

Review

Not peer-reviewed version

Renaming the ‘OS-D/CSP’ family: ‘Lipoid-Binding Proteins’—Molecular Nomenclature, Structure, Expression, Evolution, and Intracellular Functions

Guoxia Liu , [Botong Sun](#) , [Wei Fan](#) , Shousong Yue , [Qiuxia He](#) , [Jean-François Picimbon](#) *

Posted Date: 1 October 2024

doi: 10.20944/preprints202410.0043.v1

Keywords: CSP; 4 Cysteines Soluble Protein; mucin; actin skeleton regulatory complex; nuclear pore complex protein; transcription initiation factor; gene regulation; immune response; insecticide resistance



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

Renaming the 'OS-D/CSP' Family: 'Lipoid-Binding Proteins'—Molecular Nomenclature, Structure, Expression, Evolution, and Intracellular Functions

Guoxia Liu ¹, Botong Sun ^{2,3}, Fan Wei ², Shousong Yue ¹, Qiuxia He ²
and Jean-François Picimbon ^{1,2}

¹ Biotechnology Research Center, Shandong Academy of Agricultural Sciences, Jinan, Shandong, 250100, Shandong, China

² Science and Technology Service Platform of Shandong Academy of Sciences, Shandong Academy of Sciences Foreign Students Pioneer Park, Qilu University of Technology (Shandong Academy of Sciences), Jinan, 250103, Shandong, China

³ Shandong SME Productivity Promotion Center, Jinan, 250103, Shandong, China

* Correspondence: jfpicimbon@163.com & jfpicimbon@gmail.com, jfpicimbon@sdas.org;
Tel: +00-86-17661098959

Simple summary: Chemosensory proteins (CSPs) are small soluble proteins that mediate olfactory recognition at the periphery of sensory receptors in insects, similarly to odorant-binding proteins (OBPs) >> (Wikipedia). We argue that the nomenclature of this insect protein superfamily is outdated (it is based on early studies on its presence in the sensory lymph of the antennal lobe). However, the new name of the protein molecule superfamily needs to be much more accurate due to a large study from the last 20 years described here. We offered to rename this protein family to reflect the latest findings on its locations, functions, and properties made in recent years. Because of the highly conserved molecular characteristic of its four adjacent cysteine residues, we propose renaming the protein "CSP" to "4CSP" (4 Cysteines Soluble Protein). It's also recommended to use the common peptide 5-letter code ("Bommo-4CSP") for abbreviations of distinct animal proteins.

Abstract: Chemosensory proteins (CSP) are found in the olfactory sensory organs (antennae and maxillary palps) and/or gustatory sensory organs (labellum and legs) and long been accepted to function through binding 'odorants'. Yet, the same CSPs are also expressed in many tissues other than olfactory and gustatory organs, such as gut, brain, fat body, wing, epidermis, *corpora allata*, pheromone gland, prothoracic gland, etc. In this review, we suggested the rename "chemosensory protein (CSP)" to "4 Cysteines Soluble Protein (4CSP)". To support our idea, we reviewed CSPs from several aspects: their primary sequences, existence of CSP in insects, hexapods, and bacteria, expression during development and in other tissues, evolution of CSP genes, genomic mining, their ancestral functions and lipoid binding, function outside the chemosensory/binding aspect, their intracellular function, and from structure to function. The scope of the review per se is broad, and this is especially true given the volume of data that has been gathered on CSP expressed in ways that are not consistent with the olfactory paradigm. We discuss the Mp10 story in aphids, our research on insecticide resistance, and other lipid transport and immunity-related processes. There is substantial evidence to support the non-chemosensory aspects of CSPs, which is where the majority of discussions lean. There are still very few, if any, statements outlining the aspects of CSPs' chemosensory properties, specifically how they activate olfactory receptor neurons (ORNs). We present an analysis showing a stronger correlation between them and the actin complex family proteins, mucins, and translation initiation factors. An additional argument is the existence of CSPs outside of insects. Therefore, strong arguments in favor of renaming chemosensory proteins are becoming evident here, outweighing the drawbacks.

Keywords: CSP; 4 Cysteines Soluble Protein; mucin; actin skeleton regulatory complex; nuclear pore complex protein; transcription initiation factor; gene regulation; immune response; insecticide resistance

1. Introduction

A few years ago, there was a case regarding the reevaluation and/or renaming of allatostatins due to their pleiotropic properties, which go beyond their inhibition of juvenile hormone (JH) biosynthesis in the *corpora allata* [1]. This is an example of an invertebrate peptide family being named for a highly contested function (e.g. chemosensing), with the additional information that several cutting-edge investigations, fresh findings regarding evolution, molecule expression, molecule localization, and binding characteristics, as well as two decades more of documented research, strongly imply a different function that goes well beyond insect olfaction.

Odorant-Binding Proteins (OBPs) and Chemosensory Proteins (CSPs) constitute two distinct classes of small soluble proteins that have been historically been grouped under the conventional umbrella of “Odor-Binding Proteins”. The goal of this review is to revise this established convention, at least for CSPs, OS-Ds, or ‘pherokines’, as one of the multiple names given to this protein family.

They are thought to mediate the recognition of odor molecules, odorants, and ligands to olfactory receptors (ORs), at the periphery of sensory dendrites in the insect sensillum [2–5]. According to Lartigue et al. [6], CSPs are composed of six α -helices with an approximate molecular weight of 10-12 kDa (or 110-120 amino acid residues), four cysteines that form two nearby disulfide bridges, the two disulfide bridges each connect two closely spaced cysteines, and a narrow hydrophobic tunnel functional structure. Four CSP structures have so far been identified in locusts (*Schistocerca gregaria*) and moths (*Mamestra brassicae*, *Bombyx mori*, and *Spodoptera litura*) [6–10]. Jansen et al. described CSP as a “prism” that is well suited for interacting with long-chain fatty acids (FAs) [7,8]. The ligand-binding site is flexible enough to accommodate various ligand sizes [11,12]. At this point of the study, we know that the sole function of CSPs is to bind ligands, so multiple features are not heavily supported, although they are suggested by the tissue-distribution and developmental profiling of this protein gene family.

CSP molecules are present in insects at all steps of their life cycles, from eggs and larvae through immatures and adults [13–16]. In locusts (*Locusta migratoria*), CSPs are mostly expressed in the legs and antennae of both sexes, and they have been linked to phase shift (phenotypic plasticity) in that species [17–19]. Nevertheless, all CSPs in moths are up regulated in most insect body tissues after exposure to avermectin pesticide molecules, particularly in the gut and fat body [20]. This is crucial for highlighting a role in insecticide resistance as well as tissue development, as opposed to a role in olfaction. The problem is that, it is well known that CSP expression occurs in non-sensory organs and during early development, research on these proteins’ functional roles has been largely and stubbornly carried out using a set of semio-chemicals, recombinant protein testing, and competitive fluorescent binding experiments as described by Ban et al. [21]. The function of these proteins is frequently evaluated using fluorescence displacements, and the identity and purity of the protein are frequently disputed issues. Furthermore, the number of loci and genomic organization are frequently overlooked in these studies. It is common practice to use the number of transcripts (see RNA) far too quickly and when analyzing a protein gene family such as the CSPs, and to mistakenly refer to it as the number of genes (see loci and genomic organization). This is particularly significant for CSPs found in insects for which RNA and protein editing has been shown [22,23].

The experiment on *Chilo suppressalis* (Chisu, Crambidae) is one example of a study that does not show that these CSP proteins are directly involved in olfaction perception of semio-chemicals. Chisu is the rice stem borer, large enough to allow for the dissection of all internal tissues for a pertinent gene expression analysis. A far more comprehensive and pertinent study would have compared the thorax, gut, fat body, epidermis, and multiple glands instead of just using the “abdomen” as the sole non-olfactory tissue in the Chisu experiment [24]. A more accurate and

relevant analysis would use an equal number of subjects testing, rather than 200 antennae as opposed to just 20-40 abdomen or head. Inter-individual differences exist, and this is a typical experiment that ignores them and tilts the results in favor of antennae. Even when comparing the Chisu CSPs to only four tissues-samples—the abdomen, legs, head, and antennae—and comparing 100 individuals—the antennae—to just 20-40—the abdomen and head—none of them showed expression specific to the antennae [24]. The functional significance of the odor ligand molecules under test was often unclear when examining pheromones, host plant volatiles, and behaviors. It hardly ever correlated with a ‘logical’ ligand, or a ligand that can be bound by a protein that is expressed in the antennae, legs, hindguts, fat bodies, ovaries, and eggs. The protein’s tissue distribution was either never thoroughly investigated or wasn’t even taken into consideration when selecting a potentially true physiological ligand, despite the fact that this is a strong prerequisite for the selection of potential ligands. As a result, the functional significance of the protein is still up for debate, especially in cases where binding assays did not compare a wide variety of molecules or were limited to odors.

One such example is oleamide, which is well known for being both a strong inhibitor of salivary mandibular gland branching morphogenesis and a potential CSP ligand [21]. Where and when the protein molecule is expressed should be a major factor in defining the true functional ligand. Similar findings apply to OBPs from *Aedes aegypti* mosquitoes, which are expressed throughout the insect body, including the antennae, legs, and abdomen, but are only partially characterized by means of a small number of randomly selected semiochemicals [25]. There is serious uncertainty regarding the function of the protein because this ‘molecular’ study did not analyze any of the chemicals found in *Ae. aegypti*, ranging from eggs, pupae and nymphs to adults, in all many different tissues [26]. For example, methyl tetradecanoate (C15H30O2) chemical, which is very prevalent in the female thorax, should be tested. Then, since they are found at high concentration in the female wing tissue and at high levels in the male antennae, testing should be done for (Z)-9-hexadecenoic acid methyl ester (C17H32O2), hexadecanoic acid methyl ester (C17H34O2), (Z,Z,Z)-9,12,15-octadecatrienoic acid methyl ester (C19H32O2), 8,11-octadecadienoic acid methyl ester (C19H34O2), methyl stearate (C19H38O2), and ethyl oleate (C20H38O2) [26].

It is currently widely known that an elevated load of CSPs (‘pherokines’) is observed in fly hemolymph and in all insect tissues subsequent to chemical, microbial, or viral infection [20,27–29]. We can demonstrate this by examining transcriptome datasets that have been compiled from various insect species and that have examined treated and untreated samples in relation to the three challenges (chemical, microbial, and viral) [30–37]. As differentially expressed genes depending on the infectious treatment, CSPs and OBPs have been included in the majority of these studies.

Liu et al. have discussed the specific function of ‘CSPs’ in lipid transport in relation to insecticide resistance using the sweetpotato whitefly *Bemisia tabaci* (Bemta) as a model study [28,38,39]. Insecticide-mediated upregulation and interaction of the protein (Bemta-CSP1) with long chain C18-fat lipid (C18:2, linoleic acid, LA) have been demonstrated by Liu et al. [38]. When a whitefly is sprayed with neonicotinoid Thiametoxam, its expression of Bemta-CSP1 increases in a dose dependent manner. Nevertheless, it has been demonstrated that the Bemta-CSP1 protein binds LA rather than thiametoxam, suggesting that CSP functions in the insecticide response via lipid metabolic pathways and LA [38]. This was the first thorough and innovative study that focused the CSP research on lipid transport and insecticide resistance rather than olfaction.

We offer a review of the various roles of “CSPs” and CSP-like proteins in a large diversity of organisms, including insects. We claim that the sequence homology existing between CSPs, CSP-like proteins, and other proteins with diverse fundamental intracellular functions, warrants a renaming of this large group of proteins into “4CSPs”, for 4 Cysteine Soluble Proteins. We begin by rethinking the idea of the biological function of the molecules based on transcription initiation factor (TIF), mucin, and CSP analogies (sequence identity and structural attributes). We reveal that, when it comes to analyzing the tissue-distribution and developmental profiling of every member of the CSP family, one of the main tenets of the literature—the role of these molecules in chemosensing—is merely accepted without any solid supporting evidence or logical approach. For at least a few CSPs, there is some supporting evidence that they bind to relevant olfactory molecules through competitive

binding assays, eventually exhibiting some ligand specificity, but there isn’t any conclusive proof of tissue distribution. What is the purpose of a CSP expressed in the gut and fat body that binds to smells? The same kind of evidence has been gathered for OBPs in numerous insect species, and it may be time to reconsider OBPs’ function in olfaction. By emphasizing CSP’s distinct relationship to lipoids, intracellular events, DNA regulation, and the unique significance of this protein family’s role in regulating a variety of intracellular molecular complexes and metabolic pathways in multiple tissues—not just the antennae and legs—we can provide a new definition of the biological function of CSP.

We claim the renaming applies not only to locusts and moths but to other insects as well, possibly even to species that are not insects at all. Undoubtedly, this should also be applied to similar proteins found in other insect species (such as *Drosophila*’s OBP, which is widely expressed in tissues other than the gustatory and olfactory sensory organs throughout all stages of development). However, the review does not address OBP or the OBP-like proteins (i.e., lipocalins) of mammals; instead it focuses on CSP, for which there is mounting evidence of a role beyond the chemosensory paradigm.

2. History of the Discovery of “CSPs” and Their Uncertain or Pleiotropic Roles in Insects

Table 1 compiles the history of the various names used to describe this molecule protein family as early as 1992 [40]. Even though p10 was found in leg regeneration tissue rather than antennal sensory structures, it changed names three times between 1994 and 2000 (see Table 1). All of the changes ignored the first protein gene family member to be reported in the literature and had the same obvious and unambiguous meaning (chemosensory function, limited to chemosensory tuning). Despite the lack of sufficient evidence, the protein gene family was renamed to exclude any possible developmental or lipid functions and to reflect only chemosensory functions [41–43].

Table 1. The family of “ChemoSensory Proteins” (“CSPs”) goes by several names.

Name	Organism	Authors	Year	Reference
p10	<i>Periplaneta americana</i>	Nomura et al.	1992	
Ebsp-3/PebIII	<i>Drosophila melanogaster</i>	Dyanov et al.	1994	AAA87058*
A10	<i>Drosophila melanogaster</i>	Pikielny et al.	1994	[41]
OS-D	<i>Drosophila melanogaster</i>	McKenna et al.	1994	[42]
Pam	<i>Periplaneta americana</i>	Picimbon & Leal	1999	[3]
CSP	<i>Schistocerca gregaria</i>	Angeli et al.	1999	[4]
SAP	<i>Manduca sexta</i>	Robertson et al.	1999	[43]
Pherokine	<i>Drosophila melanogaster</i>	Sabatier et al.	2003	[27]
Mp10	<i>Myzus persicae</i>	Bos et al.	2010	[44]

* Dyanov, H.M.; Lyozin, G.I.; Dzitoeva, S.G.; Korochkin, L.I. A cDNA of *Drosophila melanogaster* ejaculatory bulb specific protein III (PEBme III). Unpublished.

The fourth revision to the p10 nomenclature was “CSP”, which maintains a strong emphasis on chemosensory function over developmental, neuronal growth, and brain tissue repair functions. What possible use these protein molecules may have in chemical detection is called into question by the discovery of “B-CSPs” (CSPs in bacterial species) and their presence in the fat body and gut of insects, as well as by the injection of aphid Mp10, which activates plant immune systems [39,44–46]. The specific function of CSP molecules in chemosensing or olfaction is still up for debate after more than two decades of research because there isn’t a single piece of evidence indicating they interact with sensory neurons. There is no evidence to support the theory that “CSPs attach to smells and transfer them to ORs”—in insects, crustaceans, or, most definitely, plants, bacteria, and viruses.

This CSP protein family has been shown to be pleiotropic and to function without the chemosensory system since 2003 [5,27]. The term “Pherokines” refers to highly prevalent CSP-like molecules in flies’ hemolymph that arise from bacterial, viral, or chemical infection ([27–29]; Table 1),

which presents the concept of p10 or CSP molecule in the insect defense system and immune responses. In order to emphasize these molecules' strong expression in lipid-layered chitin cell walls, first immune barriers, the exoskeleton (to which muscles are attached), the epidermis, epicuticles, major outer surfaces, antennae, eyes, legs, and other peripheral sensory organs, it was eventually suggested to use "CSP" for "Cuticular Sensory Proteins" [47]. As a result, a growing body of research has been conducted on this molecule's protein gene family, reporting functions that extend well beyond the sensory branches and the olfactory system.

What distinguishes p10, OS-D, A10, SAP, CSP, Mp10, and pherokine from one another? The most remarkable characteristic of this family of protein molecules, which is expressed throughout the insect's body [20], is entirely overlooked by many synonyms (Table 1). To account for all family members across species of arthropods, crustaceans, insects, isopods, and microbes in the bacterial prokaryotic superkingdom, the entire molecule's protein gene family should be renamed. Furthermore, it is necessary to take into account the possible connection between CSPs and intracellular processes.

3. Broad Survey and Analysis of CSP Sequences in Insects and Hexapods

Regarding the CSP gene editing seen in *Bombyx*, RNA editing, post-translational modification, and protein mutations have been connected to the diversity of the CSP protein family, specific variations in the RNA and peptide molecules, and ultimately the emergence and evolution of new organismal phenotype. The *Bombyx* CSP family exhibits clear evidence of recoding at the molecule synthesis, as evidenced by specific motif insertion, amino acid inversion, and the inclusion of a Glycine residue adjacent to a Cysteine at specific positions in protein primary sequences [22,23]. Insect CSPs and recoding mechanisms may be related, as RNA editing has also been reported for whitefly CSPs [28,38,39]. These mutation processes involving RNA and peptides may account for or be the cause of the differences in CSP primary sequences found in different insect species (see Figure 1).

By comparing the CSP primary sequences from *Acyrtosiphon pisum* (Acypi), *Apis mellifera* (Apime), *Aedes aegypti* (Aedae), *Anopheles gambiae* (Anoga), *Bombyx mori* (Bommo), *Culex pipiens* (Culpi), *Drosophila* (Droso), *Nasonia vitripennis* (Nasvi), *Pediculus humanus humanus* (Pedhu), and *Tribolium castaneum* (Trica), it was possible to determine whether these mutations can result in the emergence of a new function for the molecule and a new phenotype for the insect organism (Figure 1 & Table S1). Bommo-CSPs shared roughly 27-52% identity with CSPs from Pedhu, while Acypis showed roughly 15-52% identity with Pedhus. Looking at the identity score to Order subject, the Acypi sequences contrast with those for Bommo and Pedhu (see Table S1). This makes some sense, as both *Bombyx* and *Pediculus* are holometabolous, whereas the pea aphid is a true bug. While Pedhu-CSPs showed high identity to those from Diptera, Hymenoptera, and Coleoptera, those from Acypi show more identity to Hemipterans. On the other hand, Bommos exhibited the highest degree of similarity to their molecular orthologs in many different types of moth lepidopterans and the kingdom of bacteria (altermodales, bacillales, enterobacterales, kitasatosporales, lysobacterales, and myxococcales, among others; see Table S1 and [45,46]). Pedhus significantly grouped with Acypis when comparing the N-terminal sequences, but with Bommos when comparing the C-terminal sequences (see Figure S1). There seemed to be more branching Acypi-Pedhu than Bommo-Pedhu when comparing the entire amino acid sequences (Figure S2A). However, except for the group of Acypi005842, Acypi000093, BommoCSP19, Pedhu54410, Pedhu594420, and Acypi03368 (bootstrap value: 82%), there was no obvious orthology grouping (Figure S2B). "CSPs" from *Bacillus amyloliquefasciens* and *Klebsiella pneumoniae* were also present in this group. At least these sequences seem to have retained the "old" function of CSP, independent of the insect species, and did not produce a new function for it (Figure S2B). The 100% bootstrap relatedness between *Myzus persicae* Mp10 and Acypi000097 placed them in close proximity to this group (Figure S2).

Remarkably, the tree also revealed important relationships between allergen Thap1 and PAN1 proteins and multiple Bommo-CSPs (Table S1 & Figure S2B). Specifically, Thap1 established orthology groups consisting of Bommo-CSP14, Bommo-CSP11, and Bommo-CSP17, as well as the

three pseudogenes Bommo-CSP5, Bommo-CSP16, and Bommo-CSP18. These relationships between Bommo-CSP and Thap1 are depicted in the phylogenetic tree with remarkably high bootstrap support values (98-100%; Figure S2B). Bommo-CSPs typically exhibit a 28-29% sequence identity with Thap1. High consensus and low consensus residues are where the CSP and Thap1 proteins are similar (see sequence alignment; Figure S3). Bommo-CSPs and PAN1 have a 37% sequence identity in common. PAN1 and Thap1 also share high consensus and low consensus residues (see sequence alignment; Figure S3), including all essential residues and the four cysteines that define the “CSP” family. Bommo-CSP10, a unique CSP with a protein weight of 24.65 kDa and a 211-residue sequence that ends multiple Prolines, is specifically related to PAN1 (99% bootstrap value; Figures S2B & S3). Thap1 is an isoallergenic protein variant (Allergen Nomenclature; 15-kDa IgE-binding protein). PAN1 is a protein that has transmembrane and cytoplasmic separate domains. It is a protein that is necessary for early larval ecdysis, the completion of the molting cycle during the adult molt in worms, and developmental events that occur throughout the mid to late larval stages [48]). It is an essential control point for the development of larvae and the progression of several tissues during the transition from the larval to adult state in *C. elegans*. With PAN1 being a protein with over 1000 amino acids and CSPs having just over 100-200 amino acids, one should be extremely dubious about any conservation of function or shared ancestry. But PAN1’s sequence alignment with CSP sequences is extremely evident, as “CSP” is clearly visible in PAN1’s N-tail (Figure S3). PAN-1 and Thap1 molecules are unrelated to chemosensing, and it is very likely that the proteins that are associated with these molecules are unrelated to chemosensing as well (see Table S1 and Figures S2 and S3).

It is noteworthy that, depending on the species of insect, the protein’s polypeptide length between Cys29-Cys37 and Cys56-Cys59 varies from 6-8 to 18-19 (Figure 1). A two-residue amino acid insertion has been found between Cys29-Cys37 for Culpi and Nasvi (Figure 1). Locmi-CSPs were found to have the same sequence mutation variation [17]. An amino acid insertion (one residue) between Cys56-Cys59 has been found for Acypi, Apime, and Nasvi (Figure 1). In particular, the jewel wasp phenotype Nasvi exhibits amino acid mutation (insertion or deletion) in multiple locations within the CSP structure (Figure 1). The degree of mutation in the CSP protein molecule is highly dependent on the species of the insect, as shown by the strictly preserved intercysteine gaps in *Droso*, *Anoga*, *Aedae*, *Bommo*, *Trica*, and *Pedhu* (Figure 1). A mutation variation occurs in an insect’s CSP that increases its specificity to plant compounds [49]. Moreover, it is shown via molecular graphics and modeling that specific mutations result in the loss of alpha-turn (α2), which is directly related to α-helical profiling and functional structure. As a result of this observation, it has been proposed that insect CSP mutations be studied for cancer research [50]. Given the number of mutations found in Bommo in this region of the molecule [22,23], it would be very interesting to look into the molecular effects of the precise variation in the intercysteine gap between the species and between the tissues. In order to test a loss of binding or protein function, site-directed mutagenesis experiments aimed at adding or removing a residue or a complete residue motif should be investigated if the intercysteine gap is changed in any of these CSPs. Our findings indicate that all CSPs appear to have two disulfide bridges and four cysteines (4C) in order to preserve the stability of the molecule. However, there are differences in the length and make-up of the gap between the cysteine residues. This undoubtedly gives the molecule a new purpose and may facilitate the emergence of a novel insect phenotype (Figure 1).

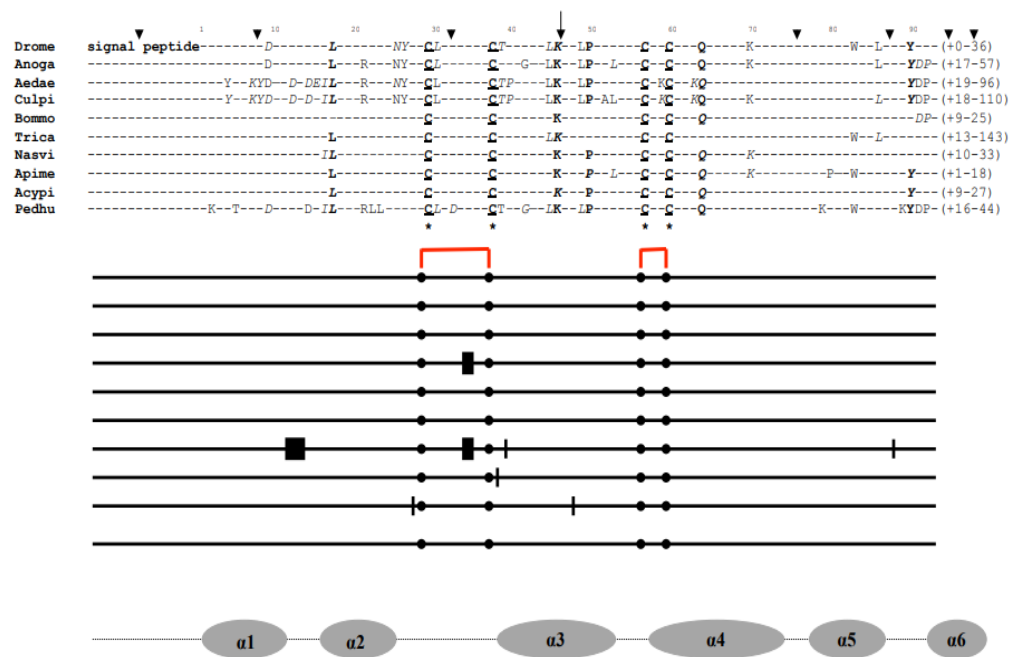


Figure 1. Consensus amino acid alignment of insect CSP sequences. Twelve *Drosophila* species are referred as to *Droso*: *D. ananassae*, *D. erecta*, *D. grimshawi*, *D. melanogaster*, *D. mojavensis*, *D. persimilis*, *D. pseudoobscura*, *D. sechellia*, *D. simulans*, *D. virilis*, *D. yakuba*, and *D. willistoni*; *Anoga*: *Anopheles gambiae*; *Aedae*: *Aedes aegypti*; *Culpi*: *Culex pipiens*; *Bommo*: *Bombyx mori*; *Trica*: *Tribolium castaneum*; *Nasvi*: *Nasonia vitripennis*; *Apime*: *Apis mellifera*; *Acypi*: *Acyrtosiphon pisum*; *Pedhu*: *Pediculus humanus humanus*. The majority of CSPs share the amino acids indicated in bold. The underlined amino acids are those that are strictly conserved across all CSPs. The stars in their respective positions indicate the four cysteine residues (4C) that are characteristic of CSPs. When comparing CSPs from the same species, amino acid residues that change in the CSP sequence are indicated by italicization. With Edman degradation [3,13], the first amino acid in the N-terminal sequence and signal peptide of insect CSPs are identified. The CSP genes' introns are consistently found after Lysine 45 (see up arrow). The codon for amino acid 46 (Glu, Ser, Lys, Asn, or Asp) is broken up by the intron. Additional introns' points of insertion are shown by the black triangles. As indicated by the squares, there are varying numbers of amino acid residues between Cys29-Cys37 and Cys56-Cys59 (6-8, 18-19, respectively). The location of functional elements (α -helices) is indicated by the grey circles beneath the alignment. The disulfide bridge and interlocked cysteines are shown in red.

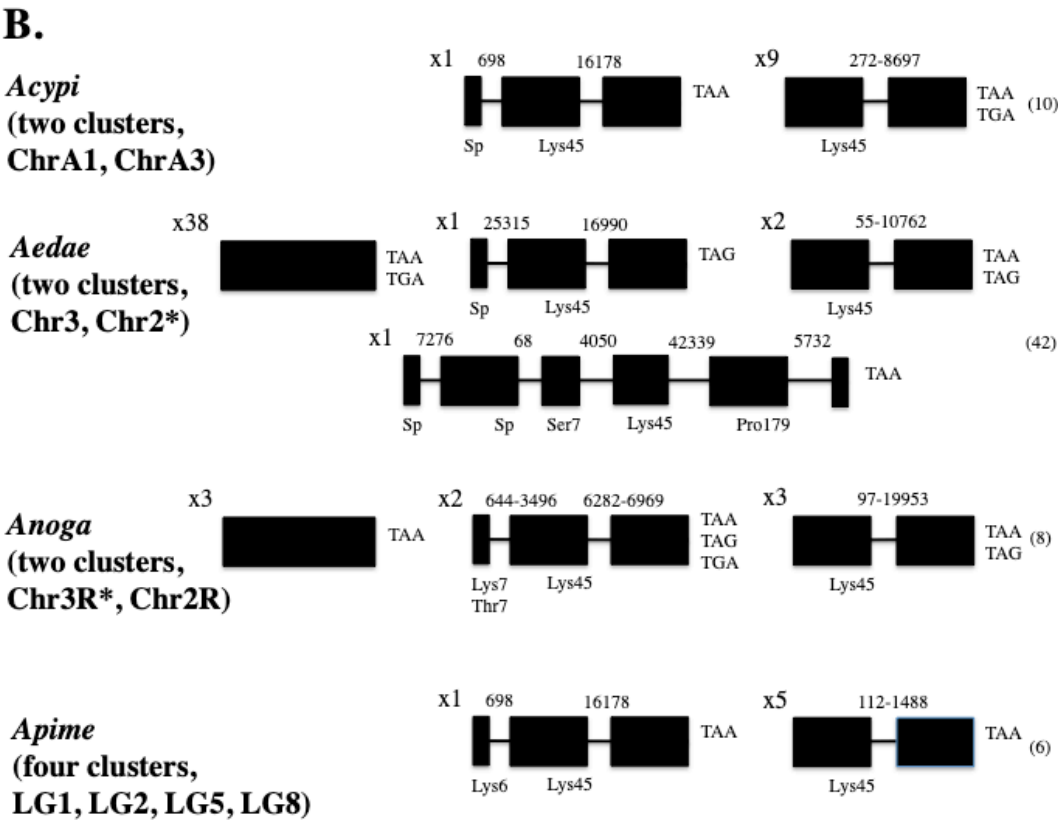
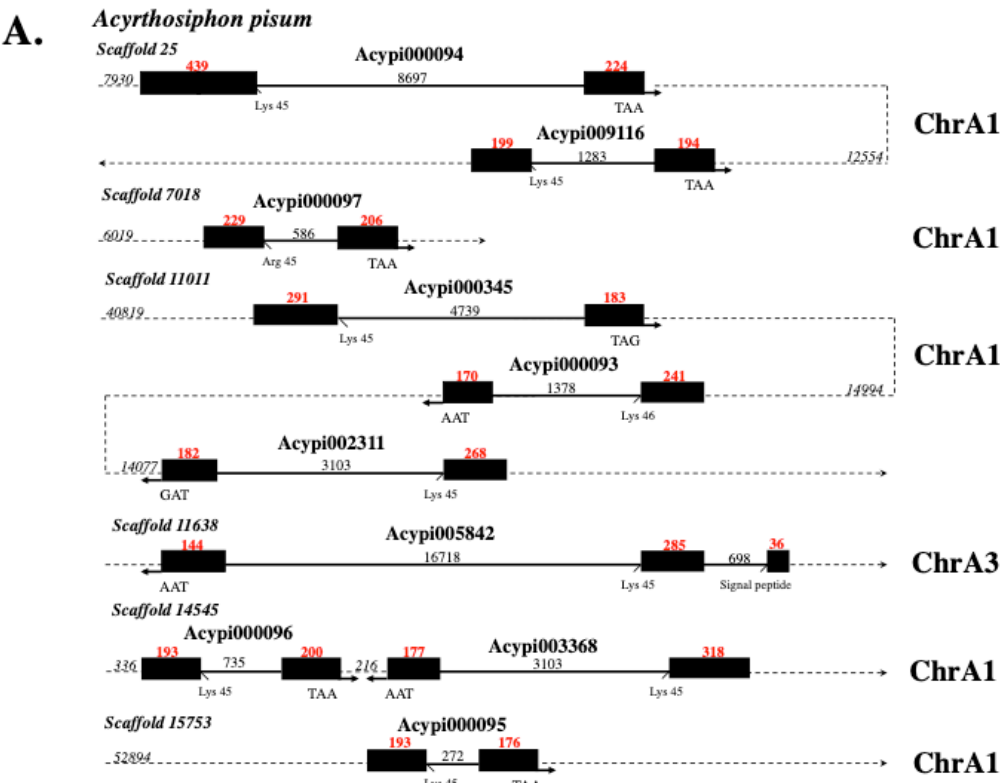
4. The Evolutionary Trajectory of CSPs

A function in chemical communication is rather at odds with the diversity, evolution, quantity, and number of CSP genes. The 4-8 CSPs found in *Apime*, *Anoga*, *Droso*, *Phedu*, and *Nasvi* show that many insect species have very few “CSP”-coding genes [16,39,45]. Does this imply that these insects have poor sense of smell or that they don't communicate chemically with one another? In insects, there is only one ORCO receptor gene, and its function in chemical communication is well established (co-receptor, chaperon molecule) [51]. The situation with CSPs is different, though, as it is thought that they transport scents to chemosensory receptors, which is not possible if there are only 4-8 “chemosensory” proteins.

For chemical communication, pheromones, and mate recognition, flies and many other arthropods are known to use a complex mixture of long chain epicuticular hydrocarbons [52]. Long chain hydrocarbons (LCHs) are used by *Drosophila* and other arthropods for chemical communication, but LCHs are not the only cues. Through gustation or smell, short chain hydrocarbons (SCHs) and other volatile chemicals or compounds (such as *cis*-11-vaccenyl acetate and

cis-4-undecenal) can also serve as chemosensory cues [52–55]. First off, the genome of *D. melanogaster* contains at least 20 “chemosensory protein” genes, but they are not members of the “CSP” family. The Pfam domain family of A10/OS-D (100-110 amino acids, 10-12 kDa, 4 cysteines) and the proteins encoded by the CheA and CheB genes (>200 amino acids, >22.4 kDa, and >4-5 cysteines) are substantially different. The Che genes, such as CheB42 and other CheBs, are abundant enough to be specifically involved in gustatory perception of pheromones, though not all Che genes are involved in pheromone detection [56,57]. Rather than being involved in pheromone detection, the CheA75a gene, which is highly transcribed in fly legs and wings and also predominantly expressed in tarsal sensilla, is able to sense the spore surface protein of fungi, suggesting a function in immune defense [58,59]. Within the ML domain superfamily of lipid-binding proteins, CheAs and CheBs represent a significant gene expansion in *Drosophila* [60]. They are most similar to human GM2-activator protein (GM2-AP), which is a necessary cofactor for α -hexosaminidase A (HexA) in the lysosomal degradation of GM2 gangliosides, rather than a pheromone factor [61]. It’s possible that the ML domain attaches to LCH and/or SCH, whereas the OS-D domain attaches to the volatile pheromone chemical compound; however, this would not make sense given the location of *Drosophila* pherokines dominant presence: hemolymph [27]. Rather, their potential for chemical communication is undermined by the small number of genes coding for this protein family, OS-Ds/CSPs/pherokines (4 cysteines, 10-12 kDa; 6 in bees where the genome does not produce CheA or CheB).

Bees, lice, and wasps only maintain 6-8 CSP genes, making it impossible for them to distinguish a variety of scents or complex chemical cuticular if CSP molecules only have one distinct function in chemosensing [39]. According to Liu et al. (2020), CSPs are distinctly arranged in pairs of duplicates on individual chromosomes in the honeybee. When examining AcyPi-CSPs (AcyPi000094-AcyPi009116, AcyPi000093-AcyPi002311, and AcyPi000096-AcyPi003368), we find a similar distribution of duplication on particular scaffolds (Figure 2A). The genes on the same scaffold are separated by relatively short distances, ranging from 216 to 14994 bps (Figure 2A). Therefore, it is possible that they are co-expressed and work together to fulfill a common purpose. These AcyPi-CSPs have an inverted DNA sequence resemblance (Figure 2A). They have a junction between them and are pointing in different directions. As such, it is expected that they will raise translocation rates [62]. All aspects of eco-evolutionary processes, including mating systems, plant-insect interactions, the transition to herbivory, growth and survival on specific host-plants, social organization, environmental adaptation, adaptation to temperature or climate change, reproductive isolation, and speciation, are largely impacted by inverted duplications (or gene inversions), i.e. the most common type of chromosome rearrangement. Inverted duplication is frequently linked to the emergence of a particular phenotype [63–66]. An understanding of the evolution of the brain requires the existence of genetically controlled duplications and deletions of neurons. The evolution of the primate brain has involved inversion [67]. The results obtained in spiders show how genetic evolution (gene inversions) can drive neuron innovation and the diversification of related sophisticated behaviors [68]. Phenotypes induced by RNA interference strongly suggest a role for OS-D/CSP in *Apime* brain integument development rather than a role in chemosensing [69]. Therefore, it is possible that the inversion observed in *Acypis* played a widespread role in adaptation to environment and host-plant, as well as in particular brain development, given the degree of inversion duplications in the *Acypis*, the various roles of inversions in adaptation, and the importance of CSPs for the brain (see Figure 2 and [69]). In this case, a thorough understanding and analysis of the tissue-distribution of CSPs during insect development are required.



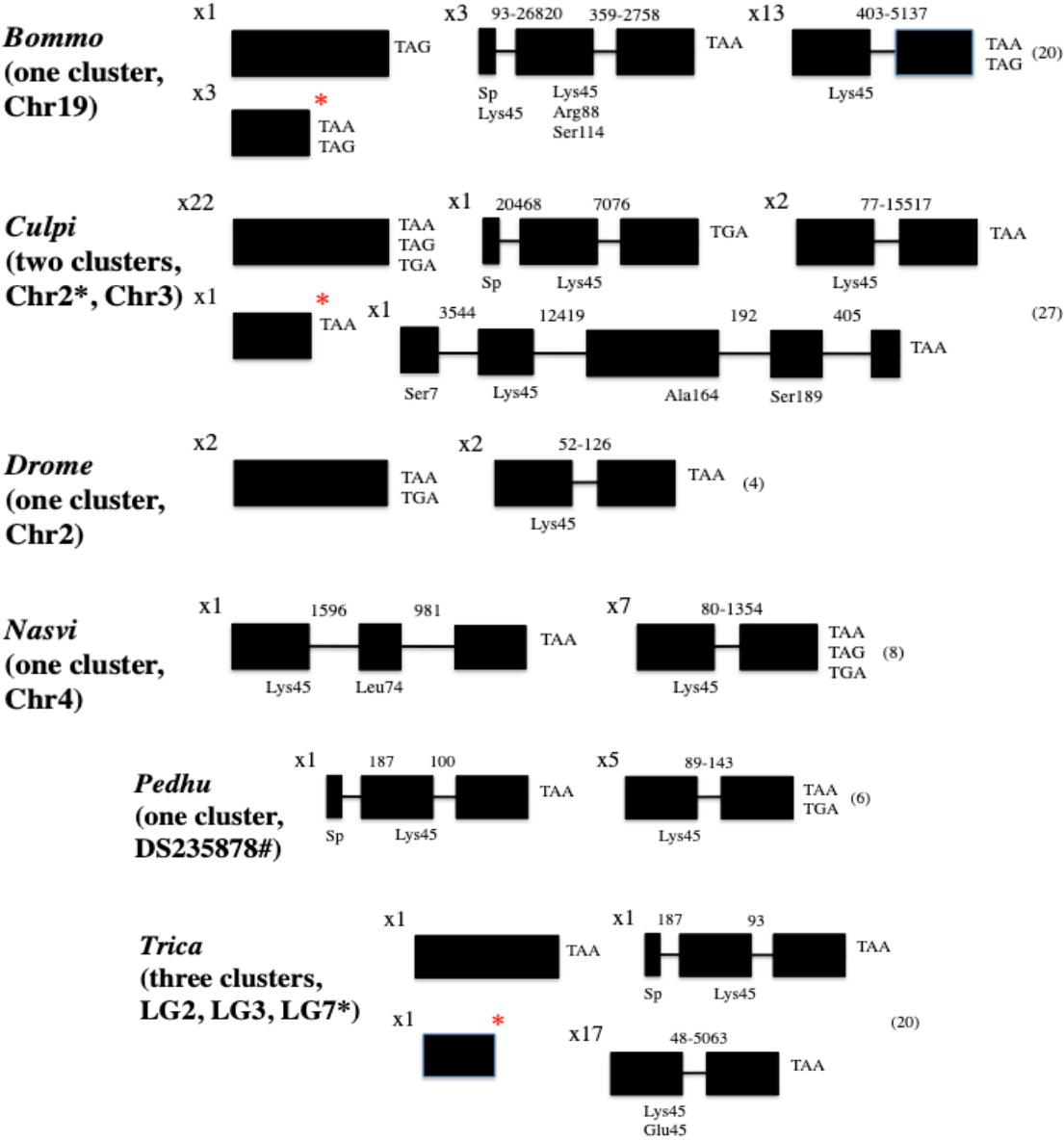


Figure 2. Ancient origin and common illustration of CSP gene structures in insects. A. organization of aphid CSPs on different scaffolds (25, 7018, 11011, 11638, 14545, and 15753) and genomic CSP gene repertoire in Acypi (Table S1). Exons are shown as black boxes and introns as plain lines. The numbers above indicate the size in base pairs of each segment of a gene or intergenic distances (italics). The exon sizes are denoted by the numbers in red. The spacing between genes is indicated by dotted lines. The direction of the arrows, either 5'-3' (right) or 3'-5' (left), denotes the orientation of the gene. Intron insertion sites are the residues at position 46 (after conserved Lysine45 or Arginine45). Acypi-CSP genes have TAA or TAG stop codons, which are indicated by the stop codons in Acypi000094, Acypi009116, Acypi000097, Acypi000093, Acypi005842, Acypi000096, Acypi003368, and Acypi0000095 (TAA), as well as Acypi000345 and Acypi002311 (TAG). The Acypi genomic DNA annotation provides the gene name, along with the gene's location on the genome (genes are plotted onto scaffolds and genome assemblies without regard to function). B. Comparative analysis of the arrangement of the CSP genes in Acypi with Aedae, Anoga, Apime, Bommo, Culpi, Drome, Nasvi, Pedhu, and Trica. Comparison in the context of gene count, gene cluster arrangement, chromosomal dispersion, intron length, intron boundary, gene structure, and stop codon. Sp: Intron insertion in the signal peptide. The black asterisk points to the main cluster. The asterisk in red indicates pseudogene. #Chromosomes have not yet been assembled from Pedhu genome. Assembling chromosomes from updated genomes (Sept. 2024).

The Acypi-CSPs are composed of two exons separated by a single, variable-length intron, just like in most other insect phenotypes (Figure 2B). Acypi-005842 gene like some other CSP genes from Aedae, Anoga, Apime, Bommo, Culpi, Pedhu, and Trica has an additional intron inserted a few nucleotides following the start codon that codes for the amino acid methionine-1 (Figure 2A,B). This intron (phase 0 intron) is inserted after the third base; it does not interfere with the codon. This supports the notion put forth by Blobel (2000) that signal peptide molecule's functional significance is directly correlated with its length and shows how tightly the splicing of the signal peptide region is regulated in CSPs (Figure 2). It is possible that the first six amino acids of CSPs are what punctures through the cellular or subcellular membrane. For correct localization and/or transit within the different cell systems, specific molecular protein partners from the Signal Recognition Particle (SRP, protein-RNA complex) may be dependent on the other amino acid residues [71,72]. This may help to clarify the nature of the "CSP" molecule, which could be a secreted protein (in hemolymph) or an intracellular protein in a multimolecular complex. Experimental data indicates that certain CSPs are stored within the cytoplasm. Similar to OBPs, CSP signals in immunocytochemistry experiments are widely dispersed from the lymph to the interior side of cells [73,74].

The fact that the introns in more than 70 CSPs and all other single-intron CSP genes vary in length but are always found in the same place or boundary (Lysine 45) in aphids, bees, beetles, flies, lice, mosquitoes, moths, and wasps suggests that all these genes share a very ancient common ancestor (Figure 2). The positioning of the intron boundaries in the CSPs from whiteflies emphasizes this [38,75]. But intron boundaries aren't the only place to find evidence of a gene's deep evolutionary conservation. It is far more crucial to conserve the primary amino acid sequence, high consensus residues, and stop codon. To produce the correct polypeptide, splice junctions are essential, but this is also connected to the function of the protein. The intron that was inserted into the signal peptide in Acypi, Aedae, Anoga, Apime, Bommo, Culpi, Pedhu, and Trica is likely of significant note as it highlights the gene's ancient function and suggests that it has a very distant high evolutionary heritage (see Figure 2). The Ordovician, when terrestrial plants first appeared, is thought to be when the class of insects first appeared on Earth 480 million years ago. Overall, the current findings would be compatible with a shared heritage for the CSPs of all insect species, including those belonging to the orders Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, and Psocodea Phthiraptera. There is an extremely long history of the gene and CSP molecule. After that, every group or order eventually started to exhibit some unique patterns, maybe as a result of the appearance of unique phenotypic and functional traits (Table S1, Figures 1-3 & S1-S3).

5. Existence of "CSPs" Outside Insects

It should be discussed here that mutation variations on CSPs may have influenced the evolution of cells in general (i.e. cells from all species, including single organisms, bacteria, and yeasts), as well as the evolution of insects and hexapods. These CSP molecules are not an insect's apnage. They are also expressed in a wide range of organisms, including many species of springtails, shrimps, crabs, lobsters, and copepods [46,76,77]. The fact that they are widely expressed at the level of the bacterial kingdom indicates that they exist also in prokaryotes as well [45,46] (see CP023456, JALMG010000101, JAAIFC010000213, JAAF01000214, JAAIFC0100000830, LYFD1000069, LYGG01000091, RBVL01000175, RBVL01000176, RBVL01000208, RBVL01000221, VUJW01000090, VUOA01000128, WP_149730592 in multi-species). Copepods have been the subject of the initial and preliminary functional studies of CSPs (lipoid binding) in crustaceans [77]. We would not go into further and too excessive detail regarding the phylogeny of all bacterial CSPs here, as functional studies and discussions about possible functions and evolutionary significance of bacterial "CSPs" have not been carried out yet. Here, we show high percent identity between insect CSPs and specific bacterial sequences (Table S1), and the presence of sequences from *Bacillus* and *Klebsellia* in a phylogenetic group containing CSPs of many different insect species (Figures S2 & S3). This paper emphasizes the non-chemosensory role of the CSP family, and the fact that they are present in numerous bacterial families provides even more evidence or proof in favor of this claim. Before assigning a CSP to a barely tuned function in olfaction, it may be necessary to take into account that

some bacterial CSPs are twins or ‘identical twins’ of moth CSPs [45,46], which is significant when discussing the function of these molecules (up to 91% identity; Table S1).

It is reasonable for us to attempt to comprehend the potential role these proteins may have in prokaryotes before examining where they are localized in the insect system, as evidence by our phylogenetic analysis showing that certain sequences from *Bacillus* and *Klebsellia* microbes group with different insect CSP sequences (Figures S2 & S3). “CSPs” have been reported on bacterial species, including Gammaproteobacteria *Lysobacter* (Xanthomonadaceae) and *Escherichia coli* (*E. coli*), *Acinetobacter baumannii* (coccobacillus), *Macrococcus caseolyticus* (formerly *Staphylococcus*), the filamentous actinomycete *Kitasatospora griseola*, the *Actinobacteria* genus in the families Enterobacteriaceae, Nocardoidaceae, Pseudonocardiaceae (*Solhabitans fulvus*), and Streptomycetaceae [45,46]. “CSPs” have also been reported to be present in Firmicutes, Aeromonadales, Alteromonadales, Eubacteriales (*Clostridium perfringens*), and Hyphomicrobiales, based on RNA and genomic reports (MDK0835621, MDK0841570; see [46]). Using Acypi, Bommo, and Pedhu CSP as a template, we performed a blastp analysis in the Microbial Protein Database, which clearly revealed “CSP” sequences (46-59% identity) in the *Sorangiinae* bacterium (MSr11954, myxobacterial strain, soil bacteria, feed on insoluble organic substrate, WXB18440, Locus LZC94_14500; Table S1). These findings from oxidase-positive nonendospore forming, motile, nonmotile, and spore-forming bacteria cast doubt on a chemosensory role for CSPs. Not every bacterium uses quorum sensing. These microbes are recognized as typical digestive tract bacteria, primary prokaryotic secondary metabolites, multi-drug resistant opportunistic pathogens, highly positive cytochrome c oxidase reactions, and multi-species symbionts in insects, and plants, but rather not for their olfactory acuity. *Acinetobacter* species are significant soil and aquatic microbes that aid in the mineralization of molecules like aromatic compounds (benzene). The smallest free-living prokaryotic cell (0.013 μm^3) with very low GC (33%) is found in marine Actinobacteria (*Candidatus actinomarinidae*) [78]. Given that their geographic distribution is similar to that of picocyanobacteria, there appears to be a strong relationship between the *Candidatus* and picocyanobacteria microbial groups. Based on the existing literature, it appears more likely that these two types of microorganisms exchange or share molecular modules and toxin-antitoxin systems rather than pheromones [79,80]. *Clostridium perfringens* (also known as *C. welchii*, or *Bacillus welchii*) is known for α -toxin, the toxin involved in clostridial myonecrosis. However, other bacteria are also known to produce volatile organic compounds (VOCs), such as formaldehyde, methyl mercaptan, isopentanol, and trimethylamine [81]. Because humans (and animals) do not produce these VOCs, they can serve as biological indicators of the presence of specific pathogens. Better yet, some VOCs have been found to be helpful in identifying particular bacteria: isovaleric acid for *Staphylococcus*; methanol, pentanol, ethyl acetate, and indole for *E. coli* [81]. This is not the case for sex pheromones, volatile carboxylic acids (VACs), pyrazines, chemosensory signals, aggregation odors, or cohesion pheromones, which have rather sporadically been linked to associations between microbes and/or insect-plant-microbe interactions [82,83]. It is therefore possible that “CSPs” bind to VOCs, isovaleric acid, indole, their precursors, or any molecules that are required for the biosynthetic machinery of VOCs. It would be interesting to investigate the binding properties and diversity of “CSP” proteins in axenic insects, as they do not have bacterial populations in their gut system. However, it is highly unlikely that fecal bacterial CSP molecules are involved in the transmission of odorant pheromone signals. Many bacteria produce CSPs that are highly similar to those found in *B. mori* (46-91% identity; Table S1), but the Bommo-CSPs are widely distributed throughout the entire body of the insect, not just found in its sense organs. For all CSPs, there is no evidence of specific binding to VACs, VOCs, *E. coli* odors or insect cohesion pheromones. In response to pesticide drug exposure, they undergo a high and striking up-regulation [20].

6. Expression Profiling in Organisms, Development, and Tissues

Chemical senses including olfaction and gustation, can be extended to chemotaxis for glucose and phenol in single-cell organisms. As such, it would not be a surprise that so many bacteria express CSP homologs. However, considering the striking similarities between Bombyx and bacterial CSPs

and the fact that all Bommo-CSPs are expressed outside of the sensory organ paradigm (primarily in the gut and fat body), referring to an entire family of proteins whose primary function is supposed to be chemical detection (i.e., proteins carrying odor molecules from the extracellular medium to olfactory receptor neurons) as “chemosensory proteins” is rather incongruous. Methyl-accepting chemotaxis proteins, or MCPs, are the class of transmembrane receptors that sense chemical gradients in bacteria. These receptors differ in the molecule they are able to detect [84]. Through distinct protein-protein interactions, thousands of MCPs and signaling proteins within a bacterial cell build a highly interconnected network. Bacteria have a few thousand MCP I and MCP II molecules per cell, but very few CSPs [85]. MCPs have the ability to directly or indirectly bind attractants or repellents through interactions with periplasmic space proteins, which can trigger a cascade of proteins that activates Che proteins [86]. The structure of MCPs, which are aspartate receptors arranged as trimers of dimers [86], differs greatly from that of insect ORs [51], and there are no intermediate soluble proteins at the interface environment-MCP. Like in insects, the process of signal transduction in bacteria is primarily dependent on protein-protein interaction, supramolecular complexes, phosphorylation, and the transport of lipids. It is highly probable that CSPs are involved in this process. The cytosol is the location where Che proteins are activated after signals from MCP receptors cross the plasma membrane [87,88]. As part of a signal transduction pathway between the “odor” and the flagellar motor switch, cytoplasmic proteins that include histidine kinases such as Che proteins change the tumbling frequency and the receptors instead of CSPs.

To call these proteins “chemosensory” — that is, proteins that carry smells to olfactory neurons — would also be highly erratic based on their tissue distribution throughout the tissues of insects and hexapods (Figures 3 & S4). As in many crustacean species, “CSPs” in Achelata, are found in phyllosoma, in larval and adult stages, and in numerous non-sensory organs in the adult stage (Figure S4, [46]). Insect pheromone gland and venom can also contain CSPs, in addition to the sensillum [22,89]. In moths, sixteen CSP molecules are expressed in the sex pheromone gland of a species that uses only one sex pheromone compound, bombykol [22]. Apart from the pheromone gland of female moths and wasp venom, other secretions such as saliva also contain CSPs. Antennal branches, mouth, mandibles, proboscis, cephalic capsula, eyes, head, thorax, abdomen, epidermis, fat body, ovary, gut, wings, and legs, are among the tissues that express CSP; these tissues include a variety of reproductive and non-reproductive, sensory and non-sensory fluids, but mostly covering tissues and metabolic tissues [20,39,90–94]. It is rather illogical to assign a single, precise role for a CSP in an organ or tissue, like the antennae, eyes, ovaries, or proboscis, since the majority of “CSPs” are expressed in fluids and enveloping and metabolic tissues.

We checked the possible expressions of CSPs in different tissues across different arthropods through reported Expressed Sequence Tags (ESTs). Some insect genomes include different genes/proteins names as either “putative pheromone-binding proteins”, OBPs, Chemical sense-related lipophilic ligand-binding proteins (CRLPBs) and/or Chemical Binding Proteins (CBPs) (in *Drosophila*: ~51 OBPs, two subfamilies of CRLPBs, and ~20 CBPs [95–98]). The ability to transport hydrophobic or hydrophilic chemical stimuli to particular odor or taste receptors is the shared characteristic of OBPs, CRLPBs, and CBPs. As a result, they are typically located in the olfactory and taste systems, with a concentration in the front legs, wings, and antennae as well as the pheromone-sensing gustatory hairs of the insect tarsi and palpi [95–98]. Here, we’re focusing on a carrier protein with a distinct structure—four cysteines, weighing about 10-12 kDa—that signals the NCBI’s A10/OS-D Pfam domain.

Finding out how OS-D/CSP molecules are expressed in multiple different tissues can provide information on how they function. By definition, the insect EST database, which contains information about the relationships of molecules and tissues of origin, contains more than 30,000 mRNA sequences from *n* tissue libraries. Similar to many other insect species, EST-base data analysis in *Pedhu* and *Acypi* species reveals distribution of CSP-RNAs outside of antennae (Figure 3). Six transcripts from first instar *Pedhu* larvae and engorged adults were used to identify the CSPs in the body louse (Table S1 & [45,99–101]). According to sequencing data from The International Aphid Genomics Consortium (2010), *Acypi*-CSPs are also found in a variety of stages, including the head,

thorax, abdomen, and the antennae of third instar nymphs, and in winged and wingless parthenogenetic females that have either received an ampicillin treatment or have been inoculated with bacteria for removal of pathogenic agents [102,103]. We “dissected” *Apime*, *Aedae*, *Anoga*, *Bommo*, *Culpi*, twelve *Drosophila* species, *Nasvi*, and *Trica* using Flybase and specific databases such as VectorBase or KAIKO/SilkDB [16,20,39,104]. The same methodology was utilized to analyze *Acypi*- and *Pedhu*-CSPs from Flybase and VectorBase (Figure 3). EST-cDNA libraries from >ten different insect species were analyzed, and the results corroborated Northern blot, Western blot, PCR, and enhanced real-time PCR results when multiple tissues were analyzed in the same combined experiment (not just reporting one clone from a single organ or tissue): CSP is extremely widely distributed throughout the insect body from the sex pheromone gland to the prothoracic gland and the brain (Figure 3).

Gustatory neurons can be found in various insect tissues, such as the labellum, pharynx (pharyngeal taste organs), legs, wings, and external reproductive system. Thus, it is likely that chemosensory neurons and/or receptors are present in many of the CSP-expressing tissues depicted on Figure 3. However, the brain, hindgut, fat body, corpora allata, salivary glands, prothoracic glands, and muscles—the primary locations where CSPs are expressed in insects—are improbable to contain chemosensory neurons or receptors. CSP-like proteins are present in both the external (claws) and internal (brain, subesophageal ganglion, abdominal muscle, heart, hepatopancreas, and molting gland) parts of shrimps and lobsters (Figure S4) [46]. The tissues in question are unrelated to chemosensing. The CSP’s family tissue distribution in insects and hexapods is depicted in the Figures 3 and S4 in a very clear way; this is not the tissue expression profile what one would anticipate from a family of “chemosensory” proteins (see Figures 3 & S4).

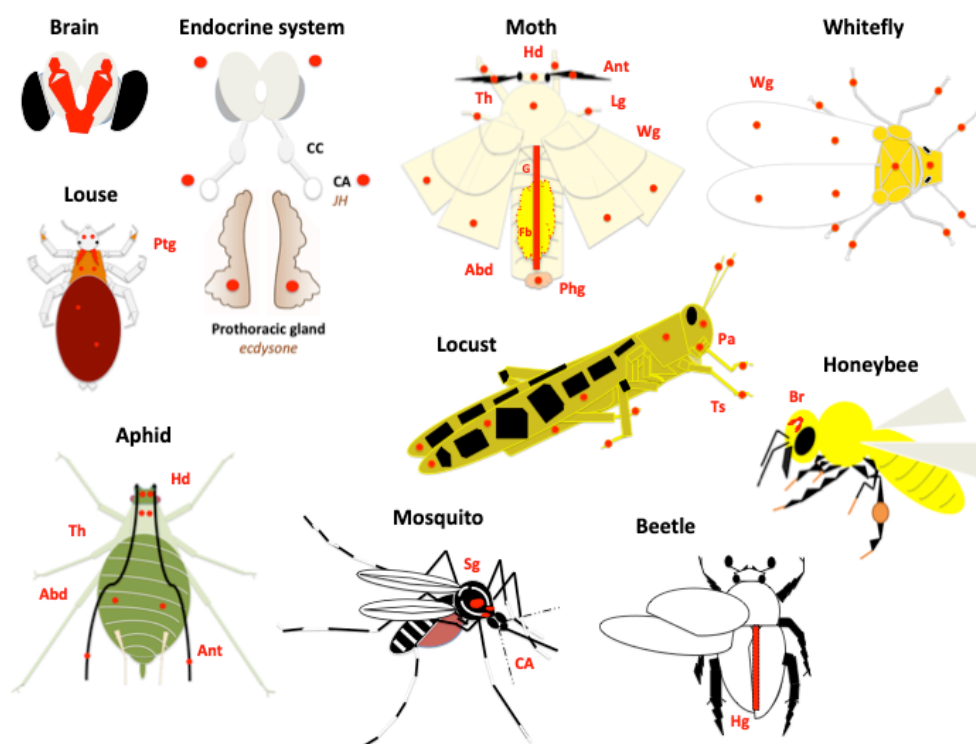


Figure 3. The general tissue expression profile of the protein family known as “chemosensory” (“CSP”) in insects. A. The brain and endocrine system of insects express “CSP”. The location of “CSPs” is indicated by the red dots in aphid, beetle, honeybee, locust, mosquito, moth, whitefly, and locust. The primary structure (brain, gut, fat body, salivary gland, wing, and corpus allatum) where CSPs are found in Coleoptera, Diptera, Hymenoptera, Heteroptera, Homoptera, and Phthiraptera is highlighted in red. Abd: Abdomen, Ant: Antennae, Br: Brain, CA: Corpora allata, CC: Corpora cardiaca, Fb: Fat body, G: Gut, Hd: Head, Hg: Hindgut, Lg: Legs, Pa: Palpi, Phg: Pheromone gland, Ptg: Prothoracic gland, Sg: Salivary gland, Ta: Tarsi, Th: Thorax, Wg: Wings.

7. Putative Function Outside the Chemosensory/Binding Aspect

Due to its widespread distribution, highly conserved sequence and gene structure, binding to lipid chains and LA, and presence in prokaryotes, hexapods, and insects alike, the protein molecule family known as “CSP/OS-D” appears to be extremely common. Its profiling of gene expression spans from glands to nerves. This compellingly suggests a fundamental role in lipid metabolism, transport, uptake, and exchange, possibly restricted to long chain fatty acids, which are necessary for development, flight, pheromonogenesis, reproduction, signal integration, signaling, and immune responses.

In insects, the antennal sensillum was labeled in immunocytochemistry experiments using polyclonal antibodies against the CSP protein; however, the labeling also spread to the cuticle, supporting cells, sensory structures, and the interior of the cell [4,73]. The situation here is strikingly similar to that of “pheromone-binding proteins” (PBPs); the extracellular sensillum lymph of the hair lumen is extensively labeled, but intracellular label is also present in the Golgi apparatus, endoplasmic reticulum (ER), and various dense granules in many different types of cells. Not only can these cells be labeled inside, but also the endocytotic pits and vesicles, multivesicular bodies, and lysosome can be labeled [74].

According to Xuan et al., every Bommo-CSP is expressed outside of the olfactory paradigm, primarily in the gut and fat body, and it participates in the moth response to avermectins [20]. According to Loftus et al., Verjovski-Almeida et al., Noriega et al., and Nene et al., CSPs are also responsive to bacterial infections in *Culpi* and infections by *Plasmodium gallinaceum*, *Brugia malayi*, and Dengue virus infections in *Aedae* [105–108]. When flies or whiteflies are exposed to bacteria, pesticides, or viruses, their response is similar (overexpression of CSPs) [27,28,38]. Ants' expression of CSP (and OBP) changes in response to fungus infection [109]. CSPs appear to be linked to thiametoxam tolerance in the cotton aphid *Aphis gossypii* Glover and the bird cherry-oat aphid *Rhopalosiphum padi*, just like in whiteflies [28,38,110,111]. Compared to ‘normal’ flies, CSP-transgenic flies appeared to have greater resistance to spirotetramat, α -cypermethrin, and thiametoxam [110]. RNA interference (RNAi)-induced suppression of CSP transcription markedly increased *A. gossypii* and *R. padi* sensitivity to thiametoxam [110,111]. These aphids exhibited elevated expression of CSPs in various body tissues, especially the midgut and fat body, and these expressions were linked to resistance against the diamide insecticide cyantraniliprole [112]. Strong CSP gene responses to viruses, toxins, insecticides, or microorganisms are probably signs of resistance rather than chemosensing in insects and hexapods. The likelihood of CSP genes being associated with chemosensing of toxins, insecticides, or even viral or bacterial products is extremely low, even if they are identified by gustatory or olfactory circuits and are chemical compounds or proteins. First off, there is little chance that non-volatile molecules like spirotetramat, α -cypermethrin, and thiametoxam will be detected by ORs. It is yet unknown if gustatory receptors (GRs) are able to detect them. Regardless, this will not account for the presence of CSPs in the hemolymph, gut, or adipose fat body tissue in response to infection or dietary modification [20,27,113]. Second, the pentacyclic lactone avermectin (C₄₈H₇₂O₁₄) has a disaccharide that is made up of methylated deoxysugar I-oleandrose. This isn't really a ligand molecule that CSPs could bind to. Furthermore, in the gut and fat body tissues, avermectin not only activates the CSP clades but also the OBP clades [20,114]. That being said, it is highly probable that the metabolites of the insecticide, or certain lipid molecules and cofactors involved in the pathways leading to the immune response, are what attach CSPs (and OBPs). It has been demonstrated that the CSPs (up-regulated by thiametoxam) interact with linoleic acid (C_{18:2}), which strongly suggests that CSPs play a role in the insecticide response via the lipid pathway, as opposed to directly interacting with the insecticide molecule [38,39]. This needs to be investigated in detail using either a robust specific ligand-binding assay employing highly pure functional protein samples or a very large comparative ligand study, not just the insecticide, in docking.

The relationship between CSP and various lipids and cofactors fits with one more tissue location of this protein gene family, in addition to the circulatory system and metabolic tract. It becomes obvious that the insect uses CSP for other purposes than chemosensing. CSPs (and OBPs) make about

more than half of the proteome extracted from the mandibular glands of moth caterpillars [115]. Although, these proteins are present in adult insects as well as larvae, they are primarily concentrated in the adult's salivary glands, prothoracic glands, ovaries, fat bodies, and corpora allata (CA, endocrine sources of JH) [16,20,94,115–118]. In parthenogenetically reproduced whitefly females, fecundity was significantly lower when CSP was knocked out via dietary RNAi, suggesting a function for CSP in reproduction [118]. CSP can be expressed in whitefly and weevil eggs [38,119]. This is not a result of herbivory. These proteins are present in adult *Aedes* as well as larvae, but they are also concentrated in the salivary glands, fat bodies, ovaries, and CA of the adult mosquito that causes dengue fever [16]. This implies that, “CSPs” have a role, direct or indirect, in the binding, transport, breakdown, and/or biosynthesis of juvenoid hormones, including the different forms of JH.

The function of “CSP” molecules in relation to the biosynthetic and/or degradative pathways of lipids and JHs (acyclic sesquiterpenoids with epoxyde moieties) is strongly suggested by the widespread tissue and developmental expression of CSPs, as well as the fact that CSP is present in other arthropod species, such as copepods, shrimps, and water fleas in addition to insects. Kim et al. suggest that this function may be comparable to that of the OBP in the hemolymphatic transport of the JH molecule [120]. It is well known that JH regulates the developmental processes in insects and crustaceans [121]. On ectodermal epithelial endocrine glands, like thoracic glands and CA, there are no sensory appendages, no OR, and no GR, despite their genesis from the first and second maxillary pouches. JH function controls almost every aspect of physiology in both larvae and adults of insects and crustaceans. JH regulates caste differentiation, pheromone production, development, male genital morphogenesis, female ovarian egg production, social and mating behaviors, and immune function in the entire arthropod kingdom [122]. Specifically, it has been shown that JH controls neuronal plasticity linked to brain structure, neurogenesis, and behavioral maturation; this is comparable to the pattern of CSP expression observed in insects and hexapods [123–125] (see Figures 3 & S4). Maleszka et al. conducted the initial knockout experiment for the CSP gene [69]. CSP gene clearly regulates how the insect's brain develops. Furthermore, according to recent EST and transcriptomic studies [16,116], not only does the CA—a source for JH—express CSP, but also the prothoracic gland—a source for ecdysone—express CSP, which is strongly inconsistent with an olfactory function.

Perhaps an important observation in this analysis is that bacteria and prokaryotes may share the JH relationship and CSPs. Phurealipids are secondary metabolites that *Photorhabdus* bacteria produce. Phurealipids and juvenoids have a similar molecular structure. JH/Phurealipids inhibit development and immunity in insects [126]. It is therefore possible that CSPs have a function related to JH in insects (hormonal control, immune response, pheromone regulation, growth, and development) as well as a function related to phurealipids in bacteria (*Photorhabdus* bacilli, for example). This would be in line with the many functions of CSPs, which include the presence of these proteins in bacilli and *Sorangium* species of myxobacteriales, and their roles in insecticide resistance, intracellular transport of endogenous fuel lipids, FAs, hormones, and secondary metabolites. Unless it can be demonstrated that CSP molecules are a part of the bacillus microbe olfactory hedonics, this would be very accurate statement.

Sifting out a subset of the CSP family that is unique to antennae or olfaction is challenging because most of the molecules are produced in the gut (hindgut) and fat body—which are believed to be the arthropod body's main organs for storing lipids and FAs as energy—. These fats and FAs are released into the hemolymph through lipolysis, allowing other organs and tissues to use them as fuel for growth, development, regeneration, and defense against infectious viruses and pesticide chemicals.

The existence of CSPs in bacteria and their crucial roles in the development of insect heads, crustacean molt, hexapod general immunity, JH, pheromone synthesis, and/or behavioral changes in specific phenotypes are consistent with a role for CSP molecules in lipid transport.

Ozaki et al. state that “CSPs” mediate the identification of chemical signatures composed of cuticular lipid FA hydrocarbons, like those found in ants; among other organisms. It is currently

unknown how each of these “CSPs” connected to FAs functions in chemical communication, growth, development, and/or other physiological processes. According to whitefly results, CSPs fulfill a metabolic role in the breakdown of insecticides in connection with C18 FA [38]. Overall, rather than just supplying cell energy, the synthesis, breakdown, absorption, and transmission of FAs like C18-LA take part in all stages of innate immune cell functions (such as activation, killing, and differentiation) through a variety of substrates from their own or other sources through metabolic pathways (e.g. FA pathway, phosphatidylcholine pathways, phenoloxidase cascade, phosphorylation cascades, hormones, proteinase inhibitors, antimicrobial peptides, degradative enzymes, mitochondrial cytochrome P450 systems, etc.). C18-lipoids are well known to alter the fluidity and composition of membranes, as well as to act through particular enzymes and receptors, to affect the effector and regulatory functions of innate responses at the tissue level [39]. Essential to the membrane of cells and organelles, internal FAs such as LA are stored in the fat body, which is comparable to the liver of vertebrates, and function as sources of energy, precursors of different waxes, secondary metabolites, and defensive secretions. The production of bactericidal peptides, including lysozyme, by the fat body (and the gut) is essential to innate humoral immunity. Hemocytes must migrate to the site of chemical, microbial, or viral infection in the meantime, using the energy from lipid metabolism to carry out phagocytosis, nodulation, and encapsulation. Therefore, a lipid molecule such as LA does not remain in the Golgi but swiftly targets the cytosol in response to infection [39,128,129]. This introduces CSP into the hindgut, fat body, hemocytes and hemolymph. This introduces CSP into insect immunity and defense within the cell.

It is possible that the presence of CSP within the cells contradicts the whole idea that they have a secretory signal. This is especially true because insect CSP sequences frequently have short signal peptides [see 3-18]. If a protein contains a signal peptide, it is fed into the endoplasmic reticulum (ER) during translation; however, it is not always the case that the protein molecule is released into the extracellular medium, lymph, or hemolymph. When the nascent protein reaches the ER, the signal peptide can be cleaved at a specific time point. Proteins destined for the hemolymph or for endomembrane system organelles like the ER, Golgi apparatus, and lysosome all need to enter the ER at this point [130]. It's possible that under some specific circumstances (bacterial infection), the signal peptide is cut off during translation, releasing the completed “CSP” protein into the cytosol where it is retained and targets distinct molecular modules.

Similar to the majority of CSPs, Bemta-CSP1 linked to LA is expressed in tissues other than the olfactory organs of insects, like the head, abdomen, thorax, and wings [38]. In insect fingerprints, LA has never been detected (in a mean of sex pheromone and/or specific cuticular signature). It is necessary to demonstrate how these long-chain FAs function as contact pheromones before claiming that LA has a semiochemical role because these compounds are hardly volatile. e12-desaturases catalyse the biosynthesis of LA by introducing a second double bond to oleic acid (C18:1); however, this enzyme or FA conversion is found in only a few species. Neither the majority of insect species nor any crustaceans can produce LA on their own [131]. No species will attempt to produce its own extremely peculiar chemical smell in dependence on a compound it cannot synthesize, such as LA. Food is the only way to absorb this compound, and food and LA are not absorbed until the gastrointestinal tract. While LA may be utilized in a number of insect species as a pheromone precursor [132], these precursors are never in contact with ORs. Last but not least, the finding that LA can be a “necromone” points to a possible area of study for chemosensory receptors and LA-binding CSPs involved in dead corpse recognition signals, such as those in ants, bees, termites, wasps, and social aphids [133]. However, only a small number of insect species—mostly those in social insects—display this behavior of cleaning the nest of the dead or track down dead prey to eat [134]. It's likely not sufficient evidence that “CSPs” play no part in semiochemical communication given that Bemta-CSP1 binds to LA and distributes it widely. However, as for phytophagous insects, like whiteflies, this strongly suggests that LA lipid transport occurs to fuel multiple physiological systems, like immunity, growth, and flight, rather than identifying a necromone (C18:2 and/or C18:1) to dispose of the bodies. With a protein that is so extensively dispersed throughout the bodies of insects and hexapods, a general regulatory role in insect physiology makes much more sense.

Two points are probably essential to understand the role of CSP molecules: 1) CSPs appear to be involved in a variety of biological roles and are expressed in both bacteria and arthropods; 2) CSP molecules in particular can activate specific innate immune pathways when injected into the phloem of plants [44,136].

8. The Myzus Persicae Mp10 Story Along with Additional Immune-Related Roles

It is not a sensory lymph, phloem. Phloem is the vascular tissue that carries and distributes water, carbohydrates, and other soluble organic substances to the various sections of the plants. The phloem transports nutrients and food that the leaves produce through photosynthesis (photosynthates) to all other parts of the plant. It functions structurally in the body of the plant and serves as a pathway for numerous hormones and other signaling chemicals. The three cell types that typically comprise this tissue are sieve elements, parenchyma, and sclerenchyma. It is through these tissues that CSP molecules, like Myzpe Mp10, can trigger an immune response instead of activating OR [44,136]. Hence, our goal is to go into more detail about the existence of "CSP"-like proteins in the salivary secretions of the aphid species *Myzus persicae*, and that serve as host plant effectors for the aphids' benefit.

First, we add *A. pisum*, *B. tabaci*, *Halyomorpha halys* (Brown Marmorated Stink Bug), and *Pachypsylla venusta* (Hackberry Petiole Gall Psyllid) to the analysis of this Mp10 CSP-like family. For these species, Genome Assembly and Transcript Annotation are available (see Table S2). In this case, we contrast Mp10 with the entire repertoire of CSPs from three distinct hemipteran species, also referred to as piercing-sucking insects or phloem-suckers that consume prokaryotes that inhabit phloem and feed on sugar. Next, we performed a blastp search in the "All Species's Protein" Database from InsectBase (insect-genome.com) using Mp10 amino acid sequence as a template. Several "Mp10" hits (length 101-145, identity 47-97%, e-value 1.80e-34-6.35e-100, score 126-287) were extracted from *A. pisum*, *Nilaparvata lugens*, the brown planthopper, and *Mayetiola destructor*, the Hessian fly. Additional Mp10-hits with varied length, identity, e-value, and score values were extracted by selectively blasting Mp10 in "All Coleoptera Protein", "All Diptera Protein", "All Hemiptera Protein", "All Hymenoptera Protein", and "All Lepidoptera Protein", in species such as *Acypi*, *Anoga*, *Bombus terrestris*, *Camponotus floridanus*, *Chilo suppressalis*, *Culpi*, *Danaus plexippus*, *Dendroctonus ponderosae*, *Diaphorina citri*, *Harpegnathos saltator*, *Heliconius melpomene*, *Linepithema humile*, *Manduca sexta*, and *Trica*. There were no hits produced by blasting Mp10 in the transcriptomes of All Odonata, All Thysanoptera, and All Phasmatodea protein. A tblastx search against all other numerous insect transcriptomes and RNAs (*Archaeopsylla*, *Blattella*, *Cataglyphis*, *Chrysopa*, *Ephemera*, *Forticula*, *Limnephilus*, *Locusta*, *Nemurella*, *Mengenilla*, *Pediculus*, *Sialis*, and *Zootermopsis*) produced the same result (no hit).

Using this set of genes, we performed a phylogenetic analysis (PAUP*10Altevec) to extend the analysis to the Order (Hemiptera) and to obtain a more comprehensive understanding of the evolutionary appearance of this Mp10 CSP-like family by contrasting these with a variety of other insect species and other aphids (Genome Assembly + RNA annotation; Table S2 & Figure 4). A particular grouping of Mp10 with "CSPs" from *Trica* (AAJJ0269C and AAJJ0269B), *D. ponderosae* (DPO006842), *B. terrestris* (XP_012166268), *H. melpomene* (HMEL010990), *C. suppressalis* (CSUOGS107535), *Bemta* (Bta06193), *P. venusta* (PVENScaf20457), *Acypi* (ACYPI000097), and *N. lugens* (BPHOGS10008228, and BPHOG10002786) was identified by the evolutionary tree built with UPGMA (Figure 4A). High consensus in the protein sequence alignment, primarily in the N-terminus and central 88-PDAL-91 motif, as well as conserved elements like Q, W, L, K-D, and the four cysteines typical of CSPs, strongly support the Mp10 grouping (Figure S5). The bootstrap-Jackknife tree, which accurately infers the relationships between sequences based on bootstrap values, supported this particular grouping (see Figure 4B). Seventy-nine percent bootstrap supported the grouping of Mp10 with ACYPI000097, AAJJ0269C and AAJJ0269B, DPO006842, XP_012166268, HMEL010990, CSUOGS107535, Bta06193, and PVENScaf20457, but the two *N. lugens* sequences were dropping out. This demonstrates that Mp10-like protein is present in other insect species and is not exclusive to Hemipteran species (Figure 4B).

It is crucial to remember that Mp10 is less closely related to the brown plant hopper and more closely related to “CSPs” from the flour beetle, mountain pine beetle, bumblebee, postman butterfly, and striped rice stem borer. It is obvious that the Mp10 protein family is connected to insects that feed on plant phloem or that need to burrow or tunnel inside plants, as evidenced by its presence in aphids, psyllids, whiteflies, and stem borers. It most likely indicates that it is a member of the insect pollinator family when it is found in bumblebees and butterflies (see Figures 4A,B & S5). The competing selection pressures of attracting pollinators and avoiding herbivores must be balanced by flowering plants. But eventually, if a pollinator visits the same flower repeatedly, causing damage to the petals or depleting the supply of nectar, it can become “invasive”. Fruit or seed set may be negatively impacted by herbivory and/or frequent visits if pollinators injure flowers while visiting them or if an abundance of pollen deposition causes developing pollen tubes to stagnate [137,138]. Pollinators’ brutal, harsh, and ruthless visits to the flowers often cause the fruits’ form to deteriorate. When plant defenses begin to affect pollinator health, foraging behavior, and reproductive success, the effects of florivory and herbivory on pollinators may not end with the first flower visit. Mp10 might be a crucial component for the pollinator to manage the flower in situations where there are frequent visits, changes in the floral outline, corolla outline, intra-floral color pattern, floral scent, and nectar volume, in addition to helping the herbivorous insect get past plant defenses.

Tenebrionidae, or flour beetles, live beneath bark, in old logs, and on most preserved food. They feed on wheat, maize, peas, beans, and cracked or broken grains, seeds, or nuts. The phloem of pine trees, especially lodgepole pine, is mined by pine beetles (Curculionidae), which burrow through the bark. Although they can consume most pine species (Western white, whitebark, and ponderosa pine), they prefer to attack lodgepole pines. Therefore, a variety of insect species appear to use Mp10 to undermine the protective layers of phloem and sapwood on plant leaves and trees. The number of Mp10s (one in aphids, one in bumblebees, one in Ponderosae, two in Tribolium) may vary according to the preferences for feeding and reproduction. More Mp10s are needed because the host plant may be more diverse. *N. lugens* feeds on a single, extremely specific diet host plant—rice—which may account for the evolutionary divergence observed between the Mp10s of this species and the rest of the Mp10 group (Figures 4 & S5). Regardless, the fact that Mp10 is present in so many diverse insect species suggests that it is a very old strategy of plant control employed by the insect. Mp10 protein injection into the phloem may have its origin in the development of angiosperms and pollinivory, the earliest known indication of plant-insect interaction, which dates back to the early Permian period (~298 Mya). That has nothing to do with smell; rather, it has to do with an insect factor that targets the plant’s immune system.

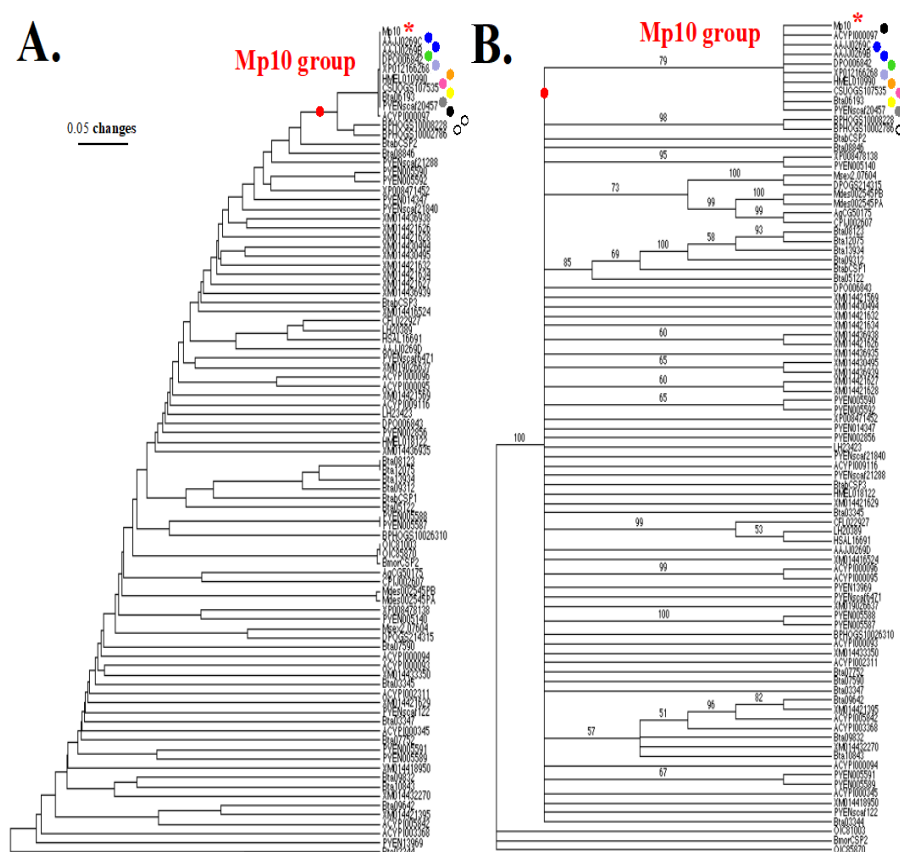
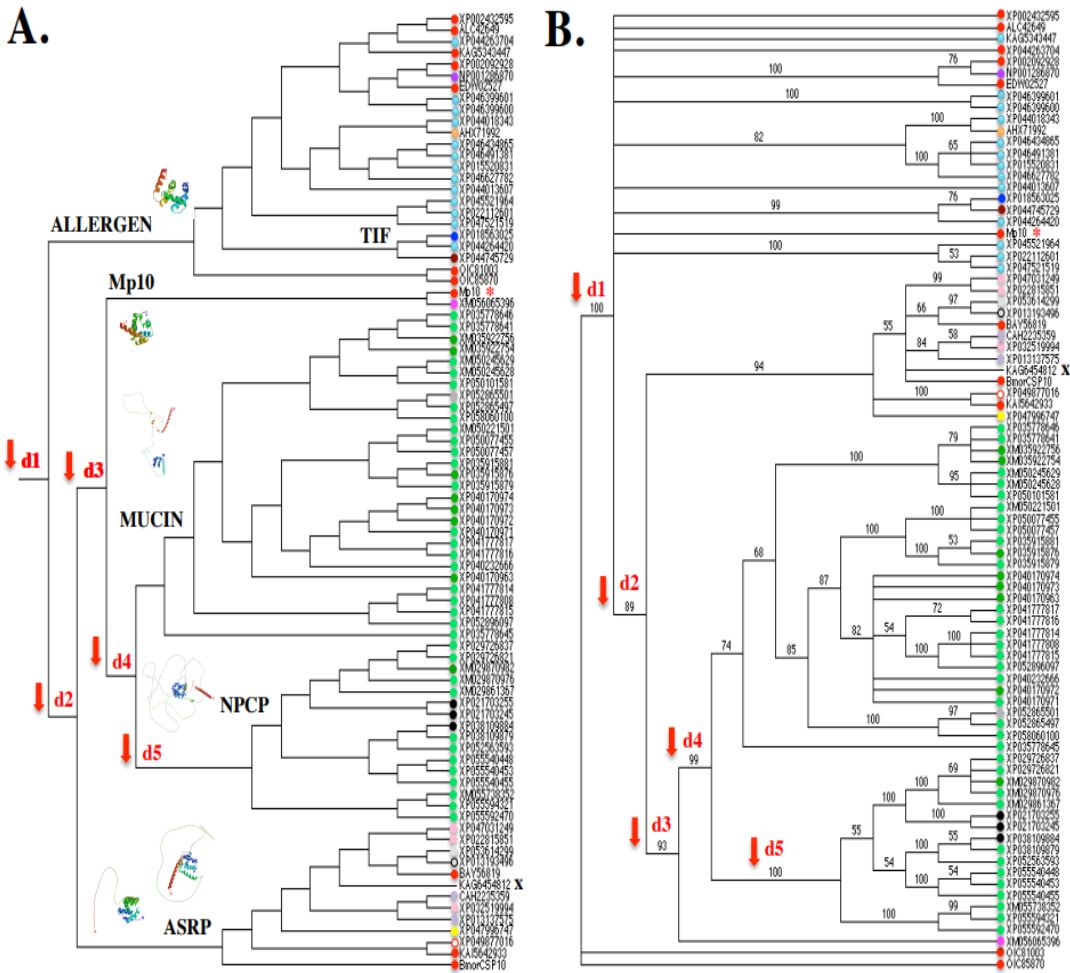


Figure 4. Molecular phylogenetic comparison (PAUP*10Altevec) of Mp10 with related proteins from various insect transcriptomes and genomes from species in the Order Hemiptera. A. UPGMA analysis of Mp10-like proteins in InsectBase and Genomic Database (Table SX): agglomerative (bottom-up) hierarchical clustering UPGMA analysis based on the distance matrix of the analysed taxa that were calculated from a multiple alignment in ClustalW. Evolution (i.e., mutation rate) across all sequences is constant. Each pair-wise distance contributes equally. The red asterisk indicates Mp10's position (*): Mp10 grouping with AAJJ0269C, AAJJ0269B, DPO006842, XP_012166268, HMEL010990, CSUOGS107535, Bta06193, PVENscaf20457, ACYPI000097, BPHOGS10008228, and BPHOG10002786 at the tree's top. B. Bootstrap/Jackknife algorithm analysis of of Mp10-like proteins in InsectBase and Genomic Database (Table SX) with bacterial *A. baumannii* CSPs (OIC81003 and OIC85870) as outgroup. Amino acid tree (data matrix: total characters 457, constant characters 48, variable parsimony-uninformative characters 144, parsimony-informative characters 265, all characters of type unord, all characters have equal weight): Length 4832, CI 0.423, RI 0.318, RC 0.134, HI 0.577, G-fit – 147.563). In red: Mp10, blue: AAJJ0269C and AAJJ0269B, green: DPO006842, purple: XP_012166268, orange: HMEL010990, pink: CSUOGS107535, yellow: Bta06193, grey: PVENscaf20457, black: ACYPI1000097, white: BPHOGS10008228, and BPHOG10002786. AAJJ: *Tribolium castaneum*, Bta: *Bemisia tabaci*, DPO: *Dendroctonus ponderosae*, XP_012166268: *Bombus terrestris*, HMEL: *Heliconius melpomene*, CSUOGS: *Chilo suppressalis*, PVEN: *Pachypsylla venusta*, ACYPI: *Acyrtosiphon pisum*, BPHOGS: *Nilaparvata lugens*. For comparative molecular analysis in the Mp10 family, the protein amino acid sequences are used. InsectBase: All Coleoptera Protein (length 96-132, identity 38-50%, e-value 5.33e-31-8.50e-34, score 108-118), All Diptera Protein (length 99-101, identity 47-49%, e-value 2.34e-34-4.82e-35, score 124-127), All Hemiptera Protein (length 94-145, identity 44-97%, e-value 3.54e-31-4.45e-101, score 109-287), All Hymenoptera Protein (length 97-121, identity 44-49%, e-value 5.19e-31-3.38e-32, score 111-113), All Lepidoptera Protein (length 109-134, identity 39-43%, e-value 2.08e-26-4.93e-34, score 98-121).

Analyzing NCBI database, it becomes significant to note that in this case, Mp10 shares a great deal of similarities with other proteins, including neural Wiskott-Aldrich Syndrome (WAS/WASL)-like proteins, actin skeleton regulatory protein (ASRP), Arp2/3 complex (Actin Related Protein 2/3 complex) activator (actin filament binding proteins and WASPs), peplos-cell cross-linker, stress response initiator, cell wall/envelope protein (CWP), Mucin, Extensin, PAN domains (these versatile domains mediate interactions between proteins and carbohydrates, fulfilling a variety of biological roles), splicing regulator, Rho GTPase (Rho) activator, Serine/Threonine-protein kinase from Social amoeba (SamkC), transcription initiation factor (TIF), nuclear pore complex protein (NPCP), Sec31, UL36 tegument protein, and many others (30.71-45.13% identity; Figures 5 & S6-S7 and Table S3). These findings raise very interesting questions about the possible roles of “CSPs” in the cytoplasmic membrane and cell surface, as well as, in the cell cytoplasm and nucleus, RNA/DNA binding, RNA/DNA regulation, interactions with multiple molecular and genetic components, transcription control, splicing regulation, and activation of specific protein complexes. That being said, it is best to regard the discussion between Mp10, Rho-activator, and other closely related proteins as an intriguing observation. Experiments have not yet shown any evidence for possible shared functional roles. A closer look at the Mp10 sequence comparison in NCBI reveals additional proteins that cluster with “CSPs” and have functional roles in a variety of organisms. It is a startling finding that is still challenging to understand. The interpretation of these relationships—which are so novel but challenging to comprehend—between CSP, Rho, and TIF should require more investigation.



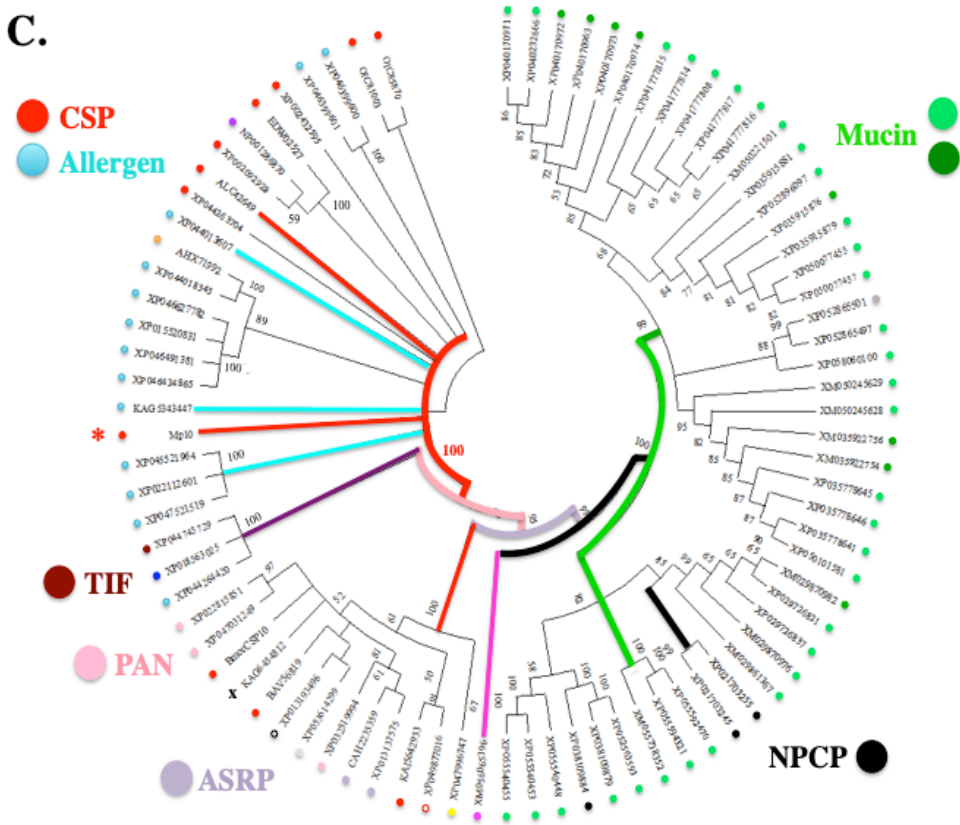


Figure 5. Molecular phylogenetic comparison (PAUP*10Altivec) of Mp10 with related proteins from the CSP, Allergen, Mucin, Rho, TIF, ASRP, and NPCP families. A. UPGMA analysis of Mp10 and counterparts (Table S3): agglomerative (bottom-up) hierarchical clustering UPGMA analysis based on the distance matrix of the analysed taxa that were calculated from a multiple alignment in ClustalW. Evolution (i.e., mutation rate) across all sequences is constant. Each pair-wise distance contributes equally. The red arrows represent gene duplication events (d1-d5) that led to Allergens ("IgE-binding protein") and the group of Mp10, Mucin, TIF, DAN4, NPCP, and Bommo-CSP10 (BmorCSP10). The red asterisk indicates Mp10's position (*): grouping with pherokine XM056065396 and Mucins. On top of the branching tree are the model protein structures that correspond to the molecular groupings. B. Bootstrap/Jackknife algorithm analysis of Mp10 and related Allergen, Mucin, TIF, ASRP, and NPCP proteins (Table S3) with bacterial *A. baumannii* CSPs (OIC81003 and OIC85870) as outgroup. The tree only displays bootstrap support values >50%. Amino acid tree (data matrix: total characters 953, constant characters 157, variable parsimony-uninformative characters 168, parsimony-informative characters 628, all characters of type unord, all characters have equal weight): Length 4366, CI 0.721, RI 0.847, RC 0.611, HI 0.279, G-fit -453.305). C. Neighbor Joining. The circular tree and the trees depicted above both read in the same order. The degree of relatedness among sequences from CSPs (red), Allergens (cyan), TIF (brown), PAN (salmon), ASRP (purple), NPCP (black), and Mucin (green) is indicated by the relative depth of the nodes supported by bootstrap values >50%. CSP (red) branches mix with Allergens, TIF, PAN, ASRP, Rho (in pink), NPCP, and Mucin branches. In red: Ebsp-3/PebIII A10/OS-D, cyan: Allergen Tha p1, purple: Pherokine-3, orange: Acid trehalase, dark blue: CWA-3, brown: TIF, pink: Rho GTPase-activator, light green: Mucin-like/Extensin-like, grey: WAS/WASL, black: DAN4/NPCP, salmon: PAN-1, light grey: Formin-1, light grey in dark circle: WASP-2, light purple: Jg5928, yellow: RickA-like, white in red circle: YLP motif protein 1, X: Hypothetical protein (unknown function). For comparative molecular analysis, the amino acid sequences are used.

9. CSPs' Intracellular Functions (Gene Regulation to Stress Response)

The fact that the Myzpe CSP Mp10, mucin, TIF, and many other “intracellular” regulatory elements are related cast doubt on the claim that CSPs are “chemosensory” molecules (see Figures 4-5 & Table S3). Mp10-like appears to even be expressed by viruses when compared to some peplos proteins and outer surface membrane proteins (see Table S3).

9.1. Evolutionary Evidence Derived from Amino Acid Sequence Phylogenetic Analysis

A preliminary phylogenetic analysis in IQ-Tree found some connections between CSPs, CWP, Rho, SamkC, Sec31, SamkC, TIF, SamkC, WAS/WASL, DNA-binding proteins (DBPs), DNA-regulatory proteins (DRPs), and several RNA-binding proteins (RBPs) sequences molecular sequences (Tables S4-S5; [45]). Transcriptional/cell division repressor, helix-turn-helix, and DRP from the Xenobiotic Response Element (XRE) family of transcriptional regulators, branched with Bommo-CSP16, Bommo-CSP8, and Bommo-CSP9 and their bacterial counterparts, whereas another orthology group included Trica-CSP AAJJ0012J and Ruminococcus DRP WP_044998036 (nucleotide binding, see [45]). This suggested that some CSPs have evolved to carry out tasks that may be associated with nucleotide binding, transcription, translation, DNA/RNA templates, DNA/RNA control, and/or intracellular gene expression mechanisms, at least in Bommo and Trica. This initial investigation encouraged us to compare deeper CSPs with intracellular components, such as transcription factors and gene expression regulators.

A phylogenetic examination of Bommo-CSPs, Trica-AAJJ0012J, BemtaCSP1, Myzpe-Mp10, its derivative sequences, Mucin-like sequences, XREs, and several RBP's sequences in PAUP (*10Altevec) provided further evidence for this (see Tables S3-S5 and Figures 5 & S6-S7). A neighbour-joining tree (BioNJ study) showed that Mp10 did not join the Bommo-CSPs, but instead seemed to be much more closely linked to the Mucins group (G1; Figure S6A), albeit not clearly forming an orthologous group (Figure S6B). In G1, the “CSPs” from *E. balteatus*, EbalDDB_G0285119X1 and CSP3 (QIS77910), linked with Bommo-CSP10 and showed high similarity to Rho-activator isoform X3 (Table S4; Figure S6A). The Trica-CSP AAJJ0012J molecule clustered with Allergen Thap1, and associated RBPs instead of BommoCSPs. The marmalade hoverfly Epiba-CSP4 (QIS77191) clearly deviated from the CSPs groups (see G2, Figure S6A). Only the G1 “EbalCSP”-group maintained a relatively high bootstrap value (94%) in the Jackknife analysis (bootstrapping calculation over 1000 repeats), strongly demonstrating the branching of “CSPs” with Rho GTPase enzyme. A common branch with a significant bootstrap value (57%) was formed by the sequences combining CSP, CWP, LRR, Mucin, SamkC, TIF, and WAS/WASL. Mp10 slipped off of this group, but AglaCWXP3 drew it in (Figure S6B). The papilionid IpodCSP (CAH2042437), DBPs, RBPs, and TIFs were drawn at the bottom of the tree in this analysis using Bommo-CSP2 and bacterial counterparts as the reference outgroup (Figure S6B). This strongly implied that there were significant and expected relationships between “CSPs” and all of these various intracellular proteins, mucins, Rho-activators, and translational regulatory factors.

Focusing on Mp10 (referred to as a very typical “CSP”, 153 amino acids, 17.2 kDa, “consensus residues”, four cysteines pattern, whole body expression), we found that this specific protein sequence (XP_022173691) constructs a relevant hierarchical clustering UPGMA analysis and a phylogenetic tree supported by a high bootstrap value when compared to Allergen, Mucin, TIF, NPCP, and ASRP amino acid sequences (Figure 5). These sequences were not selected at random. The main components of these include mucins, allergens, and phylogenetically related protein CSP, which are widely present in various insect groups (see Table S3). The blastp analysis of the Mp10 consensus sequence has extracted them all (see Table S3). The whole “Mp10” group comprises “chemosensory proteins” in addition to mucins, which have emerged as allergens that are highly specific for edible insects, much like OBPs and hexamerins [139]. However, it's interesting to note that Mp10 does not cluster with other “CSPs” on the UPGMA tree. UPGMA analysis refers to “aka” cladograms or, in very general terms, evolutionary trees. Using a distance-based approach, the UPGMA is constructed by joining the most similar pair of taxa, in this case, Mp10 and mucin proteins (Figure 5A). Here, the ASRP group—which is interpreted as a dichotomy—is the closest neighbor of

the Mp10/Mucin cluster. On the UPGMA, we find that the CSP/Allergen group is more distantly related, as neighbors construct the external nodes with the fewest mismatches. Therefore, our suggestion that CSP has no or minimal role in chemosensing is not so much based on the use of Mp10 as it is on the use of particular Mp10 counterparts. Specifically, TIF could be used to assert or imply that CSPs are not involved in smell—that is, their involvement in carrying the ligand for OR and GR—. The fact that CSP amino acid sequences are present in every group in UPGMA and Bootstrap further supports our contention that CSPs are not involved in smell (Figure 5).

Mp10 belongs to a large group that includes all Mucin taxa and TIF. It shares a close relationship with Rho-activator (XM_056065396; Figure 5A). The Rho families of small GTP-binding proteins (20-30 kDa) are not extracellular molecules. These are intracellular proteins that control the actin cytoskeleton-related Rho-GTPase signaling pathways. They act as molecular switches that control a variety of cellular processes, including gene transcription and cytoskeleton-related events. RhoGAPs, one of the main classes of Rho-regulators, are found in all eukaryotes (within the cell) and have been shown to regulate a variety of cellular functions, such as the organization of the cytoskeleton, growth, differentiation, neuronal development, and synaptic functions [140]. They have nothing to do with smell; they don't trigger transmembrane chemosensory receptors.

As for the other “CSPs”, they belong to either the ASRP group (see the position of BAY56819, KAI5642933, and Bommo-CSP10, in the UPGMA analysis on Figure 5A) or the Allergen group (which looked to be the most ancestral molecule): XP_002432595, ALC42649, KAG5343447, XP_002092928, EDW02527, OIC81003, and OIC85870 (bacterial proteins). Significant outer envelope proteins, RickA-like (Arp2/3), and particular nuclear nucleoside kinase (NNK) are found in the ASRP group, while IgE-BPs, pherokines (fly hemolymph CSP proteins), acid trehalase (AHX71992, involved in intracellular trehalose mobilization during postdiauxic growth and severe saline stress in yeasts), Cell Wall-Anchored (CWA-3, XP_018563025), and TIF sequence (XP_044745729) are found in the Allergen group (Figure 5A). RickA is the protein from rickettsiae bacteria, which are carried by ticks and lice. It is involved in bacterial host cell binding and infection as well as the actin-based motility of bacterial cells. It also triggers host cell factors related to the cytoskeleton [141]. CWA-3 molecule is a component of the cell wall integrity-signaling pathway, which is regulated by small proteins like GTP-BP Rho1. Controlling gene expression and coordinating periodic modifications to the cell wall are the primary functions of CWA molecule during the cell cycle and in response to various forms of cell stress [142]. NNKs (or nucleoside diphosphate kinases, NDPKs) catalyze the transfer of the terminal phosphate (P) from a donor triphosphate (TP) to an acceptor diphosphate (DP). Arp2/3 is a ubiquitous and essential component of the actin skeleton found in eukaryotic cells. It nucleates actin filaments, caps their sharp ends, and cross-links them to form orthogonal networks. These molecules form a large group that is associated with Mp10, a protein classified as a “chemosensory protein”. The functions of these proteins are not as closely related to smell as they are to the cytoskeleton, actin filaments, transcription of genes, phosphate transfer reaction/exchange, phosphorylation, mitochondrial energy production, TP-DP conversion, and cell regulation (Figures 5 & 6).

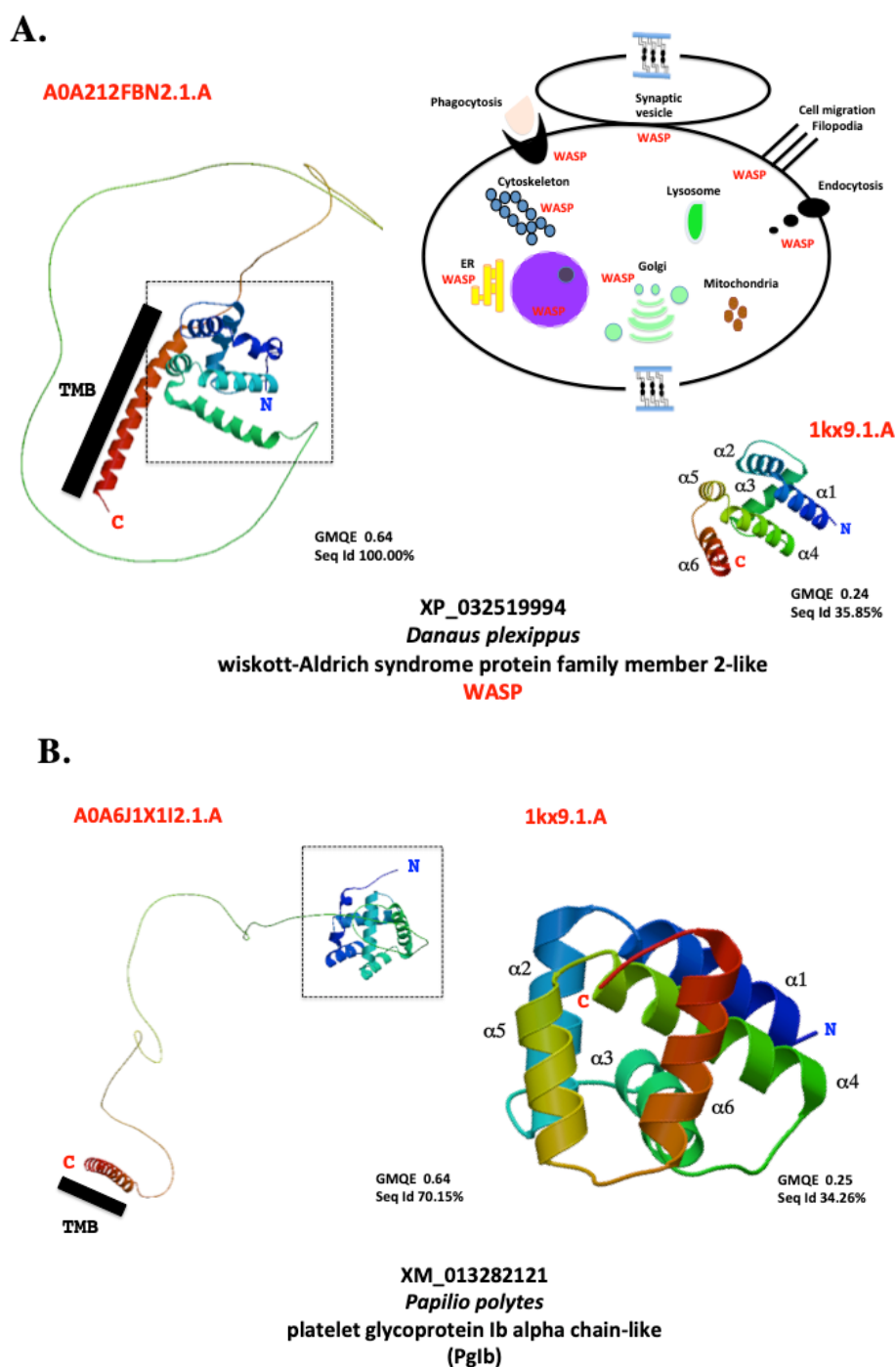


Figure 6. Molecular structure modeling of CSP, Allergen, Mucin, Rho, TIF, ASRP, and NPCP. A. Molecular structure modeling of Danpl-WASP (XP_032519994). B. Molecular structure modeling of Pappo-PgIb (XM_013282121). WASP and PgIb sequences were aligned with Bommo-CSP1 in order to identify the signal peptide and cut it off based on N-terminal sequencing [13]. The amino acid sequence of the mature protein was then subjected to molecule structure modeling using Swissmodel.expasy.org. The molecules with the highest identity score was used as template references: 1kx9.1 ("Chemosensory Protein A6", X-ray, 1.6 Å, monomer, cabbage moth, *M. brassicae*) and A0A212FBN2.1.A (WASP family member 2-like, AlphaFold DB model of A0A212FBN2_DANPL, LOC116772069 gene, monarch butterfly, *D. plexippus*) for WASP (in A); 1kx9.1 and A0A6J1X1I2.1.A (Mucin-2-like, AlphaFold DB model of A0A6J1X1I2_GALME, LOC113521739 gene, greater wax moth, *Galleria mellonella*) for PgIb (in B). For WASP (A) and PgIb (B), the Global Model Quality Estimation (GMQE) and the percentage of Sequence Identity (Seq Id) are shown. C: C-terminus, N: N-terminus. The α -helices that make up the CSP structure are numbered 1 through 6. The location and function of

WASP in the cytoskeleton, ER, Golgi, filopodia, nucleus, and near endocytotic, phagocytotic, and synaptic vesicles point to the N-terminal tail of WASP and PgIb protein molecules in the intracellular compartment as the location of the CSP structure. The CSP prism is indicated by the black square with dotted lines. The two molecules share the same model of construction inside the cell: N-terminus CSP prism, long loop, and transmembrane domain. The black bar indicates the position of the transmembrane segment (TMB).

Mp10, "CSP", Allergen, TIF, Mucin, NPCP, and ASRP have a very distant common ancestral origin according to the topology of the UPGMA tree, which is based on the assumptions of a common root and constant evolutionary rates for all lineages (i.e., it takes the "Molecular Clock Hypothesis" to account for mutation rates). They are the outcome of a series of duplication events that produced Allergen, which includes CSP and TIF proteins, prior to Mp10, Mucin, NPCP, and ASRP (Figure 5A). Mp10 (and Rho activator) emerged later in a sequence of duplication events that produced a wide range of Mucin variants, particularly in the mosquito genera *Aedes*, *Anopheles*, *Culex*, *Uranotearia*, and *Wyeomyia* (Figure 5A). The high level of duplication and variation observed in Culpi- and Aedae-CSPs makes this point noteworthy [see 16]. During its evolutionary history, the molecule protein gene family that includes Mp10, CSP, Rho, Allergen, TIF, MUCIN, NPCP, and ASRP appears to have undergone multiple duplications. Some of these duplications are specific to particular taxonomic lineages, like mosquitoes for the long Mucin precursor proteins needed for growth, development, digestion, oviposition, and control of viral infection [144,145]. Other gene duplications are more ancient and common to all lineages (see Figure 5A). This is accurate for the Allergen group, which comprises *Trica*, ants, damselflies (*Ischnura* forktails) flies, garden whites (pierids), ladybirds, lice, neodiprions, parasitoid wasps, and tuberworm moths. This also holds true for the ASRP group, which comprises taxa from multiple families of moths and butterflies (swallowtails, speckled woods, *Papilio*, *Pararge*, etc; see Figure 5A & Table S3). The Figure 5A's tree makes it very evident how far apart the ASRPs/CSPs and the Allergens are. All of these molecules have a distant common origin that is estimated to be 324–440 Mya (the latest Mississippian–Silurian or Devonian [146]), according to the UPGMA tree of amino acid sequences. This is significantly earlier than the emergence of the various flying insect species.

The relationships between the CSP, Mp10, Rho, Allergen, TIF, Mucin, NPCP, and ASRP proteins were further examined using maximum parsimony analysis (Figure 5B). CSP, Mp10, Rho, Allergen, TIF, Mucin, NPCP, and ASRP molecules are all widely expressed in different tissues, which is already a striking feature. There have been reports of allergens from the abdomen and thorax clones, and Mucin, PAN-1, DAN4, NPCP, WASP, and the YLP motif are present in all body parts, larvae, and pupae (see Tables S3-S4). There is no role for these proteins in the molecular underpinnings of chemosensing. The YLP motif is known to facilitate RNA binding activity, regulate telomere maintenance, and contribute to the decrease in telomerase activity that occurs during stem cell differentiation by attaching to the core promoter of Telomerase Reverse Transcriptase (TERT) and leading to its down-regulation [146]. DAN4 determines cellular morphology and plays a crucial role in maintaining cell integrity during cell growth and division, under stress conditions and upon cell fusion. It is a component of the molecular interactions and enzymatic activities in the cell wall in response to different growth phases and toxic signals from the environment [147]. This is remarkably similar to what it is known about CSP ontogeny, tissue distribution, and response to chemical stress (see section 6). Phylogenetic information and expression analysis are combined to link CSPs to intracellular proteins, cell walls, and gene promoter regions. The insect CSP, Mp10, Rho, Allergen, TIF, Mucin, NPCP, and ASRP molecules are closely related to each other; they formed groups with a high bootstrap value (89-100%), ranging from Pedhu-CSP (XP_002432595) to Uralo-Mucin (XP0055592470) and Eupco-Rho (XM_056065396). Together, the Mucin and NPCP molecules form a group with 99% bootstrap value. Additionally, they attach to ASRP and Rho groups, which may indicate a strong relationship between these molecule families based on their extremely high bootstrap values (93-99%; Figure 5B). Aside protein molecules implicated in actin microtubule association, cytokinesis, Arp2/3, and NDPK, PAN and the ASRP group attaches to CSPs with significantly high bootstrap value (94-100%; Figures 5B & 5C).

Bacterial CSPs in MP analysis drew Rho but not Mp10. This simply indicates that, as opposed to Mp10, the bacterial CSPs (*A. baumannii* OIC81003 and OIC85870) and Rho share the fewest evolutionary steps, branches, and common ancestors in this PAUP tree feature. We agree that the evidence for an insect CSP's potential non-chemosensation role is not very strong when it comes to Rho and bacterial CSPs. We assert that the MP does not exhibit chemo-sensation for CSPs because CSP sequences establish a robust orthology group with 100% bootstrap support for Allergens, Mucins, TIFs, NPCPs, and ASRPs (Figure 5B,C). It follows that these protein families are not all that dissimilar from CSPs found in insects. The relationships between the major groups of CSPs, allergens, TIF, PAN, ASRP, NPCP, and mucins are clearly depicted in the circle-shaped NJ phylogenetic tree where the various branches interlock (Figure 5C).

The relationship between the Allergen group—particularly TIF—, and Mp10 stayed closer, indicating that CSP and TIF have certain similarities in many parts (Figures 5B-C & S7). Coleopteran, dipteran, hemipteran, hymenopteran, lepidopteran, phthirapteran, and zygopteran (Odonata, blue-tailed damselfly) Allergens/CSPs and Mp10 protein are related, albeit they do not form a general orthology grouping, in contrast to Mucins, NPCPs, and ASRPs (Figure 5B,C). This is where the UPGMA tree and the MP tree (Bootstrap Jackknife) in PAUP4.0b10 (Altivec) study diverge significantly (Figure 5A,B). Following the evolutionary distances between insect species, Allergens/CSPs, including TIFs, segregated independently on the MP tree (see Figure 5B). Additionally, as the ASRP group bootstraps on the MP or NJ tree, it approaches Mp10 (Figures 5B & 5C), suggesting that there are many specific homologies that CSPs, TIFs, and ASRPs have in common (Figure S7). This may suggest that, despite their extreme age—the eukaryotic cell origin—the TIF-ASRP molecule family is evolving as a fast, intense, frequent, and high rate, much like CSPs [149–152]. While further experimental evidence is required to fully understand the functional relationship between CSPs and the TIF and ASRP molecular complexes, it is reasonable to assume that these complexes share a very ancient mechanism for regulating actin skeleton and/or ribosome gene expression (Figure 5).

The Mp10 orthology of the aphid, beetle, dragonfly, fly, louse, mosquito, moth, and sawfly belongs to different groups, ranging from Mucins and NPCPs/ASRPs to Allergens and TIFs, according to our phylogenetic analysis (see Figures 5 & S6-S7), regardless of anything related to olfaction, taste, and/or chemosensing (see Table S3). This suggests that, as a result of multiple local duplications, the Mucin, Allergen, TIF, and CSP families underwent common diversification before the emergence of insects (e.g. the Carboniferous Period of the Paleozoic era 299-359 Mya; Figure 5). Mosquito Mucins were grouped together in accordance with the phylogenetic distances between their genomes. The evolutionary histories of CSP and Mucin proteins are then different. Mp10 orthologs in insects seem to have undergone one to five separate duplication events (see Figure 5). Four consecutive duplication events may have given rise to Mucins: two early duplications (d1 and d2) produced Allergens/CSPs, TIFs, and ASRPs; two late duplications (d3 and d4) produced Mp10 and Mucins. If the protein amino acid tree is the right tree to explain duplication and evolution within these relatively similar molecule families, then all of the molecules in the CSP, Allergen, PAN, TIF, Mucin, NPCP, and ASRP groups have the same root (Figures 5 & S6-S7 and Table S3). The following are linked: Mp10, CSPs, pherokines, ejaculatory buld-specific proteins (Ebsps), and a multitude of immune system, coat proteins, actin-related complex, nuclear complex, cell regulation, and cytoskeleton regulation proteins. The molecular sequences of gustatory receptors and ORs do not correlate in this instance. All of these Allergen-CSP and ASRP-NPCP molecules have a very old origin that dates back to the time before flying insects even existed and even further back to the time of microbes' prokaryotes' origin (about 3.5 Bya), even though a fifth duplication that produced DAN4 and NPCP may only be found in *Culpi* (see Figure 5).

9.2. Molecular Evidence Derived from Amino Acid Sequence Modeling Analysis

The strong relationships observed between CSP, Allergen, TIF, Mucin, NPCP, and ASRP molecules can be explained by the alignment of protein amino acid sequences and structural modeling (Table S3, Figures 5-6 & S7-S8). Protein amino acid sequence alignment demonstrates that

CSP precisely matches Allergen and the N-terminus of TIF, Mucin, NPCP, and ASRP (Figure S7). When compared to 1kx9.1 as template reference (X-ray crystal structure, [6]), it appears that the majority of query molecules fold into the same tunnel or prism of six α -helices using SWISS-MODEL Workspace/GMQE [153]. This was especially true for pherokine Phk-3, immune response protein, acid trehalase, Allergen Thap1, Mp10, Bommo-CSP, and *Coccinella septempunctata* Cocse-TIF (Figure S8). The other proteins examined in the SWISS-Model are all noticeably larger molecules; the prism structure is only visible in specific regions of the molecule, such as the N-terminus (Figures 6 & S8). A long α -helical stretch (7-13 turns) forming the C-terminal tail and a CSP prism at the N-terminal region characterize the majority of representative molecules in the TIF, Mucin, NPCP, and ASRP groups (Figure S8). Swissmodel analysis predicts a transmembrane segment at the C-terminal tail for these molecular structures (see Danpl-WASP, XP_032519994; Pappo-PgIb, XM_013282121; GMQE 0.64, Sequence Identity 70-100%; Figure 6AB). The Danpl-WASP protein's gene locus and molecular structure are known (A0A212FBN2.1 AlphaFold DB model of A0A212FBN2_DANPL, LOC11677206), and they do not match the characteristics of a molecule that is categorized as "chemosensory". Danpl-WASP is a single transmembrane domain-molecule that is 297 amino acids long and 33 kDa in mass. Among the most typical uses of these kinds of molecules are their actin-bound structures and their implications for filament assembly and actin cytoskeleton regulation [154-155]. These bigger molecules, such as WASP, are found in the cytoskeleton, microtubules, integument, cell wall, nuclear pore, cytoplasm, and many other intracellular organelles. Their remarkable long loop that joins the "CSP" structure at the N-terminus and the C-terminal transmembrane domain is what distinguishes them (Figures 6 & S8). To obtain this, the AlphaFold DB model of an uncharacterized protein or proteins categorized as "chemosensory protein" or "putative insect pheromone-binding protein (PBP)" is used as a template. It is strongly suggested that "CSP" and/or "putative insect PBP" exists in viruses, at least in herpesviruses that target the Golgi apparatus (HSV-1 and HSV-2), given the presence of the 'CSP prism'-prominent loop-transmembrane domain in a Palearctic butterfly molecule (CAH2235359, "jg5928" with BLLF1; Figure S8). "CSP" sequence is associated with the herpes virus's major outer envelope glycoprotein (see "jg5928"; A0A1V1WC08, GMQE 0.64, Sequence Identity 70.31%, Figure S8). If these proteins are a part of a larger assembly that includes glycoproteins, would it really be so ridiculous to find CSP in viruses? Maybe the portion of the viral particle's glycoproteins that binds to host lipids are "CSPs"? We have found that bacteria possess the same kind of molecular structure as described by Liu et al. [45]—a CSP joined to a larger molecule with transmembrane and glycoprotein domain (see *Lysobacter capsici* WP_096417339, 491 amino acids, 50.6 kDa). Therefore, rather than olfaction, it is more likely that infection (interaction with lipids, virus or microbe binding and immune response) is the primary function of CSP.

9.3. Cellular Evidence Derived from Location, Size, Structure, and Possible Expression in Viruses and Microbes

These striking similarities between CSP, Allergen, TIF, Mucin, NPCP, and ASRP molecules suggest that "CSP" molecules function in a multitude of diverse ways inside the intracellular systems of insects and hexapods (Figure 7).

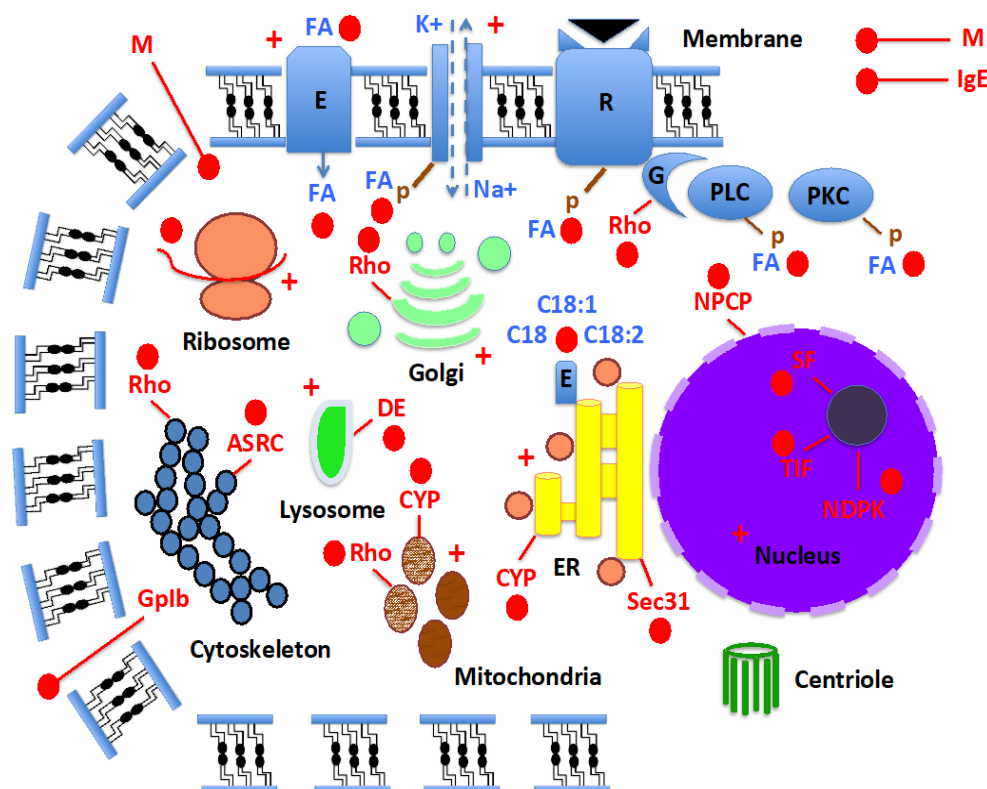


Figure 7. “CSP” molecules have several potential uses in the intracellular systems of insects and hexapods. CSP binds to fatty acid (FA), which mediates the phosphorylation (p) of different plasma membrane-bound protein molecules. E: Enzymes, K⁺ intake/Na⁺ outflow: potassium-sodium exchange pump, R: Receptors, G: G-protein, PLC: phospholipase kinase C, PKC: protein kinase C, Rho: Rho GTPase, ER: endoplasmic reticulum, E: Desaturase enzyme, which interacts with CSPs that transport FAs, such as linoleic acid (C18:2) and its precursors (stearic acid C:18 and elaidic acid C:18-1). +: Stress reactions that lead to CSP interacting with different molecules in the membrane, nucleus, ribosome, golgi, lysosome, ER and mitochondria, including cytochrome P450 (CYP), Degradative Enzyme (DE), Sec31 protein complex (Sec31), and Mucin (M). Additional molecules in the cytoskeleton, plasma membrane, and nuclear membrane that CSP proteins bind to are the actin skeleton regulatory complex (ASRC), Gplb-like protein (Gplb) and nuclear pore complex protein (NPCP). In the nucleus: splicing factor (SF), transcription initiation factor (TIF), and nucleoside diphosphate kinase (NDPK). IgE stands for Immunoglobulin E-like protein (parasitic infection, venom protection), while M stands for mucin-like fractions, secretory molecules that fortify the immune system's defense against microbial stress. The big red dots indicate each putative protein, protein assembly, or supramolecular complex in each intracellular organelle that “CSP” molecules interact with in both insects and hexapods based on phylogenetic and modeling analysis of the amino acid sequence (Mp10).

When it comes to insects and hexapods, CSP binds to LA (C18:2) which is made from arachidonic acid (AA, 20:4 ω 6) and is required for the biosynthesis of many hormones. C18:2 and 20:4 ω 6 are important modulators of intracellular processes. LA and AA phosphorylate a wide range of intracellular proteins (enzymes, pumps, receptors, and so forth), which in turn controls a multitude of signal transduction pathways and cellular processes [see 39]. They are related to the hormones leukotrienes, thromboxanes, and prostaglandins. These three classes of hormones regulate many physiological processes, including innate immune response, ion transport, egg development, and reproduction, but not smell [156-158]. In the nucleus, mitochondria, Golgi, ER, plasmic membrane, ion channels, ion pumps, ion transporters, lysosomes, and ribosomes, CSPs may participate in the lipid and phosphorylation processes that affect the biosynthesis of stress responses (see +, Figure 6). The fact that CSP binds to linoleic acid in whiteflies suggests this [see 38]. In the whitefly, the LA-

CSP is highly widely expressed throughout the entire body, not just the antennae [see 38]. Because of its interaction with LA, which helps enable the phosphorylation of many distinct cellular proteins via a variety of metabolic pathways, the CSP is able to participate in this extraordinarily broad field of cellular biology (Figure 7). In the animal model, LA activates multiple signaling pathways and receptors, regulates ion channels in metabolic cells, and induces calcium signaling and kinase phosphorylation in gustatory cells. Therefore, numerous targets for LA have been identified in a variety of cell types, including neural and pancreatic/liver cells [159-162]. Although this is primarily explained in terms of mammals, there is no reason to believe that LA, a lipid crucial to cell physiology in all living forms, will function differently in insects [163-168]. Mammal cellular fractions such as the lysosome, ribosome, proteasome, Golgi, reticulum, actin cytoskeleton, microtubules, and mitochondria are the targets of LA [164-168]. LA is essential for the assembly of parts involved in oxidative phosphorylation (OXPHOS, the electron transfer chain coupled to ATP synthesis) in the mitochondria, which is also involved in the immune response of insects [167-168]. Insect immune response is achieved when mitochondria, OXPHOS, and LA are combined [169]. The extremely broad action of LA addresses the CSP's tissue expression (extremely broad; see Figure 3). This places the CSP-LA complex mostly in the intracellular space (see Figure 7).

We wildly speculate about the potential function of CSPs and CSPs-like proteins in virtually every aspect of cellular biology, from prokaryotes to eukaryotes, taking into account the incredibly wide diversity of LA effects, LA molecular and cellular targets, and the extremely wide tissue distribution of CSP attaching LA. The "CSPs", along with LA, AA, FAs, and stress responses, could mediate the mechanisms controlling lipid metabolism in the cell cycle at the molecular translational level in the cell ribosome of insects [38-39, 170-171]. Moreover, CSP could interact with cytochrome P450 (CYP) enzyme in the ER and the mitochondrial system in response to insecticide exposure (+) and other stress responses. In response to insecticide exposure, this CSP-CYP interaction may be important for cellular metabolism, homeostasis, hormonal synthesis, toxin catabolism, and xenobiotic detoxification [20, 172]. It is possible that the CSP-Degradative Enzyme (DE) system mediates the lysosome's functions, which include defense against pathogens such as microbes, bacteria, fungi, and viruses, as well as the digestion and breakdown of macromolecules (proteins, lipids, carbohydrates, and nucleic acids) and repairs of cell membranes (Figure 6A). When food is eaten or absorbed by the cell, the lysosome releases its enzymes, which convert complex molecules—such as sugars, proteins, lipids, FAs, AA, and/or LA brought by CSPs— into energy that the cell needs to survive [173]. After that, the desaturase enzymes, the ER membrane, and the eversible vesicles that secrete the cuticle, the duct, and the pheromone compound will likely also need to be worked with the CSPs if they are to transport FAs like LA and its two lipid precursors, stearic acid (SA, C:18) and elaidic acid (EA, C:18-1). We hypothesize that the coordinated action of CSPs, FAs, and ER in the sex pheromone gland would be strictly necessary for the lipid droplet formation and pheromonogenesis. We firmly believe that the sites of metabolism that are likely regulated by CSPs are not the activation of OR on sensory neuron dendrites, but rather the ER membrane of the sex pheromone gland [174-175]. This notion is reinforced by the tissular distribution of CSPs and the precursor role of LA in several lipid biosynthesis pathways.

Here, we suggest that CSP is linked to the ER, as well as to the endomembranes from the Golgi, mitochondria, and plasma membrane via Rho GTPase signaling complex, Mucin, ASRC, and NPCP (see Figures 5-7). Perhaps Rho, Mucin, ASRC, and NPCP are connected to CSPs, allowing these molecules to protect the cell membrane, control membrane interactions, regulate the trafficking of protein precursors and lipids between the different cellular organelles, particularly the lysosomes and Golgi system [176,177]. Interestingly, we note that "CSP" consensus is the N-terminal region of "PgIb", a protein that resembles the glycoprotein receptor on the surface of human platelets (Figures 6A & S7). As a result, it may reside outside of cells in order to identify pathogens, trigger immune responses, and/or receive molecules like FA or JH and stimulate cell growth. Mechanisms linked to glycoprotein receptors are remarkably old, dating back to worms [178]. The fact that splicing factor (SF), TIF, and NDPK are linked by CSP consensus is also noteworthy (Figures 5-6 & S7-S8). This connects the CSP to nucleotide binding, nucleoside kinase, RNA metabolism, translation, splicing,

and nucleus-wide regulation of gene expression rather than odor binding. The relationship between CSPs and mucins, however, may be the most intriguing of all the Mp10 relatives that were taken from the NCBI database, as it would explain a function at both the sensory lymph level and the gut epithelial membrane (Figures 5-7 & S7-S8 and Table S3). Mucins are the most prevalent macromolecules in mucus in vertebrates and are in charge of the biochemical and physical characteristics of mucus. This role is strikingly similar to that of mucins in insects and hexapods. Similar to vertebrates, insects' mucins belong to three different families within the same tissue: transmembrane, gel-forming, and soluble, with possible strong similarities to CSPs in these categories [179,180]. In vertebrates, it is expressed in saliva, tears, glands, ear fluid, brain, lung, intestine, liver, kidney, heart, muscle, testis, ovary, and skin [181]. The transcripts of Drome mucins localize to specific organs and fluids that are similar to the sites where vertebrate mucins are expressed. It is also found that Drome mucins are present throughout the whole life cycle of the fly [182]. One prominent example is "JiangShi", an extracellular glycoprotein that bears a strong resemblance with mammalian "soluble" mucins. It is one of the most extensively distributed mucin-like proteins that has been extensively studied due to its crucial role in the growth and development of Drome [183]. Therefore, the family of mucin's rather widespread expression bears a striking resemblance to the role, ontogeny, and tissue distribution of CSPs (see Figures 3 & S4 and Section 5).

The midgut of beetles is characterized by the expression of mucus-forming mucins, which also bear a striking resemblance to the mucus found in vertebrates. It is thought that the primary functions of the mucus that surrounds food in the midgut are tissue protection and improved digestion through the facilitation of enzyme recycling [184]. Additionally, the gut of the insect has prominently enlarged posteriormost crypts that are likely specialized for vertical symbiont transmission. They have a significant impact on shielding host cells from harmful outside environments and/or invasive microbes. Like Bommo-CSP10, gut mucins are defined by highly glycosylated proteins that make up the mucin domain, which is rich in repetitive sequences of Thr, Ser, and Pro residues. These Multi-Proline mucins are primarily expressed in the gut and the female-specific symbiotic organ [185]. Eight mucin genes have been studied in locusts, and the results indicate that they are necessary for the growth and development of the species, vital for survival, involved in the formation of the gastric caecum, necessary for the formation of the peritrophic membrane in the midgut, or necessary for the cuticle of their wings to remain structurally intact [186]. These functions would effectively handle the distribution of CSP throughout the entire body of the insect (see Figure 3). Mucin, which is secreted into rice during feeding and highly expressed in salivary glands in true bugs, elicits plant immune responses, much like Mp10 [187]. An eggshell-related mucin-like protein, NIESMuc, was also identified in true bugs, specifically in the brown planthopper (Nillu). It is very likely that NIESMuc is necessary for regular oviposition and/or egg growth and development because it was highly expressed in the follicular cells of the ovarioles' egg chamber in Nillu females [188]. Lastly, gene expression analysis demonstrated that these mucins are dynamically expressed in various *Ae. albopictus* mosquito organs and throughout different developmental stages. Notably, blood feeding increased the expression of mucins in the ovary, fat body, and midgut. This clearly indicates that these mucins may serve similar developmental, growth, digestive, reproductive, and intestinal defense roles as CSPs in a multitude of insect species [189].

Thus, in many different tissues of the digestive and reproductive tracts, as well as in peripheral sensory organs with a large surface area exposed to the external environment, both Mucins and CSPs may cooperate to support the immune system against infections, pathogens, viruses, and microbial or chemical stress [190,191]. The exposed epithelia and sensory neuronal membrane of mammals are protected from environmental toxins by their mucous barrier. In insects, crustaceans, and mammals alike, this barrier serves as a lubricant during organ movements and as a mediator of substance exchanges between internal organs and the environment. Similar to mucus, 'lymph' is produced by surface cells, epithelial cells, and glands that are closely related to other components of the innate and adaptive immune systems. Lymph acts as the first line of defense against pathogens. Much like mucus, a large variety of fluids and secretions, venom, and saliva, the building blocks of the lymph are mucins and CSPs/OBPs. The capacity of mucins, CSPs, and OBPs to bind lipids in diverse

secretions and tissues is a shared characteristic [192–195]. Perhaps lipids are transported by CSPs, and lipid aggregation and droplet formation are facilitated by mucins. Gastric mucins have long been known to mediate tissue protection through lipid binding [195]. In the gut, interactions between mucin and lipid droplets have been specifically studied [196]. They may engage in direct relations with the gut microbiome through interactions with short-chain FAs there [197]. It would be interesting to look into whether the primary component of the mucin-CSP association is the interaction between mucins and short-chains of lipids, while CSPs interact with long-chain lipid molecules in the brain, gut, fat body, and sex pheromone gland. This is due to striking resemblance that we have shown here between CSPs and insect mucins (sequence identity, structural folding, response to toxins, broad tissue distribution, and developmental profiling; see Figures 3-7 & S4-S8 and Tables S3-S5).

10. To design “CSP”, New Terminology Is Needed: Structure or Function?

The relationships between p10, OS-Ds, CSPs, mucins, ASRPs, NPCPs, cytoskeleton complexes, genetic elements, nuclear pores, and intracellular processes that we discuss here were not even remotely conceivable during the p10 and OS-D era (1992-2000).

Table 2. Changing the name of “ChemoSensory Proteins” (“CSPs”) to reflect their lipid roles.

Name	Organism	Authors	Year	Reference
Lipoclistin	Insects	Steinbrecht	2003	[202]
LA-BP	Bemisia tabaci	Liu et al.	2016	[38]
Toxin-BP	Bemisia tabaci	Liu et al.	2016	[38]
B-CSP	Acinetobacter baumannii	Liu et al.	2019	[45]
Lipid-BP	Bemisia tabaci	Liu et al.	2020	[39]
JHRP	Aedes aegypti	Picimbon	2020	[16]
Mucin module	Aedes aegypti	Liu et al.	2024	This paper
TIF module	Aedes aegypti	Liu et al.	2024	This paper
ASRP module	Aedes aegypti	Liu et al.	2024	This paper
NPCP module	Aedes aegypti	Liu et al.	2024	This paper
4CSP*	Insects, crustaceans, and microbes	Liu et al.	2024	This paper

* Alejandro P. Rooney (USDA-ARS, Lubbock, Texas, USA).

Since “CSP” molecules were found in the hemolymph of infected flies [27], more research has been done on the non-chemosensory functions of this protein family and in relation to insecticide resistance [20,28,38,39,198]; however, none of the studies have looked as closely as ours at the potential intracellular role of “CSP” in relation to ASRP, NPCP, mucin, TIF, transcription initiation, RNA, nucleoside kinase, and cell regulation (see Figures 5-7 & S6-S8 and Tables S3-S5).

Typically, a neuropeptide or hormone molecule is named after the first “physiological” or “pharmacological” effect that is noted for it; however, this naming scheme often proves to be fairly inaccurate. The best course of action is probably to rename all naturally occurring peptides, hormones, carrier proteins, transporters, “CSPs”, and related molecules in a more impartial, methodical and objective manner. The location and extensive list of functions (‘pleiotropy’ and multiple phenotypic expression) linked to this molecule protein gene family defy the different labels assigned to “OS-D”, “SAP” or “CSP” (Figures 3-6 & S1-S4). It is necessary to build new terms or nomenclature, but they cannot be restricted to chemosensory or olfactory structures, take on any enigmatic roles, or be used in the lack of sufficient evidence or definitive proof. The insect physiology, organismal phylogeny, evolution of pleiotropy, number of gene loci, gene expression, developmental profiling, localization, and tissue distribution should all be represented by the name, at the very least (Table 2).

It becomes a little awkward or too inappropriate to refer to the entire protein gene family using the terms “SAP” and “CSP”, which stand for “Sensory Appendage Protein” and “Chemosensory Protein”, respectively, as they are “exclusively expressed in olfactory cilia”. This is not the case for olfactory cilia sole expression. To describe all the genes and protein molecules expressed in the antennae, mouth, and legs, gut, hemolymph, many non-sensory metabolic cell tissues, fat body, salivary gland, prothoracic gland, and/or corpora allata, as well as all those that have evolved from organisms ranging from bacteria and viruses to insects and hexapods, it would be rather inappropriate to use either term —CSP or SAP— . “JH-related proteins”, or “JHRPs”, would be a far better moniker for these proteins than CSPs, OS-Ds, OBPs, or SAPs given the tissue distribution. It is not appropriate to tune the NCBI Pfam domain A10/OS-D to a “chemosensory” protein. The EST and genome databases have been analyzed, in insects and bacteria, RNAi, gene knockout, so further discussion is not necessary. None of these “CSP/OS-D” proteins are specifically tuned to olfactory and/or taste chemosensory organs, according to EST-RNAs of bacteria, insects, and hexapods, which is consistent with the molecular expression data obtained in moths [20]. CSP gene deletion affects how the insect brain develops [69]. Our analysis shows that they are far more similar to general intracellular components than we had previously believed (Figures 3-7 & S6-S8 and Table S3).

This situation seems to be similar to that of a superfamily of widely distributed, heterogeneous proteins known as “lipocalins”, which are responsible for the transport of small hydrophobic molecules such as lipids and steroids (the term ‘lipocalin’ comes from the Greek lipos, which means fat, and Greek kálux, κάλυξ, which means cup).

Lipocalins are diverse and rather poorly conserved throughout evolution [199]. Unlike lipocalins, the “CSP” family describes a group of rather homogeneous, similar gene structures across a wide range of organisms; they share the same intron boundaries, and evolutionary well-conserved molecular structures (a hydrophobic tunnel in a prism or pyramid with a pattern of four cysteines that is always in the same position), and similar tissue profiling (ubiquitous expression; see chapters 5 & 6). Their binding properties differ somewhat (not only to odorant-like, long FAs, LA, and lipid chains, but also to small cyclic compounds), but the primary distinction between lipocalins and CSPs is that CSPs connect with TIFs, NPCPs, Mucins, DNA-regulatory proteins, SamkC, WAS/WASL-like, CWP, cytoskeleton actin-linkers, and other transcriptional factors that have never been discussed thus far (Figures 5-7 & S6-S8). It even appears that “CSPs” have undergone natural selection and evolution in viruses prior to becoming species and phenotypes (see Figures 5 & 6B), which is a stark contrast to lipocalins.

“CSPs” also designs the *Clostridium difficile* spore germination Csp family of (pseudo)proteases (which functions as a bile salt germinant receptor) and Cold Shock Proteins from *E. coli* or *Yersinia pseudotuberculosis* [200,201]. Due to the lack of any evidence supporting a CSP-OR interaction, the primary/first-identified function of CSPs in insects—which is thought to be the extracellular activation of olfactory receptor neurons— is still very poorly understood. The potential roles of CSP in lipid transport within cells are currently the subject of intense discussion. Therefore, it is probably best to use a new nomenclature to describe the peptide molecule, taking into account intracellular localization and evolution in microbes. It would be appropriate to refer to CSPs and OBPs as “lipoclistins” as proteins that enclose lipid ligands, much like lipocalins do; citing Rudolf Alexander Steinbrecht ([202]; Max-Planck Institute for Biological Intelligence, Seewiesen, Germany). The main structural component of the CSP —an open-‘air’ structure—would be best described as “lip-anoiktins” in accordance with this system of naming based on words and names’ Greek ancestry (open=ανοικτος anoiktos). Following Rudolf Alexander Steinbrecht [203], the system of naming based on the “Latin” root of words and names suggests that the CSP protein family would also be well-suited to be referred to by the Latin word “apertus”.

A generic name such as “arthrolipin” would be inappropriate for CSP molecule since it is not specific to the arthropod family; it can also be found in viruses, bacteria, and many various prokaryotic cells. Not to be forgotten, in honor of the early literature and the old name, Alejandro P. Rooney (USDA-ARS, Lubbock, Texas, USA) proposed “4CSP” (Four Cys Soluble Proteins) or

CSP/4Cys without drawing any broad conclusion regarding their physiological task, which would be the most logical name for the protein family.

Other than the zinc-binding protein of Ecl1-type and nuclear factor I's DNA binding activity, there are extremely few examples of proteins other than “chemosensory proteins” in organisms from various phyla and kingdoms that also contain four cysteines [204,205]. With regard to cytokines, things are different because there are multiple sub-families of “chemokines” that are identified by the locations of successively conserved cysteine residues (such as CXC, CC, C, and CX3C) [206]. Therefore, it wouldn't be unclear, particularly considering that the N-terminal region of these protein families contains all four of the conserved cysteines. Since the profiling of the four cysteines appears to be necessary for intracellular function and/or interaction with DNA, this is actually more intriguing than confusing.

It would be confusing if any insect CSP family members obtained six, eight, or ten conserved cysteine residues—the kind found in OBPs, for example. The A10/OS-D Pfam domain is what 4CSP refers to. This is most likely the case because larger proteins already contain this Pfam domain. The case involves ASRP, NPCP, Mucin, Rho, and TIF, all of which are examined in this study and demonstrated to be intracellular expansion of the A10/OS-D Pfam domain (renamed to 4CSP).

11. Concluding Remark

We have reassessed the function of CSPs which is extremely diverse, with the goal of shifting the paradigm from a strictly chemosensory function of CSPs, to a larger one that is reduced essentially to their putative binding function but with multiple biological roles in insects. Currently, the nomenclature used to refer to CSPs is utterly incorrect and disjointed. Even worse now, given our current level of knowledge, compared to twenty years ago, when we were limited to thinking about expression in the antennae, neglecting the expression in the gut, fat body, thoracic glands and CA. It is impossible to classify as “chemosensory” molecules this protein family that is expressed extensively throughout the insect and linked to immune responses and intracellular systems.

There is the Mp10 experiment. The results of the phylogenetic experiment demonstrate the remarkably high sequence homology between the counterparts of Mp10 and Mucin. Many insect CSPs are included in this analysis. Sequence alignment makes clear where the “A10/OS-D” Pfam domain is located in the N-tail of a large family of intracellular proteins (see Figure S7). Thus, instead of concentrating only on chemosensing or non-chemosensing, it is critical to investigate the connections among CSP, ASRP, NPCP, Mucin, TIF, RNA-BP, and other cell regulator molecules (Figure 7). Mucins, actin-related proteins, and nuclear complexes are examples of larger molecules or molecular agencies to which CSP binds. The structure of WASP and PgIb linked to CSP in the N-tail is a prediction model that requires experimental validation using NMR and/or X-ray technology. Future work on CSP should incorporate this. Using CRISPR to cut the CSP at the N-tail of TIF and observe how this affects the function of the cell or nucleus might be an interesting experiment. Nonetheless, the CSP-TIF association is currently a noteworthy and significant phenomenon with implications for both functional and evolutionary issues. The question of how and when CSP attached TIF becomes essential to comprehend how distinct structural domains or modules came together to form protein molecules with particular intracellular functions. A long list of mucins, ASRPs, actin complexes, nucleotide-binding proteins, and gene regulators includes the A10/OS-D module. This could suggest that CSP's main purpose is as a module for bigger protein molecules, which is in opposition to the protein's ability to sense odors and supports intracellular localization and multiple tissue distribution.

Many insect species, but moths in particular, have long lipid FA chains and CSPs activated for female sex pheromone biosynthesis. On the other hand, exogenous long FA chains have a significant impact on a variety of bacterial processes, such as intracellular signaling and gene expression patterns. DiRusso and Black state that the protein molecules required for the biosynthesis and breakdown of FAs are primarily affected by these transcription patterns. The current findings are consistent with this placement of the “CSP” molecules (OS-D domains), which appear to be found in insects and hexapods as well as in bacteria and viruses.

Moreover, “CSP” exhibits rather broad substrate selectivity, binding transcription regulators, nucleotides, DNA, RNA, and/or other intracellular elements in addition to lipid FA molecules. Therefore, it is imperative that this molecule protein gene family be given a new name. The new name should ideally not allude to any specific function because, even among related groups of molecules, the evidence for a given function can be shaky and subject to too rapid change and contradiction. “CSP”, “OS-D” or “SAP”, is one peptide molecule that, in many cases, would be better described or categorized according to its structure rather than just one function. This is particularly true in cases where there is long ongoing disagreement regarding the molecule’s physiological role.

Supplementary Materials: Available at Preprints.org.

Author Contributions: Conceptualization, J.F.P.; methodology, J.F.P.; software, J.F.P.; validation, J.F.P.; formal analysis, J.F.P.; investigation, G.X.L., B.S., F.W., S.Y., Q.H., J.F.P.; resources, G.X.L., S.Y., Q.H., J.F.P.; data curation, G.X.L., S.B., F.W., Q.H., J.F.P.; writing-original draft preparation, J.F.P.; writing-review and editing, J.F.P.; visualization, G.X.L., S.B., F.W., Q.H., J.F.P.; supervision, J.F.P.; project administration, J.F.P.; funding acquisition, G.X.L., S.Y., Q.H., J.F.P.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Overseas High-Level Talent Program-Taishan Scholar title [NO.tshw20091015], Ministry of Science and Technology of China-National Expert title [G2022023033L], and the APC was funded by SDAS-STSP.

Acknowledgments: Award of the most cited paper of Insect Science 2017 (Increased expression of CSP and CYP genes in adult silkworm females exposed to avermectins; Xuan et al., 2015) [20], and most cited paper of PLoS ONE-Top 10% (Biotype characterization, developmental profiling, insecticide response and binding property of Bemisia tabaci chemosensory proteins: role of CSP in insect defense; Liu et al., 2016) [38]. Overseas high-level talent program-Taishan scholar title (#NO.tshw20091015), and Ministry of Science and Technology of China-National Expert title (#G2022023033L). We would especially like to thank Profs. Drs. Alejandro Rooney (USDA) and Rudolf Alexander Steinbrecht (MPI), for their insightful and very critical conversations regarding the renaming of “CSPs” and for sharing their opinions and thoughts on the manuscript’s preliminary draft. We also express our sincere gratitude to Reviewer #4 and the Editor, Zena Yu, as well as the four reviewers for their invaluable feedback and direction during the manuscript revision process (heartfelt thanks to Reviewer #4).

Conflicts of Interest: “The authors declare no conflict of interest.”.

Data availability statements: The original contributions presented in the study are all included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Abbreviations

- 4CSP: 4 Cysteines Soluble Proteins
- AA: Arachidonic acid
- AcrR: Regulator of adjacent *acrAB* efflux genes - functional protein of the transcriptional regulation system that confers bacterial resistance to the antibiotic tetracycline (see TetR and TFTR) - Regulated by FA lipids
- Acypi: *Acyrtosiphum pisum* (pea aphid)
- Aedae: *Aedes aegypti* (dengue yellow fever mosquito)
- Anoga: *Anopheles gambiae* (malaria mosquito)
- Allergen Tha p 1: IgE-binding protein (15 kDa) and major allergen of pine processionary caterpillar (*Thaumetopoea pityocampa*, Thapi) - variant 1.0101
- Apime: *Apis mellifera* (honey bee)
- Arp2/3: Actin related protein 2/3 complex
- ASRC: Actin skeleton regulatory complex
- ASRP: Actin skeleton regulatory protein
- Avd: Accessory variability determinant
- Bacil: *Bacillus*
- B-CSP: Bacterial “Chemosensory protein”
- BioNJ: Bio (improved version) of the Neighbor Joining algorithm based on simple model of sequence data
- BLASTn: Nucleotide BLAST
- BLLF1: Epstein-Barr virus envelope glycoprotein encoded by BLLF1 gene
- Bommo: *Bombyx mori* (silkworm moth)
- BemtaCSP1: *Bemisia tabaci* “Chemosensory Protein”-1 (LA-binding protein)

CA: *Corpora allata*
 CBP: Chemical Binding Proteins
 Che: Chemotaxis proteins (histidine kinases)
 CheA, CheB: Chemosensory proteins-A, Chemosensory proteins-B (*Drosophila*)
 Cocse: *Coccinella septempunctata* (seven-spot ladybird)
 CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats
 CRLPB: Chemical sense-related lipophilic ligand binding proteins (*Drosophila*)
 CSP: Chemosensory protein
 Culpi: *Culex pipiens* (common house mosquito)
 CWA: Cell wall anchored
 CWP: Cell wall protein
 CYP: Cytochrome P450
 Cys: Cysteine
 DAN4: Cell wall mannoprotein expressed under delayed anaerobic conditions (*Saccharomyces*)
 Danpl: *Danaus plexippus* (monarch butterfly)
 Dd-Rp: DNA-dependent RNA polymerase
 DE: Degradative enzyme
 DGR: Diversity generating retroelement
 DNA-BP: Deoxyribonucleic acid-binding protein
 DNA-RP: Deoxyribonucleic acid-regulatory protein
 DP: Diphosphate
 Drome: *Drosophila melanogaster* (fruit fly)
 Droso: *Drosophila* species
 EA: Elaidic acid
 Ebsp: Ejaculatory bulb-specific protein
 Epiba: *Episyrphus balteatus* (marmalade hoverfly)
 ER: Endoplasmic reticulum
 EST: Expressed Sequence Tag
 Eupco: *Eupeodes corollae* (migrant overfly)
 FA: Fatty acid
 GC: Guanine-cytosine content
 GM2-AP: Ganglioside GM2 -activator protein
 GMQE: Global Model Quality Estimate
 GR: Gustatory receptor
 GTP: Guanosine triphosphate
 Halha: *Halyomorpha halys* (brown marmorated stink bug)
 HexA: Hexosaminidase A
 HSV: Herpesvirus
 IgE-BP: Immunoglobulin E-binding protein
 Iphpo: *Iphiclides podalirius* (scarce swallowtail)
 Jg5928: major outer envelope protein
 JH: Juvenile hormone
 JHRP: Juvenile hormone-related protein
 HSV: Herpesvirus
 LA: Linoleic acid (C18:2)
 Locmi: *Locusta migratoria* (migratory locust)
 LRR: Leucine-rich repeat protein complex
 M: Mucin
 Manse: *Manduca sexta* (tobacco hornworm, sphinx moth)
 MCP: Methyl-accepting chemotaxis protein
 ML: MD-2-related lipid-recognition domain
 MP: Maximum parsimony
 Mp10: Myzpe p10-like protein
 Myzpe: *Myzus persicae* (green peach aphid)
 NIESMuc: *Nilaparvata lugens* eggshell-related mucin
 Nillu: *Nilaparvata lugens* (brown planthopper, BPH)
 NNK: Nuclear nucleoside kinase
 NDPK: Nucleoside diphosphate kinase
 NPCP: Nuclear pore complex protein
 Nasvi: *Nasonia vitripennis* (parasitoid jewel wasp)
 OBP: Odorant-binding protein

OR: Olfactory receptor
 ORCO: Conserved Odorant Receptor type (Co-receptor)
 OS-D: Olfactory sensilla-type D protein
 OXPHOS: Oxidative phosphorylation
 P: Phosphate
 P10: Peram 10 kDa protein from regenerating legs
 Pacve: *Pachypsylla venusta* (hackberry petiole gall psyllid)
 PAN: Hexameric ATPase complex
 PAN-1: Protein encoded by the pan-1 gene in the nematode *Caenorhabditis elegans*
 Pappo: *Papilio polytes* (common Mormon)
 PAUP: Phylogenetic analysis using parsimony
 Pedhu: *Pediculus humanus humanus* (human body louse)
 Peram: *Periplaneta americana* (American cockroach)
 PBP: Pheromone-binding protein
 Pglb: Platelet glycoprotein Ib alpha chain-like
 Phk: Pherokine (hemolymph protein)
 PKC: Protein kinase C
 PLC: Phospholipase C
 Pol: Polymerase
 Ras: Family of GTPases derived from rat sarcoma virus
 Rho: Family of GTPases, family of small (~21 kDa) signaling G proteins, subfamily of the Ras superfamily
 RhoGAP: Rho GTPase-activating protein, regulator of the Rho-related protein family, crucial in many cellular processes, motility, contractility, growth, differentiation, and development
 RickA: *Rickettsia (conorii)* surface protein A (activator of Arp2/3)
 RNA-BP: Ribonucleic acid-binding protein
 RNAi: RNA interference
 SamkC: Serine/Threonine-protein kinase (from Social amoeba)
 SA: Stearic acid
 SAP: Sensory appendage protein
 Sec31: Protein transport protein encoded by the SEC31A gene (human)
 SF: Splicing factor
 SRP: Signal recognition particle
 TERT: Telomerase reverse transcriptase
 TetR: Tetracycline repressor
 TFTR: TetR-family transcriptional regulator
 TIF: Transcription initiation factor
 TP: Triphosphate
 Trica: *Tribolium castaneum* (red flour beetle)
 TSSC1: Tumor suppressing subtransferable candidate 1
 UL36: Large tegument protein denucleodysplasia encoded by UL36 gene (herpesviridae)
 UPGMA: Unweighted pair group method with arithmetic mean
 Uralo: *Uranotaenia lowii* (pale-footed Uranotaenia)
 WAS/WASL: Wiskott-Aldrich Syndrome/ Wiskott-Aldrich Syndrome-like protein
 WASP: Wiskott-Aldrich Syndrome protein
 XRE (or Xre): Xenobiotic response element family of DNA-binding transcriptional regulators
 YhbY: RNA-binding protein (folded like TIF) encoded by YhbY gene (*Escherichia coli*)
 YLPM1: YLP motif containing 1 encoded by YLPM1 gene (human)

References

- Hoffmann, K.H.; Meyerinng-Vos, M.; Lorenz, M.W. Allatostatins and allatotropins: Is the regulation of corpora allata activity their primary function? *Eur. J. Entomol.* 1999, 96, 255-266.
- Vogt, R.G.; Riddiford, L.M. Pheromone binding and inactivation by moth antennae. *Nature*, 1981, 293, 161-163.
- Picimbon, J.F.; Leal, W.S. Olfactory soluble proteins of cockroaches. *Insect Biochem. Mol. Biol.* 1999, 30, 973-978.
- Angeli, S.; Ceron, F.; Scaloni, A.; Monti, M.; Monteforti, G.; Minnocci, A.; Petacchi, R.; Pelosi, P. Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from *Schistocerca gregaria*. *Eur. J. Biochem.* 1999, 262, 745-754.
- Picimbon, J.F. Biochemistry and evolution of CSP and OBP proteins. In *Insect Pheromone Biochemistry and Molecular Biology, The Biosynthesis and Detection of Pheromones and Plant Volatiles*; Blomquist, G.J., Vogt, R.G., Eds; Elsevier Academic Press, London & San Diego, UK & USA, 2003; pp. 539-566.

6. Lartigue, A.; Campanacci, V.; Roussel, A.; Larsson, A.M.; Jones, T.A.; Tegoni, M.; Cambillau, C. X-ray structure and ligand binding study of a moth chemosensory protein. *J. Biol. Chem.* 2002, 277, 32094-32098.
7. Jansen, S.; Zidek, L.; Löfstedt, C.; Picimbon, J.F.; Sklenar, V. ¹H, ¹³C, and ¹⁵N resonance assignment of *Bombyx mori* chemosensory protein 1 (BmorCSP1). *J. Biomol. NMR* 2006, 36, 47.
8. Jansen, S.; Chmelik, J.; Zidek, L.; Padrta, P.; Novak, P.; Zdrahal, Z.; Picimbon, J.F.; Löfstedt, C.; Sklenar, V. Structure of *Bombyx mori* Chemosensory Protein 1 in solution. *Arch. Insect Biochem. Physiol.* 2007, 66, 135-145.
9. Tomaselli, S.; Crescenzi, O.; Sanfelice, D.; Ab, E.; Wechsel-berger, R.; Angeli, S.; Scaloni, A.; Boelens, R.; Tancredi, T.; Pelosi, P.; Picone, D. Solution structure of a chemosensory protein from the desert locust *Schistocerca gregaria*. *Biochemistry* 2006, 45, 1606-1613.
10. Jia, Q.; Zeng, H.; Zhang, J.; Gao, S.; Xiao, N.; Tang, J.; Dong, X.; Xie, W. The crystal structure of the *Spodoptera litura* Chemosensory Protein CSP8. *Insects* 2021, 12, 602.
11. Campanacci, V.; Lartigue, A.; Hällberg, B.M.; Jones, T.A.; Giuici-Orticoni, M.T.; Tegoni, M.; Cambillau, C. Moth chemosensory protein exhibits drastic conformational changes and cooperativity on ligand binding. *Proc. Natl. Acad. Sci. USA* 2003, 100, 5069-5074.
12. Mosbah, A.; Campanacci, V.; Lartigue, A.; Tegoni, M.; Cambillau, C.; Darbon, H. Solution structure of a chemosensory protein from the moth *Mamestra brassicae*. *Biochem. J.* 2003, 369, 39-44.
13. Picimbon, J.F.; Dietrich, K.; Angeli, S.; Scaloni, A.; Krieger, J.; Breer, H.; Pelosi, P. Purification and molecular cloning of chemosensory proteins from *Bombyx mori*. *Arch. Insect Biochem. Physiol.* 2000, 44, 120-129.
14. Picimbon, J.F.; Dietrich, K.; Krieger, J.; Breer, H. Identity and expression pattern of chemosensory proteins in *Heliothis virescens* (Lepidoptera, Noctuidae). *Insect Biochem. Mol. Biol.* 2001, 31, 1173-1181.
15. Wanner, K.W.; Isman, M.B.; Feng, Q.; Plettner, E.; Theilmann, D.A. Developmental expression patterns of four chemosensory protein genes from the Eastern spruce budworm, *Choristoneura fumiferana*. *Insect Mol. Biol.* 2005, 14, 289-300.
16. Picimbon, J.F. Chapter three—Bioinformatic, genomic and evolutionary analysis of genes: A case study in Dipteran CSPs. *Meth. Enzymol.* 2020, 642, 35-79.
17. Picimbon, J.F.; Dietrich, K.; Breer, H.; Krieger, J. Chemosensory proteins of *Locusta migratoria* (Orthoptera: Acrididae). *Insect Biochem. Mol. Biol.* 2000, 30, 233-241.
18. Guo, W.; Wang, X.; Ma, Z.; Xue, L.; Han, J.; Yu, D.; Kang, L. CSP and Takeout genes modulate the switch between attraction and repulsion during behavioral phase change in the migratory locust. *PLoS Genet.* 2011, 7, e1001291.
19. Martín-Blázquez, R.; Chen, B.; Kang, L.; Bakkali, M. Evolution, expression and association of the chemosensory protein genes with the outbreak phase of the two main pest locusts. *Sci. Rep.* 2018, 7, 6653.
20. Xuan, N.; Guo, X.; Xie, H.Y.; Lou, Q.N.; Bo, L.X.; Liu, G.X.; Picimbon, J.F. Increased expression of CSP and CYP genes in adult silkworm females exposed to avermectins. *Insect Sci.* 2015, 22, 203-219.
21. Ban, L.; Scaloni, A.; Brandazza, A.; Angeli, S.; Zhang, Y.; Pelosi, P. Chemosensory proteins of *Locusta migratoria*. *Insect Mol. Biol.* 2003, 12, 125-134.
22. Xuan, N.; Bu, X.; Liu, Y.Y.; Yang, X.; Liu, G.X.; Fan, Z.X.; Bi, Y.P.; Yang, L.Q.; Lu, Q.N.; Rajashekar, B.; Leppik, G.; Kasvandik, S.; Picimbon, J.F. Molecular evidence of RNA editing in the *Bombyx* chemosensory protein family. *PLoS ONE* 2014, 9, e86932.
23. Picimbon, J.F. Evolution of protein physical structures in insect chemosensory systems. In *Olfactory Concepts of Insect Control-Alternative to Insecticides*; Picimbon, J.F., Ed; Vol. 2, Springer Nature AG, Cham, Switzerland, 2019; pp. 231-263.
24. Khuhro, S.A.; Yan, Q.; Liao, H.; Zhu, G.H.; Sun, J.B.; Dong, S.L. Expression profile and functional characterization suggesting the involvement of three chemosensory proteins in perception of host plant volatiles in *Chilo suppressalis* (Lepidoptera: Pyralidae). *J. Insect Sci.* 2018, 18, 6.
25. Li, S.; Picimbon, J.F.; Ji, S.; Kan, Y.; Chuanling, Q.; Zhou, J.J.; Pelosi, P. Multiple functions of an odorant-binding protein in the mosquito *Aedes aegypti*. *Biochem. Biophys. Res. Commun.* 2008, 372, 464-468.
26. Wang, F.E.; Delannay, C.; Goindin, D.; Deng, L.; Guan, S.; Lu, X.; Fouque, F.; Vega-Rúa, A.; Picimbon, J.F. Cartography of odor chemicals in the dengue vector mosquito (*Aedes aegypti* L., Diptera/Culicidae). *Sci. Rep.* 2019, 9, 8510.
27. Sabatier, L.; Jouanguy, E.; Dostert, C.; Zachary, D.; Dimarcq, J.L.; Bulet, P.; Imler, J.C. Pherokine-2 and -3: Two *Drosophila* molecules related to pheromone/odor-binding proteins induced by viral and bacterial infections. *Eur. J. Biol.* 2003, 270, 3398-3407.
28. Liu, G.X. et al. Biotype expression and insecticide response of *Bemisia tabaci* chemosensory protein-1. *Arch. Insect Biochem. Physiol.* 2014, 85, 137-151.
29. Einhorn, E.; Imler, J.L. Insect immunity; from systemic to chemosensory organs protection. In *Olfactory Concepts of Insect Control-Alternative to Insecticides*; Picimbon, J.F., Ed; Vol. 2 Springer Nature AG, Cham, Switzerland, 2019; pp. 205-229.

30. Oduol, F.; Xu, J.; Niare, O.; Natajara, R.; Vernick, K.D. Genes identified by an expression screen of the vector mosquito *Anopheles gambiae* display differential molecular immune response to malaria parasites and bacteria. *Proc. Natl. Acad. Sci. USA* 2000, 97, 11397-11402.
31. Sim, S.; Ramirez, J.L.; Dimopoulos, G. Dengue virus infection of the *Aedes aegypti* salivary gland and chemosensory apparatus induces genes that modulate infection and blood-feeding behavior. *PLoS Pathog.* 2012, 8, e1002631.
32. Hou, C.X.; Qin, G.X.; Liu, T.; Mei, X.L.; Li, B.; Shen, Z.Y.; Guo, X.J. Differentially expressed genes in the cuticle and hemolymph of the silkworm, *Bombyx mori*, injected with the fungus *Beauveria bassiana*. *J. Insect Sci.* 2013, 13, 138.
33. Guo, W.; Song, J.; Yang, P.; Chen, X.; Chen, D.; Ren, D.; Kang, L.; Wang, X. Juvenile hormone suppresses aggregation behavior through influencing antennal gene expression in locusts. *PLoS Genet.* 2020, 16, e1008762.
34. Carr, A.L.; Rinker, D.C.; Dong, Y.; Dimopoulos, G.; Zwiebel, L.J. Transcriptome profiles of *Anopheles gambiae* harboring natural low-level *Plasmodium* infection reveal advantages for the mosquito. *Sci. Rep.* 2021, 11, 22578.
35. Li, Y.; Zhang, Y.; Xiang, Y.; Chen, D.; Hu, J.; Liu, F. Comparative transcriptome analysis of chemoreception organs of *Laodelphax striatellus* in response to rice stripe virus infection. *Int. J. Mol. Sci.* 2021, 22, 10299.
36. Traverso, L.; Latorre Estivalis, J.M.; da Rocha Fernandes, G.; Fronza, G.; Lobbia, P.; Mougabure Cueto, G.; Ons, S. Transcriptomic modulation in response to an intoxication with deltamethrin in a population of *Triatoma infestans* with low resistance to pyrethroids. *PLoS Negl. Trop. Dis.* 2022, 16, e0010060.
37. Wang, H.; Zhao, R.; Gao, J.; Xiao, X.; Yin, X.; Hu, S.; Zhang, Y.; Liang, P.; Gu, S. Two cuticle-enriched chemosensory proteins confer multi-insecticide resistance in *Spodoptera frugiperda*. *Int. J. Biol. Macromol.* 2024, 266, 130941.
38. Liu, G.X.; Ma, H.M.; Xie, Y.N.; Xuan, N.; Guo, X.; Fan, Z.X.; Rajashekar, B.; Arnaud, P.; Offmann, B.; Picimbon, J.F. Biotype characterization, developmental profiling, insecticide response and binding property of *Bemisia tabaci* chemosensory proteins: Role of CSP in insect defense. *PLoS ONE*, 2016, 11, e0154706.
39. Liu, G.X.; Xuan, N.; Rajashekar, B.; Arnaud, P.; Offmann, B.; Picimbon, J.F. Comprehensive history of CSP genes: Evolution, phylogenetic distribution, and functions. *Genes* 2020, 11, 413.
40. Nomura, A.; Kawasaki, K.; Kubo, T.; Natori, S. Purification and localization of p10, a novel protein that increases in nymphal regenerating legs of *Periplaneta americana* (American cockroach). *Int. J. Dev. Biol.* 1992, 36, 391-398.
41. Pikielny, C.W.; Hasan, G.; Rouyer, F.; Rosbach, M. Members of a family of *Drosophila* putative odorant-binding proteins are expressed in different subsets of olfactory hairs. *Neuron* 1994, 12, 35-49.
42. McKenna, M.P.; Hekmat-Safe, D.S.; Gaines, P.; Carlson, J.R. Putative *Drosophila* pheromone-binding-proteins expressed in a subregion of the olfactory system. *J. Biol. Chem.* 1994, 269, 16340-16347.
43. Robertson, H.M.; Martos, R.; Sears, C.R.; Todres, E.Z.; Walden, K.K.; Nardi, J.B. Diversity of odourant binding proteins revealed by an expressed sequence tag project on male *Manduca sexta* moth antennae. *Insect Mol. Biol.* 1999, 8, 501-518.
44. Bos, J.I.; Prince, D.; Pitino, M.; Maffei, M.E.; Win, J.; Hogenhout, S.A. A functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (green peach aphid). *PLoS Genet.* 2010, 6, e1001210.
45. Liu, G.X.; Yue, S.; Rajashekar, B.; Picimbon, J.F. Expression of chemosensory protein (CSP) structures in *Pediculus humanus corporis* and *Acinetobacter baumannii*. *SOJ Microbiol. Infect. Dis.* 2019, 7, 1-17.
46. Picimbon, J.F. Molecular phylogeny of "Chemosensory Proteins" in bacteria and crustaceans: CSP as an extremely ancient gene. *J. Mol. Evol.* 2024 revised.
47. Picimbon, J.F.; Regnault-Roger, C. Composés sémiochimiques volatils, phytoprotection et olfaction: Cibles moléculaires de la lutte intégrée. In *Biopesticides d'Origine Végétale*; Regnault-Roger, C., Philogène, B., Vincent, C., Eds; Lavoisier Tech & Doc, Paris, France, 2008; pp. 383-415.
48. Gissendanner, C.R.; Kelley, T.D. The *C. elegans* gene *pan-1* encodes novel transmembrane and cytoplasmic leucine-rich repeat proteins and promotes molting and the larva to adult transition. *BMC Dev. Biol.* 2013, 13, 21.
49. Gonzáles-González, A.; Yañez, O.; Ballesteros, G.I.; Palma-Millanao, R.; Figueroa, C.C.; Niemeyer, H.M.; Ramírez, C.C. A mutation increases the specificity to plant compounds in an insect chemosensory protein. *J. Mol. Graph. Model.* 2022, 114, 108191. 21.
50. Yue, S.; He, Q.X.; Picimbon, J.F. *Lactobacillus* for ribosome peptide cancer. *Clin. Transl. Oncol.* 2023, 25, 1522-1544.
51. Breer, H.; Fleischer, J.; Pregitzer, P.; Krieger, J. Molecular mechanisms of insect olfaction: Olfactory receptors. In *Olfactory Concepts of Insect Control-Alternative to Insecticides*; Picimbon, J.F., Ed; Vol. 2, Springer Nature AG, Cham, Switzerland, 2019; pp. 93-114.

52. Blomquist, G.J.; Tittiger, C.; Jurenka, R. Cuticular hydrocarbons and pheromones of arthropods. In *Oils and Lipids: Diversity, Origin, Chemistry and Fate*; Wilkes, H., Ed; Hydrocarbons, Handbook of Hydrocarbon and Lipid Microbiology, Springer, Cham, Switzerland, 2018; pp. 1-32.
53. Davis, J.S.; Percy, M.J.; Yew, J.Y.; Moyle, L.C. A shift to shorter cuticular hydrocarbons accompanies sexual isolation among *Drosophila americana* group population. *Evol. Lett.* 2021, 5, 521-540.
54. Ha, T.S.; Smith, D.P. A pheromone receptor mediates 11-cis-vaccenyl acetate-induced responses in *Drosophila*. *J. Neurosci.* 2006, 26, 8727-8733.
55. Lebreton, S.; Borrero-Echeverry, F.; Gonzales, F.; Solum, M.; Wallin, E.A.; Hedenström, E.; Hansson, B.S.; Gustavsson, A.L.; Bengtsson, M.; Birgersson, G.; Walker, W.B.3rd; Dweck, H.K.M.; Becher, P.G.; Witzgall, P.A. A *Drosophila* female pheromone elicits species-specific long-range attraction via an olfactory channel with dual specificity for sex and food. *BMC Biol.* 2017, 15, 88.
56. Starostina, E.; Xu, A.; Lin, H.; Pikielny, C.W. A *Drosophila* protein family implicated in pheromone perception is related to Tay-Sachs GM2-activator protein. *J. Biol. Chem.* 2009, 284, 585-594.
57. Pikielny, C.W. *Drosophila* CheB proteins involved in gustatory detection of pheromones are related to human neurodegeneration factor. *Vitam. Horm.* 2010, 83, 273-287.
58. Shang, J.; Tang, G.; Yang, J.; Lu, M.; Wang, C.Z.; Wang, C. Sensing of a spore surface protein by a *Drosophila* chemosensory protein induces behavioral defense against fungal parasitic infections. *Curr. Biol.* 2023, 99, 276-286.
59. Bontonou, G.; Saint-Leandre, B.; Kafle, T.; Baticle, T.; Hassan, A.; Sánchez-Alcañiz, J.A.; Arguello, J.R. Evolution of chemosensory tissues and cells across ecologically diverse *Drosophilids*. *Nat. Commun.* 2024, 15, 1047.
60. Torres-Oliva, M.; Almeida, F.C.; Sánchez-Gracia, A.; Rozas, J. Comparative genomics uncovers unique gene turnover and evolutionary rates in a gene family involved in the detection of insect cuticular pheromones. *Genome Biol. Evol.* 2016, 8, 1734-1747.
61. Wendeler, M.; Werth, N.; Maier, T.; Schwarzmann, G.; Kolter, T.; Schoeniger, M.; Hoffmann, D.; Lemm, T.; Saenger, W.; Sandhoff, K. The enzyme-binding region of human GM2-activator protein. *FEBS J.* 2006, 273, 982-991.
62. Spealman, P.; Burrelli, J.; Gresham, D. Inverted duplicate DNA sequences increase translocation rates through sequencing nanopores resulting in reduced base calling accuracy. *Nucl. Acids Res.* 2020, 48, 4940-4945.
63. Chau, L.M.; Goodisman, M.A.D. Gene duplication and the evolution of phenotypic diversity in insect societies. *Evolution* 2017, 71, 2871-2884.
64. Wellenreuther, M.; Bernatchez, L. Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol. Evol.* 2018, 33, 427-440.
65. Gloss, A.D.; Abbot, P.; Whiteman, N.K. How interactions with plant chemicals shape insect genomes. *Curr. Opin. Insect Sci.* 2019, 36, 149-156.
66. Nosil, P.; Soria-Carrasco, V.; Villoutreix, R.; De-la-Mora, M.; de Carvahó, C.F.; Parchman, T.; Feder, J.L.; Gompert, Z. Complex evolutionary processes maintain an ancient chromosomal inversion. *Proc. Natl. Acad. Sci. USA* 2023, 120, e2300673120.
67. Ding et al. Adaptive functions of structural variants in human brain development. *Science* 2024, 10, DOI: 10.1126/sciadv.adl4600.
68. Jin, P.; Zhu, B.; Jia, Y.; Zhang, Y.; Wang, W.; Shen, Y.; Zhong, Y.; Zheng, Y.; Wang, Y.; Tong, Y.; Zhang, W.; Li, S. Single-cell transcriptomics reveals the brain evolution of web-building spiders. *Nat. Ecol. Evol.* 2023, 7, 2125-2142.
69. Maleszka, J.; Forêt, S.; Saint, R.; Maleszka, R. RNAi-induced phenotypes suggest a novel role for a chemosensory protein CSP5 in the development of embryonic integument in the honeybee (*Apis mellifera*). *Dev. Genes Evol.* 2007, 217, 189-196.
70. Blobel, G. Protein targeting (Nobel lecture). *Chembiochem.* 2000, 1, 86-102.
71. Matlin, K.S. The strange case of the signal recognition particle. *Nat. Rev. Mol. Cell Biol.* 2002, 3, 538-542.
72. Doudna, J.A.; Batey, R.T. Structural insights into the signal recognition particle. *Annu. Rev. Biochem.* 2004, 73, 539-557.
73. Jin, X.; Brandazza, A.; Navarrini, A.; Ban, L.; Zhang, S.; Steinbrecht, R.A.; Zhang, L.; Pelosi, P. Expression and immunolocalization of odorant-binding and chemosensory proteins in locusts. *Cell. Mol. Life Sci.* 2005, 62, 1156-1166.
74. Steinbrecht, R.A.; Ozaki, M.; Ziegelberger, G. Immunocytochemical localization of pheromone-binding protein in moth antennae. *Cell Tissue Res.* 1992, 270, 287-302.
75. Liu, G.X.; Ma, H.M.; Xie, H.Y.; Xuan, N.; Picimbon, J.F. Sequence variation of *Bemisia tabaci* Chemosensory protein 2 in cryptic species B and Q: New DNA markers for whitefly recognition. *Gene* 2016, 576, 284-291.
76. Zhu, J.; Iovinella, I.; Dani, F.R.; Pelosi, P.; Wang, G. Chemosensory proteins: A versatile binding family. In *Olfactory Concepts of Insect Control-Alternative to Insecticides*; Picimbon, J.F., Ed; Vol. 2, Springer Nature AG, Cham, Switzerland, 2019; pp. 147-169.

77. Pino, J.; Godoy, R.; Venthur, H.; Larama, G.; Quiroz, A.; Mutis, A. Identification and ligand binding of a chemosensory protein from sea louse *Caligus rogercresceyi* (Crustacea: Copepoda). *Comp. Biochem. Physiol. B: Biochem. Mol. Biol.* 2023, 265, 110830.
78. Ghai, R.; Mizuno, C.M.; Picazo, A.; Camacho, A.; Rodriguez-Valera, F. Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Sci. Rep.* 2013, 3, 2471.
79. Zhao, Z.; Gonsior, M.; Schmitt-Kopplin, P.; Zhan, Y.; Zhang, R.; Jiao, N.; Chen, F. Microbial transformation of virus-induced dissolved organic matter from picocyanobacteria: Coupling of bacterial diversity and DOM chemodiversity. *ISME J.* 2019, 13, 2551-2565.
80. Doré, H.; Guyet, U.; Leconte, J.; Farrant, G.K.; Alric, B.; Ratin, M.; Ostrowski, M.; Ferrieux, M.; Brillet-Guéguen, L.; Hoebeke, M.; Siltanen, J.; Le Corguillé, G.; Corre, E.; Wincker, P.; Scanlan, D.J.; Eveillard, D.; Partensky, F.; Garczarek, L. Differential global distribution of marine picocyanobacteria gene clusters reveals distinct niche-related adaptive strategies. *ISME J.* 2023, 17, 720-732.
81. Bos, L.D.J.; Sterk, P.J.; Schultz, M.J. Volatile metabolites of pathogens: A systemic review. *PLoS Pathog.* 2013, 9, e1003311.
82. Taga, M.E.; Bassler, B.L. Chemical communication among bacteria. *Proc. Natl. Acad. Sci. USA* 2003, 100, 14549-14554.
83. Silva-Junior, E.A.; Ruzzini, A.C.; Paludo, C.R.; Nascimento, F.S.; Currie, C.R.; Clardy, J.; Pupo, M.T. Pyrazines from bacteria and ants: Convergent chemistry within an ecological niche. *Sci. Rep.* 2018, 8, 2595.
84. Salah Ud-Din, A.I.M.; Roujeinikova, A. Methyl-accepting chemotaxis proteins: A core sensing element in prokaryotes and archaea. *Cell Mol. Life Sci.* 2017, 74, 3293-3303.
85. Baker, M.D.; Wolanin, P.W.; Stock, J.B. Systems biology of bacterial chemotaxis. *Curr. Opin. Microbiol.* 2006, 9, 187-192.
86. Galperin, M.Y. Sensory transduction in bacteria. *Encyclopedia of Microbiology (Third Edition)*, Academic Press, 2009, 447-463.
87. Stock, A.; Chen, T.; Welsh, D.; Stock, J. CheA protein, a central regulator of bacterial chemotaxis, belongs to a family of proteins that control gene expression in response to changing environmental conditions. *Proc. Natl. Acad. Sci. USA* 1988, 85, 1403-1407.
88. Huang, Z.; Pan, X.; Xu, N.; Guo, M. Bacterial chemotaxis coupling protein: Structure, function and diversity. *Microbiol. Res.* 2019, 219, 40-48.
89. Perkin, L.C.; Friesen, K.S.; Flinn, P.W.; Oppert, B. Venom gland components of the ectoparasitoid wasp, *Anisopteromalus calandrae*. *J. Venom Res.* 2015, 6, 19-37.
90. Celorio-Mancera, Mde L.; Sundmalm, S.M.; Vogel, H.; Rutishauser, D.; Ytterberg, A.J.; Zubarv, R.A.; Janz, N. Chemosensory proteins, major salivary factors in caterpillar mandibular glands. *Insect Biochem. Mol. Biol.* 2012, 42, 796-805.
91. González-Caballero, N.; Valenzuela, J.G.; Ribeiro, J.M.C.; Cuervo, P.; Brazil, R.P. Transcriptome exploration of the sex pheromone gland of *Lutzomyia longipalpis* (Diptera: Psychodidae: Phlebotominae). *Parasit. Vectors* 2013, 6, 56.
92. Liu, Y.L.; Guo, H.; Huang, L.Q.; Pelosi, P.; Wang, C.Z. Unique function of a chemosensory protein in the proboscis of two *Helicoverpa* species. *J. Exp. Biol.* 2014, 217, 1821-1826.
93. Zhu, J.; Iovinella, I.; Dani, F.R.; Liu, Y.L.; Huang, L.Q.; Liu, Y.; Wang, C.Z.; Pelosi, P.; Wang, G. Conserved chemosensory proteins in the proboscis and eyes of Lepidoptera. *Int. J. Biol. Sci.* 2016, 12, 1394-1404.
94. Ma, C.; Cui, S.; Tian, Z.; Zhang, Y.; Chen, G.; Gao, X.; Tian, Z.; Chen, H.; Guo, J.; Zhou, Z. OcomCSP12, a chemosensory protein expressed specifically by ovary, mediates reproduction in *Ophraella communa* (Coleoptera: Chrysomelidae). *Front. Physiol.* 2019, 10, 1290.
95. Xu, A.; Park, S.K.; D'Mello, S.; Kim, E.; Wang, Q.; Pikielny, C.W. Novel genes expressed in subsets of chemosensory sensilla on the front legs of male *Drosophila melanogaster*. *Cell Tissue Res.* 2002, 307, 381-392.
96. Park, S.K.; Mann, K.J.; Lin, H.; Starostina, E.; Kolski-Andreaco, A.; Pikielny, C.W. A *Drosophila* protein specific to pheromone-sensing gustatory hairs delays males' copulation attempts. *Curr. Biol.* 2006, 16, 1154-1159.
97. Singh, S.S.; Mittal, A.M.; Chepurwar, S.; Gupta, N. Olfactory systems in insects : Similarities and differences between species In *Olfactory Concepts of Insect Control-Alternative to Insecticides*; Picimbon, J.F., Ed; Vol. 2, Springer Nature AG, Cham, Switzerland, 2019; pp. 29-48.
98. 29-48.
99. Ozaki, M. Odorant-binding proteins in taste systems: Putative roles in taste sensation and behavior. In *Olfactory Concepts of Insect Control-Alternative to Insecticides*; Picimbon, J.F., Ed; Vol. 2, Springer Nature AG, Cham, Switzerland, 2019; pp. 93-114.
100. Pedra, J.H.; Brandt, A.; Li, H.M.; Westerman, R.; Romero-Serverson, J.; Pollack, R.J.; Murdock, L.L.; Pittendrigh, B.R. Transcriptome identification of putative genes involved in protein catabolism and innate immune response in human body louse (Pediculidae: *Pediculus humanus*). *Insect Biochem. Mol. Biol.* 2003, 33, 1135-1143.

101. Kirkness, E.F. et al. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic life. *Proc. Natl Acad. Sci. USA* 2010, 107, 12168-12173.
102. Pittendrigh, B.R.; Clark, J.M.; Lee, S.H.; Yoon, K.S.; Sun, W.; Steele, L.D.; Seong, K.M. Body lice: From the genome project to functional genomics and reverse genetics. In *Short views on insect genomics and proteomics, Entomology in focus*; Raman, C., Goldsmith, M., Agunbiade, T., Eds; Springer, Cham, Switzerland, 2015 ; pp. 1-18.
103. Shigenobu, S.; Richards, S.; Cree, A.G.; Morioka, M.; Fukatsu, T.; Kudo, T.; Miyagishima, S.; Gibbs, R.A.; Stern, D.L.; Nakabashi, A. A full-length cDNA resource for the pea aphid, *Acyrtosiphon pisum*. *Insect Mol. Biol.* 2010, 19, 23-31.
104. Ollivier, M.; Legeai, F.; Rispe, C. Comparative analysis of the *Acyrtosiphon pisum* genome and expressed sequence tag-based gene sets from other aphid species. *Insect Mol. Biol.* 2019, 19, 33-45.
105. Liu, G.X.; Arnaud, P.; Offmann, B.; Picimbon, J.F. Genotyping and bio-sensing chemosensory proteins in insects. *Sensors* 2017, 17, 1801.
106. Loftus, B.J.; Utterback, T.; Perte, G.; Koo, H.; Mori, A.; Schneider, J.; Lovin, D.; de Bruyn, B.; Song, Z.; Raikhel, A.; de Fatima, B.M.; Casavant, T.; Soares, B.; Severson, D. *Aedes aegypti* cDNA sequencing. NCBI 2005, #DV263125, DV263127, DV289920, DV289921, DV297938, DV314711, DV314712, DV316589, DV316619, DV334734, DV334735, DV335057, DV335058, DV344048, DV347318, DV347319, DV365747, DV365766, DV357763, DV368339, DV368340, DV393559, DV400785, DV400787, "...".
107. Verjovski-Almeida, S.; Eiglmeier, K.; El-Dorri, H.; Gomes, S.L.; Menck, S.F.M.; Nascimento, A.L.; Roth, C.W. FAPESP and Institut Pasteur/AMSUD Network *Aedes aegypti* cDNA sequencing project. NCBI 2005, #EG001037.
108. Noriega, F.G.; Ribeiro, J.M.C.; Koener, J.F.; Valenzuela, J.G.; Hernandez-Martinez, S.; Pham, V.M.; Feyereisen, R. Comparative genomics of insect juvenile hormone biosynthesis. *Insect Biochem. Mol. Biol.* 2006, 36, 366-374.
109. Nene, V.; Worthman, J.R.; Lawson, D. et al. Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* 2007, 22, 1718-1723.
110. Wei, Z.; Ortiz-Urquiza, A.; Keyhani, N.O. Altered expression of chemosensory and odorant binding proteins in response to fungal infection in the red imported fire ant, *Solenopsis invicta*. *Front. Physiol.* 2021, 12, 596571.
111. Xu, H.; Yan, K.; Ding, Y.; Lv, Y.; Li, J.; Yang, F.; Chen, X.; Gao, X.; Pan, Y.; Shang, Q. Chemosensory proteins are associated with thiametoxam and Spirotetramat tolerance in *Aphis gossypii* Glover. *Int. J. Mol. Sci.* 2022, 23, 2356.
112. Gao, P.; Zhang, S.; Tan, J.J.; Li, X.; Chen, M. Chemosensory proteins are associated with thiametoxam tolerance in bird cherry-oat aphid *Rhopalosiphum padi*. *Pestic. Biochem. Physiol.* 2023, 192, 105393.
113. Xu, H.; Pan, Y.; Li, J.; Yang, F.; Chen, X.; Gao, X.; Wen, S.; Shang, Q. Chemosensory proteins confer adaptation to the ryanoid anthranilic diamide insecticide cyantraniliprole in *Aphis gossypii* glover. *Pestic. Biochem. Physiol.* 2022, 184, 105076.
114. Yi, X.; Qi, J.; Zhou, X.; Hu, M.Y.; Zhong, G.H. Differential expression of chemosensory-protein genes in midguts in response to diet of *Spodoptera litura*. *Sci. Rep.* 2017, 7, 296.
115. Guo, X.; Xuan, N.; Liu, G.X.; Xie, H.; Lou, Q.; Arnaud, P.; Offmann, B.; Picimbon, J.F. An expanded survey of the moth PBP/GOBP clade in *Bombyx mori*: New insight into expression and functional roles. *Front. Physiol.* 2021, 12, 712593.
116. Celorio-Mancera Mde. L.; Ytterberg, A.; Rutishauser, D.; Janz, N.; Zubarev, R.A. Effect of host plant and immune challenge on the levels of chemosensory proteins in caterpillar salivary glands. *Insect Biochem. Mol. Biol.* 2015, 61, 34-45.
117. Bian, H.X.; Chen, B.D.; Zheng, X.X., Ma, H.F.; Li, Y.P.; Li, Q.; Xia, R.X.; Wang, H.; Jiang, Y.R.; Liu, Y.Q.; Qin, L. Transcriptomic analysis of the prothoracic gland from two lepidopteran insects, domesticated silkworm *Bombyx mori* and wild silkworm *Antheraea pernyi*. *Sci. Rep.* 2019, 9, 5313.
118. Prajapati, V.K.; Varma, M.; Vadassery, J. In silico identification of effector proteins from generalist herbivore *Spodoptera litura*. *BMC genomics* 2020, 21, 816.
119. Zeng, Y.; Merchant, A.; Wu, Q.; Wang, S.; Kong, L.; Zhou, X.; Xie, W.; Zhang, Y. A chemosensory protein BtabCSP11 mediates reproduction in *Bemisia tabaci*. *Front. Physiol.* 2020, 11, 709.
120. Wang, Q.; Wen, X.; Lu, Y.; Wen, J. Comparison and functional analysis of chemosensory protein genes from *Eucryptorrhynchus scrobiculatus* Motschulsky and *Eucryptorrhynchus brandti* Harold. *Front. Physiol.* 2021, 12, 661310.
121. Kim, I.H.; Pham, V.; Jablonka, W.; Goodman, W.G.; Ribeiro, J.M.C.; Andersen, J.F. A mosquito hemolymph odorant-binding protein family member specifically binds juvenile hormone. *J. Biol. Chem.* 2017, 292, 15329-15339.
122. Laufer, H.; Borst, D.; Baker, F.C.; Reuter, C.C.; Tsai, L.W.; Schooley, D.A.; Carrasco, C.; Sinkus, M. Identification of a juvenile-hormone-like compound in a Crustacean. *Science* 1987, 235, 202-205.

123. Jindra, M.; Palli, S.R.; Riddiford, L.M. The juvenile hormone signaling pathway in insect development. *Annu. Rev. Entomol.* 2013, 58, 181-204.
124. Cayre, M.; Strambi, C.; Strambi, A. Neurogenesis in an adult insect brain and its hormonal control. *Nature* 1994, 368, 57-59.
125. Fahrbach, S.E.; Moore, D.; Capaldi, E.A.; Farris, S.M.; Robinson, G.E. Experience-expectant plasticity in the mushroom bodies of the honeybee. *Learn. Mem.* 1998, 5, 115-123.
126. Anton, S.; Rossler, W. Plasticity and modulation of olfactory circuits in insects. *Cell Tiss. Res.* 2021, 383, 149-164.
127. Ahmed, S.; Hrithik, M.T.H.; Roy, M.C.; Bode, H.; Kim, Y. Phurealipids, produced by the entomopathogenic bacteria, *Photobacterium*, mimic juvenile hormone to suppress insect immunity and immature development. *J. Invertebr. Pathol.* 2022, 193, 107799.
128. Ozaki, M.; Wada-Katsumata, A.; Fujikawa, K.; Iwasaki, M.; Yokohari, F.; Satoji, Y.; Nisimura, T.; Yamaoka, R. Ant nestmate and non-nestmate discrimination by a chemosensory sensillum. *Science* 2005, 309, 311-314.
129. Hasan, M.A.; Ahmed, S.; Kim, Y. Biosynthetic pathway of arachidonic acid in *Spodoptera exigua* in response to bacterial challenge. *Insect Biochem. Mol. Biol.* 2019, 111, 103179.
130. Vatanparast, M.; Ahmed, S.; Lee, D.H.; Hwang, S.H.; Hammock, B.; Kim, Y. EpOMes act as immune suppressors in a lepidoptera insect, *Spodoptera exigua*. *Sci. Rep.* 2020, 10, 20183.
131. Liaci, A.M.; Förster, F. Take me home, protein roads: Structural insights into signal peptide interactions during ER translocation. *Int. J. Mol. Sci.* 2021, 22, 11871.
132. Malcicka, M.; Visser, B.; Ellers, J. An evolutionary perspective on linoleic acid synthesis in animals. *Evol. Biol.* 2018, 45, 15-26.
133. Herrera, H.; Barros-Parada, W.; Bergmann, J. Linoleic acid and stearic acid are biosynthetic precursors of (7Z,10Z)-7,10-hexadecadienal, the major component of the sex pheromone of *Chilecomadia valdiviana* (Lepidoptera: Cossidae). *PLoS ONE* 2019, 14, e0215769.
134. Shibao, H.; Kutsukake, M.; Matsuyama, S.; Fukatsu, T. Linoleic acid as corpse recognition signal in a social aphid. *Zoological Lett.* 2022, 8, 2.
135. Buehlmann, C.; Graham, P.; Hansson, B.S.; Knaden, M. Desert ants locate food by combining high sensitivity to food odors with extensive crosswind runs. *Curr. Biol.* 2014, 24, 960-964.
136. Rodriguez, P.A.; Stam, R.; Warbroek, T.; Bos, J.I. Mp10 and Mp42 from the aphid species *Myzus persicae* trigger plant defenses in *Nicotiana benthamiana* through different activities. *Mol. Plant. Microbe. Interact.* 2014, 27, 30-39.
137. Yin, C.; Shan, G.; Guo, D.; Wang, S.; Ma, X.; Xiao, H.; Liu, J.; Zhang, Z.; Liu, Y.; Zhang, Y.; Yu, K.; Huang, S.; Li, F. InsectBase: A resource for insect genomes and transcriptomes. *Nucl. Acids Res.* 2016, 44, D801-D807.
138. Sáez, A.; Morales, C.L.; Ramos, L.Y.; Aizen, M.A. Extremely frequent bee visits increase pollen decomposition but reduce drupelet set in raspberry. *J. Appl. Ecol.* 2014, 51, 1603-1612.
139. Jacobsen, D.J.; Raguso, R.A. Lingering effects of herbivory and plant defenses on pollinators. *Curr. Biol.* 2018, 28, R1164-R1169.
140. Barre, A.; Pichereaux, C.; Simplicien, M.; Burlet-Schiltz, O.; Benoist, H.; Rougé, P. A proteomic- and bioinformatic-based identification of specific allergens from edible insects: Probes for future detection as food ingredients. *Foods* 2021, 10, 280.
141. Moon, S.Y.; Zheng, Y. (2003) Rho GTPase-activating proteins in cell regulation. *Trends Cell Biol.* 2003, 13, P13-22.
142. Gouin, E.; Egile, C.; Dehoux, P.; Villiers, V.; Adams, J.; Gertler, F.; Li, R.; Cossart, P. The RickA protein of *Rickettsia conorii* activates the Arp2/3 complex. *Nature* 2004, 427, 457-461.
143. Cronmiller, E.; Toor, D.; Shao, N.C.; Kariyawasam, T.; Wang, M.H.; Lee, J.H. Cell wall integrity signaling regulates cell wall-related gene expression in *Chlamydomonas reinhardtii*. *Sci. Rep.* 2019, 9, 12204.
144. Ho, S. The molecular clock and estimating species divergence. *Nat. Educ.* 2008, 1, 168.
145. Wang, P.; Granados, R.R. An intestinal mucin is the target substrate for a baculovirus enhancer. *Proc. Natl. Acad. Sci. USA* 1997, 94, 6977-6982.
146. Dias, R.O.; Cardoso, C.; Pimentel, A.C.; Damasceno, T.F.; Ferreira, C.; Terra, W.R. The roles of mucus-forming mucins, peritrophins and peritrophins with mucin domains in the insect midgut. *Insect Mol. Biol.* 2018, 27, 46-60.
147. Schachat, S.R.; Goldstein, P.Z.; Desalle, R.; Bobo, D.M.; Boyce, K.; Payne, J.L.; Labandeira, C.C. Illusion of flight? Absence, evidence and the age of winged insects. *Biol. J. Linn. Soc.* 2023, 138, 143-168.
148. Armstrong, L.; Lako, M.; van Herpe, I.; Evans, J.; Saretzki, G.; Hole, N. A role for nucleotide Zap3 in the reduction of telomerase activity during embryonic stem cell differentiation. *Mech. Dev.* 2004, 121, 1509-1522.
149. Teparic, R.; Lozancic, M.; Mrsa, V. Evolutionary overview of molecular interactions and enzymatic activities in the yeast cell walls. *Int. J. Mol. Sci.* 2020, 21, 8996.

150. Adachi, N.; Senda, T.; Horikoshi, M. Uncovering ancient transcription systems with a novel evolutionary indicator. *Sci. Rep.* 2016, 6, 217922.
151. Akil, C.; Kitaoku, Y.; Tran, L.T.; Liebl, D.; Choe, H.; Muengsaen, D.; Suginta, W.; Schulte, A.; Robinson, R.C. Mythical origins of the actin cytoskeleton. *Curr. Opin. Cell Biol.* 2021, 68, 55-63.
152. Rodriguez-Oliveira, T.; Wollweber, F.; Ponce-Toledo, R.I.; Xu, J.; Rittmann, S.K.-M.R.; Klingl, A.; Pilhofer, M.; Schleper, C. Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature* 2023, 613, 332-339.
153. Charles-Orszag, A.; Petek-Seoane, N.A.; Mullins, R.D. Archaeal actins and the origin of a multifunctional cytoskeleton. *J. Bacteriol.* 2024, 206, e00348-23.
154. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F.T.; de Beer, T.A.P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nuc. Acids Res.* 2018, 46, W296-W303. .
155. [154] Ho, H-Y.H.; Rohatgi, R.; Ma, L.; Kirschner, M.W. CR16 forms a complex with N-WASP in brain and is a novel member of a conserved proline-rich actin-binding protein family. *Proc. Natl. Acad. Sci. USA* 2001, 98, 11306-11311.
156. [155] Chereau, D.; Kerff, F.; Graceffa, P.; Grabarek, Z.; Langsetmo, K.; Dominguez, R. Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. *Proc. Natl. Acad. Sci. USA* 2005, 102, 16644-16649.
157. [156] Rogerio, A.P.; Anibal, F.F. Role of leukotrienes on protozoan and helminth infections. *Mediators Inflamm.* 2012, 2012, 595694. [157] Stanley, D.; Kim, Y. Chapter Eight – Insect prostaglandins and other eicosanoids: From molecular to physiological actions. *Adv. Insect Physiol.* 2019, 56, 283-343.
158. [158] Roy, M.C.; Nam, K.; Kim, J.; Stanley, D.; Kim, Y. Thromboxane mobilizes insect blood cells to infection foci. *Front. Immunol.* 2021, 12, 791319.
159. El-Yassimi, A.; Hichami, A.; Besnard, P.; Khan, N.A. Linoleic acid induces calcium signaling, Src kinase phosphorylation, and neurotransmitter release in mouse CD36-positive gustatory cells. *J. Biol. Chem.* 2008, 283, 12949-12959.
160. Feng, D.D.; Zhao, Y.F.; Luo, Z.Q.; Keating, D.J.; Chen, C. Linoleic acid induces Ca²⁺-induced inactivation of voltage-dependent Ca²⁺ currents in rat pancreatic β -cells. *J. Endocrinol.* 2008, 196, 377-384.
161. Yang, B.; Zhou, Y.; Wu, M.; Li, X.; Mai, K.; Ai, Q. ω -6 Polyunsaturated fatty acids (linoleic acid) activate both autophagy and antioxidant in a synergistic feedback loop via TOR-dependent and TOR-independent signaling pathways. *Cell Death Dis.* 2020, 11, 607.
162. Hidalgo, M.A.; Carretta, M.D.; Burgos, R.A. Long chain fatty acids as modulators of immune cells function: Contribution of FFA1 and FFA4 receptors. *Front. Physiol.* 2021, 12, 668330.
163. Blomquist, G.J.; Borgeson, C.E.; Vundla, M. Polyunsaturated fatty acids and eicosanoids in insects. *Insect Biochem.* 1991, 21, 99-106.
164. Marmunti, M.; Catalá, A. Incorporation of 1-(14)C linoleic acid in rat liver nuclei and chromatin fractions. *Int. J. Biochem. Cell Biol.* 2001, 33, 261-267.
165. Ando, H.; Wen, Z.M.; Kim, H.Y.; Valencia, J.C.; Costin, G.E.; Watabe, H.; Yasumoto, K.; Niki, Y.; Kondoh, H.; Ichihashi, M.; Hearing, V.J. Intracellular composition of fatty acid affects the processing and function of tyrosinase through the ubiquitin-proteasome pathway. *Biochem. J.* 2006, 394, 43-50.
166. Masner, M.; Lujala, N.; Bisbal, M.; Acosta, C.; Kunda, P. Linoleic and oleic acids enhance cell migration by altering the dynamics of microtubules and the remodeling of the actin cytoskeleton at the leading edge. *Sci. Rep.* 2021, 11, 14984.
167. Lauson, C.B.N. et al. Linoleic acid potentiates CD8⁺ T cell metabolic fitness and antitumor activity. *Cell Metabol.* 2023, 35, 633-650.
168. Vesala, L.; Basikhina, Y.; Tuomela, T.; Nurminen, A.; Siukola, E.; Vale, P.F.; Salminen, T.S. Mitochondrial perturbation in immune cells enhances cell-mediated innate immunity in *Drosophila*. *BMC Biol.* 2024, 22, 60.
169. Wrońska, A.K.; Kaczmarek, A.; Boguś, M.I.; Kuna, A. Lipids as a key element of insect defense systems. *Front. Genet.* 2023, 14, 1183659.
170. [170] Blank, H.M.; Maitra, N.; Polymenis, M. Lipid biosynthesis: When the cell cycle meets protein synthesis? *Cell Cycle* 2017, 16, 905-906.
171. [171] Blank, H.M.; Perez, R.; He, C.; Maitra, N.; Metz, R.; Hill, J.; Lin, Y.; Johnson, C.D.; Bankaitis, V.A.; Kennedy, B.K.; Aramayo, R.; Polymenis, M. Translational control of lipogenic enzymes in the cell cycle of synchronous, growing yeast cells. *EMBO Journal* 2017, 36, 487-502.
172. [172] Feyereisen, R. Insect P450 enzymes. *Annu. Rev. Entomol.* 1999, 44, 507-533.
173. [173] Fujiwara, Y.; Wada, K.; Kabuta, T. Lysosomal degradation of intracellular nucleic acids—Multiple autophagic pathways. *J. Biochem.* 2017, 161, 145-154.
174. [174] Yokoyama, N.; Fónagy, A.; Tatsuki, S.; Arie, T.; Yamashita, S.; Matsumoto, S. Ultrastructural studies on the pheromone-producing cells in the silkworm, *Bombyx mori*: Formation of cytoplasmic lipid droplets before adult eclosion. *Acta Biol. Hung.* 2003, 54, 299-311.

175. Hull, J.; Fonagy, A. Molecular basis of pheromonogenesis regulation in moths. In *Olfactory Concepts of Insect Control-Alternative to Insecticides*; Picimbon, J.F., Ed; Vol. 1, Springer Nature AG, Cham, Switzerland, 2019; pp. 115-202.
176. Ibarra, A.; Hetzer, M.W. Nuclear pore proteins and the control of genome functions. *Genes Dev.* 2015, 29, 337-349.
177. Lin, D.H.; Hoelz, A. The structure of the Nuclear Pore Complex (an update). *Annu. Rev. Biochem.* 2019, 88, 725-783.
178. Kenis, S.; Istiban, M.N.; Van Damme, S.; Vandewyler, E.; Watteyne, J.; Schoofs, L.; Beets, I. Ancestral glycoprotein hormone-receptor pathway controls growth in *C. elegans*. *Front. Endocrinol.* 2023, 14, 1200407.
179. Lang, T.; Hansson, G.C.; Samuelsson, T. Gel-forming mucins appeared early in metazoan evolution. *Proc. Natl. Acad. Sci. USA* 2007, 104, 16209-16214.
180. Lang, T.; Klasson, S.; Larsson, E.; Johansson, M.E.; Hansson, G.C.; Samuelsson, T. Searching the evolutionary origin of epithelial mucus protein components-Mucin and FCGBP. *Mol. Biol. Evol.* 2016, 33, 1921-1936.
181. McShane, A.; Bath, J.; Jaramillo, A.M.; Ridley, C.; Walsh, A.A.; Evans, C.M.; Thornton, D.J.; Ribbeck, K. Mucus. *Curr. Biol.* 2021, 31, R931-R947.
182. Syed, Z.A.; Hard, T.; Uv, A.; van Dijk-Hard, I.F. A potential role for *Drosophila* mucins in development and physiology. *PLoS ONE* 2008, 3, e3041.
183. Huang, Y.; Li, L.; Rong, Y.S. JiangShi (僵尸): A widely distributed Mucin-like protein essential for *Drosophila* development. *G3 (Bethesda)* 2022, 12, jkac126.
184. Ibrahim, S.P.; Dias, R.O.; Ferreira, C.; Silva, C.P.; Terra, W.R. Histochemistry and transcriptomics of mucins and peritrophic membrane (PM) proteins along the midgut of a beetle with incomplete PM and their complementary function. *Insect Biochem. Mol. Biol.* 2023, 162, 104027.
185. Moriyama, M.; Hayashi, T.; Fukatsu, T. A mucin protein predominantly expressed in the female-specific symbiotic organ of the stinkbug *Plautia stali*. *Sci. Rep.* 2022, 12, 7782.
186. Zhao, X.; Zhang, J.; Yang, J.; Niu, N.; Zhang, J.; Yang, Q. Mucin family genes are essential for the growth and development of the migratory locust, *Locusta migratoria*. *Insect Biochem. Mol. Biol.* 2020, 123, 103404.
187. Shanguan, X.; Zhang, J.; Liu, B.; Zhao, Y.; Wang, H.; Wang, Z.; Guo, J.; Rao, W.; Jing, S.; Guan, W.; Ma, Y.; Wu, Y.; Hu, L.; Chen, R.; Du, B.; Zhu, L.; Yu, D.; He, G. A mucin-like protein of planthopper is required for feeding and induces immunity responses in plants. *Plant Physiol.* 2018, 176, 552-565.
188. Lou, Y.H.; Shen, Y.; Li, D.T.; Huang, H.J.; Lu, J.B.; Zhang, C.X. A mucin-like protein is essential for oviposition in *Nilaparvata lugens*. *Front. Physiol.* 2019, 10, 551.
189. Deng, F.; Wu, S.; Wu, Y.; Liu, X.; Wu, P.; Zhai, Z. Identification of mucins and their expression in the vector mosquito *Aedes albopictus*. *J. Vector Ecol.* 2020, 45, 297-305.
190. Linden, S.K.; Sutton, P.; Karlsson, N.G.; Korolik, V.; McGuckin, M.A. Mucins in the mucosal barrier to infection. *Mucosal Immunol.* 2008, 1, 183-197.
191. Sheng, Y.H.; Hasnain, S.Z. Mucus and mucins: The underappreciated host defence system. *Front. Cell. Infect. Microbiol.* 2022, 12, 856962.
192. Muthusamy, N.; Sommerville, L.J.; Moeser, A.J.; Stumpo, D.J.; Sannes, P.; Adler, K.; Blackshear, P.J.; Weimer, J.M.; Ghashgaei, H.T. MARCKS-dependent mucin clearance and lipid metabolism in ependymal cells are required for maintenance of forebrain homeostasis during aging. *Aging Cell.* 2015, 14, 764-773.
193. Campanacci, V.; Krieger, J.; Bette, S.; Sturgis, J.N.; Lartigue, A.; Cambillau, C.; Breer, H.; Tegoni, M. Revisiting the specificity of *Mamestra brassicae* and *Antheraea polyphemus* pheromone-binding proteins with a fluorescence binding assay. *J. Biol. Chem.* 2001, 276, 20078-20084.
194. Picimbon, J.F. Insect chemosensory proteins against COVID-19. *Med Discov.* 2024, 3, 1152.
195. Gong, D.H.; Turner, B.; Bhaskar, K.R.; Lamont, J.T. Lipid binding to gastric mucin: Protective effect against oxygen radicals. *Am. J. Physiol.* 1990, 259, G681-686.
196. Chang, Y.; McClements, D.J. Characterization of mucin-lipid droplet interactions: Influence on potential fate of fish oil-in-water emulsions under simulated gastrointestinal conditions. *Food Hydrocoll.* 2016, 56, 425-433.
197. Hansson, G.C. Mucins and the microbiome. *Annu. Rev. Biochem.* 2020, 89, 769-793.
198. Ingham, V.A.; Anthousi, A.; Douris, V.; Harding, N.J.; Lycett, G.; Morris, M.; Vontas, J.; Ranson, H. A sensory appendage protein protects malaria vectors from pyrethroids. *Nature* 2020, 577, 376-380.
199. Diez-Hernando, S.; Ganfornina, M.D.; Skerra, A.; Gutiérrez, G.; Sanchez, D. An evolutionary perspective of the lipocalin protein family. *Front. Physiol.* 2021, 12, 718983.
200. Kevorkian, Y.; Shen, A. Revisiting the role of Csp family proteins in regulating *Clostridium difficile* spore germination. *J. Bacteriol.* 2017, 199, e00266-17.
201. Keto-Timonen, R.; Hietala, N.; Palonen, E.; Hakakorpi, A.; Lindström, M.; Korkeala, H.; Cold Shock Proteins: A mini-review with special emphasis on Csp-family of enteropathogenic *Yersinia*. *Front. Microbiol.* 2016, 7, 1151.

202. Steinbrecht, R.A. What can we learn from localizing lipoclistins? European Symposium For Insect Taste and Olfaction (ESITO VIII) 2003, July 2-7th, Harstad, Norway.
203. Steinbrecht, R.A. Fine structure immunocytochemistry—An important tool for research on odorant-binding proteins. *Meth. Enzymol.* 2020, 642, 259-278.
204. Novak, A.; Goyal, N.; Gronostajski, R.M. Four conserved cysteine residues are required for the DNA binding activity of nuclear factor I. *J. Biol. Chem.* 1992, 267, 12986-12990.
205. Shimasaki, T.; Ohtsuka, H.; Naito, C.; Azuma, K.; Tenno, T.; Hiroaki, H.; Murakami, H.; Aiba, H. Ecl1 is a zinc-binding protein involved in the zinc-limitation-dependent extension of chronological life span in fission yeast. *Mol. Genet. Genomics* 2017, 292, 475-481.
206. Miller, M.C.; Mayo, K.H. Chemokines from a structural perspective. *Int. J. Mol. Sci.* 2017, 18, 2088.
207. DiRusso, C.C.; Black, P.N. Bacterial long chain fatty acid transport: Gateway to a fatty acid-responsive signaling system. *J. Biol. Chem.* 2004, 279, 49563-49566.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.