.* 2013;8:e83984.
100. Waterhouse RM, Kriventseva EV, Meister S, Xi ZY, Alvarez KS, Bartholomay LC, Barillas-Mury C, Bian GW, Blandin S, Christensen BM, et al. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science.* 2007;316:1738–43.
101. Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Beukeboom LW, Desplan C, Elsik CG, Grimmelikhuijzen CJ, et al. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science.* 2010;327:343–8.
102. Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, De Vos M, Duncan EJ, Evans JD, Gabaldon T, Ghanim M, et al. Immunity and other defenses in pea aphids, *Acyrtosiphon pisum*. *Genome Biol.* 2010;11:R21.
103. Hanson MA, Hamilton PT, Perlman SJ. Immune genes and divergent antimicrobial peptides in flies of the subgenus *Drosophila*. *BMC Evol Biol.* 2016;16:228.
104. Obbard DJ, Welch JJ, Kim KW, Jiggins FM. Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet.* 2009;5:e1000698.
105. Bischoff V, Vignal C, Duvic B, Boneca IG, Hoffmann JA, Royet J. Downregulation of the *Drosophila* immune response by peptidoglycan-recognition proteins SC1 and SC2. *PLoS Pathog.* 2006;2:139–47.
106. Bosco-Drayon V, Poidevin M, Boneca IG, Narbonne-Reveau K, Royet J, Charroux B. Peptidoglycan sensing by the receptor PGRP-LE in the *Drosophila* gut induces immune responses to infectious bacteria and tolerance to microbiota. *Cell Host Microbe.* 2012;12:153–65.
107. Gao YF, Tang T, Gu JH, Sun LL, Gao XB, Ma XY, Wang XC, Liu FS, Wang JH. Downregulation of the *Musca domestica* peptidoglycan recognition protein SC (PGRP-SC) leads to overexpression of antimicrobial peptides and tardy pupation. *Mol Immunol.* 2015;67:465–74.
108. Gendrin M, Zaidman-Remy A, Broderick NA, Paredes J, Poidevin M, Roussel A, Lemaître B. Functional analysis of PGRP-LA in *Drosophila* immunity. *PLoS One.* 2013;8:e69742.
109. Lehane MJ, Wu D, Lehane SM. Midgut-specific immune molecules are produced by the blood-sucking insect *Stomoxys calcitrans*. *Proc Natl Acad Sci U S A.* 1997;94:11502–7.
110. Mellroth P, Karlsson J, Steiner H. A scavenger function for a *Drosophila* peptidoglycan recognition protein. *J Biol Chem.* 2003;278:7059–64.
111. Zaidman-Remy A, Herve M, Poidevin M, Pili-Floury S, Kim MS, Blanot D, Oh BH, Ueda R, Mengin-Lecreux D, Lemaître B. The *Drosophila* amidase PGRP-LB modulates the immune response to bacterial infection. *Immunity.* 2006;24:463–73.

112. Chang CI, Chelliah Y, Borek D, Mengin-Lecreux D, Deisenhofer J. Structure of tracheal cytotoxin in complex with a heterodimeric pattern-recognition receptor. *Science*. 2006;311:1761–4.
113. Robertson HM. Molecular evolution of the major arthropod chemoreceptor gene families. *Annu Rev Entomol*. 2019;64:1.
114. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res*. 2016;44:W232–5.
115. Benton R. Multigene family evolution: perspectives from insect chemoreceptors. *Trends Ecol Evol*. 2015;30:590–600.
116. Joseph RM, Carlson JR. *Drosophila* chemoreceptors: a molecular interface between the chemical world and the brain. *Trends Genet*. 2015;31:683–95.
117. Robertson HM. The insect chemoreceptor superfamily is ancient in animals. *Chem Senses*. 2015;40:609–14.
118. Robertson HM, Warr CG, Carlson JR. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 2003;100:14537–42.
119. Saina M, Busengdal H, Sinigaglia C, Petroni L, Oliveri P, Rentsch F, Benton R. A cnidarian homologue of an insect gustatory receptor functions in developmental body patterning. *Nat Commun*. 2015;6:6243.
120. Apostolopoulou AA, Mazija L, Wust A, Thum AS. The neuronal and molecular basis of quinine-dependent bitter taste signaling in *Drosophila* larvae. *Front Behav Neurosci*. 2014;8:6.
121. Choi J, van Giesen L, Choi MS, Kang K, Sprecher SG, Kwon JY. A pair of pharyngeal gustatory receptor neurons regulates caffeine-dependent ingestion in *Drosophila* larvae. *Front Cell Neurosci*. 2016;10:181.
122. Delventhal R, Carlson JR. Bitter taste receptors confer diverse functions to neurons. *Elife*. 2016;5:e11181.
123. Fan P, Manoli DS, Ahmed OM, Chen Y, Agarwal N, Kwong S, Cai AG, Neitz J, Renslo A, Baker BS, Shah NM. Genetic and neural mechanisms that inhibit *Drosophila* from mating with other species. *Cell*. 2013;154:89–102.
124. Ioannidis P, Simao FA, Waterhouse RM, Manni M, Seppey M, Robertson HM, Misof B, Niehuis O, Zdobnov EM. Genomic features of the damselfly *Calopteryx splendens* representing a sister clade to most insect orders. *Genome Biol Evol*. 2017;9:415–30.
125. Kent LB, Robertson HM. Evolution of the sugar receptors in insects. *BMC Evol Biol*. 2009;9:41.
126. Kwon JY, Dahanukar A, Weiss LA, Carlson JR. Molecular and cellular organization of the taste system in the *Drosophila* larva. *J Neurosci*. 2011;31:15300–9.
127. Lee Y, Kang MJ, Shim J, Cheong CU, Moon SJ, Montell C. Gustatory receptors required for avoiding the insecticide L-canavanine. *J Neurosci*. 2012;32:1429–35.
128. Lee Y, Moon SJ, Montell C. Multiple gustatory receptors required for the caffeine response in *Drosophila*. *Proc Natl Acad Sci U S A*. 2009;106:4495–500.
129. Lee Y, Moon SJ, Wang YJ, Montell C. A *Drosophila* gustatory receptor required for strychnine sensation. *Chem Senses*. 2015;40:525–33.
130. Liman ER, Zhang YV, Montell C. Peripheral coding of taste. *Neuron*. 2014;81:984–1000.
131. Ling F, Dahanukar A, Weiss LA, Kwon JY, Carlson JR. The molecular and cellular basis of taste coding in the legs of *Drosophila*. *J Neurosci*. 2014;34:7148–64.
132. Missbach C, Dweck HKM, Vogel H, Vilcinskas A, Stensmyr MC, Hansson BS, Grosse-Wilde E. Evolution of insect olfactory receptors. *Elife*. 2014;3:e02115.
133. Miyamoto T, Slone J, Song XY, Amrein H. A fructose receptor functions as a nutrient sensor in the *Drosophila* brain. *Cell*. 2012;151:1113–25.
134. Montell C. A taste of the *Drosophila* gustatory receptors. *Curr Opin Neurobiol*. 2009;19:345–53.
135. Moon SJ, Lee Y, Jiao Y, Montell C. A *Drosophila* gustatory receptor essential for aversive taste and inhibiting male-to-male courtship. *Curr Biol*. 2009;19:1623–7.
136. Andrews JC, Fernández MP, Yu Q, Leary GP, Leung AK, Kavanaugh MP, Kravitz EA, Certel SJ. Octopamine neuromodulation regulates Gr32a-linked aggression and courtship pathways in *Drosophila* males. *PLoS Genet*. 2014;10:e1004356.
137. Barbagallo B, Garrity PA. Temperature sensation in *Drosophila*. *Curr Opin Neurobiol*. 2015;34:8–13.
138. Benton R, Vannice KS, Gomez-Diaz C, Vossell LB. Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell*. 2009;136:149–62.
139. Bray S, Amrein H. A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron*. 2003;39:1019–29.
140. Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ, Benton R. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet*. 2010;6:e1001064.
141. Ejima A, Griffith LC. Courtship initiation is stimulated by acoustic signals in *Drosophila melanogaster*. *PLoS One*. 2008;3:e3246.
142. Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. Evolution of a complex locus: exon gain, loss and divergence at the Gr39a locus in *Drosophila*. *PLoS One*. 2008;3:e1513.
143. Kim H, Jeong YT, Choi MS, Choi J, Moon SJ, Kwon JY. Involvement of a Gr2a-expressing *Drosophila* pharyngeal gustatory receptor neuron in regulation of aversion to high-salt foods. *Mol Cells*. 2017;40:331–8.
144. Miyamoto T, Amrein H. Suppression of male courtship by a *Drosophila* pheromone receptor. *Nat Neurosci*. 2008;11:874–6.
145. Ni LN, Bronk P, Chang EC, Lowell AM, Flam JO, Panzano VC, Theobald DL, Griffith LC, Garrity PA. A gustatory receptor paralogue controls rapid warmth avoidance in *Drosophila*. *Nature*. 2013;500:580–4.
146. Park JH, Kwon JY. A systematic analysis of *Drosophila* gustatory receptor gene expression in abdominal neurons which project to the central nervous system. *Mol Cells*. 2011;32:375–81.
147. Shankar S, Chua JY, Tan KJ, Calvert MEK, Weng RF, Ng WC, Mori K, Yew JY. The neuropeptide tachykinin is essential for pheromone detection in a gustatory neural circuit. *Elife*. 2015;4:e06914.
148. Shim J, Lee Y, Jeong YT, Kim Y, Lee MG, Montell C, Moon SJ. The full repertoire of *Drosophila* gustatory receptors for detecting an aversive compound. *Nat Commun*. 2015;6:8867.
149. Thorne N, Amrein H. Atypical expression of *Drosophila* gustatory receptor genes in sensory and central neurons. *J Comp Neurol*. 2008;506:548–68.
150. Thorne N, Chromey C, Bray S, Amrein H. Taste perception and coding in *Drosophila*. *Curr Biol*. 2004;14:1065–79.
151. Wang LM, Han XQ, Mehren J, Hiroi M, Billeter JC, Miyamoto T, Amrein H, Levine JD, Anderson DJ. Hierarchical chemosensory regulation of male-male social interactions in *Drosophila*. *Nat Neurosci*. 2011;14:757–62.
152. Watanabe K, Toba G, Koganezawa M, Yamamoto D. Gr39a, a highly diversified gustatory receptor in *Drosophila*, has a role in sexual behavior. *Behav Genet*. 2011;41:746–53.
153. Weiss LA, Dahanukar A, Kwon JY, Banerjee D, Carlson JR. The molecular and cellular basis of bitter taste in *Drosophila*. *Neuron*. 2011;69:258–72.
154. Xiang Y, Yuan QA, Vogt N, Looger LL, Jan LY, Jan YN. Light-avoidance-mediating photoreceptors tile the *Drosophila* larval body wall. *Nature*. 2010;468:921–6.
155. Ai M, Blais S, Park JY, Min S, Neubert TA, Suh GS. Ionotropic glutamate receptors IR64a and IR8a form a functional odorant receptor complex in vivo in *Drosophila*. *J Neurosci*. 2013;33:10741–9.
156. Ai M, Min S, Grosjean Y, Leblanc C, Bell R, Benton R, Suh GS. Acid sensing by the *Drosophila* olfactory system. *Nature*. 2010;468:691–5.
157. Croset V, Schleyer M, Arguello JR, Gerber B, Benton R. A molecular and neuronal basis for amino acid sensing in the *Drosophila* larva. *Sci Rep*. 2016;6:34871.
158. Enjin A, Zaharieva EE, Frank DD, Mansourian S, Suh GS, Gallio M, Stensmyr MC. Humidity sensing in *Drosophila*. *Curr Biol*. 2016;26:1352–8.
159. Ganguly A, Pang L, Duong VK, Lee A, Schoniger H, Varady E, Dahanukar A, Molecular A. Cellular context-dependent role for Ir76b in detection of amino acid taste. *Cell Rep*. 2017;18:737–50.
160. Grosjean Y, Rytz R, Farine JP, Abuin L, Cortot J, Jefferis GSXE, Benton R. An olfactory receptor for food-derived odours promotes male courtship in *Drosophila*. *Nature*. 2011;478:236–U123.
161. Hussain A, Zhang M, Ucpunar HK, Svensson T, Quillery E, Gompel N, Ignell R, Kadow ICG. Ionotropic chemosensory receptors mediate the taste and smell of polyamines. *PLoS Biol*. 2016;14:e1002454.
162. Joseph RM, Sun JS, Tam E, Carlson JR. A receptor and neuron that activate a circuit limiting sucrose consumption. *Elife*. 2017;6:e24992.
163. Knecht ZA, Silbering AF, Cruz J, Yang L, Croset V, Benton R, Garrity PA. Ionotropic receptor-dependent moist and dry cells control hygrosensation in *Drosophila*. *Elife*. 2017;6:e26654.
164. Knecht ZA, Silbering AF, Ni LN, Klein M, Budelli G, Bell R, Abuin L, Ferrer AJ, Samuel ADT, Benton R, Garrity PA. Distinct combinations of variant ionotropic glutamate receptors mediate thermosensation and hygrosensation in *Drosophila*. *Elife*. 2016;5:e17879.

165. Koh TW, He Z, Gorur-Shandilya S, Menez K, Larter NK, Stewart S, Carlson JR. The *Drosophila* IR20a clade of ionotropic receptors are candidate taste and pheromone receptors. *Neuron*. 2014;83:850–65.
166. Ni L, Klein M, Svec KV, Budelli G, Chang EC, Ferrer AJ, Benton R, Samuel ADT, Garrity PA. The ionotropic receptors IR21a and IR25a mediate cool sensing in *Drosophila*. *Elife*. 2016;5:e13254.
167. Prieto-Godino LL, Rytz R, Bargeton B, Abuin L, Arguello JR, Dal Peraro M, Benton R. Olfactory receptor pseudo-pseudogenes. *Nature*. 2016;539:93–7.
168. Prieto-Godino LL, Rytz R, Cruchet S, Bargeton B, Abuin L, Silbering AF, Ruta V, Dal Peraro M, Benton R. Evolution of acid-sensing olfactory circuits in Drosophilids. *Neuron*. 2017;93:661–7.
169. Rytz R, Croset V, Benton R. Ionotropic receptors (IRs): chemosensory ionotropic glutamate receptors in *Drosophila* and beyond. *Insect Biochem Mol Biol*. 2013;43:888–97.
170. Sanchez-Alcaniz JA, Silbering AF, Croset V, Zappia G, Sivasubramaniam AK, Abuin L, Sahai SY, Munch D, Steck K, Auer TO, et al. An expression atlas of variant ionotropic glutamate receptors identifies a molecular basis of carbonation sensing. *Nat Commun*. 2018;9:4252.
171. Stewart S, Koh TW, Ghosh AC, Carlson JR. Candidate ionotropic taste receptors in the *Drosophila* larva. *Proc Natl Acad Sci U S A*. 2015;112:4195–201.
172. Zhang YLV, Ni JF, Montell C. The molecular basis for attractive salt-taste coding in *Drosophila*. *Science*. 2013;340:1334–8.
173. Ebrahim SA, Dweck HK, Stokl J, Hofferberth JE, Trona F, Weniger K, Rybak J, Seki Y, Stensmyr MC, Sachse S, et al. *Drosophila* avoids parasitoids by sensing their semiochemicals via a dedicated olfactory circuit. *PLoS Biol*. 2015;13:e1002318.
174. Gong DP, Zhang HJ, Zhao P, Lin Y, Xia QY, Xiang ZH. Identification and expression pattern of the chemosensory protein gene family in the silkworm, *Bombyx mori*. *Insect Biochem Mol Biol*. 2007;37:266–77.
175. Grillet M, Darteville L, Ferveur JF. A *Drosophila* male pheromone affects female sexual receptivity. *Proc Biol Sci*. 2006;273:315–23.
176. Hallem EA, Carlson JR. Coding of odors by a receptor repertoire. *Cell*. 2006;125:143–60.
177. Hallem EA, Ho MG, Carlson JR. The molecular basis of odor coding in the *Drosophila* antenna. *Cell*. 2004;117:965–79.
178. Hieu TT, Jung J, Kim SI, Ahn YJ, Kwon HW. Behavioural and electroantennogram responses of the stable fly (*Stomoxys calcitrans* L.) to plant essential oils and their mixtures with attractants. *Pest Manag Sci*. 2014;70:163–72.
179. Kim DH, Kim SI, Chang KS, Ahn YJ. Repellent activity of constituents identified in *Foeniculum vulgare* fruit against *Aedes aegypti* (Diptera: Culicidae). *J Agric Food Chem*. 2002;50:6993–6.
180. Leader DP, Krause SA, Pandit A, Davies SA, Dow JA. FlyAtlas 2: a new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic Acids Res*. 2018;46:D809–15.
181. Marshall B, Warr CG, de Bruyne M. Detection of volatile indicators of illicit substances by the olfactory receptors of *Drosophila melanogaster*. *Chem Senses*. 2010;35:613–25.
182. Olafson PU. Molecular characterization and immunolocalization of the olfactory co-receptor Orco from two blood-feeding muscid flies, the stable fly (*Stomoxys calcitrans*, L.) and the horn fly (*Haematobia irritans irritans*, L.). *Insect Mol Biol*. 2013;22:131–42.
183. Ray A, van Naters W, Shiraiwa T, Carlson JR. Mechanisms of odor receptor gene choice in *Drosophila*. *Neuron*. 2007;53:353–69.
184. Sánchez-Gracia A, Rozas J. Divergent evolution and molecular adaptation in the *Drosophila* odorant-binding protein family: inferences from sequence variation at the OS-E and OS-F genes. *BMC Evol Biol*. 2008;8:323.
185. Stocker RF. Design of the larval chemosensory system. In: Technau GM, editor. *Brain development in Drosophila melanogaster*. New York, NY: Springer New York; 2008. p. 69–81.
186. Termtanasombat M, Mitsuno H, Misawa N, Yamahira S, Sakurai T, Yamaguchi S, Nagamune T, Kanzaki R. Cell-based odorant sensor array for odor discrimination based on insect odorant receptors. *J Chem Ecol*. 2016;42:716–24.
187. Wanner KW, Willis LG, Theilmann DA, Isman MB, Feng Q, Plettner E. Analysis of the insect OS-D-like gene family. *J Chem Ecol*. 2004;30:889–911.
188. Vieira FG, Sanchez-Gracia A, Rozas J. Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome Biol*. 2007;8:R235.
189. Vieira FG, Rozas J. Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome Biol Evol*. 2011;3:476–90.
190. Brito NF, Moreira MF, Melo ACA. A look inside odorant-binding proteins in insect chemoreception. *J Insect Physiol*. 2016;95:51–65.
191. Graham LA, Davies PL. The odorant-binding proteins of *Drosophila melanogaster*: annotation and characterization of a divergent gene family. *Gene*. 2002;292:43–55.
192. Findlay GD, Yi X, Maccoss MJ, Swanson WJ. Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLoS Biol*. 2008;6:e178.
193. Li S, Picimbono JF, Ji SD, Kan YC, Qiao CL, Zhou JJ, Pelosi P. Multiple functions of an odorant-binding protein in the mosquito *Aedes aegypti*. *Biochem Biophys Res Commun*. 2008;372:464–8.
194. Scolari F, Benoit JB, Michalkova V, Aksoy S, Abd-Alla AMM, Malacrida AR, Aksoy S, Attardo GM. The Spermatophore in *Glossina morsitans morsitans*: insights into male contributions to reproduction. *Sci Rep*. 2016;6:20334.
195. Brand P, Robertson HM, Lin W, Pothula R, Klingeman WE, Jurat-Fuentes JL, Johnson BR. The origin of the odorant receptor gene family in insects. *Elife*. 2018;7:e38340.
196. Pitts RJ, Rinker DC, Jones PL, Rokas A, Zwiebel LJ. Transcriptome profiling of chemosensory appendages in the malaria vector *Anopheles gambiae* reveals tissue- and sex-specific signatures of odor coding. *BMC Genomics*. 2011;12:271.
197. Allan SA, Day JF, Edman JD. Visual ecology of biting flies. *Annu Rev Entomol*. 1987;32:297–316.
198. Beersma DGM, Stavenga DG, Kuiper JW, Lattice R. Visual field, and binocularities in flies - dependence on species and sex. *J Comp Physiol*. 1977;119:207–20.
199. Chayka SY, Mazokhin-Porshnyakov GA. Ultrastructure of the compound eye of the stable fly, *Stomoxys calcitrans*. *Entomol Rev*. 1986;65:148–54.
200. Salcedo E, Huber A, Henrich S, Chadwell LV, Chou WH, Paulsen R, Britt SG. Blue- and green-absorbing visual pigments of *Drosophila*: ectopic expression and physiological characterization of the R8 photoreceptor cell-specific Rh5 and Rh6 rhodopsins. *J Neurosci*. 1999;19:10716–26.
201. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
202. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 2000;17:540–52.
203. Tamura K, Stecher G, Peterson D, Filipiski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30:2725–9.
204. Wernet MF, Perry MW, Desplan C. The evolutionary diversity of insect retinal mosaics: common design principles and emerging molecular logic. *Trends Genet*. 2015;31:316–28.
205. Ni JD, Baik LS, Holmes TC, Montell C. A rhodopsin in the brain functions in circadian photoentrainment in *Drosophila*. *Nature*. 2017;545:340–4.
206. Ballard RC. Responses of *Stomoxys calcitrans* (L.) to radiant energy and their relation to absorption characteristics of the eye. *Ann Entomol Soc Am*. 1958;51:449–64.
207. Fortini ME, Rubin GM. Analysis of cis-acting requirements of the Rh3 and Rh4 genes reveals a bipartite organization to rhodopsin promoters in *Drosophila melanogaster*. *Genes Dev*. 1990;4:444–63.
208. Andere AA, li RNP, Ray DA, Picard CJ. Genome sequence of *Phormia regina* Meigen (Diptera: Calliphoridae): implications for medical, veterinary and forensic research. *Bmc Genomics*. 2016;17:842.
209. Chen ZX, Sturgill D, Qu J, Jiang H, Park S, Boley N, Suzuki AM, Fletcher AR, Plachetzki DC, FitzGerald PC, et al. Comparative validation of the *D. melanogaster* modENCODE transcriptome annotation. *Genome Res*. 2014;24:1209–23.
210. Giraldo-Calderon GI, Zanis MJ, Hill CA. Retention of duplicated long-wavelength opsins in mosquito lineages by positive selection and differential expression. *BMC Evol Biol*. 2017;17:84.
211. Armsen D, Rajakumar R, Friedrich M, Benoit JB, Robertson HM, Panfilio KA, Ahn SJ, Poelchau MF, Chao H, Dinh H, et al. The genome of the water strider *Gerris buenoi* reveals expansions of gene repertoires associated with adaptations to life on the water. *BMC Genomics*. 2018;19:832.
212. Frentiu FD, Bernard GD, Cuevas CI, Sison-Mangus MP, Prudic KL, Briscoe AD. Adaptive evolution of color vision as seen through the eyes of butterflies. *Proc Natl Acad Sci U S A*. 2007;104(Suppl 1):8634–40.
213. Frentiu FD, Bernard GD, Sison-Mangus MP, Brower AV, Briscoe AD. Gene duplication is an evolutionary mechanism for expanding spectral diversity in

- the long-wavelength photopigments of butterflies. *Mol Biol Evol.* 2007;24:2016–28.
214. Chen A, Kim H, Mayer R, Norman J. Vitellogenesis in the stable fly, *Stomoxys calcitrans*. *Comp Biochem Physiol B.* 1987;88:897–903.
  215. Barnett T, Pachel C, Gergen JP, Wensink PC. The isolation and characterization of *Drosophila* yolk protein genes. *Cell.* 1980;21:729–38.
  216. White NM, Bownes M. Cloning and characterization of three *Musca domestica* yolk protein genes. *Insect Mol Biol.* 1997;6:329–41.
  217. Martinez A, Bownes M. The sequence and expression pattern of the *Calliphora erythrocephala* yolk protein a and B genes. *J Mol Evol.* 1994;38:336–51.
  218. Hens K, Macours N, Claeys I, Francis C, Huybrechts R. Cloning and expression of the yolk protein of the tsetse fly *Glossina morsitans morsitans*. *Insect Biochem Mol Biol.* 2004;34:1281–7.
  219. Tufail M, Takeda M. Molecular characteristics of insect vitellogenins. *J Insect Physiol.* 2008;54:1447–58.
  220. Stevens L. Egg proteins: what are their functions? *Sci Prog.* 1996;79(Pt 1):65–87.
  221. Sievers F, Higgins DG. Clustal omega. *Curr Protoc Bioinformatics.* 2014;48:313.11–6.
  222. Bownes M. Why is there sequence similarity between insect yolk proteins and vertebrate lipases? *J Lipid Res.* 1992;33:777–90.
  223. Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet.* 2007;8:762–75.
  224. Meola SM. Morphology of the region of the ejaculatory duct producing the male accessory gland material in the stable fly, *Stomoxys calcitrans* L. (Diptera:Muscidae). *Int J Insect Morphol Embryol.* 1982;11:69–77.
  225. Findlay GD, MacCoss MJ, Swanson WJ. Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*. *Genome Res.* 2009;19:886–96.
  226. Sirot LK, Hardstone MC, Helinski ME, Ribeiro JM, Kimura M, Deewatthanawong P, Wolfner MF, Harrington LC. Towards a semen proteome of the dengue vector mosquito: protein identification and potential functions. *PLoS Negl Trop Dis.* 2011;5:e989.
  227. Pilch B, Mann M. Large-scale and high-confidence proteomic analysis of human seminal plasma. *Genome Biol.* 2006;7:R40.
  228. Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, Singh RS. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics.* 2007;177:1321–35.
  229. Meisel RP. Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. *Mol Biol Evol.* 2011;28:1893–900.
  230. Baldini F, Gabrieli P, Rogers DW, Catteruccia F. Function and composition of male accessory gland secretions in *Anopheles gambiae*: a comparison with other insect vectors of infectious diseases. *Pathog Glob Health.* 2012;106:82–93.
  231. Boes KE, Ribeiro JM, Wong A, Harrington LC, Wolfner MF, Sirot LK. Identification and characterization of seminal fluid proteins in the Asian tiger mosquito, *Aedes albopictus*. *PLoS Negl Trop Dis.* 2014;8:e2946.
  232. Ram KR, Wolfner MF. Sustained post-mating response in *Drosophila melanogaster* requires multiple seminal fluid proteins. *PLoS Genet.* 2007;3:e238.
  233. Sirot LK, LaFlamme BA, Sitnik JL, Rubinstein CD, Avila FW, Chow CY, Wolfner MF. Molecular social interactions: *Drosophila melanogaster* seminal fluid proteins as a case study. *Adv Genet.* 2009;68:23–56.
  234. Peng Y, Grassl J, Millar AH, Baer B. Seminal fluid of honeybees contains multiple mechanisms to combat infections of the sexually transmitted pathogen *Nosema apis*. *Proc Biol Sci.* 2016;283:20151785.
  235. Volff JN. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays.* 2006;28:913–22.
  236. Ribeiro JMC. Blood-feeding arthropods - live syringes or invertebrate pharmacologists. *Inf Agents Dis Rev Issues Comment.* 1995;4:143–52.
  237. Ribeiro JMC, Arca B. From sialomes to the sialoverse: an insight into salivary portion of blood-feeding insects. *Adv Insect Physiol.* 2009;37(37):59–118.
  238. Ribeiro JMC, Topalis P, Louis C. Anoxcel: an *Anopheles gambiae* protein database. *Insect Mol Biol.* 2004;13:449–57.
  239. Wang XY, Ribeiro JMC, Broce AB, Wilkerson MJ, Kanost MR. An insight into the transcriptome and proteome of the salivary gland of the stable fly, *Stomoxys calcitrans*. *Insect Biochem Mol Biol.* 2009;39:607–14.
  240. Ameri M, Wang X, Wilkerson MJ, Kanost MR, Broce AB. An immunoglobulin binding protein (Antigen 5) of the stable fly (Diptera: Muscidae) salivary gland stimulates bovine immune responses. *J Med Entomol.* 2008;45:94–101.
  241. Zhang D, Cupp MS, Cupp EW. Thrombostasin: purification, molecular cloning and expression of a novel anti-thrombin protein from horn fly saliva. *Insect Biochem Mol Biol.* 2002;32:321–30.
  242. Ribeiro JMC, Francischetti IMB. Role of arthropod saliva in blood feeding: Sialome and post-sialome perspectives. *Annu Rev Entomol.* 2003;48:73–88.
  243. Ribeiro JMC, Mans BJ, Arca B. An insight into the sialome of blood-feeding Nematocera. *Insect Biochem Mol Biol.* 2010;40:767–84.
  244. Lee RM, Davies DM. Feeding in the stable fly, *Stomoxys calcitrans* (Diptera: Muscidae). I. Destination of blood, sucrose solution and water in the alimentary canal, the effects of age on feeding, and blood digestion. *J Med Entomol.* 1979;15:541–54.
  245. Takac P, Nunn MA, Meszaros J, Pechanova O, Vrbjar N, Vlasakova P, Kozanek M, Kazimirova M, Hart G, Nuttall PA, Labuda M. Vasotab, a vasoactive peptide from horse fly *Hybomitra bimaculata* (Diptera, Tabanidae) salivary glands. *J Exp Biol.* 2006;209:343–52.
  246. Bloomquist JR. Chloride channels as tools for developing selective insecticides. *Arch Insect Biochem Physiol.* 2003;54:145–56.
  247. Ding Y, Ortelli F, Rossiter LC, Hemingway J, Ranson H. The *Anopheles gambiae* glutathione transferase supergene family: annotation, phylogeny and expression profiles. *BMC Genomics.* 2003;4:35.
  248. Enayati AA, Ranson H, Hemingway J. Insect glutathione transferases and insecticide resistance. *Insect Mol Biol.* 2005;14:3–8.
  249. Ffrench-Constant RH, Daborn PJ, Le Goff G. The genetics and genomics of insecticide resistance. *Trends Genet.* 2004;20:163–70.
  250. Ffrench-Constant RH, Williamson MS, Davies TG, Bass C. Ion channels as insecticide targets. *J Neurogenet.* 2016;30:163–77.
  251. Grutter T, Changeux JP. Nicotinic receptors in wonderland. *Trends Biochem Sci.* 2001;26:459–63.
  252. Jones AK, Bera AN, Lees K, Sattelle DB. The cys-loop ligand-gated ion channel gene superfamily of the parasitoid wasp, *Nasonia vitripennis*. *Heredity (Edinb).* 2010;104:247–59.
  253. Jones AK, Grauso M, Sattelle DB. The nicotinic acetylcholine receptor gene family of the malaria mosquito, *Anopheles gambiae*. *Genomics.* 2005;85:176–87.
  254. Jones AK, Raymond-Delpech V, Thany SH, Gauthier M, Sattelle DB. The nicotinic acetylcholine receptor gene family of the honey bee, *Apis mellifera*. *Genome Res.* 2006;16:1422–30.
  255. Jones AK, Sattelle DB. The cys-loop ligand-gated ion channel gene superfamily of the red flour beetle, *Tribolium castaneum*. *BMC Genomics.* 2007;8:327.
  256. Li XC, Schuler MA, Berenbaum MR. Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annu Rev Entomol.* 2007;52:231–53.
  257. Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, Unger MF, Collins FH, Feyereisen R. Evolution of supergene families associated with insecticide resistance. *Science.* 2002;298:179–81.
  258. Sattelle DB, Jones AK, Sattelle BM, Matsuda K, Reenan R, Biggin PC. Edit, cut and paste in the nicotinic acetylcholine receptor gene family of *Drosophila melanogaster*. *Bioessays.* 2005;27:366–76.
  259. Sogorb MA, Vilanova E. Enzymes involved in the detoxification of organophosphorus, carbamate and pyrethroid insecticides through hydrolysis. *Toxicol Lett.* 2002;128:215–28.
  260. Strode C, Wondji CS, David JP, Hawkes NJ, Lumjuan N, Nelson DR, Drane DR, Karunaratne SH, Hemingway J, Black WCT, Ranson H. Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. *Insect Biochem Mol Biol.* 2008;38:113–23.
  261. Temeyer KB, Chen AC. Acetylcholinesterase of *Stomoxys calcitrans* (L.) (Diptera: Muscidae): cDNA sequence, baculovirus expression, and biochemical properties. *Vet Parasitol.* 2012;184:92–5.
  262. Tu CP, Akgul B. *Drosophila* glutathione S-transferases. *Methods Enzymol.* 2005;401:204–26.
  263. Oakshott JG, Claudianos C, Campbell PM, Newcomb RD, Russell RJ. Biochemical genetics and genomics of insect esterases. In: Gilbert LI, Gill SS, editors. *Insect pharmacology: channels, receptors, toxins and enzymes.* London UK: Elsevier; 2005. p. 229–301.
  264. Niwa R, Niwa YS. Enzymes for ecdysteroid biosynthesis: their biological functions in insects and beyond. *Biosci Biotechnol Biochem.* 2014;78:1283–92.
  265. Schuler MA, Berenbaum MR. Structure and function of cytochrome P450s in insect adaptation to natural and synthetic toxins: insights gained from molecular modeling. *J Chem Ecol.* 2013;39:1232–45.
  266. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25:1972–3.

267. Barnes KG, Weedall GD, Ndula M, Irving H, Mzihalowa T, Hemingway J, Wondji CS. Genomic footprints of selective sweeps from metabolic resistance to pyrethroids in African malaria vectors are driven by scale up of insecticide-based vector control. *PLoS Genet.* 2017;13:e1006539.
268. Le Goff G, Hilliou F. Resistance evolution in *Drosophila*: the case of CYP6G1. *Pest Manag Sci.* 2017;73:493–9.
269. Seifert J, Scott JG. The CYP6D1v1 allele is associated with pyrethroid resistance in the house fly, *Musca domestica*. *Pestic Biochem Physiol.* 2002;72:40–4.
270. Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell.* 2014;158:1431–43.
271. Kaufman TC, Seeger MA, Olsen G. Molecular and genetic organization of the *antennapedia* gene complex of *Drosophila melanogaster*. *Adv Genet.* 1990;27:309–62.
272. Simeone A, Dapice MR, Nigro V, Casanova J, Graziani F, Acampora D, Avantaggiato V. *Orthopedia*, a novel homeobox-containing gene expressed in the developing CNS of both mouse and *Drosophila*. *Neuron.* 1994;13:83–101.
273. Andersson MN, Newcomb RD. Pest control compounds targeting insect chemoreceptors: another silent spring? *Front Ecol Evol.* 2017;5:5.
274. LaBrecque GC, Patterson RS, Williams DF, Weidhaas DE. Control of the stable fly, *Stomoxys calcitrans* (Diptera: Muscidae), on St. Croix, U.S. Virgin Islands, using integrated pest management measures. I. Feasibility of sterile male releases. *J Med Entomol.* 1981;18:194–6.
275. Patterson RS, LaBrecque GC, Williams DF, Weidhaas DE. Control of the stable fly, *Stomoxys calcitrans* (Diptera: Muscidae), on St. Croix, U.S. Virgin Islands, using integrated pest management measures. III. Field techniques and population control. *J Med Entomol.* 1981;18:203–10.
276. Williams DF, Patterson RS, LaBrecque GC, Weidhaas DE. Control of the stable fly, *Stomoxys calcitrans* (Diptera: Muscidae), on St. Croix, U.S. Virgin Islands, using integrated pest management measures. II. Mass rearing and sterilization. *J Med Entomol.* 1981;18:197–202.
277. Konganti K, Guerrero FD, Schilkey F, Ngam P, Jacobi JL, Umale PE, Perez de Leon AA, Threadgill DW. A whole genome assembly of the horn fly, *Haematobia irritans*, and prediction of genes with roles in metabolism and sex determination. *G3 (Bethesda).* 2018;8:1675–86.
278. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A.* 2011;108:1513–8.
279. Tsai IJ, Otto TD, Berriman M. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol.* 2010;11:R41.
280. Glossina Genomes Consortium: *Stomoxys calcitrans* genome sequencing and assembly. NCBI BioProject, PRJNA188117, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA188117>; 2015.
281. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 2009;75:5737–41.
282. Lee E, Helt GA, Reese JT, Munoz-Torres MC, Childers CP, Buels RM, Stein L, Holmes IH, Elisk CG, Lewis SE. Web Apollo: a web-based genomic annotation editing platform. *Genome Biol.* 2013;14:R93.
283. Giraldo-Calderon GI, Emrich SJ, MacCallum RM, Maslen G, Dialynas E, Topalis P, Ho N, Gesing S, VectorBase C, Madey G, et al. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Res.* 2015;43:D707–13.
284. Wheeler D, Redding AJ, Werren JH. Characterization of an ancient lepidopteran lateral gene transfer. *PLoS One.* 2013;8:e59262.
285. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–60.
286. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics.* 2018;34:867–8.
287. Benoit JB, Attardo GM, Michalkova V, Krause TB, Bohova J, Zhang Q, Baumann AA, Mireji PO, Takac P, Denlinger DL, et al. A novel highly divergent protein family identified from a viviparous insect by RNA-seq analysis: a potential target for tsetse fly-specific abortifacients. *PLoS Genet.* 2014;10:e1003874.
288. Rosendale AJ, Romick-Rosendale LE, Watanabe M, Dunlevy ME, Benoit JB. Mechanistic underpinnings of dehydration stress in the American dog tick revealed through RNA-Seq and metabolomics. *J Exp Biol.* 2016;219:1808–19.
289. Rosendale AJ, Dunlevy ME, McCue MD, Benoit JB. Progressive behavioural, physiological and transcriptomic shifts over the course of prolonged starvation in ticks. *Mol Ecol.* 2019;28:49–65.
290. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20.
291. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J. g: Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* 2016;44:W83–9.
292. Rao X, Huang X, Zhou Z, Lin X. An improvement of the 2<sup>Δ(-delta delta CT)</sup> method for quantitative real-time polymerase chain reaction data analysis. *Biostat Bioinforma Biomath.* 2013;3:71–85.
293. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007;23:2947–8.
294. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59:307–21.
295. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003;19:1572–4.
296. Miller MA, Schwartz T, Pickett BE, He S, Klem EB, Scheuermann RH, Passarotti M, Kaufman S, O’Leary MA. A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evol Bioinforma.* 2015;11:43–8.
297. Loytynoja A, Goldman N. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics.* 2010;11:579.
298. Sanchez R, Serra F, Tarraga J, Medina I, Carbonell J, Pulido L, de Maria A, Capella-Gutierrez S, Huerta-Cepas J, Gabaldon T, et al. Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Res.* 2011;39:W470–4.
299. Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. R package version 2.40.0; 2020.
300. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421.
301. Martins LA, Malossi CD, Galletti M, Ribeiro JM, Fujita A, Esteves E, Costa FB, Labruna MB, Daffre S, Fogaca AC. The transcriptome of the salivary glands of *Amblyomma aureolatum* reveals the antimicrobial peptide microplasin as an important factor for the tick protection against *Rickettsia rickettsii* infection. *Front Physiol.* 2019;10:529.
302. Esteves E, Maruyama SR, Kawahara R, Fujita A, Martins LA, Righi AA, Costa FB, Palmisano G, Labruna MB, Sa-Nunes A, et al. Analysis of the salivary gland transcriptome of unfed and partially fed *Amblyomma sculptum* ticks and descriptive proteome of the saliva. *Front Cell Infect Microbiol.* 2017;7:476.
303. Karim S, Ribeiro JM. An insight into the sialome of the Lone Star Tick, *Amblyomma americanum*, with a glimpse on its time dependent gene expression. *PLoS One.* 2015;10:e0131292.
304. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38:576–89.
305. Glossina Genomes Consortium: *Stomoxys calcitrans* RefSeq genome sequencing. NCBI BioProject PRJNA288986, <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA288986>; 2015.

## Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

