

RESEARCH ARTICLE

De Novo Transcriptome Analysis of the Common New Zealand Stick Insect *Clitarchus hookeri* (Phasmatodea) Reveals Genes Involved in Olfaction, Digestion and Sexual Reproduction

Chen Wu^{1,2*}, Ross N. Crowhurst³, Alice B. Dennis^{4,5}, Victoria G. Twort^{1,2}, Shanlin Liu⁶, Richard D. Newcomb^{2,3}, Howard A. Ross², Thomas R. Buckley^{1,2}

1 Landcare Research, Auckland, New Zealand, **2** School of Biological Sciences, The University of Auckland, Auckland, New Zealand, **3** New Zealand Institute for Plant & Food Research Ltd, Auckland, New Zealand, **4** Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland, **5** EAWAG, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland, **6** China National GeneBank, BGI-Shenzhen, Shen Zhen, China

* wuc@landcareresearch.co.nz



OPEN ACCESS

Citation: Wu C, Crowhurst RN, Dennis AB, Twort VG, Liu S, Newcomb RD, et al. (2016) *De Novo* Transcriptome Analysis of the Common New Zealand Stick Insect *Clitarchus hookeri* (Phasmatodea) Reveals Genes Involved in Olfaction, Digestion and Sexual Reproduction. PLoS ONE 11(6): e0157783. doi:10.1371/journal.pone.0157783

Editor: J Joe Hull, USDA-ARS, UNITED STATES

Received: March 15, 2016

Accepted: June 3, 2016

Published: June 23, 2016

Copyright: © 2016 Wu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Raw data are available from the NCBI database (accession numbers: SRR3080264, SRR3080247 and SRR3080266). Transcriptome assembly and genomic scaffolds are available from Dryad (doi:10.5061/dryad.p52sq).

Funding: This work was supported by award number Massey University RM13790/24947, Allan Wilson Centre (<http://www.allanwilsoncentre.ac.nz/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Phasmatodea, more commonly known as stick insects, have been poorly studied at the molecular level for several key traits, such as components of the sensory system and regulators of reproduction and development, impeding a deeper understanding of their functional biology. Here, we employ *de novo* transcriptome analysis to identify genes with primary functions related to female odour reception, digestion, and male sexual traits in the New Zealand common stick insect *Clitarchus hookeri* (White). The female olfactory gene repertoire revealed ten odorant binding proteins with three recently duplicated, 12 chemosensory proteins, 16 odorant receptors, and 17 ionotropic receptors. The majority of these olfactory genes were over-expressed in female antennae and have the inferred function of odorant reception. Others that were predominantly expressed in male terminalia ($n = 3$) and female midgut ($n = 1$) suggest they have a role in sexual reproduction and digestion, respectively. Over-represented transcripts in the midgut were enriched with digestive enzyme gene families. *Clitarchus hookeri* is likely to harbour nine members of an endogenous cellulase family (glycoside hydrolase family 9), two of which appear to be specific to the *C. hookeri* lineage. All of these cellulase sequences fall into four main phasmid clades and show gene duplication events occurred early in the diversification of Phasmatodea. In addition, *C. hookeri* genome is likely to express *γ*-proteobacteria *pectinase* transcripts that have recently been shown to be the result of horizontal transfer. We also predicted 711 male terminalia-enriched transcripts that are candidate accessory gland proteins, 28 of which were annotated to have molecular functions of peptidase activity and peptidase inhibitor activity, two groups being widely reported to regulate female reproduction through proteolytic cascades. Our study has yielded new insights into the genetic basis of odour detection, nutrient

Competing Interests: The authors have declared that no competing interests exist.

digestion, and male sexual traits in stick insects. The *C. hookeri* reference transcriptome, together with identified gene families, provides a comprehensive resource for studying the evolution of sensory perception, digestive systems, and reproductive success in phasmids.

Introduction

Stick insects (Phasmatodea) are a diverse group of insects with over 3000 described species. They occur throughout much of the world, with the greatest species diversity occurring in tropical and subtropical regions [1]. They are experts in camouflage, as both their morphology and movements resemble twigs and foliage. Phasmids across the globe have been the subject of studies on a variety of traits, including the patterns of geographic variation and evolution, sensory control of movement, variation of reproductive modes, evolution of body form, and intestinal digestion [2–8]. Most recently the genus *Timema*, endemic to North America, has become a successful model for addressing evolutionary questions, especially surrounding the genomic basis of speciation [9, 10]. However, compared with insects from other orders, the molecular basis of phasmid biological function has been poorly studied, impeding our understanding of interesting phasmid traits.

Next generation sequencing (NGS) methods such as RNA-Seq now provide the opportunity to greatly enrich molecular resources for non-model species. The relative ease with which RNA-Seq can be utilized in both *de novo* assembly and read mapping to assembled transcripts facilitates the exploration of gene diversity and expression in non-model organisms. Such technology has already been applied to stick insects in a few instances [5, 8, 11–16] with only four of them focusing on the study of phasmid functional traits. Midgut transcriptomes were sequenced to examine plant cell wall degrading enzymes (PCWDEs) in *Peruphasma schultei*, *Sipyloidea sipyilus*, *Aretaon asperrimus*, *Extatosoma tiaratum*, *Medauroidea extradentata* and *Ramulus artemis* [5]. The presence of cellulase (GH9), cellobiase (GH1), pectinase (GH28) and β -1,3-glucanase (GH16) transcripts in the stick insect anterior midgut demonstrated that, unlike some insects that rely on bacterial symbionts to digest some components of PCWs, these stick insects have the ability to independently digest cellulose into glucose and also other components such as pectin and chitin substrates to obtain carbon [5, 17, 18]. More recently, exploration of RNA-Seq from *Phyllium siccifolium* antennae identified olfactory receptors (ORs), ionotropic receptors (IRs) and gustatory receptors (GRs), revealing a fully developed olfactory receptor-based system in phasmids [14]. Also, to investigate expression of the genes associated with cold temperature adaptation in the alpine New Zealand stick insect species, RNA-Seq was used to sequence heads and prothoraxes of both alpine and lowland species including those belonging to the genera *Micrarchus*, *Tectarchus*, *Niveaphasma*, *Asteliaphasma*, *Argosarchus*, *Spinotectarchus*, *Acanthoxyla* and *Clitarchus* [12, 13, 15]. Despite these studies, transcriptomic data from stick insects has still not been collected comprehensively across tissues, sexes, and species.

The New Zealand stick insect fauna includes 24 species from 10 genera, and are thought to have originated from New Caledonia approximately 24–45 million years ago [19, 20]. They now occupy habitats from the cold alpine zone to dry coastal scrub across the country [15, 21]. *Clitarchus hookeri* is one of the most common New Zealand stick insect species, and is widely distributed throughout the North Island and across large areas of the South Island [4, 22]. It is an ecological generalist, found in both low altitude forest and scrub habitat [22, 23]. Compared with other New Zealand species, *C. hookeri* has a relatively large body size, with adult females

ranging in size from 80 to 106 mm while males are slimmer and shorter (approximately 70 mm) [24]. Their bodies are narrow and mostly cylindrical with smooth or slightly wrinkled exoskeletons, and are often bright green or brown [24]. The main native host plants of *C. hookeri* are *Leptospermum scoparium* (Manuka) and *Kunzea sp.* (Kanuka), but they can also be found on other plants, from the families of *Myrtaceae*, *Polygonaceae*, *Laxmanniaceae* and *Rubiaceae* [4, 22], as well as the introduced species of *Rubus*.

The most genetically varied populations of *C. hookeri* occur in the northern half of the North Island, and this genetic variation decreases dramatically from north to south [4, 22]. It has been suggested that during the Last Glacial Maximum, this species was restricted to several refugia located on the northern half of North Island, followed by subsequent population expansion across the country [22]. Molecular resources from *C. hookeri* are lacking, restricted to a few neutral gene markers that were sequenced and used for phylogenetic construction [4, 19, 20, 22, 25] and a head and prothorax transcriptome that was constructed for the study of cold temperature adaptation [15].

In order to examine digestive enzymes, olfaction-associated proteins and accessory gland proteins in a New Zealand stick insect species, we carried out an RNA-Seq experiment by sequencing the transcriptomes of *C. hookeri* female antennae, midgut and male terminalia. To construct a more complete reference transcriptome for the species for future use, we also integrated the *C. hookeri* composite leg, head and prothorax transcriptomes that have been constructed for the use in other studies.

Materials and Methods

Sample information

The *C. hookeri* transcriptome presented in this paper was constructed from five different types of tissue derived from nine adults collected in February 2013. These insects were all from the same colony inhabiting on Manuka/Kanuka trees in Totara Park, Auckland, New Zealand (37° 0.111 S, 174°55.039 E) (Table 1). This collection site was on Auckland Council administered land and we collected samples under permits issued by the Auckland Council. We use the term ‘tissue’ to describe all of these samples in the following text for simplicity; although some samples are more accurately described as ‘organs’ or ‘regions of the body’. Insects were directly snap-frozen and stored at –80°C after collection.

RNA isolation and deep sequencing

Tissue for RNA extraction was removed using sterilised scalpel blades, with the midgut dissection being carried out in 100% ethanol followed by removal of the gut contents. The RNA extraction and sequencing from head and prothorax is described elsewhere [15]. The leg transcriptome was previously sequenced for the 1KITE project on the ‘evolution of insects’ (NCBI accession number: PRJNA286351); however, the RNA extraction and sequencing have not been described elsewhere. We therefore describe them below.

Table 1. Sample summary.

Tissue type	Sample code	Sex	Assembly available
Antenna	CLI602	Female	Current study
Midgut	CLI602	Female	Current study
Terminalia	CLI601	Male	Current study
Head and prothorax	CLI303-305, CLI308-310	Female	Yes [15]
Leg	CLI524	Female	Yes (For 1KITE project)

doi:10.1371/journal.pone.0157783.t001

Antennae, midgut and terminalia. RNA extraction was conducted separately on the antenna, midgut and terminalia tissues using the same method. Total RNA was extracted using TRIzol RNA extraction reagent (Life Technologies), following grinding in liquid nitrogen, according to the manufacturer's protocol. Subsequent RNA clean-up was performed using the RNeasy Mini Kit (Qiagen). Quality and concentration of RNA extractions were assessed using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). cDNA libraries were constructed using the Illumina TruSeq RNA Sample Preparation Kit v2 (Illumina San Diego, CA, USA) according to the manufacturer's protocol. Individual libraries had randomly assigned barcodes ligated to the cDNA fragments and were amplified with 12 cycles of PCR. Library quality was assessed using an Agilent 2100 Bioanalyzer. Extractions were pooled together for sequencing on the Illumina HiSeq2000 platform at New Zealand Genomics Limited (<http://www.nzgenomics.co.nz/>) to generate 100 bp paired end (PE) reads. All reads have been deposited in the Sequence Read Archive (SRA) at NCBI (accession numbers: SRR3080264, SRR3080247, SRR3080266).

Leg. Forelegs were preserved in RNAlater and sent to BGI Shenzhen for RNA extraction, library preparation and sequencing. mRNA was isolated using the Dynabeads mRNA Purification Kit (Invitrogen, Grand Island, NY, USA) and subsequently sheared using RNA fragmentation reagent (Ambion, Austin, Texas, US) at 72°C. These cleaved RNA fragments were transcribed into first-strand cDNA using SuperScript™II Reverse Transcriptase (Invitrogen, Grand Island, NY, USA) and random N6 primer (IDT). The second-strand cDNA was then synthesized using RNase H (Invitrogen, Grand Island, NY, USA) and DNA polymerase I (New England BioLabs, Ipswich, MA, USA). The double-stranded cDNA then underwent end-repair, a single 'A' base addition, adapter ligation, and size selection on agarose gels (250 ± 20 bp). The product was then indexed and PCR amplified to finalize library preparation for the paired-end cDNA. Verification of the cDNA fragment size and concentration was accomplished using an Agilent 2100 Bioanalyzer and an ABI StepOnePlus Real-Time PCR machine. The cDNA library was subsequently sequenced on an Illumina HiSeq2000 with 150 bp PE reads, generating approximately 2.5 Gb raw data.

De novo transcriptome assembly and assessment

The *C. hookeri* transcriptome assembly was produced by combining three *de novo* assemblies: the pooled antennae, midgut and terminalia assembly, and the two previously assembled leg, and head and prothorax assemblies. The head and prothorax assembly is described elsewhere [15].

Antennae, midgut and terminalia. Illumina sequencing reads were pooled together and quality assessed with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). To reduce the effects from inherent sequencing biases resulting from library preparation [26, 27], raw reads were trimmed by removing eight nucleotides (nt) from the 5' end using PRINSEQ (v.0.20.3) [28]. To maximise quality, reads were then processed to remove ambiguities using PRINSEQ, trim adapters and low quality regions (Phred < 30) using cutadapt (v1.2.1) [29]. Remaining reads that were less than 50 bp in length, and those no longer paired were removed using PRINSEQ.

To optimise the completeness of the transcriptome and the number of full-length transcripts generated from the pooled three libraries, the *de novo* assembly was obtained by combining assembled contigs derived from Trinity (v20140413) [30], SOAPdenovo-trans (v1.03) [31], and Velvet/Oases (v0.2.08) [32]. The Kmer sizes 21, 25, 29, and sizes from 35 to 85 with stepwise increase of ten were used in SOAPdenovo-trans and Velvet/Oases, and scaffold gaps

of each assembly were filled by GapCloser (<http://sourceforge.net/projects/soapdenovo2/files/GapCloser/>). Each final transcript was given an ID consisting of species, assembly version, assembler, Kmer size if applied, and an unique seven digit ID, eg. Chv1VELVK21_0000001.

Leg. Raw data were preprocessed by removing reads of poor quality including: 1) reads with adapter; 2) reads with a total number of >10 Ns; 3) reads with >50 base pairs of low quality (Phred quality score = 2, ASCII 66 “B”, Illumina 1.5+ Phred+64). Transcripts were assembled using SOAPdenovo-trans (v1.01) with a Kmer of 31. The gapped sequences were filled using GapCloser. Each contig ID was replaced by the code ChvKITESOAP with a 7 digit ID, which vKITE stands for the assembly version produced for the project of ‘evolution in insects’ (GenBank bioproject accession: PRJNA286351) conducted by 1K Insect Transcriptome Evolution (<http://www.1kite.org/>).

Final assembly and assessment. The final transcriptome of coding transcripts was created by combining a total of 19 assemblies described above which were then run through the EvidentialGene tr2aacds pipeline (<http://arthropods.eugenes.org/EvidentialGene/>). This pipeline employs CD-HIT-EST (v3.1.1) [33] to cluster transcripts with identity of 90% or greater and then select the ‘best’ contig based on the open reading frame (ORF) and UTRs present from each cluster as the final coding transcript. The resulting transcript set was then quality evaluated with Core Eukaryotic Genes Mapping Approach (CEGMA: v2.4) [34] software, which detects a core protein set of 248 highly conserved proteins that are found in a wide range of eukaryotes. The cleaned reads were also mapped back to both the raw and final transcriptomes using Bowtie2 (v2.1.0) [35] to produce a coverage profile of the assemblies from short reads. To estimate levels of xenobiotic sequence contamination, transcripts were assigned to different taxonomic categories (eg. arthropod, fungi, bacteria, and virus) with Assemblage (<https://github.com/sujaikumar/assemblage>).

Gene functional annotation

De novo transcripts were annotated using Blastx (v2.2.28) [36] (e-value < $1e^{-5}$) to search against the GenBank non-redundant (*nr*) protein database (downloaded in April 2015). Transcripts lacking Blast hits were searched for conserved protein domains with InterProScan [37] within Blast2GO (v2.8) [38]. Gene Ontology (GO) terms were assigned using Blast2GO.

Identification of chemosensory proteins

As many chemosensory genes are often duplicated in a lineage-specific manner and are divergent from previously annotated chemosensory genes at both the nucleotide and amino acid levels, particularly in understudied lineages (eg. phasmids), we employed additional identification steps to optimise the completeness of these proteins in *C. hookeri*. First, the predicted *C. hookeri* initial set of chemosensory transcripts were generated using proteins from closely related insect lineages, if available, to search the homologues against the *C. hookeri* transcriptome assembly (tBlastn, evalue < 10^{-5}). The searching set included 29 *Zootermopsis nevadensis* OBPs [39], three *Eurycantha calcarata* CSPs [40], all the *P. siccifolium* ORs, IRs and GRs [14], and two *Epiphyas postvittana* SNMPs [41]. Second, the resulting *C. hookeri* homologous transcripts were translated and used to additionally search the homologues against the transcriptome assembly with the same Blast search. Third, all the predicted *C. hookeri* chemosensory transcripts were searched against the GenBank *nr* protein database (Blastx; e-value < 10) to confirm their protein identities. Putative OBPs and CSPs were additionally confirmed based on the presence of conserved cysteines. Finally, the identified transcripts were aligned to a *C. hookeri* genome assembly (Wu et al. in prep for publication) to rule out mis-assembled sequences.

Transcript abundance and differential expression analysis

Transcript quantification for each tissue, based on the combined transcriptome, was carried out with RSEM (v1.2.11) [42]. Transcript abundances were measured by mapping short reads to the assembly using Bowtie2 and calculating a maximum likelihood abundance with a credibility interval for counts of genes or isoforms to determine levels of differential expression. Estimates of transcript abundance were produced in the form of two measures, the expected counts (ECS) and transcripts per million (TPM). The ECS is an estimate of the number of fragments that can be derived from an isoform or gene, which is calculated with the maximum likelihood expression levels using an Expectation-Maximisation (EM) algorithm. In this study, transcripts were considered present in the library if ECS was larger than one. TPM estimates the fraction of transcripts within the sample that are represented by the given isoform or gene [43]. A data matrix containing ECS from all libraries was imported into EBSeq (v1.10.0) [44] to find significantly enriched transcripts in one tissue library relative to the others. EBSeq is an R package utilizing an empirical Bayesian approach to model the features observed in RNA-Seq data. It was used as it (i) allows multi-class testing among RNA-Seq datasets and can generate all possible expression patterns given by the conditions rather than only calculating pair-wise comparisons, (ii) estimates cross-condition variance as the transcript variance by pooling similar genes together for the comparison among samples without replicates, and (iii) has been shown to have a relatively low false discovery rate and calculate expression of isoforms more accurately compared with other algorithms that have been used to detect differentially expressed genes [45]. Here, we considered a gene to have tissue-biased expression if it was detected with posterior probability of being differentially expressed (PPDE) greater than 0.95 (this is equivalent to false discovery rate < 0.05). The gene was also considered as significantly over-represented in a tissue library if it achieved a expression fold change of at least one when compared with every other tissue libraries. GO enrichments for each gene set with tissue-biased expression were calculated using Fisher's Exact Test with multiple testing correction of false discovery rate (Benjamini and Hochberg) less than 0.05 on Blast2GO.

Phylogenetic analysis

Protein alignments of OBP and olfactory receptor co-receptor (ORCO) were generated using MUSCLE [46] with the default options in Geneious (v7.1.7) [47] and curated using Gblocks (v0.91b) [48]. Rooted phylogenetic trees were constructed using PhyML (v2.0) [49] under the best fit models (OBP: VT+I+G; ORCO: LG+G+F) selected by Prottest (v3.4) [50]. A non-parametric bootstrap analysis was conducted to generate support for branching topology ($n = 500$ bootstrap pseudoreplicates).

The GH9 cellulase gene phylogeny was constructed using protein coding nucleotide sequences. The GH9 sequences were aligned using MUSCLE translation alignment method with default gap penalty parameters in Geneious. Rooted phylogenetic trees were constructed using PhyML under the best fitting substitution model (TIM2+I+G) selected by JModelTest2 (v2.1.6) [51] with 500 bootstrap pseudoreplicates.

Tests for branch selection in ORCO

Tests for branch selection were implemented in the CODEML program from PAML (v4.8) [52] using a branch-site model available for an arbitrary number of ratios (model = 2, Nsites = 2). The null and alternative hypotheses for this test were inferred by giving an initial ω value of one and an option of estimate respectively. The inference of positive selection on the given phylogenetic branch was conducted using likelihood ratio tests (LRTs) between nested

models. The ORCO protein phylogenetic tree generated above using *P. siccifolium* and *C. hookeri* branches as foreground lineages was used as the tree topology imported into PAML.

Identification of putative seminal fluid proteins (SFPs)

ORFs of transcripts over-represented in the male terminalia were identified with TransDecoder [53], a plugin from the Trinity package. Signal peptides and sub-cellular locations were identified using SignalP (v4.1) [54] and ProtComp (v9.0) (<http://linux1.softberry.com>), respectively.

Results and Discussion

De novo transcriptome assembly

The *C. hookeri* transcriptome constructed in the current study integrates multiple assemblies derived from five different tissues. The pooled set of approximately 113 million PE reads (Table 2) sequenced from antenna, midgut and terminalia were first quality controlled and then subjected to *de novo* assembly to construct an initial transcriptome assembly. The more than two million contigs generated from three *de novo* assemblers under the use of various K-mers were then merged together to produce a set of contigs. The leg transcriptome assembly was produced from 25,969,776 PE reads (150 bp; GenBank accession: SRR2230521) and contained 146,050 contigs. This assembly, after filtering cross-species contamination, has been deposited at DDBJ/EMBL/GenBank as a Transcriptome Shotgun Assembly project under the accession GDVG00000000 (vGDVG01000000). The Head and prothorax transcriptome was assembled from 56,112,518 single reads with the read length of 50 bp and the assembly contained 84,307 contigs (Data available on the Dryad database: <http://datadryad.org/resource/doi:10.5061/dryad.kc826>) [15].

The final *C. hookeri* transcriptome assembly was generated from merging the three assemblies described above to create a super-set of contigs, and then redundancy, potentially miss-assembled and non-coding transcripts were removed. This yielded a consensus transcript set comprising 77,140 contigs made up of 69.8 Mbp (Table 3) (Dryad doi:10.5061/dryad.p52sq).

Assessment of assembly

The quality and completeness of the combined *C. hookeri* transcript set was assessed using CEGMA by searching for the presence of highly conserved eukaryotic core orthologues. Out of the 248 core proteins, 235 (94.76%) were completely present (alignment length defined as longer than 70%) and four proteins were found with partial sequences (alignment length defined between 50% and 70%). In addition, there was an average of 2.8 orthologues per core protein and 76.2% of the CEGMA proteins detected in the *C. hookeri* transcriptome were present in multiple transcripts. Aligning raw data against the assembly revealed 92.3% of PE reads and 96.5% of head and prothorax reads mapped back to the raw assembly (assembly pre Evidential-Gene pipeline). In addition, the percentage of Blast matches to fungal, plant, protist, bacterial and viral accessions was about 2.5%, indicating very low levels of xenobiotic RNA

Table 2. Transcriptome assembly summary.

	Raw pairs	Clean pairs
Antenna	14,016,766	13,614,596
Midgut	19,781,777	19,269,565
Terminalia	22,538,169	21,815,983
Total	56,336,712	54,700,144

doi:10.1371/journal.pone.0157783.t002

Table 3. Summary of the transcriptome assembly.

	Contigs
Total (bp)	69,789,602
Number	77,140
Number (> 1000 bp)	17,117
N50 (bp)	1,879
Shortest (bp)	124
Longest (bp)	38,180
Mean (bp)	905
Median (bp)	407
GC%	46.91
N%	0.01

doi:10.1371/journal.pone.0157783.t003

contamination in our samples. All sets of mapping results highlight that the *C. hookeri* transcriptome encompasses the majority of expressed genes present in the assembly and therefore, is a comprehensive resource for analysing genes of interest in this species.

Gene functional annotation

More than half (56.1%) of transcripts had matching records (e-value < 10^{-5}) from GenBank (*nr*). As expected, the vast majority of the Blast top-hits (83.8%) were with known arthropod proteins, from which 78.6% of the matches came from the eusocial termite *Z. nevadensis*, followed by 9.3% and 5.9% from the red flour beetle *Tribolium castaneum* and the spider *Stegodyphus mimosarum*, respectively.

Among all the contigs with significant Blast hits, 16,518 can be assigned to at least one GO term and they in turn were further refined to GO slim terms. This analysis resulted in 9,271 (12.5%), 14,905 (19.3%), and 5,586 (7.2%) of the transcripts having an assigned term from the categories of biological process (BP; GO: 0008150), molecular function (MF; GO: 0003674) and cellular component (CC; GO: 0005575), respectively (Fig 1).

Gene expression repertoire

The distribution of shared and unique transcripts present in the libraries are shown in Fig 2. The analysis of the over-representation of library specific sequences focused on the antenna, midgut and terminalia. The number of transcripts over-represented in these three libraries were 1,286, 915 and 2,045, respectively, and 488, 416 and 713 of these were assigned with GO terms (S1 Table).

Distributions of transcripts with tissue-biased expression among the three tissue libraries and their enriched GO terms are given in Fig 3. Over-represented GO terms from MF category that were identified as being enriched in the antennae were ‘transmembrane transporter activity’, ‘signal transducer activity’ and ‘oxidoreductase activity’. These terms contain transcripts annotated as OBPs, ORs and odorant degrading enzymes largely associated with olfactory functions. In comparison, transcripts that were over-represented in the midgut library included many belonging to the families of glycoside and proteolytic hydrolase, resulting in the enriched GOs such as ‘hydrolase activity acting on glycosyl bonds’ and ‘peptidase activity’, corresponding to the function of intestinal digestion. The ‘transferase activity transferring acyl group’ was the only enriched MF term in the library of male terminalia. This group contains γ -glutamyl-transpeptidase, a protease has been reported to be enriched in *Drosophila melanogaster* seminal fluid and is involved in maintaining a protective redox environment for sperm [56].

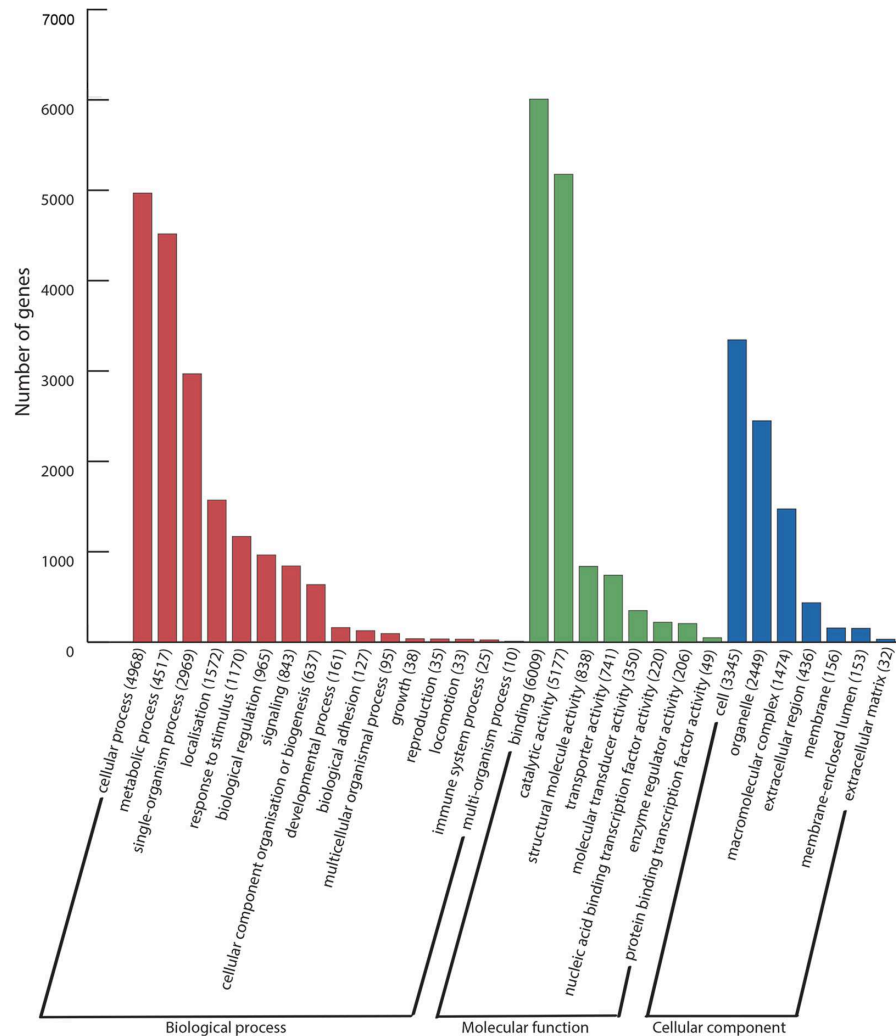


Fig 1. Distribution of the annotated transcripts into GO slim categories. Gene ontology terms of level 2 are shown.

doi:10.1371/journal.pone.0157783.g001

Female olfactory gene repertoire

Olfaction plays an important role in the life of an insect, being used to detect and respond to environmental odours, thereby locating potential mates, and finding food [57]. Gene families involved in olfaction, including OBPs, Chemosensory Binding Proteins (CSPs), ORs, IRs, and in some cases GRs, have been detected across insect orders, many of which are predominantly expressed in antennae [14, 41, 58–60]. In Phasmatodea, eight OBP N-termini from the species of *E. tiaratum*, *E. calcarata*, *Acrophylla wuelfingi*, *S. sipylus* and *Carausius morosus* [61, 62], and three partial coding sequences of CSPs from *E. calcarata* [40] have been characterised. In addition, transcripts of 12 ORs, 13 IRs and three GRs from *P. siccifolium* [14] were predicted from RNA-Seq. These studies suggest phasmids have a fully developed repertoire of olfactory gene families.

OBP and CSP. With high concentrations in the sensillum lymph of antennae, OBPs are small hydrophilic carrier proteins that are capable of binding small molecules [63]. This protein family is large, and contains subfamilies such as General Odorant Binding Proteins

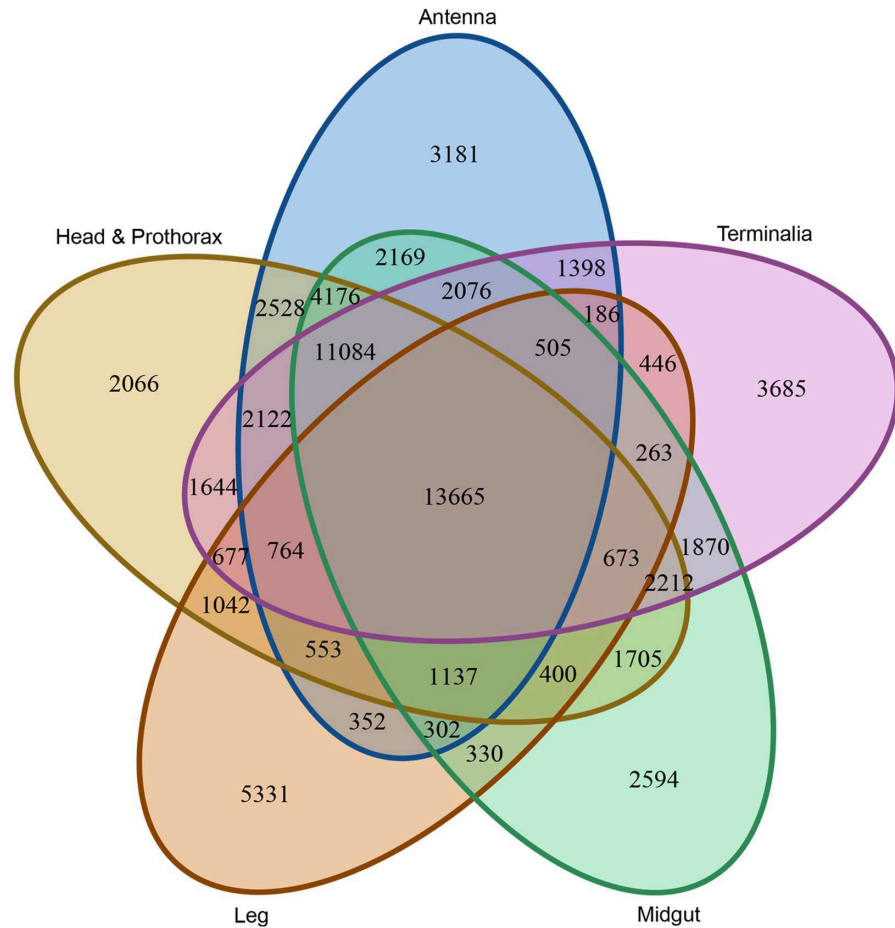


Fig 2. Venn diagram showing the distribution of shared and unique transcripts that were represented among the libraries of the antenna, midgut, terminalia, head and prothorax and leg. A transcript is considered as present in the library if it achieved an ECS value of larger than one. The Venn diagram was created using VennDiagram [55], an R package.

doi:10.1371/journal.pone.0157783.g002

(GOBPs) and Pheromone Binding Proteins (PBPs) that specialise in carrying general odorants and sex pheromone components respectively [63–65]. CSPs are smaller carrier proteins, and similar to OBPs, they are also found in the sensillum lymph, and have been shown to bind molecules including odorants and pheromones [63]. We identified ten olfactory and 12 chemosensory binding proteins in *C. hookeri*. Sequence information and alignments showing conserved cysteines are shown in Table 4 and Fig 4, respectively.

Most OBP proteins (*OBP1*–*OBP8* and *OBP10*) contain six conserved cysteines with the spacing pattern of C₁-X_{26–29}-C₂-X₃-C₃-X_{41–43}-C₄-X_{8–10}-C₅-X₈-C₆ (where X is any amino acid). The exception is *OBP9*, missing two conserved cysteines at positions two and five. The 12 *C. hookeri* CSPs all contain four conserved cysteines, with *CSP1* and *CSP3* missing amino acids at their 5' and 3' ends, respectively.

The expression level of the identified *C. hookeri* OBPs varied significantly among tissues (S1 Table). Analysis showed six OBPs (*OBP1*, *OBP2*, *OBP3*, *OBP7*, *OBP8*, and *OBP10*) and one CSP (*CSP3*) were over-represented in the female antenna library. Among these, *OBP2*, *OBP3*, and *OBP8* were most likely to be antennae-specific, as they were very lowly expressed in every

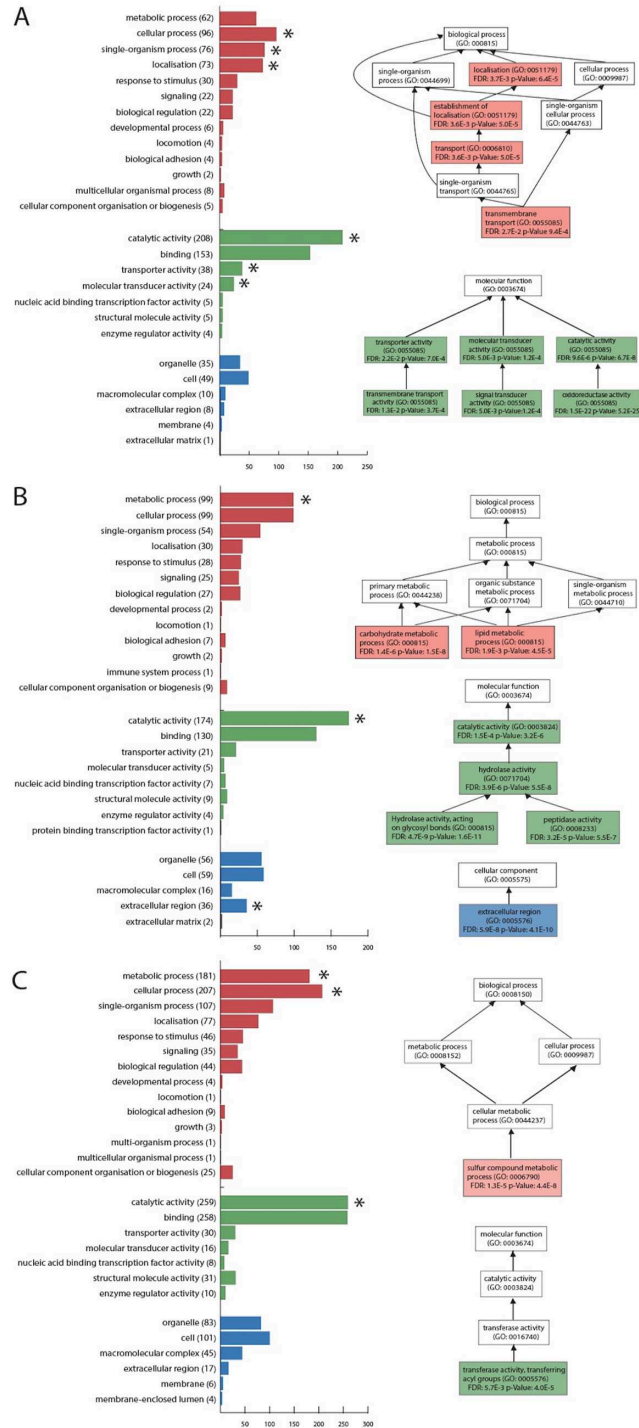


Fig 3. Distribution of the three GO categories in the (A) antenna, (B) midgut and (C) terminalia libraries and the significantly enriched GO terms detected from the transcripts with tissue-biased expression. Red, green and blue indicate the level two BP, MF and CC terms respectively, which were predicted from the transcripts showing tissue-biased expression. * indicates the term is enriched or it contains enriched terms from higher levels. The enriched GOs and their hierarchical relationships with the lower level GOs are shown on the right side.

doi:10.1371/journal.pone.0157783.g003

Table 4. Summary of *C. hookeri* putative CSP and OBP transcripts.

Category	Transcript ID	Name	Length (bp)	Length (aa)	Hit Accession	E-value
OBP	Chv0TRINITY_0005088	OBP1 ^Φ	653	158	KDR09909	2.68E-41
	Chv1SOAPK29_0141332	OBP2 ^Φ	504	155	AAN15921	1.38E-35
	Chv1SOAPK35_0145937	OBP3 ^Φ	654	150	KDR09910	6.52E-43
	Chv1VELVK45_0023357	OBP4	611	146	AID61301	2.82E-08
	Chv1VELVK45_0023358	OBP5	499	146	AID61301	9.43E-08
	Chv1VELVK45_0023359	OBP6	562	146	ABM05970	5.63E-07
	Chv1VELVK75_0000982	OBP7 ^Φ	732	148	ACR39388	4.96E-09
	Chv1VELVK85_0005072	OBP8 ^Φ	542	141	KDR13658	2.97E-35
	ChvKITESOAP_0129398	OBP9	410	135	AEP27187	2.11E-21
	ChvKITESOAP_0139181	OBP10 ^Φ	603	145	AI100991	1.00E-11
CSP	Chv1SOAPK25_0127093	CSP1*	376	94	NP_001039280	1.11E-18
	Chv1SOAPK65_0069724	CSP2	783	129	AAM77025	2.00E-36
	ChvKITESOAP_0138844	CSP3* ^Φ	590	137	AEP27186	8.34E-30
	Chv0TRINITY_0001321	CSP4	821	116	AIW65103	3.07E-12
	Chv0TRINITY_0001337	CSP5	556	118	AIW65100	3.52E-12
	Chv0TRINITY_0009092	CSP6	483	135	AAM77025	4.04E-12
	Chv1SOAPK21_0133512	CSP7	762	134	AAD30552	1.57E-37
	Chv1SOAPK25_0128356	CSP8	394	122	AIW65103	1.01E-22
	Chv1SOAPK85_0010512	CSP9	545	128	AEP95756	2.12E-19
	Chv1TRINITY_0124810	CSP10	565	124	AID61322	1.41E-30
	Chv1VELVK45_0002302	CSP11	832	130	AAM77025	6.80E-30
	Chv1VELVK65_0001291	CSP12	525	127	AGZ04911	1.13E-22

“*” indicate partial sequences and “Φ” indicates the transcript was over-represented in the antenna library when it was compared with the other four tissue libraries.

doi:10.1371/journal.pone.0157783.t004

other tissue library (FPKM < 1). In contrast, *OBP1*, *OBP7*, and *OBP10* were detected in antennae, head and prothorax, and *OBP1* and *OBP10* were also expressed in the composite leg.

OBP1 was the most abundant OBP in the *C. hookeri* female antennae (FPKM: 26,978.3), and had a highest scoring blast match to a *Z. nevadensis* PBP (Table 4). In the phylogeny, *OBP1* is closely clustered with the two *Z. nevadensis* PBPs (*OBP1* and *OBP2*) (Fig 5), indicating the *C. hookeri* *OBP1* is most likely a PBP. A recent study has demonstrated that male *C. hookeri* detects females using airborne chemical cues, most likely pheromones emitted by the females [7]. The presence of a higher density of basiconic sensilla on male than female antennae further indicates that males are responsible for long range pheromone reception [7]. Therefore, a high abundance of a PBP transcript in a *C. hookeri* female antennae seems unexpected. Interestingly, a PBP (N-terminal amino acids) has been also detected in the antennae of female *C. morosus* [62], an obligate parthenogenetic stick insect species, suggesting that the presence of the PBP was not associated with mate searching. Rather, it might be useful for females to space themselves out in the environment through detecting at least some components of sex-pheromones released from other females [66] or it could be utilised in sensing a subset of plant volatiles [67]. Furthermore, *OBP1*, *OBP2* and *OBP3* were shown to cluster phylogenetically with the two *Z. nevadensis* PBPs (Fig 5), indicating the possible presence of multiple PBPs in female *C. hookeri*. We also found *C. hookeri* *OBP8* was nested in the *Z. nevadensis* *GOBP19a* clade (*OBP5*—*OBP8* with branches coloured by blue in Fig 5), suggesting that the protein is a general odorant binding protein.

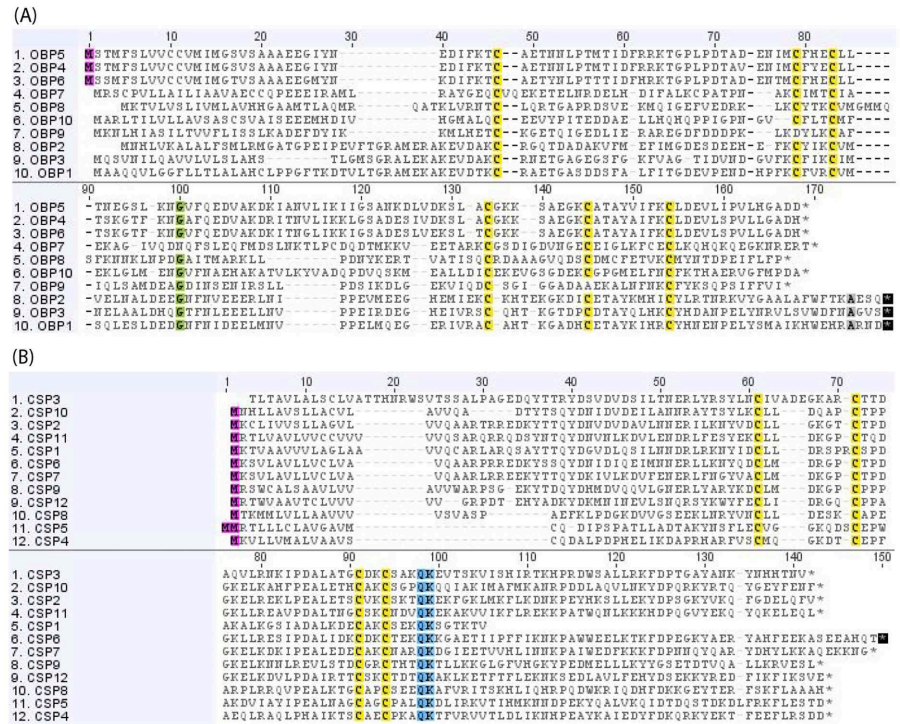


Fig 4. Amino acid alignments of predicted *C. hookeri* OBPs (A) and CSPs (B). Conserved cysteines are highlighted in yellow.

doi:10.1371/journal.pone.0157783.g004

OBPs containing less than six conserved cysteines are referred to a subgroup named non-classic Minus-C [68]. These proteins have been demonstrated to exist in the genomes of a range of insect species, such as *Ceratitis capitata*, *D. melanogaster*, *Tenebrio molitor*, and *Anopheles gambiae*, with functions always inferred as non-olfaction due to the fact that the majority are actually more highly expressed in non-olfactory organs, such as male accessory gland and salivary gland [69–71]. As a member of Minus-C subgroup, *C. hookeri* OBP9 was highly abundant in the male terminalia library (FPKM: 1,377.6), but its expression was also detected in the female antenna (FPKM: 299.8), leg (FPKM: 15.4) and head and prothorax (FPKM: 2.1), suggesting that this protein might be involved in multiple biological processes. The predominant expression in the male accessory gland suggests that OBP9 plays a role in sexual reproduction that is likely to improve offspring survival. It has been reported that an OBP detected in the seminal fluid of *Helicoverpa armigera* and *H. assulta* can be transferred to females through mating, eventually being found on the surface of fertilised eggs [72]. The presence of this protein was suggested to favour the spreading of the eggs to alleviate cannibalism among larvae [72]. Phylogenetic analysis shows OBP9 is closely related to both the PBP and GOBP clades, suggesting potential functions involved in the detection of pheromone components or general odors (Fig 5).

OBP5, OBP6, and OBP7, found to be adjacent to each other on one genomic scaffold (scaffold1029; Dryad database: <http://datadryad.org/resource/doi:10.5061/dryad.p52sq>), are present as paralogues forming a single clade in the phylogeny (Fig 5). Their high sequence similarity (83–89% amino acid similarity) may lead to spurious estimates of the expression level based on RNA-Seq methods because it is possible that the reads derived from one of the genes could have aligned with one of the other two. However, all of these genes appear to be transcribed at

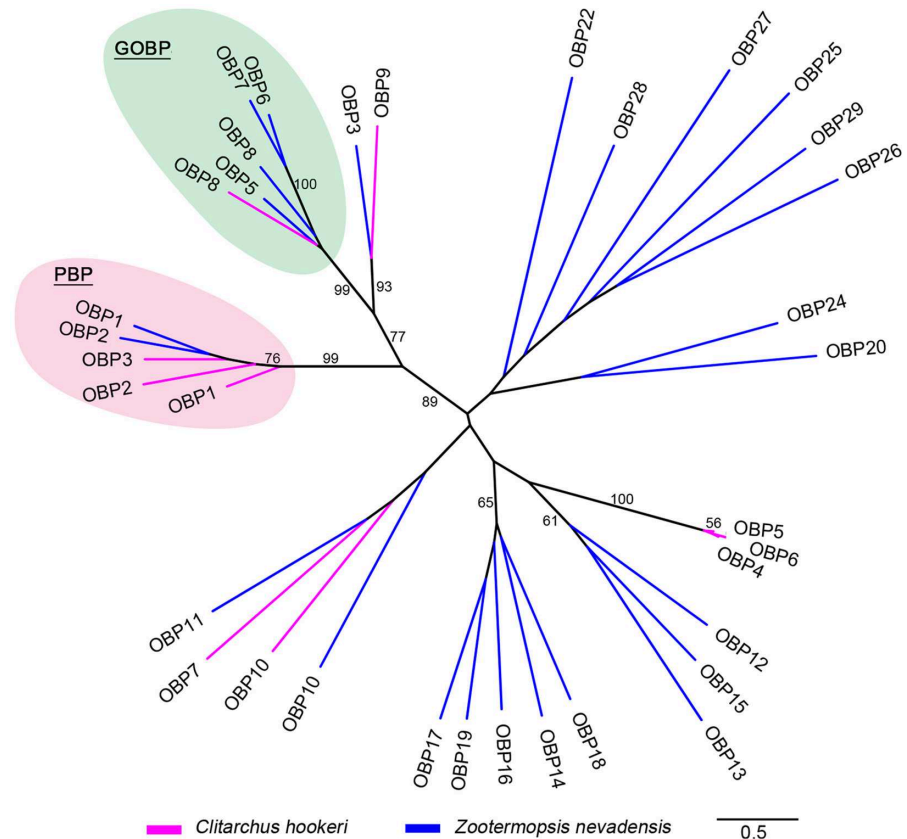


Fig 5. Phylogenetic tree for *C. hookeri* and *Z. nevadensis* OBP protein sequences. The accession IDs of the sequences downloaded from GenBank that are used for this phylogeny are in [S3 Text](#). The other *Z. nevadensis* OBP sequences from the same species were obtained from the supplementary material of Terrapon et al. [39]. Pink and blue branches represent *C. hookeri* and *Z. nevadensis* OBPs respectively. The bootstrap values larger than 50 are written next to each branch. The potential PBP and GOBP groups from the two species are circled with light pink and light blue respectively. The scale bar represents branch lengths.

doi:10.1371/journal.pone.0157783.g005

much lower levels than most of the other OBPs. *OBP5* and *OBP6* show expression in the head and prothorax library (FPKM: 24.3 and 63.5) and *OBP6* may also be expressed in antennae (FPKM: 8.2) and leg (10.6), whereas *OBP4* was only detected with a small number of transcripts from the leg library (FPKM: 6.7). Phylogenetic analysis also shows these three proteins form a single clade more closely related to *Z. nevadensis* *OBP12*, *OBP13* and *OBP15*. It is unknown which type of molecules these OBPs might be able to detect, but their recent duplication indicates they may underlie key adaptive events in the evolution of *C. hookeri*. However, low expression raises the possibility that all or some of these genes are non-functional, such as pseudogenes.

All of the identified CSPs were found to be present, but not over-represented in the antennae, except for *CSP3*, indicating they may also function as general carriers across multiple tissue types. *CSP6* was over-expressed in male terminalia (FPKM: 83.1), indicating that it potentially plays a role in sexual reproduction, while *CSP10* was over-expressed in midgut (FPKM: 9729.6), most likely functioning as a carrier in digestion.

OR, IR, SNMP, and GR. ORs and IRs present on the dendrites of olfactory sensory neurons detect a range of odorants shuttled by OBPs and CSPs from the environment, the activation of which leads to ion channel activation and receptor neuron depolarization [73–75]. The

identified *C. hookeri* putative protein sequences of these genes are shown in [S1 Text](#). Of the 16 ORs, four (*OR2*, *OR5*, *OR11* and *OR16*) are likely to encode full-length proteins. *OR16* is annotated as an odorant receptor co-receptor (ORCO), a highly conserved receptor across insects that is required for ion channel function and dendritic localisation of other ORs [76]. *Timema cristinae* ORCO was identified by performing blast searches using *C. hookeri* amino acid sequence against the genome (downloaded from http://nosil-lab.group.shef.ac.uk/?page_id=25). A phylogenetic tree of ORCO amino acid sequences shows *C. hookeri* is nested within Phasmatodea, grouping with another phasmid *P. siccifolium*, which in turn is closely related to Orthoptera (Fig 6). The branch leading to *C. hookeri* and *P. siccifolium* is considerably longer than the

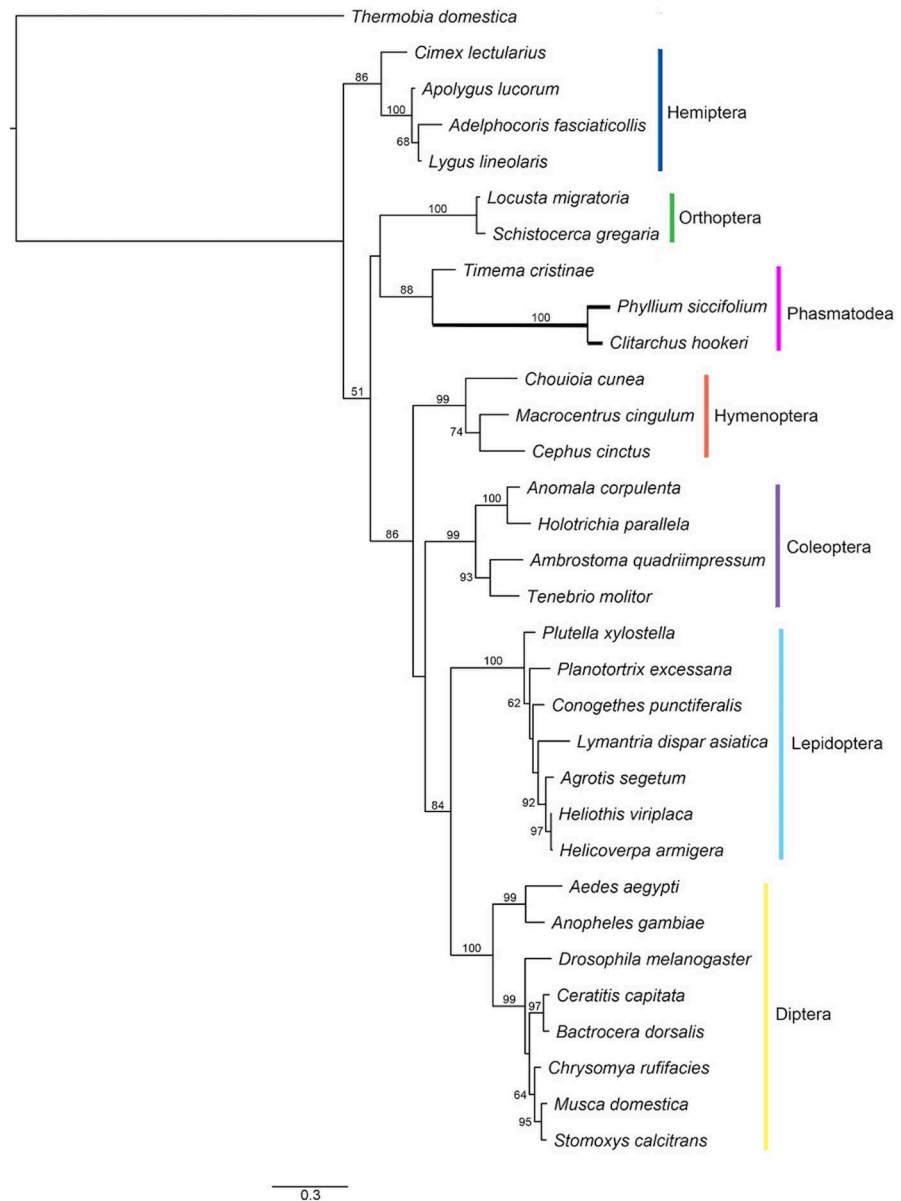


Fig 6. Phylogenetic tree of ORCOs. The coloured vertical bars indicate the sequence clades corresponding to the insect orders. The subclade branches containing *P. siccifolium* and *C. hookeri* are in bold. Scale bar represents branch length. The accession IDs of the sequences used are in [S3 Text](#).

doi:10.1371/journal.pone.0157783.g006

Table 5. Summary of *C. hookeri* putative OR, IR and SNMP OBP transcripts.

Category	Transcript ID	Name	Length (bp)	Length (aa)	Hit Accession	E-value
OR	Chv1SOAPK35_0009687	OR1*	532	168	XP_001120660	1.23E-10
	Chv1VELVK29_0021980	OR2	1567	448	KDR14144	2.16E-21
	Chv1VELVK29_0015825	OR3*	630	134	EFN78840	2.96E-14
	Chv1VELVK29_0033146	OR4*	404	106	XP_008207129	1.04E-07
	Chv1VELVK35_0019903	OR5 ^Φ	1722	452	XP_011166508	1.77E-14
	Chv1VELVK45_0034664	OR6*	594	151	AIIO1061	4.16E-24
	Chv1VELVK21_0035864	OR7* ^Φ	1072	292	BAG12817	1.99E-14
	Chv1VELVK21_0058449	OR8* ^Φ	2466	388	KDR17467	1.56E-17
	Chv1VELVK21_0062558	OR9* ^Φ	951	259	KDR14144	1.65E-21
	Chv1VELVK21_0073095	OR10*	719	193	KDR14144	9.37E-20
	Chv1VELVK21_0077746	OR11 ^Φ	1533	424	KDR14144	9.20E-16
	Chv1VELVK21_0088043	OR12*	993	291	KDR11214	8.44E-18
	Chv1VELVK25_0038317	OR13*	941	226	KDR16425	2.03E-21
	Chv1VELVK35_0042637	OR14*	880	233	AEE62637	9.90E-21
	Chv1VELVK25_0067528	OR15*	573	146	KDR14144	4.20E-21
	Chv1VELVK21_0029367	OR16 ^Φ	3087	477	KDR12002	3.11E-180
	IR	Chv1SOAPK25_0135053	IR1*	538	178	AFC91761
Chv1VELVK21_0063588		IR2* ^Φ	1092	309	AIIO1119	1.19E-76
Chv1VELVK25_0038786		IR3*	618	205	AHA80144	3.05E-47
Chv1VELVK21_0041817		IR4* ^Φ	2156	606	AHA80144	0
Chv1VELVK21_0061997		IR5	1992	644	XP_004932732	9.08E-179
Chv1VELVK21_0029672		IR6*	937	281	XP_001655464	4.70E-99
Chv1VELVK25_0009676		IR7	2419	446	XP_001661131	6.39E-23
Chv1VELVK65_0046897		IR8	2731	838	XP_008201635	0
Chv1VELVK21_0053197		IR9* ^Φ	2216	618	KDR23344	9.30E-151
Chv1VELVK21_0109764		IR10*	722	234	KDR23343	6.40E-53
Chv1VELVK25_0018305		IR11*	3370	386	XP_002431127	1.97E-101
Chv1VELVK25_0043548		IR12* ^Φ	3173	299	KDR23344	3.19E-51
Chv1VELVK55_0002278		IR13	2316	614	KDR16875	2.17E-41
Chv1VELVK21_0047941		IR14 ^Φ	2810	639	KDR14986	1.26E-163
Chv1VELVK21_0052279		IR15 ^Φ	2262	598	KDR23343	6.69E-142
Chv1VELVK21_0091436		IR16*	1720	515	KDR24341	6.42E-176
Chv1VELVK45_0004605		IR17* ^Φ	1357	335	KDR22647	2.09E-64
GR	Chv1SOAPK35_0133587	GR*	434	108	KDR10758	5.16E-14
SNMP	Chv1VELVK55_0003575	SNMP	5437	691	XP_011195945	4.55E-89

“*” indicate partial sequences and “^Φ” indicates the transcript was over-represented in the antenna library when it was compared with the other four tissue libraries.

doi:10.1371/journal.pone.0157783.t005

other branches, suggesting an accelerated rate of evolution. However, a branch selection test implemented in CODEML (from PAML (v4.8) [52]) did not show evidence of positive selection on this branch ($p = 1$). The *C. hookeri* IRs have higher sequence similarity to homologous IRs identified in *Z. nevadensis* compared to the homologous ORs identified in the two genomes (e-values in Table 5). This finding suggests the ORs likely evolve faster than the IRs in phasmids. Apart from five IRs likely to encode full-length proteins, the others only contain partial coding sequences.

Transcript abundance analysis showed six ORs (*OR5*, *OR7*, *OR8*, *OR9*, *OR11* and *OR16*) and seven IRs (*IR2*, *IR4*, *IR9*, *IR12*, *IR14*, *IR15* and *IR17*) were over-represented in the antennae library, with expression levels significantly lower than the OBPs and CSPs. Except for the antennae, short reads that were aligned to the putative OR transcripts and most of the IR transcripts from the other four tissue libraries were rare, indicating the main function of the proteins encoded by these transcripts is olfactory reception. *IR11* and *IR13* were shown to be expressed in all tissue libraries indicating they are likely to be involved in multiple roles. *IR6* and *IR7* were over-expressed in head and prothorax and male terminalia, respectively.

In addition to the two major receptor families described above that are essential for olfaction, other proteins such as sensory neuron membrane protein (SNMP) and GR may also play a role. SNMP is suggested to act in concert with ORs to capture pheromone molecules in *D. melanogaster* and the moth *Heliothis virescens* [64, 77, 78], while GR is known to be associated with taste [79]. A complete *C. hookeri* SNMP (Table 5) encoding 691 amino acids was identified, which had pairwise similarities of 24.7% and 25.4% with the two moth SNMPs, respectively. The expression of this *C. hookeri* SNMP was observed in all tissues, with the most transcripts coming from the midgut (FPKM: 65.9). A partial *C. hookeri* GR transcript encoding 108 amino acids with the highest scoring blast alignment (non-hypothetical protein) of a *Pediculus humanus corporis* GR (GenBank:XP_002426236; e-value = 2.73e-14) was also identified. The sequence information of these two transcripts are shown in Table 5 and S1 Text.

The numbers of putative *C. hookeri* OR, IR and GR are comparable to the repertoire of these gene families in *P. siccifolium* [14], with slightly higher numbers of OR and IR occurring in *C. hookeri*. It is possible that OR and IR have undergone multiple gene duplications in *C. hookeri*; however, in the absence of a complete genome sequence, it is difficult to determine if we have obtained the total number of family members. Also, the *C. hookeri* transcriptome assembly only contains transcripts derived from female antennae, while potential male-biased ORs and IRs, such as *OR7* and *OR6* in *E. postvittana* [41], may be missing. Finally, sequencing to a greater depth may identify further ORs.

In summary, the antennae showed expression of the majority of expected olfactory proteins. The total number of identified *C. hookeri* olfactory genes from our transcriptome are generally lower than the number of genes identified from the genomes of *Locusta migratoria*, *Acyrtosiphon pisum*, *T. castaneum*, and *Bombyx mori* (Fig 4a from [80]). Only IRs are more numerous than the number present in *L. migratoria* and *A. pisum*.

Digestive enzymes

Endogenous cellulase. Sequencing the *C. hookeri* midgut transcriptome enables us to examine its digestive enzymes, especially GH9 endogenous cellulase, which has previously been found to have undergone gene copy expansion [5] and have divergent functions in phasids, such as being capable of breaking down xylan, xyloglucan, and/or glucomannan [18]. The first insect gene from GH9, also called endo-beta-1,4-glucanase (EG), was initially found in a termite, and is now believed to have existed in the ancestor of all metazoans [81, 82]. High GH9 activity in the anterior midgut of stick insects has been recently confirmed in both *E. calcarata* and *Entoria okinawaensis* [6], while the prediction of two to seven cellulase transcripts from the midgut transcriptomes of six phasid species, respectively, indicate the presence of multiple gene copies in Phasmatodea genomes [5].

We identified five putative *C. hookeri* EG transcripts, named *EG1* to *EG5*, with similarity to the termite and cockroach EG proteins (percentage amino acid identity between 56.5% and 66.6%). These transcripts contain Pfam GH9 protein domains and show moderately high

Table 6. Summary of *C. hookeri* and *T. cristinae* putative GH9 genes.

Transcript ID or Genomic position	Name	Length (bp)	Length (aa)	Hit Accession	E-value
Chv1VELVK35_0068512	ChooEG1* ^Φ	1344	448	BAA34050	0
Chv1VELVK29_0092141	ChooEG2* ^Φ	1347	449	CAD54729	0
Chv1VELVK21_0117909	ChooEG3 ^Φ	1347	449	AFD33365	0
Chv1VELVK21_0101553	ChooEG4 ^Φ	1353	451	AAF80584	0
Chv1VELVK21_0022349	ChooEG5* ^Φ	945	315	CAD54729	1.55E-166
Scaffold1797 (8 exons: 75383-57750 bp)	ChooEG6	1353	451	CAD54729	0
Scaffold1797 (8 exons: 111023-85577 bp)	ChooEG7	1311	437	CAD54728	0
Scaffold1797 (9 exons: 157301-132988 bp)	ChooEG8	1317	439	CAD54728	0
Scaffold1797 (8 exons: 280626-265133 bp)	ChooEG9	1356	452	CAD54729	0
Scaffold24 (8 exons: 453210-441732 bp)	TcriEG1*	1185	395	CAD54730	9.82E-166
Scaffold24 (9 exons: 468621-458748 bp)	TcriEG2*	1311	437	CAD54730	0
Scaffold24 (9 exons: 481252-473897 bp)	TcriEG3	1353	451	AAF80584	0
Scaffold24 (9 exons: 489799-500978 bp)	TcriEG4	1347	449	CAD54728	0
Scaffold551 (299934-288976 bp)	TcriEG5*	381	127	XP_002426465	44E-50

“*” indicate partial sequences and “^Φ” indicates the transcript was over-represented in the antenna library when it was compared with the other four tissue libraries. The nucleotide length in this table indicates the length of ORF predicted from either the transcript if available or the genomic scaffolds (Dryad: doi:[10.5061/dryad.p52sq](https://doi.org/10.5061/dryad.p52sq)).

doi:10.1371/journal.pone.0157783.t006

similarity (> 57%) to known EG coding sequences on Genbank (*nt*). A comparison between these transcripts and the preliminary *C. hookeri* genome assembly indicates all the putative EG proteins are encoded from five gene copies. As expected, these cellulase genes show over-expression in the female midgut, suggesting they function in *C. hookeri* intestinal digestion and most likely, like other phasmids, play a role in degrading multiple plant cell wall components into sugar monomers [18].

Surprisingly, we also found four additional cellulase copies encoding full-length protein sequences present on the same genomic scaffold that contains *EG2*, *EG4* and *EG5* (Table 6). Three of these, *EG6*, *EG7*, and *EG8*, show pairwise amino acid identities between 84% and 92%, significantly higher than those between any two of the other putative EGs, indicating a recent duplication. The protein sequence of the other gene, *EG9*, had 80.4% identity to *EG2*. Homology searches using these EG coding sequences against both *C. hookeri* raw and final transcriptome assemblies did not yield any similar hits, indicating they were most likely absent from the transcriptome data. There are several possibilities that might lead to the absence of these EGs in our *C. hookeri* midgut transcriptome. First, the genes are silenced or transcribed at extremely low levels and could not be captured by RNA-Seq with the sequencing depth conducted by our study. Second, the construction of an RNA-Seq library may introduce library bias that leads to the false negatives. Third, the genes may be transcribed in different tissues of the body rather than midgut. It is also possible that some of them are pseudogenes, thereby lacking expression. Overall, we conclude that there are at least nine cellulase gene duplicates of ORFs present within the *C. hookeri* genome, with a minimum of five expressed in the midgut of female adults (Table 6 and Fig 7). This estimated copy number of GH9 cellulases is higher than the total number of its transcripts/genes detected from other insect lineages [5, 81, 83], indicating a complex evolutionary history of GH9 in stick insects.

EG sequences of the stick insect *T. cristinae*, sister group to all other phasmids, were identified by the use of the *C. hookeri* amino acid sequences to search against the *T. cristinae* genome. Four EGs were found from one scaffold and an additional partial sequence was detected on a

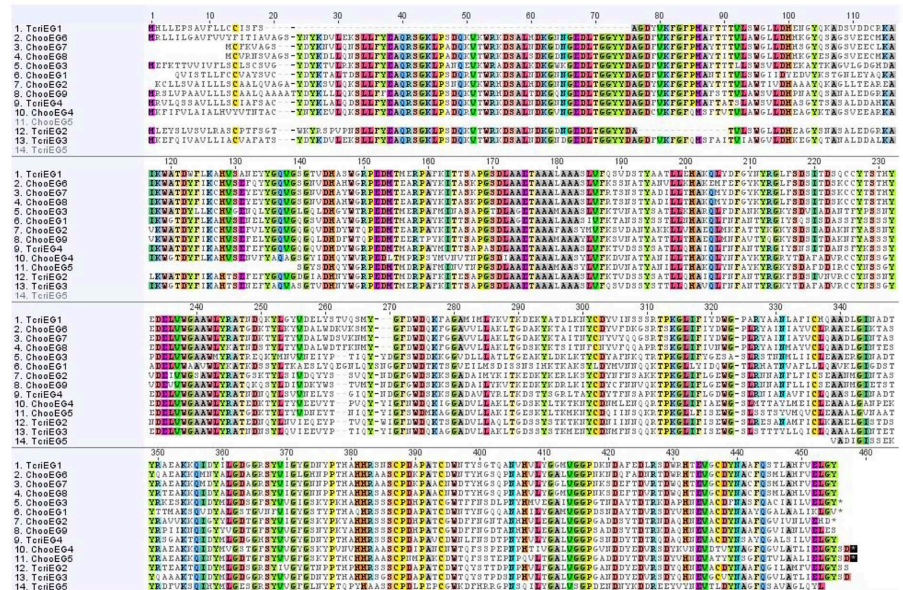


Fig 7. Amino acid alignments of predicted *C. hookeri* and *T. cristinae* EGs. The conserved regions are highlighted with colours.

doi:10.1371/journal.pone.0157783.g007

different genomic sequence, bringing the total of *T. cristinae* EGs to five (Table 6). Compared with nine *C. hookeri* gene copies, this suggests that there has been a dramatic GH9 family expansion during the evolution of phasmids.

Symbiont-independent degradation of cellulose in phasmids could have selected for gene duplications within the endogenous cellulase gene family. It has been suggested that phasmid digestion is most likely to be symbiont-independent, although molecular evidence for this is restricted to the species of *R. artemis* and *P. schultei* [17, 84]. Without the assistance of symbionts, stick insects must maintain the ability to digest cellulose independently, such as secreting their own cellulase to break down plant cell walls, and this digestive efficiency might have increased through evolving more gene copies. Having additional copies of GH9 cellulase genes may increase dosage and overall expression of these genes and enhance the efficiency of cellulose digestion without relying on enzymes produced by symbionts. Additionally, duplicated gene copies could have evolved to acquire additional functions to facilitate the digestion of other polysaccharides or oligosaccharides present in plant cell walls. On the other hand, it is also possible that some of the duplicated cellulase genes have evolved different functions and are expressed in other tissues or organs [85].

To investigate evolutionary relationships among *C. hookeri* GH9 genes and their orthologues, a maximum likelihood phylogeny was built from coding sequences (CDS) collected from a variety of insect species (Fig 8). The phylogenetic tree shows *C. hookeri* cellulases are closely related to those derived from other phasmids. The stick insect and cricket clade is nested within the cockroach clade, which is sister to a large clade containing predominantly termite sequences. In addition, phasmid sequences exhibit greater divergence among one another compared with those present in any termite and most cockroaches. Their gene duplication events occurred much earlier in the evolution of the stick insects comparing with the termite and the majority of the cockroach clades, and might be even earlier than the divergence of Phasmatoidea from other insects because the cricket (*Teleogryllus emma*) sequence is nested within the phasmid clade. The nine *C. hookeri* EGs fall into four sub-clades, with each of them rooted by

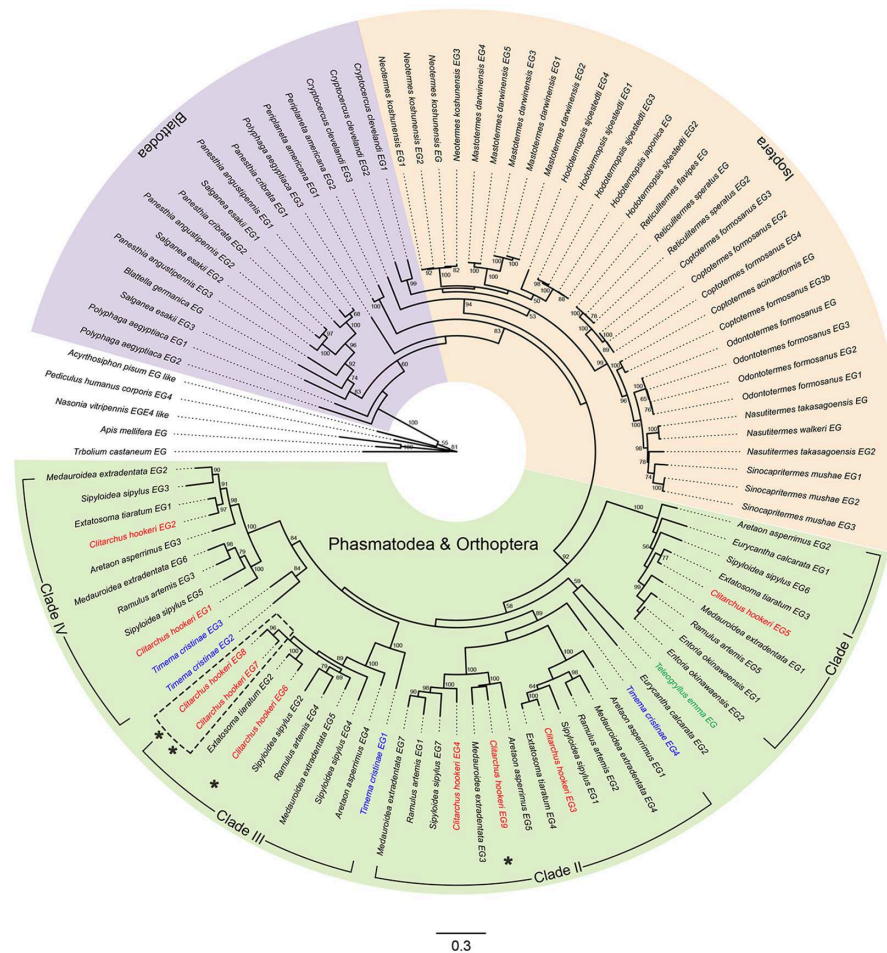


Fig 8. Phylogenetic tree of GH9. Maximum likelihood tree for insect derived GH9 family nucleotide sequences. Nodes are annotated with species and GH9 paralogue numbers that were either obtained from NCBI or given in this study. Species belonging to Phasmatodea and Orthoptera clade, Isoptera clade and Blattodea clade are coloured with backgrounds in light green, yellow and purple respectively. The branch labels of *C. hookeri*, *T. cristinae* and *Teleogryllus emma* sequences are in red, blue and green respectively. Dashed box highlights the two *C. hookeri* specific cellulase lineages. The four *C. hookeri* cellulases identified from the genome assembly are marked with '*'. Numbers above branches are bootstrap proportions and values above 50 are shown. Scale bar represents branch lengths. The accession IDs of the sequences that are used for this phylogeny are in [S3 Text](#).

doi:10.1371/journal.pone.0157783.g008

at least one *T. cristinae* sequence, except for clade I. We did not include *T. cristinae* EG5 in the phylogeny because the identified sequence was only about one third of the length of other full-length sequences. However, preliminary analysis suggests this gene is likely to be clustered into clade IV. Clade I contains *C. hookeri* EG5 as sister to the rest of clades and clade II is sister to clades III and IV. Except for clade I, each of the other clades contains two to three *C. hookeri* duplicates, whereas only clade IV contains at least two *T. cristinae* EG paralogues, which likely duplicated after the divergence of *T. cristinae* from the rest of phasmids. Clade III contains two *C. hookeri* EG paralogues (EG7 and EG8) which likely duplicated in the *C. hookeri* lineage. Those *C. hookeri* EGs that were not detected with expression in midgut all fall into the clade III, with the exception of EG9, which falls into a sub-clade of the clade II. The future discovery of more cellulase gene copies from other stick insects may reshape this phylogeny and reveal a different pattern of gene family evolution.

The EG transcripts identified from the study of six phasmid midgut transcriptomes [5], together with the nine and five genes identified from *C. hookeri* and *T. cristinae* in this study suggest gene gain or loss events are common occurrence during phasmid evolution. However, the absence of complete genomes makes it difficult to fully address this question. Further studies to characterise the expression patterns of all nine cellulases in *C. hookeri*, and confirmation of the cellulase gene copy number within other stick insect lineages may shed light on the evolutionary history of this essential plant cell wall digestive enzyme system in phasmids.

Other enzymes. Consistent with the findings from other stick insect species [5], pectinase transcripts (GH28) with the highest scoring blast alignment of γ -proteobacteria sequences were also present in the *C. hookeri* midgut library (S1 Table). These enzymes hydrolyse glycosidic bonds in pectin and are essential to plant pathogenic and saprotrophic fungi [86]. It has been found that, similar to beetle pectinases [87], pectinase transcripts present in phasmid midgut transcriptomes are endogenic and stick insects have obtained the genes through horizontal transfer from bacteria [8]. Comparison between the pectinase transcripts and the *C. hookeri* genome assembly reveals at least three genomic scaffolds (scaffold7807, scaffold1326 and scaffold5845) harbour pectinase-like genes (results not shown). However, none of these genes contains an intron to support the contention that these genes are true genome integrations. Nonetheless, two of these scaffolds (scaffold1326 and scaffold5845) harbour genes annotated to Arthropoda proteins, including those having the highest scoring blast alignments to the proteins of the species *Z. nevadensis*, *Solenopsis invicta*, *Haplopelma schmidti*, *Tribolium castaneum*, and *Megachile rotundata*. This may indicate that the *C. hookeri* pectinase genes are indeed endogenous. Future work involving sequencing of genomic regions containing a predicted pectinase gene contiguous with another *C. hookeri* gene will rule out potential mis-assemblies and confirm this finding.

The midgut over-represented transcripts also included glycosidases belonging to other gene families, such as β -galactosidase, chitinase, β -hexosaminidase subunit beta and glucosylceramidase, α -L-fucosidase, mannosyl-oligosaccharide1, 2- α -mannosidase, maltase, and trehalase (S1 Table). The expression of chitinase indicates that *C. hookeri* may have the ability to recycle chitinous cell walls of fungi as carbohydrate and nitrogen resources, similar to the wood-feeding beetle *A. glabripennis* [88]. Furthermore, the midgut over-expressed maltase (S1 Table), which hydrolyses maltose to the simple sugar glucose, and, together with the presence of an amylase in the transcriptome raises the prospect that *C. hookeri* is likely to be able to use starch as an energy source.

Seminal fluid proteins

Insect seminal fluid proteins (SFPs) are a group of secreted proteins produced from male reproductive tract tissues, predominantly accessory glands, and stored in the accessory gland lumen until required. These proteins are particularly interesting for the study of sexual reproduction because they are essential to male reproductive success and their rapid evolution likely contributes to speciation [89, 90]. The specific functions of these proteins is that they are transferred to females with sperm during mating and are then involved in regulating female reproductive functions, such as inducing antimicrobial activities, increasing ovulation and egg-laying rate, and reducing the likelihood of re-mating [91]. Hitherto, proteins identified from insect seminal fluids include proteases and protease inhibitors, lipases, oxidative stress proteins, immune proteins and cell adhesion-related proteins [56, 92, 93], but none of which have been investigated in phasmids to date.

From the transcripts that were over-represented in the library of *C. hookeri* male terminalia, we predicted 711 candidate SFP transcripts by searching for signal peptides and their protein localisations in the plasma membrane or extracellularly (S1 Table). Of these, 140 had no

Table 7. Summary of terminalia-enriched proteases or peptidases and protease inhibitors.

Name	Transcripts ID	Length (bp)	Length (aa)	FPKM	Hit Accession	E-value
Serine protease	Chv1VELVK55_0057832	1971	261	216.5	XP_001121121	2.85E-55
	Chv1VELVK35_0071739	1241	203	6.88	XP_001121121	6.22E-42
	Chv1VELVK45_0064057	1240	260	21.16	XP_317172	7.24E-57
	Chv1VELVK29_0052877	2167	285	22.71	ACX54054	9.2E-110
Aminopeptidase	Chv1VELVK21_0040553	1990	548	122.9	KDR21177	0
	Chv1VELVK25_0028117	1644	359	3.23	KDR21177	5.02E-81
	Chv1VELVK25_0028115	1901	514	9.1	KDR21177	1.8E-168
MME	Chv1VELVK55_0007423	2731	765	108.49	KDR18716	0
	Chv1VELVK21_0124071	3503	797	5.7	KDR09552	0
GGT	Chv1VELVK21_0017697	2375	695	3.12	XP_974407	0
	Chv1VELVK85_0006040	1907	580	189.19	KDR15762	0
	Chv1VELVK75_0011772	1948	597	85.17	KDR15762	0
	Chv1VELVK25_0015659	3680	942	0.51	KDR18594	0
	Chv1VELVK21_0024934	2697	252	2.95	KDR15762	1.7E-125
	Chv1VELVK21_0024932	6370	235	5.56	KDR15762	1.06E-84
FURIN	Chv1VELVK45_0035041	2290	632	1.23	XP_008545163	2.44E-07
	Chv1VELVK21_0083474	4574	1029	1.29	XP_002422846	3.05E-41
	Carboxypeptidase					
	Chv1VELVK21_0032458	2769	547	22.71	KDR18068	0
	Chv1VELVK65_0051556	1517	445	1365.21	AEI58649	9.4E-105
	Chv1VELVK65_0051554	1208	342	24.07	AFZ78840	9.3E-07
	Chv1VELVK55_0060004	1363	339	95.56	AEI58657	9.23E-63
SPCS2	Chv1VELVK65_0011192	963	193	251.15	KDR10695	5.6E-115
Protease inhibitor	Chv1SOAPK55_0010469	414	99	1441.73	XP_005304994	2.32E-09
Serpin	Chv1VELVK65_0051330	1791	387	104.68	KDR09404	1.73E-48
	Chv1VELVK65_0051329	1605	325	49.43	KDR09404	1.84E-19
	Chv1VELVK55_0006994	1300	410	67.27	KDR09404	6.16E-69
	Chv1VELVK35_0011937	1165	323	19.7	KDR09404	1.62E-56
	Chv1VELVK75_0000424	1586	407	102.13	KDR19769	6.1E-99

MME: Membrane metallo-endorpeptidase; GGT: Gamma-glutamyltranspeptidase; FURIN: Endoprotease FURIN; SPCS2: Signal peptidase complex subunit 2.

doi:10.1371/journal.pone.0157783.t007

Blast hits, indicating they might be novel proteins or proteins highly diverged in phasmids. There are a total of 28 *C. hookeri* transcripts classified with GO terms of peptidase activity and peptidase regulator activity (S2 Text), two groups of proteins that are believed to be essential and regulate reproduction through proteolytic cascades [94]. The predicted members of 22 proteases/peptidases included serine protease (including trypsin), furin protease, carboxypeptidase, γ -glutamyl transpeptidase, membrane metallo-endorpeptidase, signal peptidase complex subunit 2, angiotensin converting enzyme (ACE) and thrombospondin type-1, and six protease inhibitors (Table 7).

In *D. melanogaster*, most seminal fluid proteases are serine proteases, which break down proteins into smaller peptides by cutting the protein serine-protease cleavage sites [94]. Serine proteases are required for the localisation of the sex peptide (SP) to regulate sperm storage parameters in *D. melanogaster*, one of which was inferred to act upstream of an astacin protease to process SFPs involved in ovulation and sperm entry into storage [94, 95]. Trypsin is also

a significant factor involved in releasing the mature or activated SPs in the hemolymph of females, and the mature forms of SPs contribute to both short-term and long-term post-mating responses such as inducing egg-laying and inhibition of receptivity [96]. Carboxypeptidases hydrolyse the carboxy-terminal of a peptide bond, and together with serine proteases, they are thought to be important in spermiogenesis and possibly post-coital stimulation of females [92]. The presence of ACE in seminal fluids is also essential as feeding ACE inhibitors to a male *Anopheles stephensi* resulted in a reduced egg-laying rate of its mate [97]. In addition, these proteases are mostly regulated by protease inhibitors that specifically target the recipients to avoid causing damage to normal cells or tissues [98].

We also found other *C. hookeri* transcripts annotated as proteins that have been reported to be essential in spermiogenesis and/or regulating of female reproductive function. Three transcripts were annotated as serine/threonine-protease kinases, which are believed to be functionally involved in sperm capacitation in humans [99] and the recognition of sperm chromatin during spermiogenesis in vertebrates [100]. The abundance of serine/threonine-protease kinases in the testis-vas deferens and male accessory glands but not in the spermatophore of *Dermacentor variabilis* (tick) suggests they are likely to execute a similar function in insects [92]. We also found a transcript annotated as transmembrane emp24 domain-containing protein that might be involved in ovulation. In *D. melanogaster*, a similar protein is present in both the central nervous system and vitellogenic egg chambers and is believed to regulate female ovulation through the sex-determination hierarchy to control female egg-laying behaviour [101]. In addition, transcripts annotated as lipases were also present, whose activity was detected in *Drosophila* male accessory glands and transferred into female bodies during copulation, providing energy to sperms through hydrolysing triglycerides [102].

In summary, male *C. hookeri* produces similar SFPs when compared with other insects. However, there are numerous transcripts lacking Blast hits, indicating the presence of novel proteins. These proteins, together with those SFPs classified above, provide large resource for the study of sexual reproduction, speciation and parthenogenesis in the New Zealand stick insects.

Conclusion

We report the transcriptome of the stick insect *C. hookeri*. The assembly comprising RNA-Seq reads sequenced from female antennae, head and prothorax, leg, midgut and male terminalia, was assessed to include a broad representation of expressed genes, and provides a comprehensive resource for analysing genes of interest for this species.

In the mining of gene families involved in olfaction, we conclude that *C. hookeri* has a fully developed repertoire of olfactory gene families. The majority of these olfactory genes are over-expressed in the olfactory organ of female antennae. Some are enriched in male terminalia (*OBP9*, *CSP6* and *IR7*) and female midgut (*CSP10*), suggesting their roles in sexual reproduction and digestion. Highly abundant PBP in female antennae, in contrast to their usual function for male insects to detect mates at long range, may be useful for females to detect pheromones released from other females in order to space themselves out in the environment. In addition, recent gene duplications among three olfactory binding proteins (*OBP4*, *OBP5* and *OBP6*) may underlie key adaptive events during sensory evolution in *C. hookeri*.

Identification of the members belonging to GH9 reveals a total of nine gene duplicates in *C. hookeri*. Comparison with five gene copies detected from the divergent phasmid *T. cristinae*, suggests there has been a dramatic cellulase gene family expansion during the evolution of phasmids. The increased copy number of GH9 likely allows this insect to endogenously degrade cellulose without assistance from symbionts and some of them may be able to break down xylan, xyloglucan, and/or glucomannan as has been observed in several other stick

insects [18]. However, some of these duplicates might also have evolved new roles rather than digesting plant cell walls as they were not detected in the midgut. We also identified *C. hookeri* pectinase sequences from the transcriptome, belonging to the GH28 family, the group of enzymes that were recently demonstrated to have transferred from bacteria to stick insects [8]. However, these enzymes are most likely to be endogenic in *C. hookeri*, because the genes encoding them are present in the genome and the two genomic scaffolds containing these genes also harbour other insect genes.

Our identification of accessory gland proteins in *C. hookeri* represents the first molecular genetic investigation of sexual reproduction in the phasmids. The male terminalia containing predominantly accessory glands were enriched with SFPs. The predicted transcripts of protease/peptidase and protease inhibitors that are over-expressed in male terminalia suggest they are essential to sexual reproduction, and most likely regulate processes through proteolytic cascades based on findings from *D. melanogaster* [94]. Many putative SFPs lacking Blast matches indicate at least some of them are either novel proteins or highly diverged.

In conclusion, our *C. hookeri* transcriptome reference, together with identified gene families, provides a comprehensive resource for studying the evolution of olfaction, digestion and sexual reproduction in phasmids. The duplication events in olfactory binding and cellulase genes in *C. hookeri* may underlie key adaptive events in the evolution of the stick insects.

Supporting Information

S1 Table. Table of over-represented transcripts from libraries of antenna, midgut and terminalia. Transcripts were ranked based on their abundance (TPM) in the sample library. Transcripts from the terminalia library were added by signal peptide check and sub-cellular localisation.

(XLSX)

S1 Text. Amino acid sequences of predicted *C. hookeri* ORs, IRs, GR and SNMPs.

(DOCX)

S2 Text. Amino acid sequences of predicted *C. hookeri* SFPs of protease and protease inhibitors.

(DOCX)

S3 Text. Species and accession numbers used for constructing ORCO and GH9 phylogenies.

(DOCX)

Acknowledgments

We thank New Zealand eScience Infrastructure (NeSI) for high-performance computing facilities; P. Maxwell, X. Cheng, K. Rutherford, J. Blasco, M. Feller, H. Liu, M. Zhang and Y. Wang for analytic help; S. Myers, K. King and M. McDonald for field assistance; L. Dunning for his contribution in the construction to the *C. hookeri* head and prothorax transcriptome; R. Howitt, J. Allwood, A. Ramon-Laca, and D. Park for laboratory support. We thank K. Meusemann, A. Donath, X. Zhou, K. Kjer, B. Misof, L. Podsiadlowski, B. Wipfler and the 1kite Polyneopetra subgroup (<http://www.1kite.org/subprojects.html>) for providing the leg RNA-Seq data and especially A. Donath for submission to and data release on NCBI. We also thank the Allan Wilson Centre for funding, collaboration and facility support. We thank J. Joe Hull and three anonymous reviewers for giving us valuable suggestions to strengthen this paper.

Author Contributions

Conceived and designed the experiments: CW TRB RNC. Performed the experiments: VGT CW. Analyzed the data: CW SL ABD. Contributed reagents/materials/analysis tools: ABD SL TRB. Wrote the paper: CW RNC ABD VGT RDN HAR TRB. Contributed wonderful ideas to data analysis: RNC.

References

1. Bragg PE. Phasmids of Borneo. Natural History Publications (Borneo), Kota Kinabalu; 2001.
2. Buckley TR, Attanayake D, Bradler S. Extreme convergence in stick insect evolution: phylogenetic placement of the Lord Howe Island tree lobster. *Proceedings of the Royal Society of London B: Biological Sciences*. 2009; 276(1659):1055–1062. doi: [10.1098/rspb.2008.1552](https://doi.org/10.1098/rspb.2008.1552)
3. Bässler U. Sensory control of leg movement in the stick insect *Carausius morosus*. *Biological cybernetics*. 1977; 25(2):61–72. doi: [10.1007/BF00337264](https://doi.org/10.1007/BF00337264) PMID: [836915](https://pubmed.ncbi.nlm.nih.gov/836915/)
4. Morgan-Richards M, Treweek SA, Stringer IA. Geographic parthenogenesis and the common tea-tree stick insect of New Zealand. *Molecular Ecology*. 2010; 19(6):1227–1238. doi: [10.1111/j.1365-294X.2010.04542.x](https://doi.org/10.1111/j.1365-294X.2010.04542.x) PMID: [20163549](https://pubmed.ncbi.nlm.nih.gov/20163549/)
5. Shelomi M, Jasper WC, Atallah J, Kimsey LS, Johnson BR. Differential expression of endogenous plant cell wall degrading enzyme genes in the stick insect (Phasmatodea) midgut. *BMC genomics*. 2014; 15(1):1. doi: [10.1186/1471-2164-15-917](https://doi.org/10.1186/1471-2164-15-917)
6. Shelomi M, Watanabe H, Arakawa G. Endogenous cellulase enzymes in the stick insect (Phasmatodea) gut. *Journal of insect physiology*. 2014; 60:25–30. doi: [10.1016/j.jinsphys.2013.10.007](https://doi.org/10.1016/j.jinsphys.2013.10.007) PMID: [24216471](https://pubmed.ncbi.nlm.nih.gov/24216471/)
7. Myers SS, Buckley TR, Holwell GI. Mate detection and seasonal variation in stick insect mating behaviour (Phasmatodea: *Clitarchus hookeri*). *Behaviour*. 2015; 152(10):1325–1348. doi: [10.1163/1568539X-00003281](https://doi.org/10.1163/1568539X-00003281)
8. Shelomi M, Danchin EG, Heckel D, Wipfler B, Bradler S, Zhou X, et al. Horizontal Gene Transfer of Pectinases from Bacteria Preceded the Diversification of Stick and Leaf Insects. *Scientific reports*. 2016; 6:26388. doi: [10.1038/srep26388](https://doi.org/10.1038/srep26388) PMID: [27210832](https://pubmed.ncbi.nlm.nih.gov/27210832/)
9. Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, et al. Stick insect genomes reveal natural selection's role in parallel speciation. *Science*. 2014; 344(6185):738–742. doi: [10.1126/science.1252136](https://doi.org/10.1126/science.1252136) PMID: [24833390](https://pubmed.ncbi.nlm.nih.gov/24833390/)
10. Gompert Z, Comeault AA, Farkas TE, Feder JL, Parchman TL, Buerkle CA, et al. Experimental evidence for ecological selection on genome variation in the wild. *Ecology letters*. 2014; 17(3):369–379. doi: [10.1111/ele.12238](https://doi.org/10.1111/ele.12238) PMID: [24354456](https://pubmed.ncbi.nlm.nih.gov/24354456/)
11. Comeault AA, Sommers M, Schwander T, Buerkle C, Farkas TE, Nosil P, et al. *De novo* characterization of the *Timema cristinae* transcriptome facilitates marker discovery and inference of genetic divergence. *Molecular ecology resources*. 2012; 12(3):549–561. doi: [10.1111/j.1755-0998.2012.03121.x](https://doi.org/10.1111/j.1755-0998.2012.03121.x) PMID: [22339780](https://pubmed.ncbi.nlm.nih.gov/22339780/)
12. Dunning LT, Dennis AB, Park D, Sinclair BJ, Newcomb RD, Buckley TR. Identification of cold-responsive genes in a New Zealand alpine stick insect using RNA-Seq. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*. 2013; 8(1):24–31.
13. Dunning LT, Dennis AB, Sinclair BJ, Newcomb RD, Buckley TR. Divergent transcriptional responses to low temperature among populations of alpine and lowland species of New Zealand stick insects (*Micrarchus*). *Molecular ecology*. 2014; 23(11):2712–2726. doi: [10.1111/mec.12767](https://doi.org/10.1111/mec.12767) PMID: [24762129](https://pubmed.ncbi.nlm.nih.gov/24762129/)
14. Missbach C, Dweck HK, Vogel H, Vilcinskas A, Stensmyr MC, Hansson BS, et al. Evolution of insect olfactory receptors. *Elife*. 2014; 3:e02115. doi: [10.7554/eLife.02115](https://doi.org/10.7554/eLife.02115) PMID: [24670956](https://pubmed.ncbi.nlm.nih.gov/24670956/)
15. Dennis AB, Dunning LT, Sinclair BJ, Buckley TR. Parallel molecular routes to cold adaptation in eight genera of New Zealand stick insects. *Scientific reports*. 2015; 5. doi: [10.1038/srep13965](https://doi.org/10.1038/srep13965)
16. Morgan-Richards M, Hills SF, Biggs PJ, Treweek SA. Sticky Genomes: Using NGS Evidence to Test Hybrid Speciation Hypotheses. *PloS one*. 2016; 11(5):e0154911. doi: [10.1371/journal.pone.0154911](https://doi.org/10.1371/journal.pone.0154911) PMID: [27187689](https://pubmed.ncbi.nlm.nih.gov/27187689/)
17. Shelomi M, Lo WS, Kimsey LS, Kuo CH. Analysis of the gut microbiota of walking sticks (Phasmatodea). *BMC research notes*. 2013; 6(1):368. doi: [10.1186/1756-0500-6-368](https://doi.org/10.1186/1756-0500-6-368) PMID: [24025149](https://pubmed.ncbi.nlm.nih.gov/24025149/)
18. Shelomi M, Heckel DG, Pauchet Y. Ancestral gene duplication enabled the evolution of multifunctional cellulases in stick insects (Phasmatodea). *Insect Biochemistry and Molecular Biology*. 2016; 71:1–11. doi: [10.1016/j.ibmb.2016.02.003](https://doi.org/10.1016/j.ibmb.2016.02.003) PMID: [26855199](https://pubmed.ncbi.nlm.nih.gov/26855199/)

19. Buckley TR, Attanayake D, Nylander JA, Bradler S. The phylogenetic placement and biogeographical origins of the New Zealand stick insects (Phasmatodea). *Systematic Entomology*. 2010; 35(2):207–225. doi: [10.1111/j.1365-3113.2009.00505.x](https://doi.org/10.1111/j.1365-3113.2009.00505.x)
20. Dunning LT, Dennis AB, Thomson G, Sinclair BJ, Newcomb RD, Buckley TR. Positive selection in glycolysis among Australasian stick insects. *BMC evolutionary biology*. 2013; 13(1):215. doi: [10.1186/1471-2148-13-215](https://doi.org/10.1186/1471-2148-13-215) PMID: [24079656](https://pubmed.ncbi.nlm.nih.gov/24079656/)
21. Dennis A, Dunning L, Dennis C, Sinclair B, Buckley T. Overwintering in New Zealand stick insects. *New Zealand Entomologist*. 2014; 37(1):35–44. doi: [10.1080/00779962.2013.784235](https://doi.org/10.1080/00779962.2013.784235)
22. Buckley TR, Marske K, Attanayake D. Phylogeography and ecological niche modelling of the New Zealand stick insect *Clitarchus hookeri* (White) support survival in multiple coastal refugia. *Journal of Biogeography*. 2010; 37(4):682–695. doi: [10.1111/j.1365-2699.2009.02239.x](https://doi.org/10.1111/j.1365-2699.2009.02239.x)
23. Buckley TR, Myers SS, Bradler S. Revision of the stick insect genus *Clitarchus* Stål (Phasmatodea: Phasmatidae): new synonymies and two new species from northern New Zealand. *Zootaxa*. 2014; 3900(4):451–482. doi: [10.11646/zootaxa.3900.4.1](https://doi.org/10.11646/zootaxa.3900.4.1) PMID: [25543751](https://pubmed.ncbi.nlm.nih.gov/25543751/)
24. Salmon JT. The stick insects of New Zealand. Raupo; 1991.
25. Trewick SA, Goldberg J, Morgan-Richards M. Fewer species of *Argosarchus* and *Clitarchus* stick insects (Phasmida, Phasmatinae): evidence from nuclear and mitochondrial DNA sequence data. *Zoologica Scripta*. 2005; 34(5):483–491. doi: [10.1111/j.1463-6409.2005.00204.x](https://doi.org/10.1111/j.1463-6409.2005.00204.x)
26. Hansen KD, Brenner SE, Dudoit S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic acids research*. 2010; 38(12):e131–e131. doi: [10.1093/nar/gkq224](https://doi.org/10.1093/nar/gkq224) PMID: [20395217](https://pubmed.ncbi.nlm.nih.gov/20395217/)
27. Li W, Jiang T. Transcriptome assembly and isoform expression level estimation from biased RNA-Seq reads. *Bioinformatics*. 2012; 28(22):2914–2921. doi: [10.1093/bioinformatics/bts559](https://doi.org/10.1093/bioinformatics/bts559) PMID: [23060617](https://pubmed.ncbi.nlm.nih.gov/23060617/)
28. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011; 27(6):863–864. doi: [10.1093/bioinformatics/btr026](https://doi.org/10.1093/bioinformatics/btr026) PMID: [21278185](https://pubmed.ncbi.nlm.nih.gov/21278185/)
29. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011; 17(1):pp–10. doi: [10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200)
30. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*. 2011; 29(7):644–652. doi: [10.1038/nbt.1883](https://doi.org/10.1038/nbt.1883) PMID: [21572440](https://pubmed.ncbi.nlm.nih.gov/21572440/)
31. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. *Bioinformatics*. 2014; 30(12):1660–1666. doi: [10.1093/bioinformatics/btu077](https://doi.org/10.1093/bioinformatics/btu077) PMID: [24532719](https://pubmed.ncbi.nlm.nih.gov/24532719/)
32. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*. 2012; 28(8):1086–1092. doi: [10.1093/bioinformatics/bts094](https://doi.org/10.1093/bioinformatics/bts094) PMID: [22368243](https://pubmed.ncbi.nlm.nih.gov/22368243/)
33. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006; 22(13):1658–1659. doi: [10.1093/bioinformatics/btl158](https://doi.org/10.1093/bioinformatics/btl158) PMID: [16731699](https://pubmed.ncbi.nlm.nih.gov/16731699/)
34. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*. 2007; 23(9):1061–1067. doi: [10.1093/bioinformatics/btm071](https://doi.org/10.1093/bioinformatics/btm071) PMID: [17332020](https://pubmed.ncbi.nlm.nih.gov/17332020/)
35. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012; 9(4):357–359. doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923) PMID: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/)
36. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*. 1997; 25(17):3389–3402. doi: [10.1093/nar/25.17.3389](https://doi.org/10.1093/nar/25.17.3389) PMID: [9254694](https://pubmed.ncbi.nlm.nih.gov/9254694/)
37. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. *Nucleic acids research*. 2005; 33(suppl 2):W116–W120. doi: [10.1093/nar/gki442](https://doi.org/10.1093/nar/gki442) PMID: [15980438](https://pubmed.ncbi.nlm.nih.gov/15980438/)
38. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005; 21(18):3674–3676. doi: [10.1093/bioinformatics/bti610](https://doi.org/10.1093/bioinformatics/bti610) PMID: [16081474](https://pubmed.ncbi.nlm.nih.gov/16081474/)
39. Terrapon N, Li C, Robertson HM, Ji L, Meng X, Booth W, et al. Molecular traces of alternative social organization in a termite genome. *Nature communications*. 2014; 5. doi: [10.1038/ncomms4636](https://doi.org/10.1038/ncomms4636) PMID: [24845553](https://pubmed.ncbi.nlm.nih.gov/24845553/)

40. Marchese S, Angeli S, Andolfo A, Scaloni A, Brandazza A, Mazza M, et al. Soluble proteins from chemosensory organs of *Eurycantha calcarata* (Insects, Phasmatodea). *Insect biochemistry and molecular biology*. 2000; 30(11):1091–1098. doi: [10.1016/S0965-1748\(00\)00084-9](https://doi.org/10.1016/S0965-1748(00)00084-9) PMID: [10989296](https://pubmed.ncbi.nlm.nih.gov/10989296/)
41. Corcoran JA, Jordan MD, Thrimawithana AH, Crowhurst RN, Newcomb RD. The peripheral olfactory repertoire of the lightbrown apple moth, *Epiphyas postvittana*. *PloS one*. 2015; 10(5):e0128596. doi: [10.1371/journal.pone.0128596](https://doi.org/10.1371/journal.pone.0128596) PMID: [26017144](https://pubmed.ncbi.nlm.nih.gov/26017144/)
42. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*. 2011; 12(1):1. doi: [10.1186/1471-2105-12-323](https://doi.org/10.1186/1471-2105-12-323)
43. Langmead B, Trapnell C, Pop M, Salzberg SL, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10(3):R25. doi: [10.1186/gb-2009-10-3-r25](https://doi.org/10.1186/gb-2009-10-3-r25) PMID: [19261174](https://pubmed.ncbi.nlm.nih.gov/19261174/)
44. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BM, et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*. 2013; p. btt087.
45. Sonesson C, Delorenzi M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC bioinformatics*. 2013; 14(1):1. doi: [10.1186/1471-2105-14-91](https://doi.org/10.1186/1471-2105-14-91)
46. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*. 2004; 5(1):1. doi: [10.1186/1471-2105-5-113](https://doi.org/10.1186/1471-2105-5-113)
47. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012; 28(12):1647–1649. doi: [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199) PMID: [22543367](https://pubmed.ncbi.nlm.nih.gov/22543367/)
48. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic biology*. 2007; 56(4):564–577. doi: [10.1080/10635150701472164](https://doi.org/10.1080/10635150701472164) PMID: [17654362](https://pubmed.ncbi.nlm.nih.gov/17654362/)
49. Guindon S, Dufayard J, Hordijk W, Lefort V, Gascuel O. PhyML: fast and accurate phylogeny reconstruction by maximum likelihood. *Infection Genetics and Evolution*. 2009; 9(3):384–385.
50. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011; 27(8):1164–1165. doi: [10.1093/bioinformatics/btr088](https://doi.org/10.1093/bioinformatics/btr088) PMID: [21335321](https://pubmed.ncbi.nlm.nih.gov/21335321/)
51. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*. 2012; 9(8):772–772. doi: [10.1038/nmeth.2109](https://doi.org/10.1038/nmeth.2109) PMID: [22847109](https://pubmed.ncbi.nlm.nih.gov/22847109/)
52. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*. 2007; 24(8):1586–1591. doi: [10.1093/molbev/msm088](https://doi.org/10.1093/molbev/msm088) PMID: [17483113](https://pubmed.ncbi.nlm.nih.gov/17483113/)
53. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*. 2013; 8(8):1494–1512. doi: [10.1038/nprot.2013.084](https://doi.org/10.1038/nprot.2013.084) PMID: [23845962](https://pubmed.ncbi.nlm.nih.gov/23845962/)
54. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods*. 2011; 8(10):785–786. doi: [10.1038/nmeth.1701](https://doi.org/10.1038/nmeth.1701) PMID: [21959131](https://pubmed.ncbi.nlm.nih.gov/21959131/)
55. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC bioinformatics*. 2011; 12(1):35. doi: [10.1186/1471-2105-12-35](https://doi.org/10.1186/1471-2105-12-35) PMID: [21269502](https://pubmed.ncbi.nlm.nih.gov/21269502/)
56. Walker MJ, Rylett CM, Keen JN, Audsley N, Sajid M, Shirras AD, et al. Proteomic identification of *Drosophila melanogaster* male accessory gland proteins, including a pro-cathepsin and a soluble γ -glutamyl transpeptidase. *Proteome science*. 2006; 4(1):1.
57. Leal WS. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annual review of entomology*. 2013; 58:373–391. doi: [10.1146/annurev-ento-120811-153635](https://doi.org/10.1146/annurev-ento-120811-153635) PMID: [23020622](https://pubmed.ncbi.nlm.nih.gov/23020622/)
58. Leitch O, Papanicolaou A, Lennard C, Kirkbride KP, Anderson A. Chemosensory genes identified in the antennal transcriptome of the blowfly *Calliphora stygia*. *BMC genomics*. 2015; 16(1):1–17. doi: [10.1186/s12864-015-1466-8](https://doi.org/10.1186/s12864-015-1466-8)
59. Rinker DC, Zhou X, Pitts RJ, Rokas A, Zwiebel LJ, et al. Antennal transcriptome profiles of anopheline mosquitoes reveal human host olfactory specialization in *Anopheles gambiae*. *BMC genomics*. 2013; 14(1):749. doi: [10.1186/1471-2164-14-749](https://doi.org/10.1186/1471-2164-14-749) PMID: [24182346](https://pubmed.ncbi.nlm.nih.gov/24182346/)
60. Zhou SS, Sun Z, Ma W, Chen W, Wang MQ. *De novo* analysis of the *Nilaparvata lugens* (Stål) antenna transcriptome and expression patterns of olfactory genes. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*. 2014; 9:31–39.
61. Mameli M, Tuccini A, Mazza M, Petacchi R, Pelosi P. Soluble proteins in chemosensory organs of phasmids. *Insect biochemistry and molecular biology*. 1996; 26(8):875–882. doi: [10.1016/S0965-1748\(96\)00055-0](https://doi.org/10.1016/S0965-1748(96)00055-0) PMID: [9014332](https://pubmed.ncbi.nlm.nih.gov/9014332/)

62. Tuccini A, Maida R, Rovero P, Mazza M, Pelosi P. Putative odorant-binding protein in antennae and legs of *Carausius morosus* (Insecta, Phasmatodea). *Insect biochemistry and molecular biology*. 1996; 26(1):19–24. doi: [10.1016/0965-1748\(95\)00051-8](https://doi.org/10.1016/0965-1748(95)00051-8) PMID: [8673076](https://pubmed.ncbi.nlm.nih.gov/8673076/)
63. Pelosi P, Maida R. Odorant-binding proteins in insects. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*. 1995; 111(3):503–514. doi: [10.1016/0305-0491\(95\)00019-5](https://doi.org/10.1016/0305-0491(95)00019-5)
64. Xu P, Atkinson R, Jones DN, Smith DP. *Drosophila* OBP LUSH is required for activity of pheromone-sensitive neurons. *Neuron*. 2005; 45(2):193–200. doi: [10.1016/j.neuron.2004.12.031](https://doi.org/10.1016/j.neuron.2004.12.031) PMID: [15664171](https://pubmed.ncbi.nlm.nih.gov/15664171/)
65. Damberger FF, Michel E, Ishida Y, Leal WS, Wüthrich K. Pheromone discrimination by a pH-tuned polymorphism of the *Bombyx mori* pheromone-binding protein. *Proceedings of the National Academy of Sciences*. 2013; 110(46):18680–18685. doi: [10.1073/pnas.1317706110](https://doi.org/10.1073/pnas.1317706110)
66. Schneider D, Schulz S, Priesner E, Ziesmann J, Francke W. Autodetection and chemistry of female and male pheromone in both sexes of the tiger moth *Panaxia quadripunctaria*. *Journal of Comparative Physiology A*. 1998; 182(2):153–161. doi: [10.1007/s003590050166](https://doi.org/10.1007/s003590050166)
67. Callahan F, Vogt R, Tucker M, Dickens J, Mattoo A. High level expression of “male specific” pheromone binding proteins (PBPs) in the antennae of female noctuid moths. *Insect biochemistry and molecular biology*. 2000; 30(6):507–514. doi: [10.1016/S0965-1748\(00\)00027-8](https://doi.org/10.1016/S0965-1748(00)00027-8) PMID: [10802242](https://pubmed.ncbi.nlm.nih.gov/10802242/)
68. Hekmat-Scafe DS, Scafe CR, McKinney AJ, Tanouye MA. Genome-wide analysis of the odorant-binding protein gene family in *Drosophila melanogaster*. *Genome research*. 2002; 12(9):1357–1369. doi: [10.1101/gr.239402](https://doi.org/10.1101/gr.239402) PMID: [12213773](https://pubmed.ncbi.nlm.nih.gov/12213773/)
69. Paesen GC, Happ GM. The B proteins secreted by the tubular accessory sex glands of the male mealworm beetle, *Tenebrio molitor*, have sequence similarity to moth pheromone-binding proteins. *Insect biochemistry and molecular biology*. 1995; 25(3):401–408. doi: [10.1016/0965-1748\(94\)00085-V](https://doi.org/10.1016/0965-1748(94)00085-V) PMID: [7773257](https://pubmed.ncbi.nlm.nih.gov/7773257/)
70. Thymlanou S, Mavroidis M, Kokolakis G, Komitopoulou K, Zacharopoulou A, Mintzas A. Cloning and characterization of a cDNA encoding a male-specific serum protein of the Mediterranean fruit fly, *Ceratitis capitata*, with sequence similarity to odourant-binding proteins. *Insect molecular biology*. 1998; 7(4):345–353. doi: [10.1046/j.1365-2583.1998.740345.x](https://doi.org/10.1046/j.1365-2583.1998.740345.x)
71. Arcà B, Lombardo F, de Lara Capurro M, Della Torre A, Dimopoulos G, James AA, et al. Trapping cDNAs encoding secreted proteins from the salivary glands of the malaria vector *Anopheles gambiae*. *Proceedings of the National Academy of Sciences*. 1999; 96(4):1516–1521. doi: [10.1073/pnas.96.4.1516](https://doi.org/10.1073/pnas.96.4.1516)
72. Sun YL, Huang LQ, Pelosi P, Wang CZ. Expression in antennae and reproductive organs suggests a dual role of an odorant-binding protein in two sibling *Helicoverpa* species. *PloS one*. 2012; 7(1): e30040. doi: [10.1371/journal.pone.0030040](https://doi.org/10.1371/journal.pone.0030040) PMID: [22291900](https://pubmed.ncbi.nlm.nih.gov/22291900/)
73. Sato K, Pellegrino M, Nakagawa T, Nakagawa T, Vosshall LB, Touhara K. Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature*. 2008; 452(7190):1002–1006. doi: [10.1038/nature06850](https://doi.org/10.1038/nature06850) PMID: [18408712](https://pubmed.ncbi.nlm.nih.gov/18408712/)
74. Smart R, Kiely A, Beale M, Vargas E, Carraher C, Kralicek AV, et al. *Drosophila* odorant receptors are novel seven transmembrane domain proteins that can signal independently of heterotrimeric G proteins. *Insect biochemistry and molecular biology*. 2008; 38(8):770–780. doi: [10.1016/j.ibmb.2008.05.002](https://doi.org/10.1016/j.ibmb.2008.05.002) PMID: [18625400](https://pubmed.ncbi.nlm.nih.gov/18625400/)
75. Rytz R, Croset V, Benton R. Ionotropic receptors (IRs): chemosensory ionotropic glutamate receptors in *Drosophila* and beyond. *Insect biochemistry and molecular biology*. 2013; 43(9):888–897. doi: [10.1016/j.ibmb.2013.02.007](https://doi.org/10.1016/j.ibmb.2013.02.007) PMID: [23459169](https://pubmed.ncbi.nlm.nih.gov/23459169/)
76. Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB. Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron*. 2004; 43(5):703–714. doi: [10.1016/j.neuron.2004.08.019](https://doi.org/10.1016/j.neuron.2004.08.019) PMID: [15339651](https://pubmed.ncbi.nlm.nih.gov/15339651/)
77. Große-Wilde E, Gohl T, Bouché E, Breer H, Krieger J. Candidate pheromone receptors provide the basis for the response of distinct antennal neurons to pheromonal compounds. *European Journal of Neuroscience*. 2007; 25(8):2364–2373. doi: [10.1111/j.1460-9568.2007.05512.x](https://doi.org/10.1111/j.1460-9568.2007.05512.x) PMID: [17445234](https://pubmed.ncbi.nlm.nih.gov/17445234/)
78. Kurtovic A, Widmer A, Dickson BJ. A single class of olfactory neurons mediates behavioural responses to a *Drosophila* sex pheromone. *Nature*. 2007; 446(7135):542–546. doi: [10.1038/nature05672](https://doi.org/10.1038/nature05672) PMID: [17392786](https://pubmed.ncbi.nlm.nih.gov/17392786/)
79. Montell C. A taste of the *Drosophila* gustatory receptors. *Current opinion in neurobiology*. 2009; 19(4):345–353. doi: [10.1016/j.conb.2009.07.001](https://doi.org/10.1016/j.conb.2009.07.001) PMID: [19660932](https://pubmed.ncbi.nlm.nih.gov/19660932/)
80. Wang X, Fang X, Yang P, Jiang X, Jiang F, Zhao D, et al. The locust genome provides insight into swarm formation and long-distance flight. *Nature communications*. 2014; 5.

81. Davison A, Blaxter M. Ancient origin of glycosyl hydrolase family 9 cellulase genes. *Molecular Biology and Evolution*. 2005; 22(5):1273–1284. doi: [10.1093/molbev/msi107](https://doi.org/10.1093/molbev/msi107) PMID: [15703240](https://pubmed.ncbi.nlm.nih.gov/15703240/)
82. Watanabe H, Noda H, Tokuda G, Lo N. A cellulase gene of termite origin. *Nature*. 1998; 394(6691):330–331. doi: [10.1038/28527](https://doi.org/10.1038/28527) PMID: [9690469](https://pubmed.ncbi.nlm.nih.gov/9690469/)
83. Calderón-Cortés N, Quesada M, Watanabe H, Cano-Camacho H, Oyama K. Endogenous plant cell wall digestion: a key mechanism in insect evolution. *Annual Review of Ecology, Evolution, and Systematics*. 2012; 43:45–71. doi: [10.1146/annurev-ecolsys-110411-160312](https://doi.org/10.1146/annurev-ecolsys-110411-160312)
84. Cazemier AE, den Camp HJO, Hackstein JH, Vogels GD. Fibre digestion in arthropods. *Comparative Biochemistry and Physiology Part A: Physiology*. 1997; 118(1):101–109. doi: [10.1016/S0300-9629\(96\)00443-4](https://doi.org/10.1016/S0300-9629(96)00443-4)
85. Koenig D, Weigel D. Beyond the thale: comparative genomics and genetics of *Arabidopsis* relatives. *Nature Reviews Genetics*. 2015; 16(5):285–298. doi: [10.1038/nrg3883](https://doi.org/10.1038/nrg3883) PMID: [25854181](https://pubmed.ncbi.nlm.nih.gov/25854181/)
86. Sprockett DD, Piontkivska H, Blackwood CB. Evolutionary analysis of glycosyl hydrolase family 28 (GH28) suggests lineage-specific expansions in necrotrophic fungal pathogens. *Gene*. 2011; 479(1):29–36. doi: [10.1016/j.gene.2011.02.009](https://doi.org/10.1016/j.gene.2011.02.009) PMID: [21354463](https://pubmed.ncbi.nlm.nih.gov/21354463/)
87. Kirsch R, Gramzow L, Theißen G, Siegfried BD, Heckel DG, Pauchet Y, et al. Horizontal gene transfer and functional diversification of plant cell wall degrading polygalacturonases: key events in the evolution of herbivory in beetles. *Insect biochemistry and molecular biology*. 2014; 52:33–50. doi: [10.1016/j.ibmb.2014.06.008](https://doi.org/10.1016/j.ibmb.2014.06.008) PMID: [24978610](https://pubmed.ncbi.nlm.nih.gov/24978610/)
88. Scully ED, Hoover K, Carlson JE, Tien M, Geib SM. Midgut transcriptome profiling of *Anoplophora glabripennis*, a lignocellulose degrading cerambycid beetle. *BMC genomics*. 2013; 14(1):1. doi: [10.1186/1471-2164-14-850](https://doi.org/10.1186/1471-2164-14-850)
89. Swanson WJ, Vacquier VD. The rapid evolution of reproductive proteins. *Nature Reviews Genetics*. 2002; 3(2):137–144. doi: [10.1038/nrg733](https://doi.org/10.1038/nrg733) PMID: [11836507](https://pubmed.ncbi.nlm.nih.gov/11836507/)
90. Panhuis TM, Clark NL, Swanson WJ. Rapid evolution of reproductive proteins in abalone and *Drosophila*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 2006; 361(1466):261–268. doi: [10.1098/rstb.2005.1793](https://doi.org/10.1098/rstb.2005.1793) PMID: [16612885](https://pubmed.ncbi.nlm.nih.gov/16612885/)
91. Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF. Insect seminal fluid proteins: identification and function. *Annual review of entomology*. 2011; 56:21. doi: [10.1146/annurev-ento-120709-144823](https://doi.org/10.1146/annurev-ento-120709-144823) PMID: [20868282](https://pubmed.ncbi.nlm.nih.gov/20868282/)
92. Sonenshine DE, Bissinger BW, Egekwu N, Donohue KV, Khalil SM, Roe RM. First transcriptome of the testis-vas deferens-male accessory gland and proteome of the spermatophore from *Dermacentor variabilis* (Acari: Ixodidae). *PloS one*. 2011; 6(9):e24711. doi: [10.1371/journal.pone.0024711](https://doi.org/10.1371/journal.pone.0024711) PMID: [21949745](https://pubmed.ncbi.nlm.nih.gov/21949745/)
93. 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(6):e2946. doi: [10.1371/journal.pntd.0002946](https://doi.org/10.1371/journal.pntd.0002946) PMID: [24945155](https://pubmed.ncbi.nlm.nih.gov/24945155/)
94. LaFlamme BA, Ram KR, Wolfner MF. The *Drosophila melanogaster* seminal fluid protease “seminase” regulates proteolytic and post-mating reproductive processes. *PLoS Genet*. 2012; 8(1):e1002435. doi: [10.1371/journal.pgen.1002435](https://doi.org/10.1371/journal.pgen.1002435) PMID: [22253601](https://pubmed.ncbi.nlm.nih.gov/22253601/)
95. Ram KR, Wolfner MF. A network of interactions among seminal proteins underlies the long-term post-mating response in *Drosophila*. *Proceedings of the National Academy of Sciences*. 2009; 106(36):15384–15389. doi: [10.1073/pnas.0902923106](https://doi.org/10.1073/pnas.0902923106)
96. Pilpel N, Nezer I, Applebaum SW, Heifetz Y. Mating-increases trypsin in female *Drosophila* hemolymph. *Insect biochemistry and molecular biology*. 2008; 38(3):320–330. doi: [10.1016/j.ibmb.2007.11.010](https://doi.org/10.1016/j.ibmb.2007.11.010) PMID: [18252246](https://pubmed.ncbi.nlm.nih.gov/18252246/)
97. Sirot LK, Poulson RL, McKenna MC, Girnary H, Wolfner MF, Harrington LC. Identity and transfer of male reproductive gland proteins of the dengue vector mosquito, *Aedes aegypti*: potential tools for control of female feeding and reproduction. *Insect biochemistry and molecular biology*. 2008; 38(2):176–189. doi: [10.1016/j.ibmb.2007.10.007](https://doi.org/10.1016/j.ibmb.2007.10.007) PMID: [18207079](https://pubmed.ncbi.nlm.nih.gov/18207079/)
98. Lung O, Tram U, Finnerty CM, Eipper-Mains MA, Kalb JM, Wolfner MF. The *Drosophila melanogaster* seminal fluid protein Acp62F is a protease inhibitor that is toxic upon ectopic expression. *Genetics*. 2002; 160(1):211–224. PMID: [11805057](https://pubmed.ncbi.nlm.nih.gov/11805057/)
99. Naz RK. Involvement of protein serine and threonine phosphorylation in human sperm capacitation. *Biology of reproduction*. 1999; 60(6):1402–1409. doi: [10.1095/biolreprod60.6.1402](https://doi.org/10.1095/biolreprod60.6.1402) PMID: [10330099](https://pubmed.ncbi.nlm.nih.gov/10330099/)
100. Sabeur K, Ball BA, Corbin CJ, Conley A. Characterization of a novel, testis-specific equine serine/threonine kinase. *Molecular reproduction and development*. 2008; 75(5):867–873. doi: [10.1002/mrd.20792](https://doi.org/10.1002/mrd.20792) PMID: [18246530](https://pubmed.ncbi.nlm.nih.gov/18246530/)

101. Carney GE, Taylor BJ. logjam encodes a predicted EMP24/GP25 protein that is required for *Drosophila* oviposition behavior. *Genetics*. 2003; 164(1):173–186. PMID: [12750330](#)
102. Smith G, Rothwell K, Wood S, Yeaman S, Bownes M. Specificity and localization of lipolytic activity in adult *Drosophila melanogaster*. *Biochemical journal*. 1994; 304(3):775–779. doi: [10.1042/bj3040775](#) PMID: [7818480](#)