

RESEARCH ARTICLE

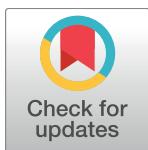
Comparative genomics of chemosensory protein genes (CSPs) in twenty-two mosquito species (Diptera: Culicidae): Identification, characterization, and evolution

Ting Mei[✉], Wen-Bo Fu, Bo Li, Zheng-Bo He, Bin Chen ^{*✉}

Chongqing Key Laboratory of Vector Insects; Chongqing Key Laboratory of Animal Biology; Institute of Entomology and Molecular Biology, Chongqing Normal University, Chongqing, P.R. China

[✉] These authors contributed equally to this work.

* c_bin@hotmail.com, bin.chen@cqnu.edu.cn



OPEN ACCESS

Citation: Mei T, Fu W-B, Li B, He Z-B, Chen B (2018) Comparative genomics of chemosensory protein genes (CSPs) in twenty-two mosquito species (Diptera: Culicidae): Identification, characterization, and evolution. PLoS ONE 13(1): e0190412. <https://doi.org/10.1371/journal.pone.0190412>

Editor: J Joe Hull, USDA Agricultural Research Service, UNITED STATES

Received: June 4, 2017

Accepted: December 14, 2017

Published: January 5, 2018

Copyright: © 2018 Mei et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the Par-Eu Scholars Program (20136666), the National Natural Science Foundation of China (31672363, 31572332, 31372265), Science and Technology Major Special Project of Guangxi (GKAA17129002), the National Key Program of Science and Technology Foundation Work of China

Abstract

Chemosensory proteins (CSP) are soluble carrier proteins that may function in odorant reception in insects. CSPs have not been thoroughly studied at whole-genome level, despite the availability of insect genomes. Here, we identified/reidentified 283 CSP genes in the genomes of 22 mosquitoes. All 283 CSP genes possess a highly conserved OS-D domain. We comprehensively analyzed these CSP genes and determined their conserved domains, structure, genomic distribution, phylogeny, and evolutionary patterns. We found an average of seven CSP genes in each of 19 *Anopheles* genomes, 27 CSP genes in *Cx. quinquefasciatus*, 43 in *Ae. aegypti*, and 83 in *Ae. albopictus*. The *Anopheles* CSP genes had a simple genomic organization with a relatively consistent gene distribution, while most of the Culicinae CSP genes were distributed in clusters on the scaffolds. Our phylogenetic analysis clustered the CSPs into two major groups: CSP1-8 and CSE1-3. The CSP1-8 groups were all monophyletic with good bootstrap support. The CSE1-3 groups were an expansion of the CSP family of genes specific to the three Culicinae species. The Ka/Ks ratios indicated that the CSP genes had been subject to purifying selection with relatively slow evolution. Our results provide a comprehensive framework for the study of the CSP gene family in these 22 mosquito species, laying a foundation for future work on CSP function in the detection of chemical cues in the surrounding environment.

Introduction

Over time, insects have developed a complete chemosensing system to perceive chemical cues from external environment [1]. Discerned olfactory stimuli are conveyed into the central nervous system via electrical signal transduction, producing a series of behavioral responses such as feeding, courtship, and avoidance to adapt to the external environment [1,2]. The lymph of the insect sensillum houses all processes and interactions by which the environmental odorant molecules reach the nerve membrane receptors [1]. Several classes of olfactory proteins have

(2015FY210300), and the Science and Technology Research Project of Chongqing Municipal Education Commission (KJ1500328). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

been reported to be involved in chemosensory perception, including odorant-binding proteins (OBPs), chemosensory proteins (CSPs), odorant receptors (ORs), ionotropic receptors (IRs), and sensory neuron membrane proteins (SNMPs) [3]. Generally, OBPs and CSPs represent two functionally similar classes of carrier proteins, dissolving and transporting chemical signals or other stimuli of lipophilic compounds to the chemosensory receptors. CSPs are major binding proteins in insects, and they are primarily differentiated from OBPs in that they bind and carry non-volatile odorants and semiochemicals [4,5]. The earliest identified members of the CSP family, isolated from the antennae of *Drosophila melanogaster*, were the olfactory specific protein D (OS-D), the OS-D-like protein [6,7], and the pheromone-binding protein A-10 (A-10) [8]. Subsequently identified CSPs include CLP-1 [9], p10 [10], the sensory appendage protein (SAP) [7,11,12], and the CSPs themselves [13,14]. While these proteins have been early described in relation with the insect olfactory system [13], there is no solid evidence to support what remains an odd assumption. An increasing number of studies are rather in strong agreement with a role of CSPs in immune responses not only in moths, but also in flies and white-flies [15–18]. This is also supported by ubiquitous tissue distribution and ontogeny in expression of CSP genes, such as the labial palp, maxilla, pheromone gland, wing, and leg [5,13,19]. CSPs are all small, globular, soluble proteins that possess a conserved cysteine CSP motif (C1-X_{6–8}-C2-X_{16–21}-C3-X₂-C4) containing two disulfide bonds (C1-X_{6–8}-C2, C3-X₂-C4) [20,21]. The disulfide bonds in the CSP motif are inter-helical with two small loops, forming a rigid hydrophobic pocket involved in ligand binding [22–24]. All CSPs shuttle in the aqueous lymphatic fluid of chemosensilla [1]. Some CSPs are involved in chemosensory signal transduction, in the solubilization of pheromone components [5,14]. Other CSPs affect the physiological processes and behavior of insects, e.g., moulting [25], tissue formation or regeneration [10,15,26], reproduction [27], and resistance reactions [28].

Recently, CSPs have been identified in insect species besides *D. melanogaster*, including *Camponotus japonicas* [29], *Heliothis virescens* [30], *Apis mellifera* [31], *Bombyx mori* [32], and *Anopheles gambiae* [33]. Forêt *et al.* (2007) analyzed the members of the CSP gene family in *Ap. mellifera* using both bioinformatic annotation and expression profiling, and compared these CSPs with those in other arthropods to better understand their evolution and function [31]. Pelletier and Leal (2011) identified different families of olfactory proteins (OBPs, CSPs, SNMPs) in *Culex quinquefasciatus*, and analyzed the characterization and expression of genes in these three families [34]. Kulmuni *et al.* (2013) identified and annotated the CSP genes in the genomes of seven ant species, and studied their evolution [35]. Vieira and Rozas (2011) conducted an exhaustive comparative genomic analysis of the CSP and OBP gene families in 20 Arthropoda species, giving insight into the origin and evolutionary history of these two gene families [36]. Neafsey *et al.* (2015) sequenced and assembled the genomes and transcriptomes of 16 anopheline mosquitoes [37]. However, the genes in the CSP family have not been identified in these mosquito species, nor have they been thoroughly analyzed at the whole-genome level.

Mosquitoes are regarded as the deadliest animals to humans due to their capacity for infectious disease transmission [38]. At present, 22 mosquito genomes are available, providing a foundation for a comparative genomic study of mosquito CSPs. Here, we identified and annotated the CSP genes of these 22 mosquito species at the whole-genome level, and analyzed the characteristics and conserved domains of these CSPs using bioinformatic techniques. We constructed a phylogeny to inform mosquito classification. We also determined the rate of sequence evolution (non-synonymous to synonymous changes, Ka/Ks), and the putative orthologous CSPs across the mosquito species. Finally, we investigated species-specific expansions in the Culicinae subfamily using comparative genomic methods. This work provides a framework for study of the CSP family of genes in mosquitoes, and provides a basis for further

study of the specific role of CSPs in the detection of chemical signals in the surrounding environment by insects.

Materials and methods

Genome sequence sources

We downloaded 21 previously published assembled genome and transcriptome mosquito from VectorBase (<https://www.vectorbase.org>) [37,39–51]. We obtained the annotated genome (unpublished) and transcriptome of an additional species, *An. sinensis*, from the Institute of Entomology and Molecular Biology of Chongqing Normal University, China (see Table 1 for details of all data stated above, [52]). We also downloaded four CSP gene sequences from *D. melanogaster* (GenBank accession ID: CAG26928, NP_001286809, NP_001286871, and AAF49381) from GenBank (<http://www.ncbi.nlm.nih.gov/>).

Genome-wide identification of CSPs

To find all putative CSPs in the 22 mosquito species, we first performed several rounds of exhaustive BlastP respectively searches against the amino acid (aa) databases of these mosquitoes using characterized CSP sequences from GenBank (<http://www.ncbi.nlm.nih.gov/>) as queries (with an E-value threshold of 10^{-5}). Second, we built Hidden Markov Model (HMM) profiles downloaded from the Pfam database (<http://pfam.xfam.org/>) [53] using OS-D (for CSP; Pfam ID: PF03392), and performed an HMM search against the aa database of each mosquito species. Third, we performed a tBlastN search against each the genome assembly of each mosquito species using the corresponding aa sequences obtained in the previous two steps (with an E-value threshold of 10^{-5}). Using this procedure, we repeatedly identified candidate CSPs until no new hits were found. We manually removed duplicate sequences. We obtained the raw nucleotide sequences of all candidate CSPs directly from the genome. Finally, we used Fgenesh (<http://www.softberry.com>) [54] to predict the candidate CSP genes, and then confirmed all candidate CSP genes against CSP conserved domain information (OS-D, Pfam03392) with SMART (<http://smart.embl-heidelberg.de/>) [55]. We only considered sequences belonging to the OS-D superfamily as putative CSPs. All CSP genes were verified by BlastN searches against the EST database of each species (with E-value threshold of 1×10^{-15}). All putative CSP genes were classified and named according to their characteristics and phylogenetic relationships.

Analysis of CSP characteristics

For each predicted aa sequence encoded by our putative CSP genes, we calculated the molecular weight and the theoretical isoelectric point (pI) using ExPASy ProtParam (<http://web.expasy.org/protparam/>). We predicted the subcellular location with TargetP (v1.1; <http://www.cbs.dtu.dk/services/TargetP/>) [56]; the N-terminus signal peptides with PrediSi (<http://www.predisi.de>) [57]; and the secondary structure with PSIPRED (v3.3; <http://bioinf.cs.ucl.ac.uk/psipred>) [58]. We observed the structures of the putative CSP genes, including the intron phase, with GSDS (<http://gsds.cbi.pku.edu.cn/index.php>) [59], based on the corresponding coding and the genomic sequences of the putative CSP genes obtained.

The multiple alignments of CSP aa sequences of the 22 mosquito species were conducted by ClustalX [60], and the alignment parameters were set to the defaults with gap-opening and gap extension penalties 10.0 and 0.2, respectively. Each group of CSP aa sequences (CSP1–8 and CSE1–8, established based on the phylogeny below and previously published works) were further separated, aligned, and displayed with Genedoc [61]. The conserved domains of the

Table 1. Source of genomes used in the study and numbers of putative CSPs found in each genome.

Subfamily Genus/ Subgenus/ Series	Species	Abbreviations of species	Genome version	Genome assembly ID ^a	Genome size (Mb)	Putative CSPs	Genome assembly reference	Transcriptome assembly reference
Anophelinae						130		
Anopheles/ Nyssorhynchus	<i>An. darlingi</i>	Ada	AdarC3	GCA_000211455.3	173.92	4	Marinotti <i>et al.</i> (2013)	unpublished
	<i>An. albimanus</i>	Aal	AalbS1	GCA_000349125.1	165.33	8	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
Anopheles/ Anopheles	<i>An. sinensis</i>	Asi	unpublished	unpublished	327.22	8	unpublished	Chen <i>et al.</i> (2014)
	<i>An. atroparvus</i>	Aat	AatrE1	GCA_000473505.1	217.57	7	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
Anopheles/ Cellia/ Neomyzomyia	<i>An. farauti</i>	Afa	AfarF2	GCA_000473445.2	175.52	8	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
	<i>An. dirus</i>	Adi	AdirW1	GCA_000349145.1	209.79	8	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
Anopheles/ Cellia/Myzomyia	<i>An. funestus</i>	Afu	AfunF1	GCA_000349085.1	218.45	8	Neafsey <i>et al.</i> (2015)	Crawford <i>et al</i> (2010)
	<i>An. minimus</i>	Ami	AminM1	GCA_000349025.1	195.70	6	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
	<i>An. culicifacies</i>	Acu	AculA1	GCA_000473375.1	198.03	7	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
Anopheles/ Cellia/Neocellia	<i>An. maculatus</i>	Ama	AmacM1	GCA_000473185.1	141.20	6	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
	<i>An. stephensi</i>	Ast	AsteS1	GCA_000349045.1	216.26	6	Jiang <i>et al.</i> (2014)	Gokhale <i>et al.</i> (2013)
Anopheles/ Cellia/ Pyretophorus	<i>An. epiroticus</i>	Aep	AepiE1	GCA_000349105.1	216.83	7	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
	<i>An. christyi</i>	Ach	AchrA1	GCA_000349165.1	169.04	6	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
	<i>An. gambiae</i>	Aga	AgamP4/ AgamS1	GCA_000005575.2 GCA_000150785.1	268.44/ 229.28	8	Holt <i>et al.</i> (2002) Lawniczak <i>et al.</i> (2010)	Neafsey <i>et al.</i> (2015)
	<i>An. arabiensis</i>	Aar	AaraD1	GCA_000349185.1	239.13	7	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
	<i>An. quadriannulatus</i>	Aqu	AquaS1	GCA_000349065.1	275.35	6	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
	<i>An. merus</i>	Amr	AmerM2	GCA_000473845.2	244.34	7	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
	<i>An. melas</i>	Aml	AmelC2	GCA_000473525.2	222.01	6	Neafsey <i>et al.</i> (2015)	Neafsey <i>et al.</i> (2015)
	<i>An. coluzzii</i>	Aco	AcolM1	GCA_000150765.1	218.22	7	Lawniczak <i>et al.</i> (2010)	Cassone <i>et al.</i> (2014)
Culicinae						153		
Culex	<i>Cx. quinquefasciatus</i>	Cqu	CpipJ2	GCA_000209185.1	574.57	27	Arensburger <i>et al.</i> (2010)	Lv <i>et al.</i> (2016)
Aedes	<i>Ae. aegypti</i>	Aae	AaegL3	GCA_000004015.1	1342.21	43	Nene <i>et al.</i> (2007)	Vogel <i>et al.</i> (2017)
	<i>Ae. albopictus</i>	Aao	AaloF1	GCA_001444175.1	1868.07	83	Chen <i>et al.</i> (2015)	Esquivel <i>et al.</i> (2016)

^a All assembled mosquito genomes and transcriptomes were downloaded from VectorBase (<https://www.vectorbase.org>), except for those of *An. sinensis*. The genome and transcriptome of *An. sinensis* were sequenced and annotated by the Chongqing Normal University with the genome data not yet published, and the transcriptome sequences published in Chen *et al.* (2014).

<https://doi.org/10.1371/journal.pone.0190412.t001>

CSP aa sequences were identified through the NCBI conserved domain database (CDD) with default setting (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). We checked our multiple alignments against these conserved domains. We interrogated the occurrence frequencies of each CSP aa over the 22 mosquito species, displayed with WebLogo 3 (<http://weblogo.threplusone.com/>).

To locate the CSP genes within each of the 22 mosquito genomes, we performed tBlastN search against current assembly of each genome sequence of 22 mosquito species using the corresponding aa sequence of CSP gene. We considered genes located within 20Kb of each other a gene cluster [62]. We drew plots displaying the genomic distribution of the CSP genes on the chromosomes of the 22 mosquitos, as well as scaffolds and contigs manually with Photoshop CS8. We compared the distributions of CSP genes across a subgroup of 19 *Anopheles* species.

Phylogenetics and evolution analysis

The best-fit models (JTT) produced by using automatically generated trees in MEGA5 [63] for the putative CSP aa sequences of the 22 mosquitos and *D. melanogaster*, and we used maximum likelihood (ML) to construct unrooted phylogenies of these sequences with MEGA5 [63]. Bootstrap values were calculated with 1000 replicates; bootstrap values $\geq 50\%$ were marked on the branches of the unrooted ML trees constructed. We used the Interactive Tree of Life (<http://itol.embl.de/>) to display the trees [64].

To investigate selection pressures on the CSP genes, we calculated the Ka/Ks ratio (non-synonymous (Ka) to synonymous (Ks) substitution rates) for the gene group of CSP1-8 and CSE1-3, as well as all CSP genes for each mosquito species. We aligned the sequences, and constructed the tree topology of each CSP group with MEGA5 using default parameters. The consensus sequence of each gene group was used as reference in the Ka/Ks ratio calculation. We compared different substitution models in HyPhy [65] and in PAML (the program codeml) [66,67], and estimated Ka/Ks for each CSP group using the most conservative model chosen using HyPhy and PAML.

Results and discussion

Identification of CSP genes and divergence across mosquito species

We identified a total of 130 putative CSP genes from the genomes of these 19 *Anopheles* species, with an average of seven CSP genes per species (4 CSP genes in 1 species; 6 in 6 species; 7 in 6 species, and 8 in 6 species). We identified 153 putative CSP genes from the genomes of the three Culicinae species: 27 in *Cx. quinquefasciatus*, 43 in *Aedes aegypti*, and 83 in *Ae. albopictus* (Table 1, Fig 1). Among the 283 CSP genes identified across the 22 mosquito species, 269 genes are complete protein-coding sequences extracted from the genome assemblies. The remaining 14 genes (*AalCSP4*, *AalCSP5*, *AchCSP5*, *AcuCSP5*, *AdiCSP5*, *AfuCSP5*, *AmaCSP4*, *AmaCSP5*, *AstCSP5*, *CquCSP27*, *AaeCSP3*, *AaeCSP37*, *AaoCSP23* and *AaoCSP67*) were not complete due to incomplete genomic sequences. Out of the 283 genes, 266 were supported by transcriptome data but 17 not (*AcoCSP8*, *AepCSP5*, *AfaCSP7*, *CquCSP1*, *AaeCSP1*, and 12 *AaoCSPs* for *Ae. albopictus*, see S1 File for details). The aa sequences corresponding to these 269 genes all had four conserved cysteines as well as a CSP family domain (OS-D). We therefore considered all 269 genes as members of CSP family (i.e. insect pheromone-binding family or A10/OS-D, Pfam number: PF03392). An additional eight short gene fragments (coding length < 54 aa) with high similarities to CSPs, and three pseudogenes (*AmiCSP6*, *AepCSP4*, *AchCSP6*) lacking the characteristic CSP domain were not used in any analysis (Fig 1).

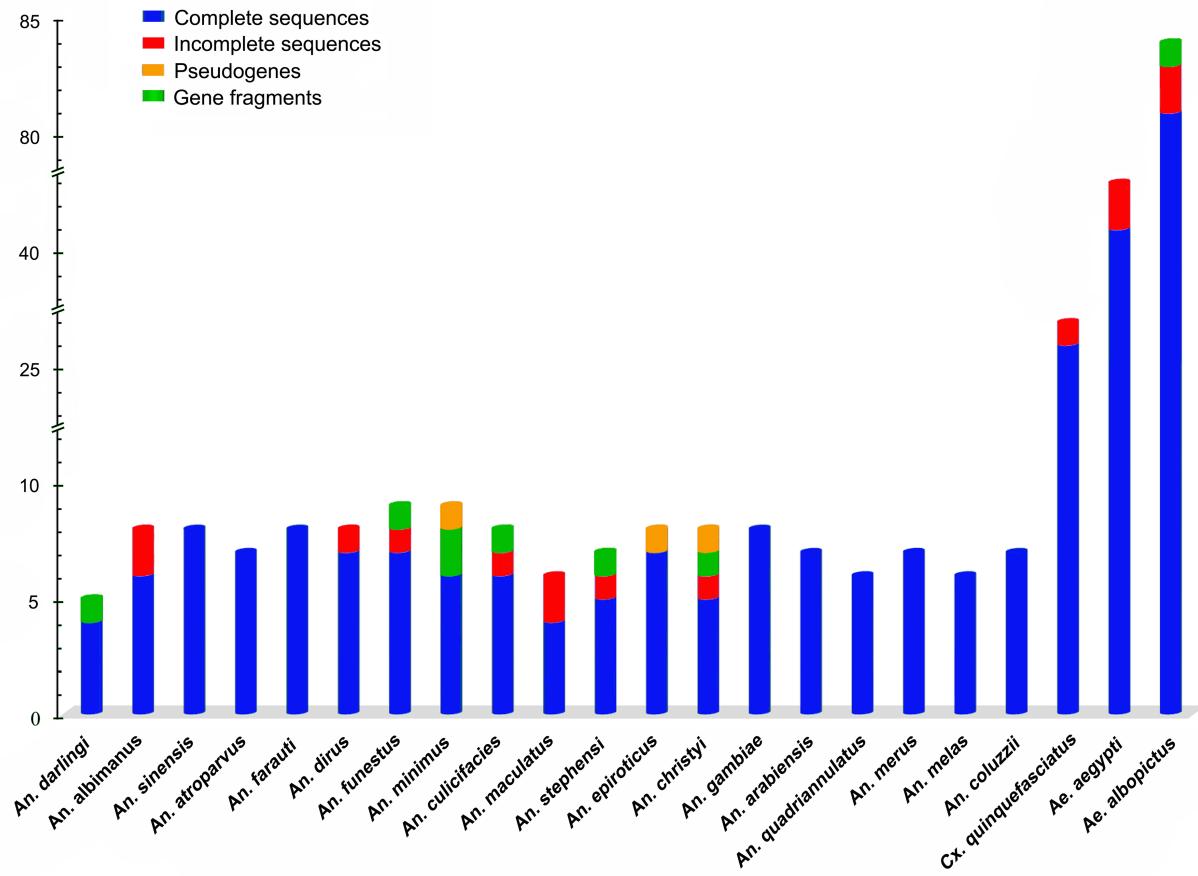


Fig 1. Numbers of CSP genes in the 22 mosquito species investigated. Gene numbers include complete, incomplete, and gene fragments as well as pseudogenes. Genes with complete and incomplete sequences are considered putatively functional CSP genes. Gene fragments sequences and pseudogenes are not used in further analysis.

<https://doi.org/10.1371/journal.pone.0190412.g001>

In the 22 mosquito species investigated, the CSP number of each species ranged from four genes in *An. darlingi* to 83 genes in *Ae. albopictus*. Based on the greater number of CSP genes found in the Culicinae species, we suggest that the CSP family expanded largely in the subfamily Culicinae mosquitoes in terms of the gene numbers. Various numbers of CSP genes have been reported in other insect species: one CSP gene is known in *Thricolepisma aurea* and *Fol-somia candida* [68]; six are known in *Pediculus humanus* [36] and *Ap. mellifera* [31]; 13 are known in *Acyrthosiphon pisum* [69]; 20 are known in *Tribolium castaneum* [70] and *B. mori* [5]; and 70 are known in *Locusta migratoria* [71]. The gene numbers could be also divergent within the same genus or family: the number of CSP genes varies from 3 to 4 across *Drosophila* [36] and from 11 to 21 among ants [35]. CSP genes have also been identified in very limited number of non-insect arthropods, although the number of CSPs is much less on average: only one CSP gene has been found *Ixodes scapularis* and *Artemia franciscana*; two in *Archispirostreptus gigas* and *Triops cancriformis*; and three in *Daphnia pulex* [36,68]. All of these results showed that the numbers of the CSP family of genes were highly variable and their evolution was divergent and dynamic among species [36,68]. CSP genes might be subject to different evolutionary pressures in different species through gene lost or gain, based on the need for different molecular mechanisms with which to detect and identify complex odor components. It

is likely that some insects, such as *Ae. albopictus* (83 CSP genes) and *L. migratoria* (70 CSP genes), might require large numbers of different carrier proteins for chemosensing.

Characterization of identified CSP genes

Across the 130 CSPs of the 19 *Anopheles* species, the theoretical isoelectric points (pI) ranged from 4.5 to 10.4; the aa sequence length ranged from 68 to 181 aa; and the molecular weight (Mw) ranged 8.2 to 20.2 kDa. We do not include CSP5s in these ranges as it had an unusually long aa sequence (187–331 aa) due to C-terminal extension, and a consequently high Mw (21.1–36.4 kDa). Across the 153 CSPs of the three Culicinae species, the pI ranged from 4.4 to 10.4; the aa sequence length ranged from 71 to 25; and the Mw ranged from 7.7 to 26.9. The CSPs of these two groups of mosquitoes were generally comparable. Most of these 283 CSPs (82.3%) encoded secretory pathway proteins with a 14 to 33 aa N-terminus signal peptide. The remaining CSP genes lacked a signal peptide: six genes in the shared CSP group (*AfaCSP7*, *CquCSP1*, *AaoCSP3*, *AfuCSP5*, *AmaCSP5* and *AaeCSP3*, although the sequences of the last three might be incomplete) and all 40 AaeCSPs and AaoCSPs genes in CSE3 group (the Culicinae-specific expansion CSP groups) (see phylogenetic analysis section). The subcellular location showed that two genes (*AfaCSP7* and *AaeCSP3*) belonged to mitochondria targeting peptides. Analysis of the predicted secondary structure of the CSPs suggested that almost all full-length CSPs were characterized by six α-helix structures (S1 Fig).

We investigated the frequency of conserved regions based on the alignment of CSPs across all species tested. A total of 269 CSPs (95.1% of the total CSPs) had complete conserved OS-D domain sequences (Pfam motif: Pfam03392). Of these, 243 had a 93 aa domain, while in an additional 26 the domain length ranged from 75 aa to 125 aa. The remaining 14 CSPs (4.9%) had a conserved OS-D domain but lacked at least one aa due to sequence incompleteness (see phylogenetic analysis section). Within the OS-D domain of complete sequences, four cysteine residues were completely conserved with two disulfide bonds linking each pair of neighboring cysteines (C1-X₆-C2, C3-X₂-C4) (Fig 2A). This pattern of conserved cysteines (C-Pattern, C1-X₆-C2-X₁₈-C3-X₂-C4) was comparable to those identified in other groups of insects (Fig 2B). Such high conservation across all mosquito species examined means that these genes carry out critical conserved functions. The insect CSP C-Pattern was: the first and second cysteines were separated by 5 to 8 non-cysteine aa residues; the third and fourth cysteines were separated by two non-cysteine aa residues; and the second and third cysteines were separated by 18 or 19 non-cysteine aa residues [72]. This C-Pattern might be different in non-insects: there were 12 and 1 or 3 residues between the first and second pair of cysteines in the *Julida* genus of millipedes [33].

Structure, location, and expansion of CSP genes

The eight groups of CSPs shared by the 22 mosquito species all had relatively consistent genomic structures (S1 File). Sequences belonging to the CSP1, CSP2, and CSP3 groups had a single exon from 381 bp to 390 bp long; the CSP4, CSP6, and CSP7 sequences had two exons; the CSP8 sequences had two or three exons; and the CSP5 sequences had four or five exons. Four CSP sequences did not follow this pattern (*AmaCSP5*, *AfuCSP5*, *AcuCSP5*, and *AaeCSP3*), but these might have been incomplete. Both the CSP5 and CSP7 groups differed with respect to the number and size of exons between the *Anopheles* and the Culicinae species. Three of the CSE groups of genes (CSE1, CSE2, and CSE3) had only one exon, except for two AaoCSPs (*AaoCSP14* and *AaoCSP44*) that had two exons. All 283 genes had 0 to 4 introns varying widely in size (from 29 bp to 24,165 bp). Of these, eight had introns longer than 10,000 bp (*AarCSP7*,

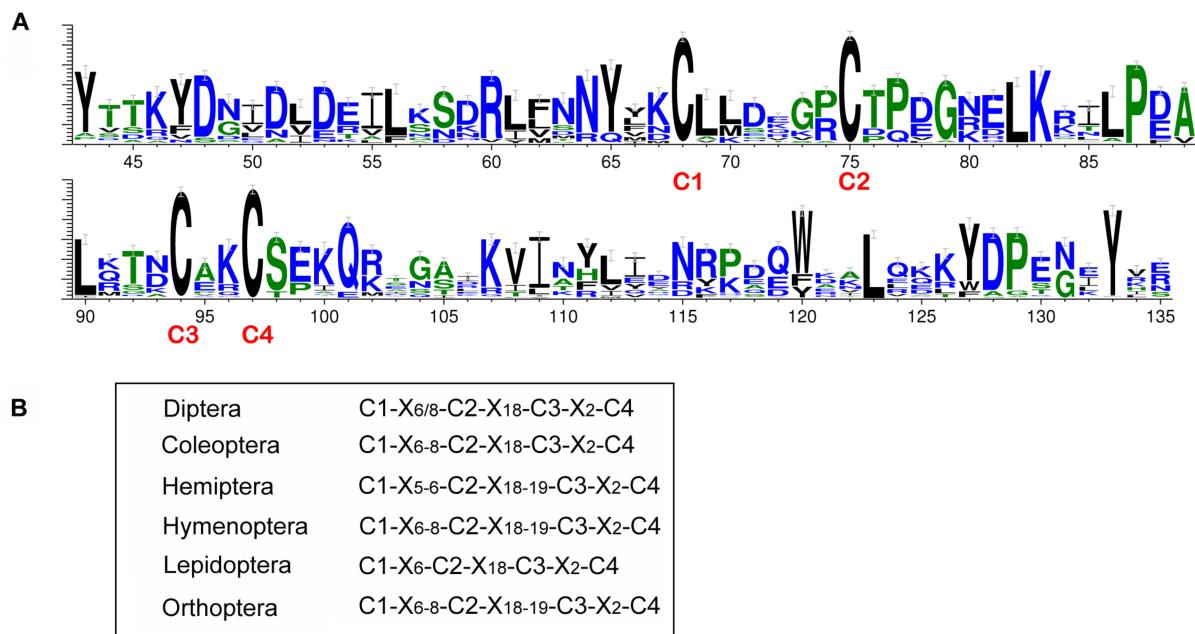


Fig 2. Conserved amino acid sequences and domains of the CSP family. A) The conserved aa sequences with frequency of aa occurrence for the 22 mosquito genomes. B) The conserved four cysteines domain (C1-C4) in different orders of insects. X_{6/8} indicates 6 or 8 non-cysteine aa; X₁₈ is 18 aa; X₆₋₈ is 6–8 aa.

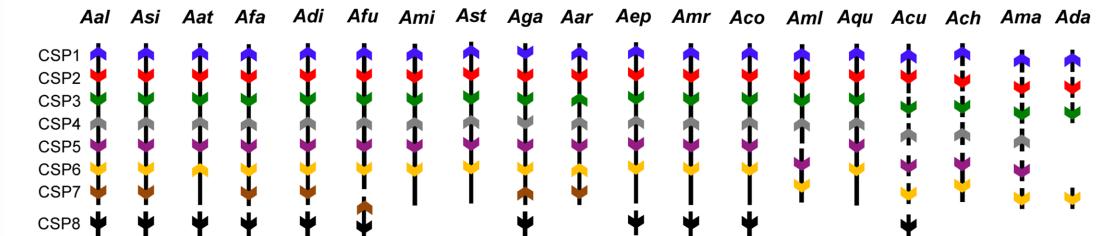
<https://doi.org/10.1371/journal.pone.0190412.g002>

AgaCSP7, AsiCSP5, AsiCSP7, CquCSP2, AaeCSP2, AaeCSP37 and AaoCSP2). All characteristics of the 283 CSP genes are listed in S1 File.

We located the CSP genes of each of the 22 mosquito species on the corresponding genomes (Fig 3). In 18 of *Anopheles* species, groups CSP1–6 are distributed in 1 to 6 scaffolds with the same relative positions. The exception is *An. darlingi* (Ada), which lacks CSP4 and CSP5 possibly due to genomic sequence incompleteness. All of the genes in groups CSP1–6 have the same gene direction except for *AgaCSP1*, *AarCSP3*, *AgaCSP4*, *AatCSP6*, and *AarCSP6* (Fig 3A). Only seven species (Aal, Asi, Afa, Adi, Afu, Aga and Aar) have the CSP7 gene. All CSP7 genes are adjacent to CSP6 in a single scaffold (except for *AfuCSP7* that is in another scaffold) and all have the same gene direction (except for *AfuCSP7* and *AgaCSP7*). Eleven species have the CSP8 gene, all with same gene direction, but located in separate scaffolds. These results suggest that the genomic organization of CSP1–7 in *Anopheles* is relatively simple, with these CSPs located close together, while CSP8 is rather more distant. Compared with the CSP genes of *B. mori*, *D. melanogaster*, *Ap. mellifera* and *T. castaneum*, the genomic organization shows clear genetic differentiation between different species [7,17,18]. For instance, all *BmorCSP* genes (20 *BmorCSP*, except *BmorCSP10*, 13 and 19) sit close to each other in the same genomic region that span over 104337 bps separated in average by about 2000–20000 bps [18]. Four *DmelCSP* genes, two of them (*DmelAAM68292* and *DmelPhk3*) are located within 5 Kb of each other, and approximately 900 Kb from *DmelPebIII* on chromosome 2R, the rest of the *DmelCSP* gene (*Dmelos-d*) is located on chromosome 3L [7].

In *Cx. quinquefasciatus*, the 27 *CquCSPs* were located on three different scaffolds (supercont3.42, 3.1044, and 3.2685, supercont indicating some large contig). Most of these (89%; *CquCSP1-24*) were distributed in the same scaffolds of supercont3.42, with *CquCSP4-24* gathered into a large clusters (Fig 3B). In *Ae. aegypti*, 83.7% of the CSPs (36 of 43) were located in supercont1.47; 14.0% in supercont1.170 (6 of 43) and 2.3% in supercont1.687 (1 of 43; *AaeCSP37*). In this species, 30 CSP genes were gathered into eight clusters. In *Ae. albopictus*,

A 19 Anopheles species:



B 3 Culicinae species:

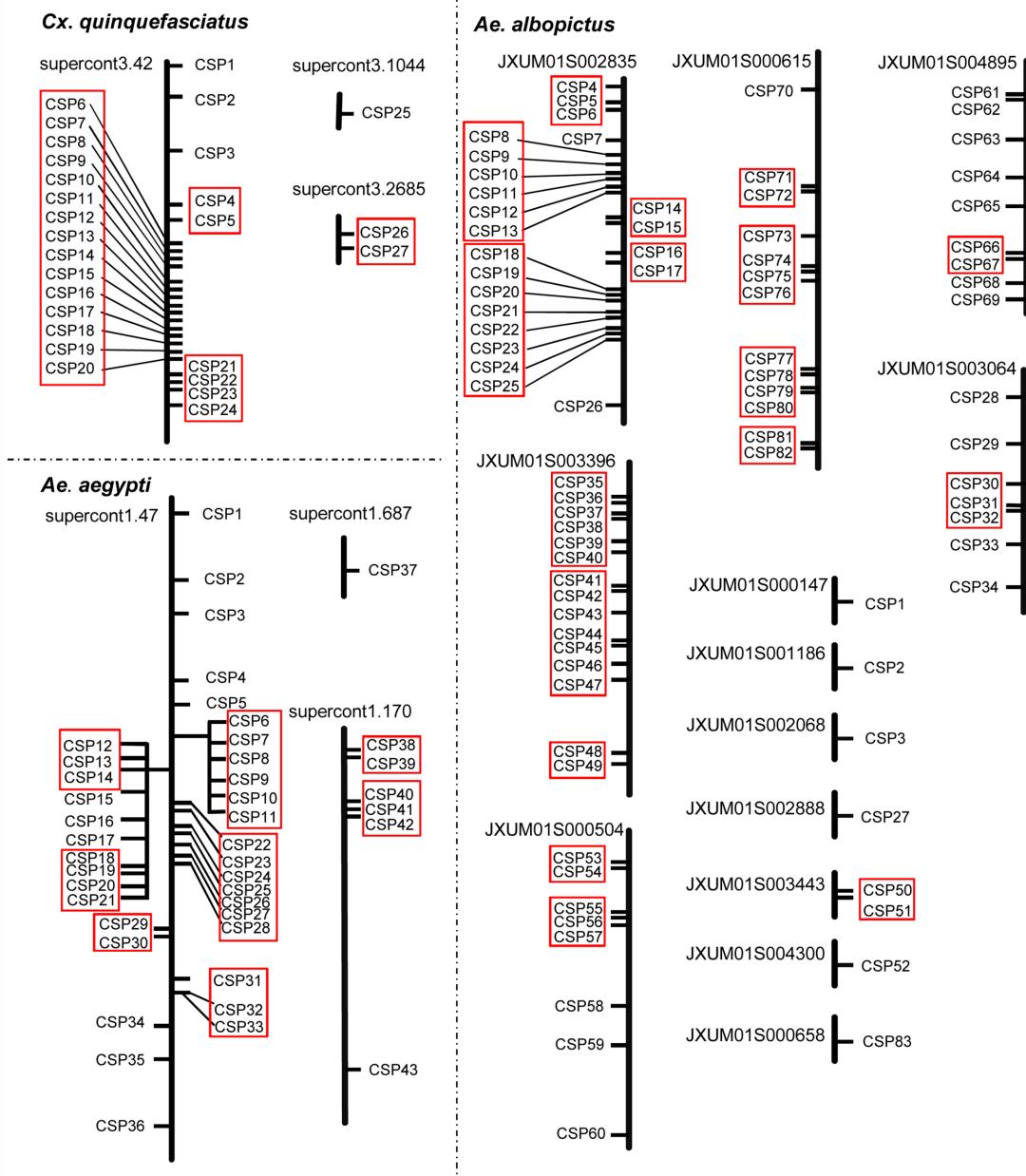


Fig 3. Genomic localization of CSP genes in 22 mosquito species. A) Relative positions of CSP genes in the 19 *Anopheles* species. The eight horizontal bars indicate eight different genes, marked in different colors with arrows showing

the 5'-3' direction of the sequences. B) Location of CSP genes in the three Culicinae species, with the 27–83 genes of each species marked in relative genomic distance. Gene clusters are represented by a red box.

<https://doi.org/10.1371/journal.pone.0190412.g003>

the 75 of the 83 CSP genes were located on 13 different scaffolds: 23 in JXUM01S002835, 15 in JXUM01S003396, 13 in JXUM01S000615, 9 in JXUM01S004895, 8 in JXUM01S000504, and seven JXUM01S003064. The remaining eight CSPs were located in seven different contigs. Of the 83 CSP genes, 24 genes only have 10 different aa sequences. That is, *AaoCSP8*, *AaoCSP10*, *AaoCSP11*, and *AaoCSP39* with identical aa sequences, and *AaoCSP9*, *AaoCSP12*, *AaoCSP37*, and *AaoCSP40* as well. In contrast to the *Anopheles* species, most of CSP genes of the three Culicinae species were distributed in clusters on the scaffolds, and were obviously expanded, likely through a series of gene duplication events. This was particularly noticeable in *Ae. albopictus*, where the CSP genes seemed to be the most expanded, and those genes with same aa sequences might experience most recent gene duplications through transposition of transcripts.

Phylogenetics of CSP genes in the 22 mosquito species

We constructed two unrooted ML trees based on the CSP aa sequences, using the best-fit model of evolution (WAG) as selected by ModTest. One ML tree (the "partial" tree) included six representative *Anopheles* species, three Culicinae species, and *D. melanogaster* (Fig 4A). The other ML tree included the 19 *Anopheles* species (Fig 4B). In the former tree, CSP genes were clearly separated into 11 groups with high bootstrap support ($\geq 67\%$): the shared groups CSP1-8 in all mosquitoes and the CSE1-3 for Culicinae-specific expansion groups (CSE being abbreviated for the Culicinae-specific expansion CSP groups). We named CSP1-8 based on conventions used in previous studies [31,34,70]. The names of CSE1-3 are novel, as earlier naming conventions were neither unified nor meaning-specific (e.g. OS-D in *D. melanogaster* [6] and *L. migratoria* [73]; and SAP in *Manduca sexta* [74] and *An. gambiae* species [11]). In the CSP1, CSP3, and CSP5 groups of *Cx. quinquefasciatus* and *Ae. albopictus*, three sister pairs of CSP genes might each have experienced a single gene duplication event (*CquCSP23* and *CquCSP24*; *AaoCSP26* and *AaoCSP58*; and *AaoCSP1* and *AaoCSP60*). Three CSPs of *D. melanogaster* (*DmeCSP4*, *DmeCSP3*, and *DmeCSP1*) grouped into CSP4, CSP6, and CSP7 groups, respectively, suggesting that the origin of these CSP groups predates the divergence of the mosquito and *Drosophila* lineages. The CSP5 and CSP8 groups have no homologs in *D. melanogaster*, suggesting that these groups originated after the split of mosquito and *Drosophila* lineages; alternatively, the homologs of *D. melanogaster* may have been lost. In addition, our results suggest that *DmeCSP2* showed to be a sister group with complexes of other groups (i.e. groups CSP1-3 and groups CSE1-3), suggesting that these groups may have expanded in mosquitoes in order to adapt to environmental change during evolution. There are 18 *CquCSPs*, 35 *AaeCSPs*, and 73 *AaoCSPs* in the groups CSE1-3. In the three Culicinae species, there are many more CSPs in these groups than there are in the groups CSP1-8. The CSE1 and CSE3 gene groups are present in all three Culicinae species investigated, but the CSE2 group was missing in *Cx. quinquefasciatus*, suggesting that CSE1-3 did not originate for all Culicinae mosquitoes.

In the CSP phylogenetic tree for all 19 *Anopheles* species (Fig 4B), the 130 CSP genes were also classified into eight groups (CSP1-8) with high bootstrap support ($\geq 99\%$; except for CSP1 with only 72% support). In each of the groups CSP1, CSP2, CSP3, and CSP6, we found a single CSP gene in each of the 19 *Anopheles* species, suggesting that these groups of genes are conserved in *Anopheles*, and that each group of genes may have a similar chemosensing function. In *An. gambiae*, it has been suggested that four groups (*AgaCSP1*, *AgaCSP2*, *AgaCSP3*,

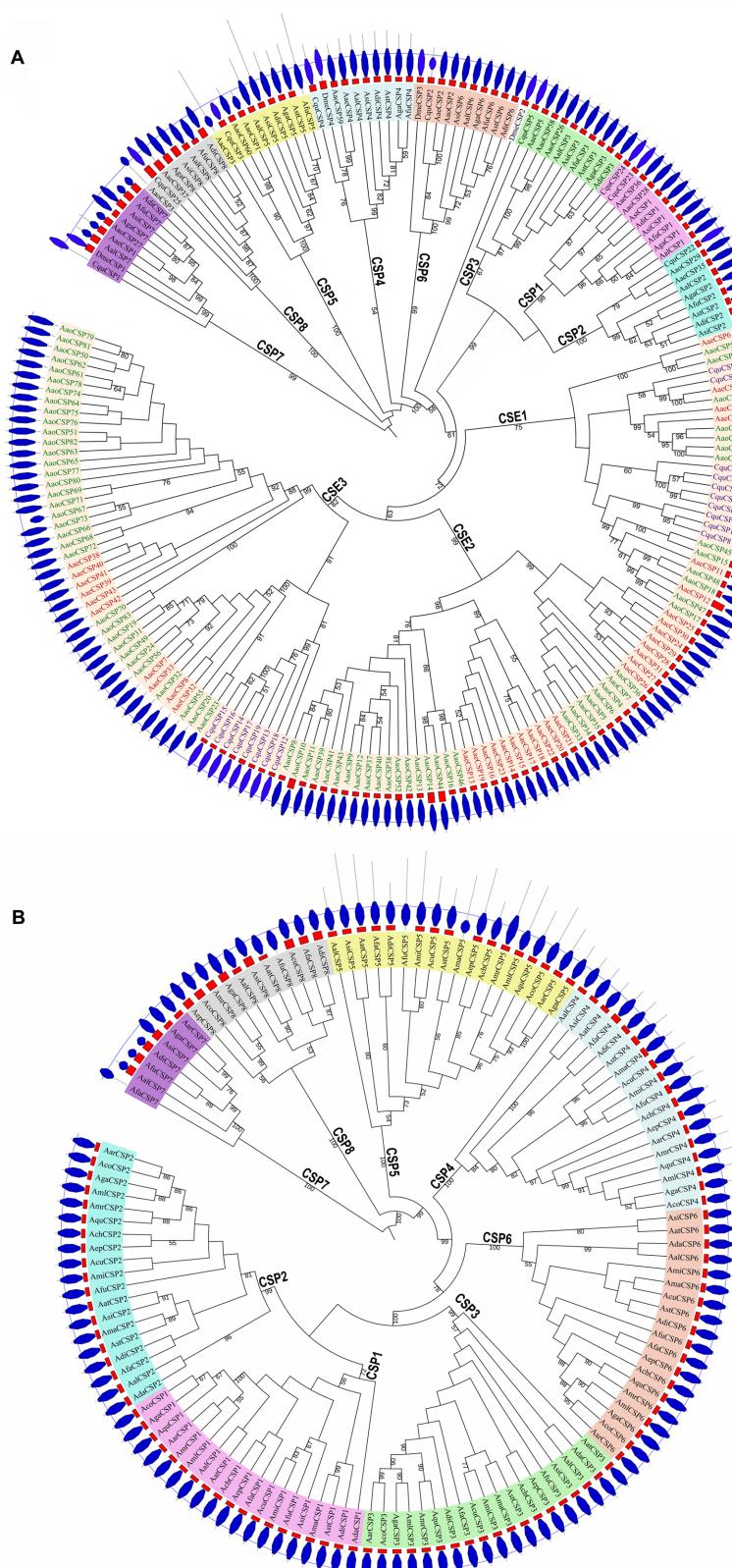


Fig 4. Phylogeny of the CSP aa sequences of 22 mosquito species and *D. melanogaster*. A) Phylogeny of CSP genes in six representative *Anopheles* species, three Culicinae species, and *D. melanogaster*. B)

Phylogeny of CSP genes in all 19 *Anopheles* species. ML phylogenetic trees were constructed using the WAG (Whelan and Goldman) model, as selected by ModTest. Bootstrap values were calculated using 1000 replications. Bootstrap values $\geq 50\%$ are marked on branches. The most outer layer of blue octagon denotes the position and size of the domain OS-D. The second outer layer of red rectangles shows the signal peptide and its relative size. CSP: Shared CSP groups; CSE: Culicinae-specific expanded CSP groups. In the CSE groups, CquCSPs are shown in purple, AaeCSPs in red, AaoCSPs in green. In the CSP groups, *Anopheles* CSPs are shown in black.

<https://doi.org/10.1371/journal.pone.0190412.g004>

and *AgaCSP6*, corresponding to *AgaSAP1*, *AgaSAP2*, *AgaSAP3*, and *AgaCSP1*) have independently evolved a common function, these groups are phylogenetically close to honeybee CSP3, a protein known to be involved in the binding of brood pheromone components [33]. The CSP4 and CSP5 genes are present in all *Anopheles* species except for *An. darlingi*. In *An. gambiae*, *AgaCSP4* (as *AgaCSP3*) showed a completely different spectrum of binding, and suggested to play a more specific role [33]. The function of *AgaCSP5* was unknown. Only eight of the *Anopheles* species investigated have CSP7 genes, and only 13 have CSP8. *AgaCSP7* (as *AgaCSP4*) and *AgaCSP8* (as *AgaCSP6*) in *An. gambiae* are phylogenetically close to the honeybee gene *AmelCSP5*, which encodes a protein previously reported to be involved in embryo development [33].

Amino acid sequence identities was high within the groups CSP1 (65–100%), CSP2 (70–100%), CSP3 (75–100%), CSP4 (45–99%), CSP6 (73–100%), and CSP8 (74–98%; S1 Fig, S2 File). The CSP7 group was present in only seven *Anopheles* species and three Culicinae species; aa sequence identities among those species was 33–100% slightly lower than those in the CSP1-CSP4, CSP6 and CSP8 groups (S2 File). Overall, aa sequence identities within the CSP5 group was low (19–99%), although we found higher aa sequence identities among the *Anopheles* species alone (57–99%), suggesting that there is great variation in CSP5 between *Anopheles* species and Culicinae species. Sequence identities among the three Culicinae species was higher (CSE1, 56–100%; CSE2, 77–100%; and CSE3, 56–100%), and the groups were more conserved (S2 File, S2 Fig).

Two major families of proteins in the chemosensory system, OBPs and CSPs, were thought to belong to a superfamily of general binding proteins, and to share a common ancestor near the origin of the arthropods. We searched the sequences of these two families in GenBank to investigate their origins [36]. We found CSPs in the arthropod subphylums Hexapoda, Crustacea, Myriapoda, and Chelicerata; CSPs were especially widespread in the insects. In contrast, OBPs were only present in the subphylum Hexapoda. These results suggest that CSPs originated earlier than OBPs, consistent with other studies [36,68,75,76]. We also found fewer, more variable CSP genes as compared to OBP genes, again consistent with previous work [36,68]. We were unable to construct a well-supported phylogenetic tree of all known CSP genes to confidently show their phylogenetic relationship. Finally, we only investigated the phylogenetic relationship of CSPs of mosquito species in this study.

Evolution of CSP genes

The comparison tests of three different substitution models (M0 (one-ratio), Branch model and Site model) using the program codeml in PAML and three models (SLAC, FEL and REL) in HyPhy with default parameters showed that the M0 and SLAC (single likelihood ancestor counting) were more conservative than other two, respectively. Therefore, the M0 and SLAC were applied in the subsequent analyses with the program codeml and HyPhy, respectively. The Ka/Ks ratios for CSP1-8 for all 22 mosquito species tested varied from 0.08 to 0.38 when estimated with HyPhy, and from 0.09 to 0.36 when estimated with PAML. The Ka/Ks ratios for CSE1-3 for the three Culicinae species ranged from 0.07 to 0.16 when estimated with

Table 2. Ka/Ks ratios of each group of CSP genes in the 22 mosquito species, as calculated by HyPhy and PAML.

	HyPhy Ka/Ks	PAML Ka/Ks
Shared CSP groups		
CSP1	0.15	0.23
CSP2	0.09	0.11
CSP3	0.12	0.19
CSP4	0.30	0.31
CSP5	0.38	0.36
CSP6	0.08	0.09
CSP7	0.34	0.28
CSP8	0.24	0.23
Culicinae-specific expansion CSP groups		
CSE1	0.16	0.17
CSE2	0.07	0.07
CSE3	0.11	0.10

<https://doi.org/10.1371/journal.pone.0190412.t002>

HyPhy, and from 0.07 to 0.17 when estimated with PAML (Table 2). The Ka/Ks ratios of all CSP genes as a whole for 22 mosquito species ranged from 0.19 to 0.47 when estimated with HyPhy, and from 0.13 to 0.53 when estimated with PAML. The average Ka/Ks ratio across all mosquito species was 0.29 (HyPhy) and 0.30 (PAML) (S3 File). The rates of Ka/Ks change are important for the understanding of the environmental selection pressures on genes in the dynamics species evolution [77,78]. Purifying selection is indicated when the value of Ka/Ks < 1, neutral evolution when Ka/Ks = 1, and positive selection when Ka/Ks > 1 [66]. Generally, as the Ka/Ks ratio decreases, the tolerated selection pressure of the gene increases; small Ka/Ks ratios indicate a highly conserved gene [79]. The largest Ka/Ks ratio we calculated was 0.53, suggesting that these CSP genes are likely subject to purifying selection with relatively slow evolution and high conservation. The results are consistent with orthologous groups of CSP genes in ant species, which are also highly conserved and thought to be under purifying selection [35].

Conclusion

Here we make the first comprehensive genome-wide analysis of the CSP gene family in 22 mosquito species, identifying and naming 283 CSP genes. Characteristic comparisons and phylogenetic analyses of these 283 genes suggested that eight groups of genes (CSP1-8) are shared across almost all mosquito species. Within each CSP group, gene structure is similar and aa sequence identity is high. We thus propose that genes within a group are homologous and perform similar functions across different mosquito species. In the Culicinae species, three additional group genes (CSE1-3) were found, in much larger numbers than those in CSP1-8. Most of CSP genes in Culicinae were distributed in clusters on the scaffolds. We suggest that the CSP family of genes is highly conserved in mosquitos, and that they evolve slowly with purifying selection. This comparative genomic study on the CSPs of 22 mosquito species provides a comprehensive framework for further investigation of possible CSP functions.

Supporting information

S1 Fig. Alignment of groups of CSP genes (CSP1-8) in the 22 mosquito species analyzed. Conserved cysteines are represented by a red box. The alpha helical domains ($\alpha 1-\alpha 6$) identified in the chemosensory protein are marked by spiral line above the alignment. CSP5 is

shown only up to the aa 187 as the C-terminal of this gene had long aa sequence extension.
(TIF)

S2 Fig. Alignment of each group of CSP genes (CSE1-3) in the three Culicinae species.
(TIF)

S1 File. Detailed information and intron-exon organization of the CSP genes in the 22 mosquito species.
(XLS)

S2 File. Percent identity and coverage between amino acid sequences in each CSP group (CSP1-8, CSE1-3) of the 22 mosquito species.
(XLS)

S3 File. The average Ka/Ks ratio in each of the 22 mosquito species, as estimated by HyPhy and PAML.
(XLS)

Author Contributions

Conceptualization: Ting Mei, Bin Chen.

Formal analysis: Ting Mei, Wen-Bo Fu, Zheng-Bo He, Bin Chen.

Funding acquisition: Bin Chen.

Investigation: Ting Mei, Wen-Bo Fu, Bo Li, Zheng-Bo He, Bin Chen.

Methodology: Wen-Bo Fu, Zheng-Bo He, Bin Chen.

Supervision: Bin Chen.

Writing – original draft: Ting Mei, Bin Chen.

Writing – review & editing: Ting Mei, Bin Chen.

References

1. Schneider D (1969) Insect olfaction: deciphering system for chemical messages. *Science* 163: 1031–1037. PMID: [4885069](#)
2. Getchell TV, Margolis FL, Getchell ML (1984) Perireceptor and receptor events in vertebrate olfaction. *Progress in Neurobiology* 23: 317–345. PMID: [6398455](#)
3. Leal WS (2013) Odorant Reception in Insects: Roles of Receptors, Binding Proteins, and Degrading Enzymes. *Annual Review of Entomology* 58: 373–391. <https://doi.org/10.1146/annurev-ento-120811-153635> PMID: [23020622](#)
4. Campanacci V LA, Hällberg B M, Jones TA, Giudici-Orticoni MT, Tegoni M, et al (2003) Moth chemo-sensory protein exhibits drastic conformational changes and cooperativity on ligand binding. *Proceedings of the National Academy of Sciences* 100: 5069–5074. <https://doi.org/10.1073/pnas.0836654100> PMID: [12697900](#)
5. Dani FR, Michelucci E, Francese S, Mastrobuoni G, Cappelozza S, La Marca G, et al. (2011) Odorant-binding proteins and chemosensory proteins in pheromone detection and release in the silkmoth *Bombyx mori*. *Chemical Senses* 36: 335–344. <https://doi.org/10.1093/chemse/bjq137> PMID: [21220518](#)
6. McKenna MP, Hekmat-Scafe DS, Gaines P, Carlson JR (1994) Putative *Drosophila* pheromone-binding proteins expressed in a subregion of the olfactory system. *Journal of Biological Chemistry* 269: 16340–16347. PMID: [8206941](#)
7. Wanner KW, Willis LG, Theilmann DA, Isman MB, Feng Q, Plettner E (2004) Analysis of the insect os-d-like gene family. *Journal of Chemical Ecology* 30: 889–911. PMID: [15274438](#)
8. Pikielny C, Hasan G, Rouyer F, Rosbash M (1994) Members of a family of *Drosophila* putative odorant-binding proteins are expressed in different subsets of olfactory hairs. *Neuron* 12: 35–49. PMID: [7545907](#)

9. Maleszka R, Stange G (1997) Molecular cloning, by a novel approach, of a cDNA encoding a putative olfactory protein in the labial palps of the moth *Cactoblastis cactorum*. *Gene* 202: 39–43. PMID: 9427543
10. Kitabayashi AN, Arai T, Kubo T, Natori S (1998) Molecular cloning of cDNA for p10, a novel protein that increases in the regenerating legs of *Periplaneta americana* (American cockroach). *Insect Biochemistry and Molecular Biology* 28: 785–790. PMID: 9807224
11. Biessmann H, Walter MF, Dimitratos S, Woods D (2002) Isolation of cDNA clones encoding putative odourant binding proteins from the antennae of the malaria-transmitting mosquito, *Anopheles gambiae*. *Insect Molecular Biology* 11: 123–132. PMID: 11966877
12. Biessmann H, Nguyen Q, Le D, Walter M (2005) Microarray-based survey of a subset of putative olfactory genes in the mosquito *Anopheles gambiae*. *Insect Molecular Biology* 14: 575–589. <https://doi.org/10.1111/j.1365-2583.2005.00590.x> PMID: 16313558
13. Angeli S, Ceron F, Scaloni A, Monti M, Monteforti G, Minnoccia A, et al. (1999) Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from *Schistocerca gregaria*. *European Journal of Biochemistry* 262: 745–754. PMID: 10411636
14. Jacquin-Joly E, Vogt RG, François M-C, Nagnan-Le Meillour P (2001) Functional and expression pattern analysis of chemosensory proteins expressed in antennae and pheromonal gland of *Mamestra brassicae*. *Chemical senses* 26: 833–844. PMID: 11555479
15. Sabatier L, Jouanguy E, Dostert C, Zachary D, Dimarcq J-L, Bulet P, et al. (2003) Pherokine-2 and -3. Two *Drosophila* molecules related to pheromone/odor-binding proteins induced by viral and bacterial infections. *European Journal of Biochemistry* 270: 3398–3407. <https://doi.org/10.1046/j.1432-1033.2003.03725.x> PMID: 12899697
16. Liu GX, Xuan N, Chu D, Xie HY, Fan ZX, Bi YP, et al. (2014) Biotype expression and insecticide response of *Bemisia tabaci* Chemosensory Protein-1. *Archives of Insect Biochemistry and Physiology* 85: 137–151. <https://doi.org/10.1002/arch.21148> PMID: 24478049
17. Liu G, Ma H, Xie H, Xuan N, Guo X, Fan Z, et al. (2016) Biotype characterization, developmental profiling, insecticide response and binding property of *Bemisia tabaci* chemosensory proteins: role of CSP in insect defense. *PloS One* 11: e0154706. <https://doi.org/10.1371/journal.pone.0154706> PMID: 27167733
18. Xuan N, Guo X, Xie HY, Lou QN, Lu XB, Liu GX, et al. (2015) Increased expression of CSP and CYP genes in adult silkworm females exposed to avermectins. *Insect Science* 22: 203–219. <https://doi.org/10.1111/1744-7917.12116> PMID: 24677614
19. Jin X, Brandazza A, Navarrini A, Ban L, Zhang S, Steinbrecht RA, et al. (2005) Expression and immuno-localisation of odorant-binding and chemosensory proteins in locusts. *Cellular and Molecular Life Sciences* 62: 1156–1166. <https://doi.org/10.1007/s0018-005-5014-6> PMID: 15928808
20. Zhou JJ, Kan Y, Antoniw J, Pickett JA, Field LM (2006) Genome and EST analyses and expression of a gene family with putative functions in insect chemoreception. *Chem Senses* 31: 453–465. <https://doi.org/10.1093/chemse/bjj050> PMID: 16581978
21. Briand L, Swasdiwan N, Nespolous C, Bézirard V, Blon F, Huet J-C, et al. (2002) Characterization of a chemosensory protein (ASP3c) from honeybee (*Apis mellifera* L.) as a brood pheromone carrier. *European Journal of Biochemistry* 269: 4586–4596. PMID: 12230571
22. Lartigue A, Campanacci V, Roussel A, Larsson AM, Jones TA, Tegoni M, et al. (2002) X-ray structure and ligand binding study of a moth chemosensory protein. *Journal of Biological Chemistry* 277: 32094–32098. <https://doi.org/10.1074/jbc.M204371200> PMID: 12068017
23. Tomaselli S, Crescenzi O, Sanfelice D, Ab E, Wechselberger R, Angeli S, et al. (2006) Solution structure of a chemosensory protein from the desert locust *Schistocerca gregaria*. *Biochemistry* 45: 10606–10613. <https://doi.org/10.1021/bi060998w> PMID: 16939212
24. Jansen S, Chmelík J, Žídek L, Padrtá P, Novák P, Zdráhal Z, et al. (2007) Structure of *Bombyx mori* chemosensory protein 1 in solution. *Archives of Insect Biochemistry and Physiology* 66: 135–145. <https://doi.org/10.1002/arch.20205> PMID: 17966128
25. Wanner K, Isman M, Feng Q, Plettner E, Theilmann D (2005) Developmental expression patterns of four chemosensory protein genes from the Eastern spruce budworm, *Chroistoneura fumiferana*. *Insect Molecular Biology* 14: 289–300. <https://doi.org/10.1111/j.1365-2583.2005.00559.x> PMID: 15926898
26. Nomura A, Kawasaki K, Kubo T, Natori S (1992) Purification and localization of p10, a novel protein that increases in nymphal regenerating legs of *Periplaneta americana* (American cockroach). *The International Journal of Developmental Biology* 36: 391–398. PMID: 1445782
27. Gong L, Luo Q, Rizwan-UI-Haq M, Hu MY (2012) Cloning and characterization of three chemosensory proteins from *Spodoptera exigua* and effects of gene silencing on female survival and reproduction. *Bulletin of Entomological Research* 102: 1.

28. Guo Z, Cheng Zhu Y, Huang F, Luttrell R, Leonard R (2012) Microarray analysis of global gene regulation in the Cry1Ab-resistant and Cry1Ab-susceptible strains of *Diatraea saccharalis*. Pest Manag Sci 68: 718–730. <https://doi.org/10.1002/ps.2318> PMID: 22228544
29. Hojo MK, Ishii K, Sakura M, Yamaguchi K, Shigenobu S, Ozaki M (2015) Antennal RNA-sequencing analysis reveals evolutionary aspects of chemosensory proteins in the carpenter ant, *Camponotus japonicus*. Scientific Reports 5: 13541. <https://doi.org/10.1038/srep13541> PMID: 26310137
30. Picimbon J-F, Dietrich K, Krieger J, Breer H (2001) Identity and expression pattern of chemosensory proteins in *Heliothis virescens* (Lepidoptera, Noctuidae). Insect Biochemistry and Molecular Biology 31: 1173–1181. PMID: 11583930
31. Forêt S, Wanner KW, Maleszka R (2007) Chemosensory proteins in the honey bee: Insights from the annotated genome, comparative analyses and expressional profiling. Insect Biochemistry and Molecular Biology 37: 19–28. <https://doi.org/10.1016/j.ibmb.2006.09.009> PMID: 17175443
32. Gong DP, Zhang HJ, Zhao P, Lin Y, Xia QY, Xiang ZH (2007) Identification and expression pattern of the chemosensory protein gene family in the silkworm, *Bombyx mori*. Insect Biochemistry and Molecular Biology 37: 266–277. <https://doi.org/10.1016/j.ibmb.2006.11.012> PMID: 17296501
33. Iovinella I, Bozza F, Caputo B, Della Torre A, Pelosi P (2013) Ligand-binding study of *Anopheles gambiae* chemosensory proteins. Chemical Senses 38: 409–419. <https://doi.org/10.1093/chemse/bjt012> PMID: 23599217
34. Pelletier J, Leal WS (2011) Characterization of olfactory genes in the antennae of the Southern house mosquito, *Culex quinquefasciatus*. Journal of Insect Physiology 57: 915–929. <https://doi.org/10.1016/j.jinsphys.2011.04.003> PMID: 21504749
35. Kulmuni J, Wurm Y, Pamilo P (2013) Comparative genomics of chemosensory protein genes reveals rapid evolution and positive selection in ant-specific duplicates. Heredity 110: 538–547. <https://doi.org/10.1038/hdy.2012.122> PMID: 23403962
36. Vieira FG, Rozas J (2011) Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. Genome Biology and Evolution 3: 476–490. <https://doi.org/10.1093/gbe/evr033> PMID: 21527792
37. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al. (2015) Highly evolvable malaria vectors: The genomes of 16 *Anopheles* mosquitoes. Science 347: 1258522. <https://doi.org/10.1126/science.1258522> PMID: 25554792
38. Sinka ME, Bangs MJ, Manguin S, Rubio-Palis Y, Chareonviriyaphap T, Coetzee M, et al. (2012) A global map of dominant malaria vectors. Parasites & Vectors 5: 69. <https://doi.org/10.1186/1756-3305-5-69> PMID: 22475528
39. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, et al. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 298: 129. <https://doi.org/10.1126/science.1076181> PMID: 12364791
40. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, et al. (2007) Genome Sequence of *Aedes aegypti*, a major arbovirus vector. Science 316: 1718–1723. <https://doi.org/10.1126/science.1138878> PMID: 17510324
41. Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, et al. (2010) Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. Science 330: 86–88. <https://doi.org/10.1126/science.1191864> PMID: 20929810
42. Crawford JE, Guelbeogo WM, Sanou A, Traoré A, Vernick KD, Sagnon NF, et al. (2010) De novo transcriptome sequencing in *Anopheles funestus* using Illumina RNA-Seq technology. PLoS One 5: e14202. <https://doi.org/10.1371/journal.pone.0014202> PMID: 21151993
43. Lawniczak MK, Emrich SJ, Holloway AK, Regier AP, Olson M, White B, et al. (2010) Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. Science 330: 512. <https://doi.org/10.1126/science.1195755> PMID: 20966253
44. Gokhale K, Patil DP, Dhotre DP, Dixit R, Mendki MJ, Patole MS, et al. (2013) Transcriptome analysis of *Anopheles stephensi* embryo using expressed sequence tags. Journal of Biosciences 38: 301–309. PMID: 23660664
45. Marinotti O, Cerqueira GC, de Almeida LG, Ferro MI, Loreto EL, et al. (2013) The Genome of *Anopheles darlingi*, the main neotropical malaria vector. Nucleic Acids Research 41: 7387–7400. <https://doi.org/10.1093/nar/gkt484> PMID: 23761445
46. Cassone BJ, Kamdem C, Cheng C, Tan JC, Hahn MW, Costantini C, et al. (2014) Gene expression divergence between malaria vector sibling species *Anopheles gambiae* and *An. coluzzii* from rural and urban Yaoundé Cameroon. Molecular Ecology 23: 2242–2259. <https://doi.org/10.1111/mec.12733> PMID: 24673723

47. Jiang X, Peery A, Hall AB, Sharma A, Chen XG, Waterhouse RM, et al. (2014) Genome analysis of a major urban malaria vector mosquito, *Anopheles stephensi*. *Genome Biol* 15: 459. <https://doi.org/10.1186/s13059-014-0459-2> PMID: 25244985
48. Chen XG, Jiang X, Gu J, Xu M, Wu Y, Deng Y, et al. (2015) Genome sequence of the Asian Tiger mosquito, *Aedes albopictus*, reveals insights into its biology, genetics, and evolution. *Proceedings of the National Academy of Sciences of the United States of America* 112: E5907. <https://doi.org/10.1073/pnas.1516410112> PMID: 26483478
49. Esquivel CJ, Cassone BJ, Piermarini PM (2016) A de novo transcriptome of the Malpighian tubules in non-blood-fed and blood-fed Asian tiger mosquitoes *Aedes albopictus*: insights into diuresis, detoxification, and blood meal processing. *PeerJ* 4: e1784 <https://doi.org/10.7717/peerj.1784> PMID: 26989622
50. Lv Y WW, Hong S, Lei Z, Fang F, Guo Q, et al. (2016) Comparative transcriptome analyses of deltamethrin-susceptible and -resistant *Culex pipiens pallens* by RNA-seq. *Mol Genet Genomics* 291: 309–321. <https://doi.org/10.1007/s00438-015-1109-4> PMID: 26377942
51. Vogel KJ, Valzania L, Coon KL, Brown MR, Strand MR (2017) Transcriptome sequencing reveals large-scale changes in axenic *Aedes aegypti* larvae. *Plos Neglected Tropical Diseases* 11: e0005273. <https://doi.org/10.1371/journal.pntd.0005273> PMID: 28060822
52. Chen B, Zhang Y-J, He Z, Li W, Si F, Tang Y, et al. (2014) De novo transcriptome sequencing and sequence analysis of the malaria vector *Anopheles sinensis* (Diptera: Culicidae). *Parasites & vectors* 7: 1. <https://doi.org/10.1186/1756-3305-7-314> PMID: 25000941
53. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* 44: D279–D285. <https://doi.org/10.1093/nar/gkv1344> PMID: 26673716
54. Nasiri J, Naghavi M, Rad SN, Yolmeh T, Shirazi M, Naderi R, et al. (2013) Gene identification programs in bread wheat: a comparison study. *Nucleosides Nucleotides & Nucleic Acids* 32: 529–554. <https://doi.org/10.1080/15257770.2013.832773> PMID: 24124688
55. Letunic I, Doerks T, Bork P (2015) SMART: recent updates, new developments and status in 2015. *Nucleic Acids Research* 43: D257–D260. <https://doi.org/10.1093/nar/gku949> PMID: 25300481
56. Emanuelsson O, Nielsen H, Brunak S, Von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* 300: 1005–1016. <https://doi.org/10.1006/jmbi.2000.3903> PMID: 10891285
57. Hiller K, Grote A, Scheer M, Munch R, Jahn D (2004) PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Research* 32: W375–379. <https://doi.org/10.1093/nar/gkh378> PMID: 15215414
58. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16: 404–405. PMID: 10869041
59. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. (2015) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31: 1296–1297. <https://doi.org/10.1093/bioinformatics/btu817> PMID: 25504850
60. Thompson JD, Gibson T, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*: 2.3. 1–2.3. 22.
61. Caffrey DR, Dana PH, Mathur V, Ocano M, Hong EJ, Wang YE, et al. (2007) PFAAT version 2.0: a tool for editing, annotating, and analyzing multiple sequence alignments. *BMC Bioinformatics* 8: 381. <https://doi.org/10.1186/1471-2105-8-381> PMID: 17931421
62. He X, He ZB, Zhang YJ, Zhou Y, Xian PJ, Qiao L, et al. (2016) Genome-wide identification and characterization of odorant-binding protein (OBP) genes in the malaria vector *Anopheles sinensis* (Diptera: Culicidae). *Insect science* 23: 366–376. <https://doi.org/10.1111/1744-7917.12333> PMID: 26970073
63. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739. <https://doi.org/10.1093/molbev/msr121> PMID: 21546353
64. Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* 44: W242–245. <https://doi.org/10.1093/nar/gkw290> PMID: 27095192
65. Pond SL, Frost SD (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21: 2531–2533. <https://doi.org/10.1093/bioinformatics/bti320> PMID: 15713735
66. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24: 1586–1591. <https://doi.org/10.1093/molbev/msm088> PMID: 17483113

67. Pond SLK, Frost SD (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular biology and evolution* 22: 1208–1222. <https://doi.org/10.1093/molbev/msi105> PMID: 15703242
68. Pelosi P, Iovinella I, Felicioli A, Dani FR (2014) Soluble proteins of chemical communication: an overview across arthropods. *Frontiers in Physiology* 5: 320. <https://doi.org/10.3389/fphys.2014.00320> PMID: 25221516
69. Zhou JJ, Vieira FG, He XL, Smadja C, Liu R, et al. (2010) Genome annotation and comparative analyses of the odorant-binding proteins and chemosensory proteins in the pea aphid *Acyrtosiphon pisum*. *Insect Molecular Biology* 19: 113–122. <https://doi.org/10.1111/j.1365-2583.2009.00919.x> PMID: 20482644
70. Dippel S, Oberhofer G, Kahnt J, Gerischer L, Opitz L, Schachtner J, et al. (2014) Tissue-specific transcriptomics, chromosomal localization, and phylogeny of chemosensory and odorant binding proteins from the red flour beetle *Tribolium castaneum* reveal subgroup specificities for olfaction or more general functions. *BMC Genomics* 15: 1141. <https://doi.org/10.1186/1471-2164-15-1141> PMID: 25523483
71. Zhou XH, Ban LP, Iovinella I, Zhao LJ, Gao Q, Felicioli A, et al. (2013) Diversity, abundance, and sex-specific expression of chemosensory proteins in the reproductive organs of the locust *Locusta migratoria manilensis*. *Biological chemistry* 394: 43–54. <https://doi.org/10.1515/hzs-2012-0114> PMID: 23096575
72. Xu YL, He P, Zhang L, Fang SQ, Dong SL, Zhang YJ, et al. (2009) Large-scale identification of odorant-binding proteins and chemosensory proteins from expressed sequence tags in insects. *BMC Genomics* 10: 632. <https://doi.org/10.1186/1471-2164-10-632> PMID: 20034407
73. Picimbon JF, Dietrich K, Breer H, Krieger J (2000) Chemosensory proteins of *Locusta migratoria* (Orthoptera: Acrididae). *Insect Biochemistry & Molecular Biology* 30: 233–241. PMID: 10732991
74. Robertson (1999) Diversity of odourant binding proteins revealed by an expressed sequence tag project on male *Manduca sexta* moth antennae. *Insect Molecular Biology* 8: 501–518. PMID: 10620045
75. Pelosi P, Zhou JJ, Ban LP, Calvello M (2006) Soluble proteins in insect chemical communication. *Cellular and Molecular Life Sciences* 63: 1658–1676. <https://doi.org/10.1007/s0018-005-5607-0> PMID: 16786224
76. Sanchez-Gracia A, Vieira FG, Rozas J (2009) Molecular evolution of the major chemosensory gene families in insects. *Heredity* 103: 208–216. <https://doi.org/10.1038/hdy.2009.55> PMID: 19436326
77. Gillespie JH (1994) The causes of molecular evolution: Oxford University Press.
78. Tomoko O (1995) Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *Journal of Molecular Evolution* 40: 56–63. PMID: 7714912
79. Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 267: 275–276. PMID: 865622