

Essay

Not peer-reviewed version

---

# Renaming ‘Chemosensory’ Proteins (CSPs): Lipid/Nucleotide-Binding Proteins – Molecular Nomenclature, Structure, Expression, Function, Evolutionary Networks, Clinical Diseases and Associated Molecular Medicine

---

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

Posted Date: 27 December 2023

doi: [10.20944/preprints202312.1863.v1](https://doi.org/10.20944/preprints202312.1863.v1)

Keywords: 4 cysteines soluble proteins; transcription initiation factor; actin skeleton regulatory complex; nuclear pore complex protein; rho gtpase activator; mucin; lipid metabolism; immune response



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.

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.

Essay

# Renaming 'Chemosensory' Proteins (CSPs): Lipid/Nucleotide-Binding Proteins – Molecular Nomenclature, Structure, Expression, Function, Evolutionary Networks, Clinical Diseases and Associated Molecular Medicine

Guoxia Liu <sup>1</sup>, Botong Sun <sup>2,3</sup>, Fan Wei <sup>2</sup>, Shousong Yue <sup>1</sup>, Qiuxia He <sup>2</sup>  
and Jean-François Picimbon <sup>1,2,\*</sup>

<sup>1</sup> Biotechnology Research Center, Shandong Academy of Agricultural Sciences, Jinan, Shandong, 250100, Shandong, China.

<sup>2</sup> 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.

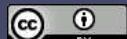
<sup>3</sup> Shandong SME Productivity Promotion Center, Jinan, 250103, Shandong, China.

\* Correspondence: High Level Talent (R), National Expert of China, STSP-SDAS, Shandong Academy of Sciences Foreign Students Pioneer Park, Jinan, Shandong, China;  
jfpicimbon@163.com & jfpicimbon@gmail.com, jfpicimbon@sdas.org; Tel: 00-86-17661098959

**Abstract:** This is a brief critique of the functions—particularly olfactory functions—specified for the “Chemosensory Protein” (CSPs) molecule family. On the basis of these proteins’ presence in the sensory antennal lymph of locusts, odor chemosensory ligand binding functions have been hypothesized. According to this hypothesis, the entire protein molecule superfamily is referred to as “CSPs”. However, new information and developments in the field of CSP molecular research, such as the expression of CSP genes in the gut, brain, fat body, epidermis, and pheromone gland, as well as gene expression profiling from most early developmental stages—that is, CSP expression well in advance of the appearance of chemical sense nerve cells—strongly suggest that the protein molecule has other roles that are unrelated to chemosensing. Moreover, CSPs are found in bacterial microbial prokaryote organisms in addition to insects. Thus, we examine the molecule’s name, definition, RNA editing, protein structure, lipid binding properties, DNA interaction, and evolutionary characteristics in brief before referring to this protein family as “Chemosensory Proteins”. This article tries to compel and discuss the most recent information as a way to rename this protein family. Because of its highly conserved molecular distinctive feature (four adjacent cysteines), we propose renaming “CSPs” as “4CSPs” (*4 Cysteines Soluble Proteins*).

**Keywords:** 4 cysteines soluble proteins; transcription initiation factor; actin skeleton regulatory complex; nuclear pore complex protein; rho gtpase activator; mucin; lipid metabolism; immune response

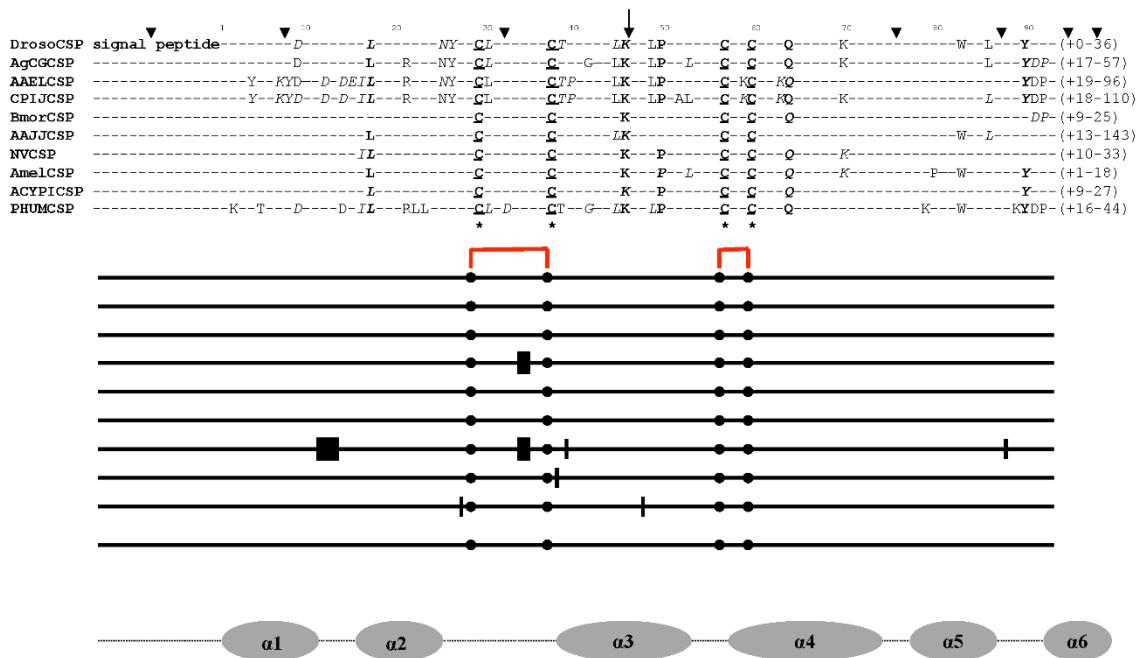
The topic of this essay is whether the biological function of “CSP” can be expanded in view of the fact that insect organisms inherit these molecules for the purpose of binding DNA and fatty acid lipids. We begin by rethinking the idea of the biological function of the molecules based on transcription initiation factor (TIF), mucin, and CSP analogies. We show that one of the central tenets of the literature—the role of these molecules in chemosensing—is simply accepted without any supporting data or logic. We can offer a new definition of the biological function of CSP by highlighting its unique relationship to lipids, intracellular events, DNA regulation, and the particular significance of this cell concept in biology. This is accomplished by examining several distinctions and facts, such as their extensive distribution across tissues, their pervasive expression during growth and development, their reaction to stress, and their strong resemblance to the N-terminus of the



nuclear, endoplasmic reticulum, ribosome, mitochondrial, cytoskeleton, and plasma membrane proteins, and thus their localization within the cells in numerous different compartments, as well as the novel insights and lessons that can be derived from them.

## 1. Definition

The term “CSPs” stands for “Chemosensory Proteins”. They typically refer to small, water-soluble binding proteins, also known as odor-binding proteins (obps), that are strongly 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* (Vogt and Riddiford, 1981; Picimbon and Leal, 1999; Angeli et al., 1999; Picimbon, 2003). According to Lartigue et al. (2002), CSPs are composed of six  $\alpha$ -helical chains with an appropriate molecular weight of 10-12 kDa (or 110-120 amino acid residues), four cysteines that form two tiny loops, two nearby disulfide bridges, and a globular “prism-like” functional structure (Figure 1). Four CSP structures have so far been identified in locusts (*Schistocerca gregaria*; Tomaselli et al., 2006) and moths (*Mamestra brassicae*, *Bombyx mori*, and *Spodoptera litura*; Lartigue et al., 2002; Jansen et al., 2006, 2007; Jia et al., 2021). *Drosophila ananassae*, *D. erecta*, *D. grimshawi*, *D. melanogaster*, *D. mojavensis*, *D. persimilis*, *D. pseudoobscura*, *D. sechellia*, *D. simulans*, *D. virilis*, *D. yakuba*, *D. willistoni* (Droso), *Aedes aegypti* (AAEL), *Anopheles gambiae* (AgCG), *Culex pipiens* (CPIJ), *Bombyx mori* (Bmor), *Apis mellifera* (Amel), *Nasonia vitripennis* (NV), *Pediculus humanis corporis* (Phum), and *Tribolium castaneum* (AAJJ) genomic sequences are reported in Xuan et al. (2015), Liu et al. (2017, 2019, 2020), and Picimbon (2020). Sequences for *Acyrthosiphon pisum* (ACYPI) are presented in Table S1 (Genome Assembly and Protein Identity) and contrasted with those for *Pediculus* (Phum) and *Bombyx* (Bmor). While *Pediculus* CSPs show higher identity to molecular orthologs in Diptera, Hymenoptera, and Coleoptera, *Acyrthosiphum* CSPs show more identity to Hemipteran orthologs. *Bombyx* CSPs, in contrast, show the greatest degree of resemblance to their molecular orthologs in the kingdoms of bacteria and lepidopterans. (Table S1). They are highly similar to Allergen Tha p 1 and PAN-1-like protein molecules (BmorCSP5, BmorCSP14, BmorCSP17, BmorCSP18, BmorCSP11, BmorCSP10, and BmorCSP16; following the arrangement of the genes in the genome, Xuan et al., 2015, Table S1), which is an interesting finding. Allergen Tha p 1 is connected to the molecular sequence of three truncated pseudogenes (BmorCSP5, BmorCSP16, and BmorCSP18). The protein gene family’s ancestral function, which has nothing to do with olfaction, may have been immune system activation.



**Figure 1.** Consensus amino acid alignment of CSP molecules from dipterans, lepidopterans, coleopterans, hymenopterans, homopterans, and neopterans. *Drosophila ananassae*, *D. erecta*, *D. grimshawi*, *D. melanogaster*, *D. mojavensis*, *D. persimilis*, *D. pseudoobscura*, *D. sechellia*, *D. simulans*, *D. virilis*, *D. yakuba*, *D. willistoni* (Droso), *Aedes aegypti* (AAEL), *Anopheles gambiae* (AgCG), *Culex pipiens* (CPIJ), *Bombyx mori* (Bmor), *Apis mellifera* (Amel), *Nasonia vitripennis* (NV), *Pediculus humanis corporis* (Phum), and *Tribolium castaneum* sequences are reported in Xuan et al. (2015), Liu et al. (2017, 2019, 2020), and Picimbon (2020). *Acyrthosiphon pisum* (ACYPI) sequences are reported in Table S1 (Acc. Num.). The amino acids in bold are those that are common to most CSPs. The amino acids that are strictly conserved throughout all CSPs are underlined. The four cysteine residues (C) that are unique to CSPs are marked by stars in their respective locations. The amino acid residues highlighted in italics are those that change in particular CSPs. The numbering of amino acid residues is determined by the identification of the N-terminal sequence of CSP proteins by Edman degradation (Picimbon and Leal, 1999; Picimbon et al., 2000b). The CSP genes' introns are consistently found after Lysine 45 (see arrow). The codon for amino acid 46 (Glu, Ser, Lys, Asn, or Asp) is broken up by the intron sequence. Additional introns' points of insertion are shown by the other arrows. As indicated by the squares, there are varying numbers of amino acid residues between Cys29-Cys37 and Cys56-Cys59 (6-8, 18-19, respectively). For *Culex* CPIJ002628, and *Nasonia* NV16080, an amino acid insertion (two residues) between Cys29-Cys37 has been found (Liu et al., 2020; Picimbon, 2020). *Locusta migratoria* OSDs have a similar amino acid insertion (Picimbon et al., 2000a,b; Picimbon, 2005). For *Nasonia* NV16076 and NV16077, *Apis* GB17875, and *Acyrthosiphon* ACYPI000345, an amino acid insertion (one residue) between Cys56-Cys59 has been found. The intercysteine gaps in *Drosophila*, *Anopheles*, *Aedes*, *Bombyx*, *Tribolium*, and *Pediculus* are strictly preserved. *Nasonia* in particular exhibits amino acid insertion (or deletion) in various regions of the CSP protein. The location of functional elements ( -helices) is indicated by the grey circles beneath the alignment. The disulfide bridge and interlocked cysteines are shown in red.

It's also noteworthy that the molecule contains different amounts of amino acid residues between Cys29-Cys37 and Cys56-Cys59 (6-8, 18-19, respectively; Figure 1). For *Culex* CPIJ002628, and *Nasonia* NV16080, an amino acid insertion (two residues) between Cys29-Cys37 has been found (Liu et al., 2020; Picimbon, 2003, 2005, 2019, 2020). *Locusta migratoria* OSDs have a similar amino acid insertion (Picimbon et al., 2000a; Picimbon, 2005). For *Nasonia* NV16076 and NV16077, *Apis* GB17875, and *Acyrthosiphon* ACYPI000345, an amino acid insertion (one residue) between Cys56-Cys59 has been found. The intercysteine gaps in *Drosophila*, *Anopheles*, *Aedes*, *Bombyx*, *Tribolium*, and *Pediculus* are strictly preserved. *Nasonia* in particular exhibits amino acid insertion (or deletion) in many various regions of the CSP protein (Figure 1). As with *B. mori* ribosome peptide changes, these motif insertion mutations in jewel wasp and *Culex* primarily affect the N-terminal region and the loop between  $\alpha 2$  and  $\alpha 3$  (Figure 1; Xuan et al., 2014, 2016; Picimbon, 2019; Yue et al., 2023). All CSPs have two disulfide bridges and four cysteines, which keep the molecule stable (Figure 1). However, mutations in the ribosome and genome are probably what give the CSP protein molecule family its versatility for a variety of biological functions.

## 2. Protein structure and RNA sequence adaptability

Because of the CSP structure's extreme flexibility, multifunctional features are heavily supported. RNA editing and/or post-translational changes, which were found in the silkworm moth *B. mori* are characteristics of CSPs (Xuan et al., 2014, 2015, 2016, 2019; Picimbon, 2017, 2019, 2023). The presence of recoding at the level of protein synthesis in the CSP family is clearly supported by the inclusion of a Glycine residue next to a Cysteine residue at certain positions, amino acid inversion, and motif insertion in protein sequence (Xuan et al., 2014, 2015, 2016, 2019; Picimbon, 2017, 2019, 2023). They can also 'breathe' or modify their conformation in certain ways in response to ligand binding, which may be another important characteristic of the ancestor multifunctional soluble binding protein (Campanacci et al., 2003; Mosbah et al., 2003).

## 3. Expression profiling in development, organisms, and tissues

CSP molecules are present in insects at all steps of their life cycles, from eggs and larvae through nymphs and adults (Picimbon, 2003, 2020; Picimbon et al., 2000a,b, 2001; Wanner et al., 2005). They are mostly expressed in the antennae, mouth, pedipalps, and legs of locusts, and they have been connected to phase shift (phenotypic plasticity) in those insects (Angeli et al., 1999; Picimbon et al., 2000a; Guo et al., 2011; Martín-Blázquez et al., 2018).

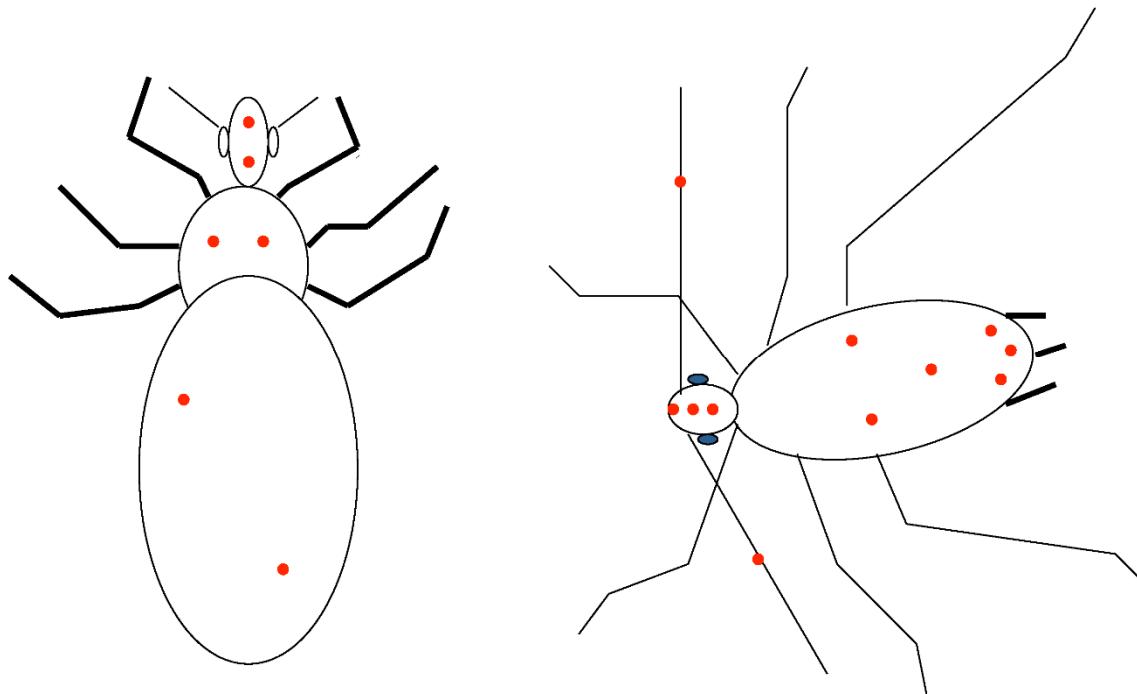
CSP molecules are not an insect's apnage. They are also expressed in a wide range of organisms, including many species of arthropods, crustaceans, shrimp, crab, lobster, and copepods (Zhu et al., 2019; Picimbon, 2023). They don't, however, only exist in arthropod species. They are clearly seen to exist in both eukaryotes and prokaryotes by the fact that they are very widely expressed at the level of the bacteria superkingdom (Liu & Picimbon, 2017; Picimbon, 2019; Liu et al., 2019; Picimbon, 2023; also see WP\_149730592 in multi-species). Prokaryote CSPs are twins or identical twins to insect CSPs (Liu et al., 2019; Picimbon, 2023; WP\_149730592), which is significant when discussing the function of these molecules. They have been reported on bacterial species, including *Lysobacter* and *Escherichia coli* (*E. coli*), *coccobacillus* *Acinetobacter baumannii*, *Macrococcus/Staphylococcus caseolyticus*, the filamentous actinomycete *Kitasatospora griseola*, the Actinobacteria genus in the families Entereobacteriaceae, Nocardioidaceae, Pseudonocardiaceae (*Solhabitans fulvus*), and Streptomycetaceae (Picimbon, 2023). According to RNA and genomic reports, CSP molecules are also found in firmicutes, aeromonadales, alteromonadales, eubacteriales (*Clostridium perfringens*), and hyphomicrobiales (MDK0835621, MDK0841570; Picimbon, 2023). Based on this, a chemosensing-related function is strongly disputed. 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. Acinetobacteria 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  $\mu\text{m}^3$ ) with very low GC (33%) is found in marine Actinobacteria (*Candidatus actinomarinidae*; Ghai et al. 2013). 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 (Zhao et al., 2019; Doré et al., 2023). *Clostridium perfringens* (formerly known as *C. welchii*, or *Bacillus welchii*) is known for  $\alpha$ -toxin, the toxin involved in gas gangrene (clostridial myonecrosis). It's possible that CSPs participate in quorum sensing, a process where actinobacterial microbes produce and react to signaling molecules (autoinducers) to detect the presence of picocyanobacteria in their environment. But in this case, quorum-sensing controlled activities like bioluminescence, virulence factor secretion (toxins, hemolysins, proteases), biofilm formation, sporulation, conjugation, secondary metabolism, plasmid transfer, and/or pigment production would only work if a number of bacteria carried the others in synchrony. This might not include molecules like volatile organic compounds (VOCs), volatile carboxylic acids (VACs), pyrazines, chemosensory signals, aggregation odors, cohesion or sex pheromones, which have only seldom been linked to insect-bacteria associations (Taga and Bassler, 2003; Silva-Junior et al., 2018). Even while it would be intriguing to investigate the variety and binding property of CSP proteins in axenic insects (which lack bacterial populations in the gut system), it is still improbable that fecal bacterial CSP molecules are involved in the transmission of pheromone signals. The specific binding of VOCs, VACs, or locust cohesion pheromones has not been demonstrated for any CSP proteins (Dillon et al., 2000; Wada-Katsumata et al., 2015). Numerous bacteria produce CSPs that are strictly identical to those found in silkworms, however, the silkworm CSPs are all widely distributed throughout the entire body of the insect, not just found in its sense organs. They experience a striking up-regulation in response to pesticide drug exposure (Xuan et al., 2015; Picimbon, 2023).

Similarly to fungus, some *Candidatus* bacterial species express rhizobial *Nod* factors such as the canonical *nod* gene *nodABC* in symbiosis with their host plant (Persson et al., 2015). These bacteria also fix nitrogen-producing *Nod* for the plant. Thus, a multi-purpose activator and transporter molecule like CSP, which could bind to DNA and lipids and transport lipids and carbohydrates to

bacteria, could be very helpful in the two organismal systems, activating rhizobial genes like *nod* through bacterial factors as multiple mechanisms of symbiotic interaction for adaptation and co-evolution (Siebers et al., 2016). Similar to this, Actinobacteria engage in numerous endosymbiotic interactions with insects in addition to interacting with plants and/or other microorganisms (Gopikrishnan et al., 2023). The insect gut bacterial communities may perform a variety of tasks, including cellulose degradation, nitrogen fixation, detoxification of protective plant compounds, pesticide degradation, inhibition of pathogens, and transport of nutrients, as shown by the gut bacteriome of *Ips* (Coleoptera, Curculionidae, Scolytinae; Chakraborty et al., 2020; Siddiqi et al., 2022). As a result, it is highly likely that bacteria, such as actinobacteria, share and exchange CSP molecule in the rhizobium of plants and the guts of insects for signaling, cell-cell communication, gene expression regulation, and/or the degradation of various toxins, all of which are necessary for the successful adaptation to specific ecological niches. Although this has not yet been experimentally demonstrated, the theory of horizontal transfer of DNA/RNA from bacteria to hosts strongly suggests that CSPs exist in plants (Han and Luan, 2015; Liu et al., 2016a; Zhu et al., 2016a; Lacroix and Citovsky, 2019). However, the widespread demonstration of CSPs in bacterial strains and insect guts across a wide range of species and genera suggests multi-duplication at a very early stage of evolution (Liu et al., 2020; Picimbon, 2023). To refer to an entire family with functions which are solely focused on chemical detection as 'CSP' is very incongruous.

'CSPs' are found in phyllosoma, in larval and adult stages, and in the adult stage, in many different non-sensory organs of many different crustacean species, including Antarctic copepod, crab, crayfish, lobster, salmon louse, prawn, shrimp, and water flea (Picimbon, 2023). This situation is pretty similar to how insect CSPs are being described right now. Insect venom can also contain CSPs, in addition to the *sensillum* (Angeli et al., 1999; Perkin et al., 2015). The sex pheromone gland in moths is where the majority of CSP molecules are expressed (Xuan et al., 2014, 2015; Picimbon, 2017). However, in addition to the female moth pheromone gland, CSP-expressing tissues and secretions also include the antennal branches, mouth, mandibles, salivae, proboscis, cephalic capsula, eyes, head, thorax, abdomen, epidermis, fat body, gut, wings and legs, i.e., a variety of reproductive and non-reproductive, sensory and non-sensory fluids, but in particular the majority of metabolic tissues (Celorio-Mancera et al., 2012; González-Caballero et al., 2013; Liu et al., 2014; Zhu et al., 2016b; Liu et al., 2020). EST-base data analysis in *Pediculus* and *Acyrthosiphon* (Figure 2) reveals distribution of CSP-RNAs throughout the entire body, in the antennae but also in the head, thorax, and abdomen, similar to Coleoptera, Diptera, and Hymenoptera (Liu et al., 2017, 2020; Picimbon, 2020). Six transcripts from first instar *P. humanus corporis* larvae and engorged adults were used to identify the CSPs in the body louse (Pedra et al., 2003; Kirkness et al., 2010; Liu et al., 2019). The International Aphid Genomics Consortium (2010) reported that the CSPs from the aphid *A. pisum* are also found at mixed stages, in the head and the antennae of third instar nymphs, and in winged and wingless parthenogenetic females inoculated with bacteria or treated with ampicillin for removal of pathogenic bacteria (Hunter et al., 2004; Stern et al., 2005; Richards et al., 2007; Shigenobu et al., 2010). Finding out how CSPs are expressed in different tissues can provide information on how they function. Pioneering Northern blot experiments, Western blotting, PCR, and advanced real-time PCR analysis reveal that moth CSP molecules are widely distributed throughout the body of the insect, including the head (epidermis, brain, and eyes), thorax, abdomen (gut and fat body), and wings in addition to the antennae, pheromone gland, and legs (Picimbon et al., 2000a,b, 2001; Xuan et al., 2014, 2015). 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 (Pittendrigh et al., 2015; Ollivier et al., 2019). Therefore, we "dissected" the honey bee, *Apis mellifera*, the silkworm moth, *Bombyx mori*, the jewel wasp, *Nasonia vitripennis*, the red flour beetle, *Tribolium castaneum*, the drosophilid dipterans *Drosophila ananassae*, *D. erecta*, *D. grimshawi*, *D. melanogaster*, *D. mojavensis*, *D. persimilis*, *D. pseudoobscura*, *D. sechellia*, *D. simulans*, *D. virilis*, *D. willistoni*, *D. yakuba*, and the mosquitoes, *Aedes aegypti*, *Anopheles gambiae*, and *Culex pipiens* Flybase and specific databases such as VectorBase or KAIKO/SilkDB (<http://flybase.org/blast>, <http://www.vectorbase.org>, <http://sgp.dna.affrc.go.jp>, <http://silkworm.genomics.org.cn>; Xuan et al., 2015; Liu et al., 2017, 2020;

Picimbon, 2020). The same methodology was utilized to analyze the aphid and *Pediculus* CSPs from Flybase and VectorBase (Figure 2). The EST-cDNA library from 21 insect species was analyzed, and the results revealed that the CSPs are widely expressed and are present in both chemosensory and non-chemosensory organs throughout the whole insect body (antennae, head, thorax, and abdomen, Figure 2). Following exposure to avermectin pesticide molecules, nearly all CSPs are up regulated in the majority of insect body tissues, especially in the gut, and fat body, which is crucial for repelling an olfactory function (Xuan et al., 2015).



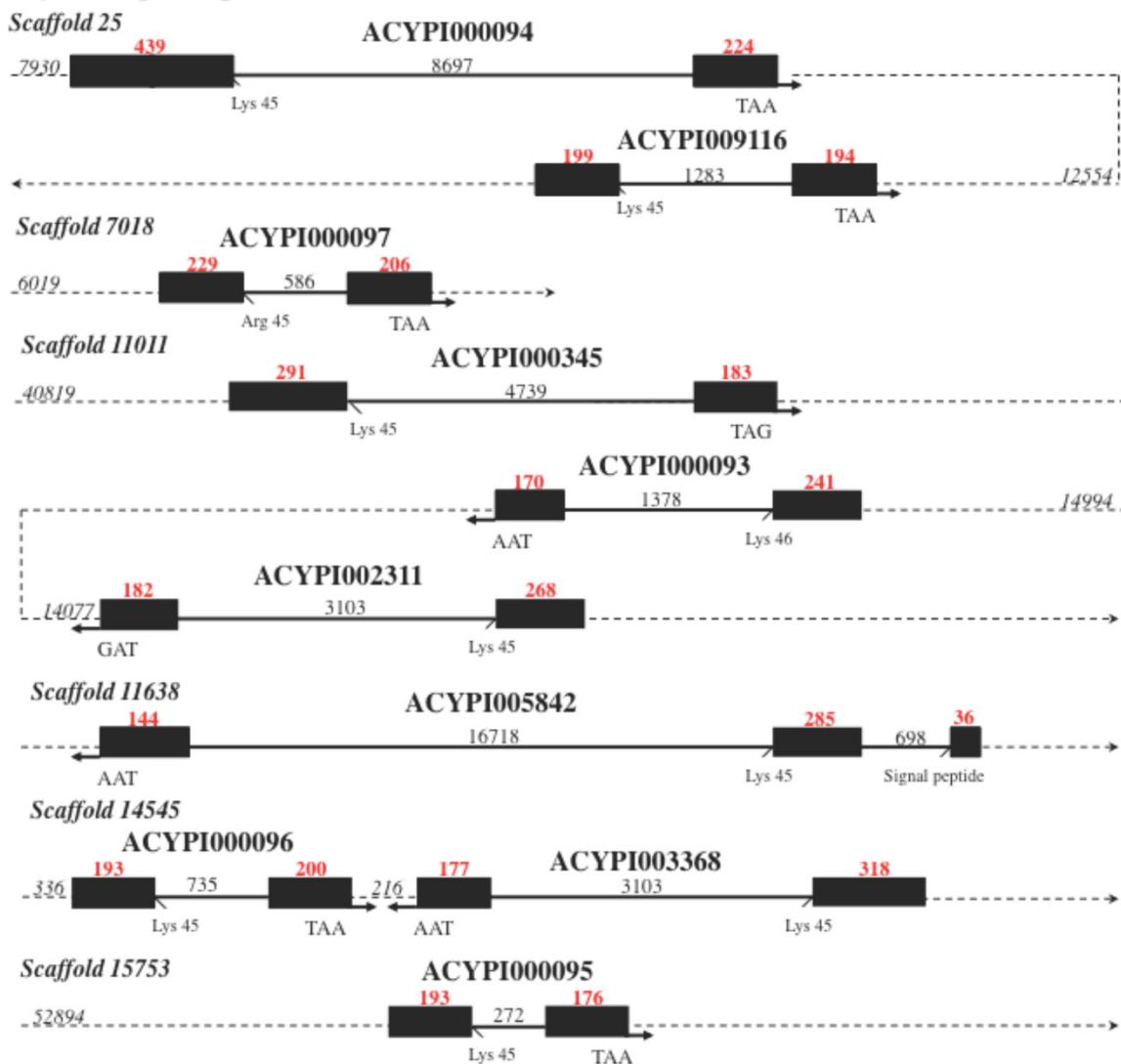
**Figure 2.** EST-based tissue expression of aphid and louse CSPs. From a BLASTn search of *Acyrthosiphon* and *Pediculus* EST database using FlyBase and VectorBase, EST-cDNAs encoding CSP are sorted (Table S1). The main location for CSPs is not the antennae. The head (cephalic capsula, brain, and eyes), thorax (prothoracic glands), and abdomen (gut and fat body) are the three primary site of expression for “CSPs”.

#### 4. The evolution and diversity of CSP genes

A function in chemical communication is also at odds with the diversity, evolution, quantity, and number of CSP genes. The 4-8 CSPs found in pediculus lice, honeybees, emerald cockroach (jewel) wasps, drosophila flies, and *anopheles* mosquitoes show that insects generally have very few CSP-coding genes (Picimbon, 2003, 2020, 2023; Wanner et al., 2004; Forêt et al., 2007; Liu et al., 2019, 2020). Does this imply that these insects have poor sense of smell or that they don't communicate chemically with one another? Instead, the small number of CSPs disproves their potential for chemical communication (4 in *Drosophila*). For chemical communication, pheromones, and mate recognition, flies are known to use a complex mixture of long chain epicuticular hydrocarbons (Savarit et al., 1999; Wicker-Thomas et al., 2015; Yew and Chung, 2015; Blomquist et al., 2018). 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 (Liu et al., 2019, 2020). According to Liu et al. (2020), CSPs are distinctly arranged in pairs of duplicates on individual chromosomes in bees. When examining the aphid CSP genes (ACYPI000094-ACYPI009116, ACYPI000093-ACYPI002311, and ACYPI000096-ACYPI003368), we find a similar distribution of duplication on particular scaffolds (Figure 3). CSPs may therefore cooperate to cause cellular function. These CSPs in aphids resemble DNA sequences that have been copied invertedly (Figure 3). They have a junction between them and are directed in opposing

directions. They are therefore anticipated to increase translocation rates (Spealman et al., 2020). The most prevalent type of chromosome rearrangement, inverted duplications, also known as foldback inversions, is frequently seen in the emergence of new phenotypes, including particular developmental features in the brain. Chromosomal rearrangement and the degree of synteny play a role in pesticide resistance in aphids, which has long been known (Blackman et al., 1978; Mandrioli et al., 2019). As seen in aphids and ants, chromosome rearrangements are undoubtedly the most dramatic type of mutation, frequently resulting in rapid evolution, adaptation, and speciation (Mathers et al., 2021; Vela et al., 2022). Although the aphid CSP genes lack any specific retroposon elements, the ACYPI000095 intron resembles the *Solanum lycopersicum* cultivar I-3 on chromosome 8 (CP023764). The I-3 gene is responsible for tomato's resistance to *fusarium* and increased susceptibility to bacterial spots, which may indicate that CSPs were transferred from aphids to plants, and boosted host resistance.

### *Acyrthosiphon pisum*



**Figure 3.** Organization of aphid CSPs on different scaffolds (25, 7018, 11011, 11638, 14545, and 15753). Genomic CSP gene repertoire in the pea aphid *Acyrthosiphon pisum* (see Table S1). Exons are shown as black boxes and introns as plain lines. The spacing between genes is indicated by dotted 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 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). Aphid CSPs have TAA or TAG stop codons, which are indicated by the stop codons in *A. pisum* ACYPI000094, ACYPI009116, ACYPI000097,

ACYPI000093, ACYPI005842, ACYPI000096, ACYPI003368, and ACYPI003368 (TAA), as well as ACYPI000345 and ACYPI002311 (TAG). .

The number of *CSPs* and retroposon elements in the genomes of butterflies, moths, and beetles is significantly higher (about 19-20 containing *SINE Bm1*, *Bm2*, *BMC1*, *BmRTE*, *L1Bm*, *MLE*, *Taguchi*, *Kendo*, and/or *Woot* retroposons; Ozaki et al., 2008; Xuan et al., 2015; Liu et al., 2017). With *Feilai* and *Wujin-Aa4/Wuneng* retroposons, *Culex* and *Aedes* mosquitoes have even more *CSP* genes, indicating a high rate of gene duplication (Picimbon, 2020). The number of *CSP* genes varies between 27 and 83 in different *Culex* mosquito species (Mei et al., 2018; Picimbon, 2020). *Aedes aegypti* mosquitoes still carry about 69 *CSP* genes (Picimbon, 2020). This will eventually lead to discussions about pheromones, chemical communication, and long chain epicuticular hydrocarbon fatty acid diversity in these mosquito species (Wang et al., 2019). However, it is important to remember that the majority of these genes are strictly identical copies of the same gene due to tandem duplication, transposition to new chromosomes, or whole-genome duplication (polyploidy; Picimbon, 2020). As a result, this probably reinforces the same phenotype or function at the molecular level rather than introducing a new function like the binding of a new fatty acid. Several copies of the same gene that encode for the same molecule and are located on the same chromosome surely have an impact on how the cell functions.

The *CSP* protein may only be able to achieve "chemosensing" from such a limited set of genes through splicing, RNA editing, and protein recoding (Picimbon, 2019). Through retrotransposition, translocation, duplication, post-translational modifications and/or RNA + peptide editing mutation, *CSPs* can produce over hundreds of protein variants, as shown in *Dscam* and cochlear sensory genes of the bird auditory system (Neves et al., 2004; Xuan et al., 2014, 2015, 2016, 2019; Picimbon, 2017, 2019, 2023; Hou et al., 2022; Liu et al., 2023; Yue et al., 2023). Therefore, assuming that these mechanisms are limited to the antennae or sensory organs, which remains to be proven (see Xuan et al., 2014), it would be interesting to investigate the change of function in relation to chemosensing after gene manipulation, DNA splicing, RNA editing, and/or protein recoding and molecular changes through ribosome peptide mutations.

The *CSP* genes in aphids are composed of two exons separated by a single, variable-length intron, just like in most other insect species (Figure 3; Xuan et al., 2015, Liu et al., 2016a, 2017, 2019, 2020; Picimbon, 2019, 2020). A few nucleotides following the start codon that codes for the amino acid methionine, one more intron is inserted in ACYPI005842 and genes like *AAJJ1196A*, *BmorCSP19*, *GB19453* and *PHUM594410*. This intron (phase 0 intron) is inserted after the third base and does not disrupt the codon. This demonstrates the tight regulation of the splicing of the signal peptide region and supports the notion put forth by Blobel (2000) that the functional significance of a signal peptide molecule is directly related to its length. The first six amino acids could be the needle tip that breaks through the cellular or subcellular membrane. The other amino acid residues might be crucial for the localization and/or transit of particular molecular partners from the Signal Recognition Particle complex within the cell (Matlin, 2002; Doudna and Batey, 2004). It's interesting to note that ACYPI genes like ACYPI000094, ACYPI000345 and ACYPI003368 that have large signal peptides (large exon1 of 292-439 nucleotides) are connected to large intron (3103-8697 nucleotides; Figure 3). The length of exon 1 and its flanking intron in aphids vary together. This is a fundamental pattern regarding exon and intron motif differences in eukaryotic genomes (Zhu et al., 2009). This is a fundamental pattern regarding the arrangement of exons and introns in insect *CSP* genes (Xuan et al., 2015; Liu et al., 2016a, 2017, 2019, 2020; Picimbon, 2020; Figure 3). However, in *CSP* genes, the length of the flanking intron is not always correlated with the length of exon2. While the exon2 of the gene *AAJJ0269C* is enormous at 532 nucleotides, its intron length is 5063 bps. In contrast, the exon2 of the genes *CPIJ002608*, which has 312 nucleotides, and *AgCG50174*, which has 317 nucleotides, has intron lengths of only 77 and 97 bps, respectively, in their introns (Liu et al., 2017; Picimbon, 2020). It seems that in flies, an increase in average exon2 length is correlated with a decrease in intron length. The exon2 of the *Drosophila* genes *DmelOSD*, *GD12430*, *GM24353*, *GE22174*, *GG15834*, *GA19747*, *GL15312*, *GK24936*, *GI14872*, *GJ19536*, and *GH12519* measures 214-299 bps, and the intron measures 52-191. The *Drosophila ananassae* *GF20186* gene that is associated with these fly genes is composed of a single exon only. A conserved exon2 (143 bps) and an intron length of 52-436 bps are found in a number of genes,

including CG30172, GD24980, GM18223, GE11495, GG19963, GA10970, GL10431, GK19383, GI20684, GJ20437, GH20166, and GF12039. This indicates that *Drosophila* CSP genes develop through exon1/intron deletion and exon2 expansion (Picimbon, 2020). Whether structural and functional properties of the molecule DanaGF20186 (exon2) remain, and whether these properties can be expanded through RNA editing and/or peptide mutation, would be an intriguing avenue to investigate.

Three exons and two introns are found to be common among genes encoding moth pheromone-binding proteins (Vogt et al., 2002; Picimbon, 2003; Abraham et al., 2005), in the *Acyrthosiphon* gene ACYPI005842, AAEL012383 (*Aedes*), AgCG50175 (*Anopheles*), GB19453 (*Apis*), NV16079 (*Nasonia*), PHUM594410 (*Pediculus*), AAJJ1796 (*Tribolium*), and three *Bombyx* genes, *BmorCSP4*, *BmorCSP10*, and *BmorCSP19* (Figure 3; Xuan et al., 2015; Liu et al., 2017, 2019, 2020; Picimbon, 2020). Different positions are used to introduce the extra intron. The additional intron follows Ala82 in PBPs (Abraham et al., 2005). It is inserted into the nucleotide region that codes for the signal peptide in ACYPI003368, AAJJ1796A, *BmorCSP19*, CPIJ017094, GB19453, and PHUM594410. It is inserted at the codon that codes for Arg87 in *BmorCSP4* while it is inserted at the codon that codes for Arg8, Arg76, and Ser115 in NV16079, AgCG50175, and *BmorCSP10*, respectively. This implies that the two exons and three introns genes in the “CSP” gene family, despite their structural resemblance, did not originate from the duplication of a common ancestor but rather from independent intron insertions that happened in different insect species during their evolutionary history. The *Aedes* gene AAEL001957 and the *Culex* gene CPIJ002607 both have notable arrangements and dizzying sizes. At the codons for the signal peptide residues, Arg9, Lys46, Ala165, Ser190, and Pro298 in that order, introns build up (Picimbon, 2020).

The fact that the introns in ACYPI000094, ACYPI009116, ACYPI000097, ACYPI000345, ACYPI000093, ACYPI002311, ACYPI000096, ACYPI003368, ACYPI000095, AAJJ0012A-I, AAJJ0283A, AAJJ0283B, ASP3c, GB10389, GB13325, GB19242, GB17875, PHUM594420, PHUM594430, PHUM594540, PHUM594550, PHUM594660, OS-D, CG30172, GD24980, GD12430, GM18283, GM24353, GE11495, GE22174, GG19963, GG15834, GF12039, GA15697, GA19747, GL10431, GL15312, GK19383, GK24936, GI20684, GI14872, GJ20437, GJ19536, GH20166, GH12519, AgCG50174, AgCG50200, AgCG50208, AAEL001963, AAEL001985, CPIJ002608, CPIJ002605, NV16075, NV16076, NV16077, NV16080, NV16079, NV16108, NV16109, *BmorCSP1*, *BmorCSP2*, *BmorCSP3*, *BmorCSP6*, *BmorCSP7*, *BmorCSP8*, *BmorCSP9*, *BmorCSP11*, *BmorCSP12*, *BmorCSP13*, *BmorCSP15*, *BmorCSP17*, *BmorCSP20*, and all other single-intron CSP genes vary in length but are always found in the same place in aphids, beetles, bees, body louses, flies, mosquitoes, moths, and wasps suggests that all these genes share a very ancient common ancestor (Xuan et al., 2015; Liu et al., 2017, 2019, 2020; Picimbon, 2019, 2020, 2023; Figure 3). 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, Homoptera, Hymenoptera, Lepidoptera, and Neoptera. There is a very long history of the CSP molecule. After that, every grouping started to exhibit unique patterns, maybe as a result of the appearance of unique phenotypic and/or functional traits (Liu et al., 2020).

Intron loss and gain, retrotransposition, RNA editing + retrotransposition, and duplication were all factors in the evolution of CSP genes (Picimbon, 2003, 2019, 2020, 2023; Forêt et al., 2007; Kulmuni et al., 2013; Xuan et al., 2015, 2019; Liu et al., 2017). In moths, a single unifying hypothesis of RNA editing and retrotransposition-driven evolution of CSPs has recently been put out (Xuan et al., 2019). This hypothesis states that new CSP protein motifs are initially produced via DNA- and RNA-dependent RNA polymerization before edited CSP-RNA variations are retrotransposed in the genome. This evolutionary route might be able to explain the new abilities, skills, and tasks that the CSP molecule developed.

## 5. Multiple functions and binding properties

The specific binding to linoleic acid and the broad pattern of gene expression across a wide range of sensory and non-sensory fluids or tissues strongly suggest that this protein gene family plays a very general basic role in long chain fatty acid transport, lipid exchange, and fat metabolism (Liu et al., 2016b, 2017, 2020; Picimbon, 2023).

Liu et al. (2014, 2016b) and Xuan et al. (2015) demonstrated a drastic and remarkable up-regulation of *CSP* genes in many different tissues over exposure to abamectin/neonicotinoid insecticide molecule, raising the possibility that *CSP* proteins play a role in general immunity, insecticide resistance, and xenobiotic degradation. Following microbial or viral infection, an increased load of *CSPs* ('*pherokines*') is seen in fly hemolymph (Sabatier et al., 2003). Liu et al. (2014, 2016b, 2020) have discussed the specific function of *CSPs* in lipid transport in relation to insecticide resistance using the whitefly *Bemisia tabaci* as a model study (Liu et al., 2016b). Insecticide-mediated upregulation and interaction of the protein (BtabCSP1) with long chain C18-lipid (C18:2, linoleic acid, LA) were demonstrated by Liu et al. (Liu et al., 2016b), pointing to a metabolic role for *CSP* in insect defense and immunity as opposed to olfaction or chemical communication. BtabCSP1 linked to LA is expressed in many diverse tissues of the insect body (Liu et al., 2016b). LA has never been found in insect fingerprints, which is proof that *CSPs* are not involved in pheromone chemical communication.

Nomura et al. (1982) described the first member of this small soluble protein family as an up-regulated factor (*p10*) in the regenerating legs of the American cockroach *Periplaneta americana*. The same protein (*p10*) was found in the antennae and legs of *P. americana* at the adult sexually mature stage, with some differences between males and females. This finding rather suggests a "chemodevol" function for this protein, contributing to tissue development and recognition of sex-specific signals like odor/odorant sexual pheromones (Picimbon and Leal, 1999). One (polyclonal) antibody against *CSP* labeled the antennal sensillum in immunocytochemistry experiments, but the labeling was diffused to the cuticle and supporting cells as well as sensory structures (Angeli et al., 1999; Jin et al., 2005). It must be clarified how signals travel from supporting cells heated internally to sensory dendrites bathing the sensory lymph, and/or how these *CSP* protein molecules can develop in an extracellular environment from an intracellular one. The purpose of the labeled sensilla also needs to be made clear. The dendritic branching of coeloconica sensilla, peg-in-pit sensilla, thermoreceptor, hygroreceptor, infrared receptor, and acoustic receptor is comparable to that of chemosensory pheromone sensilla (Schneider et al., 2018).

A role for *CSPs* in lipid transport is consistent with their critical involvement in honeybee head development, insect general immunity, moth pheromone synthesis or behavioral changes in locusts (Maleszka et al., 2007; Guo et al., 2011; Liu et al., 2014, 2016b, 2020; Xuan et al., 2014, 2015; Picimbon, 2023).

According to Ozaki et al. (2005), 'CSP proteins' have been proposed to mediate the recognition of chemical signatures made up of cuticular lipids, such as those found, for example, in ants. The role of each of these *CSPs* linked to lipids in chemical communication, development, or other physiological processes is not yet apparent. Fatty acid lipids (FA) like linoleic acid (LA) bind to the functional *CSP* structure (Lartigue et al., 2002; Liu et al., 2016b). Chemical communication is not a part of LA. Most insect species cannot generate LA on their own, much like crustaceans cannot. Insecticide chemical compounds (cinnamaldehydes) generated from plant oils, such as "dangerous" toxic chemicals, can interact directly with other functional *CSP* structures, as has been shown (Liu et al., 2016b). To further describe the role of *CSPs*, two further points are crucial: *CSPs* are expressed in both arthropods and bacteria and seem to have a variety of biological roles. When injected into the plant's phloem, *CSPs* in particular can activate specific innate immune pathways (Bos et al., 2010; Rodriguez et al., 2014). Water, carbohydrates, and other soluble organic substances are transported and distributed through the vascular tissue known as phloem. Food and nutrients produced by photosynthesis in the leaves are transported through the phloem. It has a structural purpose in the plant body and is a channel for many signaling chemicals (hormones). Sieve elements, parenchyma, and sclerenchyma, are the three cell types that normally make up this tissue, and it is through these tissues that *CSPs* such as Mp10 can elicit an immunological response rather than OR activation. It becomes significant to note that in this case, Mp10 shares a great deal of similarities with other

proteins, including actin skeleton regulatory protein, Arp2/3 complex activator, peplos-cell cross-linker, stress response initiator, cell wall/envelope protein, Mucin-like, Extensin-like, PAN, splicing regulator, nuclear pore complex protein, WAS/WASL-like, Rho GTPase activator, Sec31, UL36, and many others (30.71-45.13% identity; Table 1, Figures 4 & S1-S2). These results raise interesting questions about the possible functions of "CSPs" in intracellular compartments (cytoplasm and nucleus) and/or near the cytoplasmic membrane, RNA/DNA binding, interactions with genetic components, splicing regulation, transcription control, and activation of specific protein complexes.

The fact that LA binds to whitefly CSPs, that CSPs are present in many non-sensory tissues, that CSPs are found in bacteria and prokaryotic microbial cells, and that the *Myzus* CSP Mp10 and many other "intracellular" regulatory elements are related cast doubt on the claim that CSPs are "chemosensory" proteins. If we compare Mp10 to peplos proteins and outer surface membrane proteins, they even appear to be expressed by viruses (see Table 1).

**Table 1.** Sequences producing significant alignments with Mp10 protein (query length 153 amino acids; e-value 2e-107 to 3e-31). From highest identity % protein to query to lowest identity percentage protein, the sequences are listed (NCBI BLAST, blastp, nr database, 24<sup>th</sup> August 2023). OIB81003 and OIC85870 are from microbial protein blast (Picimbon, 2023). \* The asterisk denotes that an antennal clone was used to obtain the sequence. There was no tissue-distribution research done. f Results of the real-time PCR tissue distribution analysis test (Xuan et al., 2015). Lack of information leaves a void under Tissue and Function.

| Locus        | mRNA bps | Access Number (R) | Protein aa (kDa) | Access Number (P) | Species                    | Identity                         | Tissue  | Function                          |
|--------------|----------|-------------------|------------------|-------------------|----------------------------|----------------------------------|---|-----------------------------------|
| LOC111036095 | 459      | XM_022317999      | 153 (17.2)       | XP_022173691      | <i>Myzus persicae</i>      | Mp10<br>"CSP4" Ebsp-3            | Whole body<br>Asexual adult                             | Innate induction<br>Immunity      |
| BABH01021709 | 633      | NM_001043604      | 211 (24.6)       | AK368835          | <i>Bombyx mori</i>         | "BmorCSP10"                      | Antennae Legs<br>Fat body Gut<br>Epidermis <sup>f</sup> | Immune response <sup>f</sup>      |
| LOC126581753 | 2492     | XM_050245629      | 351 (37.7)       | XP_050101586      | <i>Anopheles aquasalis</i> | Mucin-2<br>Variant X6            |   | Translation initiation factor     |
| LOC109621340 | 2802     | XM_029870982      | 408 (43.4)       | XP_029726842      | <i>Aedes albopictus</i>    | SamkC<br>Variant X9              | All body<br>Pupae                                       | Serine-Threonine protein kinase   |
| LOC129745337 | 1986     | XM_055738352      | 364 (38.8)       | XP_055594327      | <i>Uranotaenia lowii</i>   | Mucin C.1<br>Variant X6          | Integument<br>Whole body                                | Translation initiation factor     |
| LOC115260417 | 1257     | XM_029861367      | 418 (45.8)       | XP_029717227      | <i>Aedes albopictus</i>    | Mucin-2                          | All body<br>Pupae                                       | Translation initiation factor     |
| LOC126581753 | 2693     | XM_050245628      | 418 (45.6)       | XP_050101585      | <i>Anopheles aquasalis</i> | Mucin-2<br>Variant X5            |   | WAS-WASL-interacting protein      |
| LOC118459415 | 2720     | XM_035922756      | 418 (46.2)       | XP_035778649      | <i>Anopheles albimanus</i> | Extensin-like<br>Variant X5      | Whole body<br>Pupae                                     | Cell wall component Defense       |
| LOC118459415 | 2909     | XM_035922754      | 418 (52.7)       | XP_035778647      | <i>Anopheles albimanus</i> | Extensin-like<br>Variant X7      | Whole body<br>Pupae                                     | Leucine-rich repeat Extensin-like |
| LOC129952650 | 2022     | XM_056065396      | 361 (39.4)       | XP_055921371      | <i>Eupeodes corollae</i>   | Rho GTPase-activating protein X3 |   | Rho GTPase-activator              |
| LOC109621340 | 3093     | XM_029870976      | 505 (54.4)       | XP_029726836      | <i>Aedes albopictus</i>    | Mucin-2 Variant X3               | All body<br>Pupae                                       | Nuclear pore protein complex      |
| LOC126564444 | 1245     | XM_050221501      | 401 (44.3)       | XP_050077458      | <i>Aedes maculipalpis</i>  | Mucin-2<br>Variant X4            |   | Translation initiation factor     |

|              |      |              |            |              |                             |                          |                       |                                       |            |
|--------------|------|--------------|------------|--------------|-----------------------------|--------------------------|-----------------------|---------------------------------------|------------|
| LOC118513797 | 2611 | XM_036059988 | 400 (43.9) | XP_035915881 | <i>Aedes stephensi</i>      | Mucin-2 Variant X3       | Whole body            | Translation factor                    | initiation |
| LOC118459415 | 2461 | XM_035922752 | 547 (56.2) | XP_035778645 | <i>Aedes albimanus</i>      | Mucin-5AC Variant X2     | Whole body Pupae      | Extensin-like                         |            |
| LOC109621340 | 3084 | XM_029870977 | 502 (54.8) | XP_029726837 | <i>Aedes albopictus</i>     | Mucin-2 Variant X4       | All body Pupae        | Nuclear pore protein complex          |            |
| LOC129744111 | 2632 | XM_055736495 | 496 (50.1) | XP_055592470 | <i>Uranotaenia lowii</i>    | Mucin-2 Variant X3       | Whole body            | Nuclear pore protein complex          |            |
| LOC109621340 | 3318 | XM_029870961 | 580 (62.3) | XP_029726821 | <i>Aedes albopictus</i>     | Mucin-2 Variant X1       | All body Pupae        | Nuclear pore protein complex          |            |
| LOC120904750 | 2951 | XM_040315040 | 493 (54.4) | XP_040170974 | <i>Aedes arabiensis</i>     | Extensin-like Variant X5 |                       | Extensin-like                         |            |
| LOC118459415 | 2921 | XM_035922753 | 493 (51.4) | XP_035778646 | <i>Aedes albimanus</i>      | Mucin-2 Variant X3       | Whole body Pupae      | Extensin-like                         |            |
| LOC121596701 | 3034 | XM_041921881 | 452 (48.7) | XP_041777815 | <i>Anopheles merus</i>      | Mucin-5AC Variant X7     | Whole body            | Protein transport protein Sec31-like  |            |
| LOC121596701 | 3371 | XM_041921880 | 564 (59.6) | XP_041777814 | <i>Anopheles merus</i>      | Mucin-2 Variant X2       | Whole body            | Mucin5AC-like                         |            |
| LOC121596701 | 3377 | XM_041921874 | 566 (59.8) | XP_041777808 | <i>Anopheles merus</i>      | Mucin-2 Variant X1       | Whole body            | Mucin5AC-like                         |            |
| LOC121596701 | 3022 | XM_041921883 | 448 (47.8) | XP_041777817 | <i>Anopheles merus</i>      | Mucin-2 Variant X5       | Whole body            | Mucin5AC-like                         |            |
| LOC121596701 | 3028 | XM_041921882 | 450 (47.9) | XP_041777816 | <i>Anopheles merus</i>      | Mucin-5AC Variant X8     | Whole body            | Mucin5AC-like                         |            |
| LOC120904750 | 3137 | XM_040315039 | 555 (60.8) | XP_040170973 | <i>Anopheles arabiensis</i> | Extensin-like Variant X4 | Male                  | Extensin-like                         |            |
| LOC129745337 | 2346 | XM_055738346 | 483 (51.0) | XP_055594321 | <i>Uranotaenia lowii</i>    | Mucin-C1 Variant X1      | Integument Whole body | Mucin-C1-like                         |            |
| LOC131210824 | 2638 | XM_058204117 | 386 (42.7) | XP_058060100 | <i>Anopheles bellator</i>   | Mucin-2 Variant X1       | Tegument Whole body   | Protein UL36                          |            |
| LOC124638351 | 1860 | XM_047175293 | 292 (32.1) | XP_047031249 | <i>Helicoverpa zea</i>      | PAN1-like                | Whole body Pupae      | Actin cytoskeleton-regulatory complex |            |
| LOC128303246 | 2787 | XM_053040137 | 472 (51.2) | XP_052896097 | <i>Anopheles moucheti</i>   | Mucin-2                  |                       | Mucin-like                            |            |

| Variant X1    |      |              |            |              |                                 |                          |                                   |
|---------------|------|--------------|------------|--------------|---------------------------------|--------------------------|-----------------------------------|
| LOC118513797  | 2824 | XM_036059983 | 470 (51.2) | XP_035915876 | <i>Anopheles stephensi</i>      | Extensin-like Variant X1 | Whole body                        |
| Scaffold00198 | 878  | JH668476     | 288 (32.1) | KAG6454812   | <i>Manduca sexta</i>            | Hypothetical             | Leucine-rich repeat Extensin-like |
| LOC118513797  | 2774 | XM_036059986 | 454 (49.5) | XP_035915879 | <i>Anopheles stephensi</i>      | Mucin-like Variant X2    | Whole body                        |
| LOC126374415  | 3912 | XM_050021059 | 298 (32.7) | XP_049877016 | <i>Pectinophora gossypiella</i> | YLP motif-1              | proteinWhole body Larvae          |
| LOC126564444  | 2774 | XM_050221498 | 554 (59.7) | XP_050077455 | <i>Anopheles maculipalpis</i>   | Mucin-2 Variant X1       | Mucin-2-like                      |
| LOC118459415  | 3326 | XM_035922748 | 628 (66.0) | XP_035778641 | <i>Anopheles albimanus</i>      | Mucin-2 Variant X1       | Whole body Pupae                  |
| LOC126564444  | 1410 | XM_050221500 | 456 (50.0) | XP_050077457 | <i>Anopheles maculipalpis</i>   | Mucin-like Variant X2    | Mucin-2-like                      |
| Taxon93504    | 651  | LC027715     | 217 (24.0) | BAV56819     | <i>Ostrinia furnacalis</i>      | “CSP15”                  | Antennae*                         |
| LOC128271871  | 1721 | XM_053009541 | 396 (43.6) | XP_052865501 | <i>Anopheles cruzii</i>         | WAS/WASL                 | Actin skeleton-regulatory complex |
| LOC128677457  | 1206 | XM_053758324 | 301 (33.6) | XP_053614299 | <i>Plodia interpunctella</i>    | Formin-1-like            | Whole body Larvae                 |
| LOC120904750  | 3488 | XM_040315029 | 672 (72.6) | XP_040170963 | <i>Anopheles arabiensis</i>     | Extensin-like Variant X1 | Extensin-like                     |
| LOC111349097  | 1268 | XM_022960083 | 287 (31.8) | XP_022815851 | <i>Spodoptera litura</i>        | PAN-1                    | Whole body                        |
| LOC5573135    | 2771 | XM_021847563 | 482 (51.4) | XP_021703255 | <i>Aedes aegypti</i>            | DAN4 Variant X4          | Actin skeleton-regulatory complex |
| LOC126581753  | 2885 | XM_050245624 | 482 (52.0) | XP_050101581 | <i>Anopheles aquasalis</i>      | Mucin-2 Variant X1       | Whole body Pupae                  |
| LOC120904750  | 3470 | XM_040315038 | 666 (71.5) | XP_040170972 | <i>Anopheles arabiensis</i>     | Extensin-like Variant X9 | Nuclear pore protein complex      |
| LOC120904750  | 3476 | XM_040315037 | 668 (71.7) | XP_040170971 | <i>Anopheles arabiensis</i>     | Mucin-2 Variant X2       | Male                              |
| LOC5573135    | 3059 | XM_021847553 | 576 (71.5) | XP_021703245 | <i>Aedes aegypti</i>            | NPCP Variant X1          | Mucin-like                        |

|                |      |                 |            |              |                                    |                            |                                |                                     |
|----------------|------|-----------------|------------|--------------|------------------------------------|----------------------------|--------------------------------|-------------------------------------|
| LOC120955660   | 3533 | XM_040376732    | 667 (71.4) | XP_040232666 | <i>Anopheles coluzzii</i>          | Mucin-2 Variant X1         | Mucin-like                     |                                     |
| LOC106137243   | 1976 | XM_021847553    | 305 (34.0) | XP_013193496 | <i>Amyelois transitella</i>        | WASP-2                     | Larvae                         | Actin skeleton-regulatory complex   |
| LOC6034358     | 2296 | XM_038253956    | 402 (43.3) | XP_038109884 | <i>Culex quinquefasciatus</i>      | DAN4 Variant X4            | Whole body                     | Cell wall protein                   |
| LOC106102595   | 1190 | XM_013282121    | 294 (32.9) | XP_013137575 | <i>Papilio polites</i>             | GP Ib-like -chain          | Outer surface membrane protein |                                     |
| LOC120427056   | 2533 | XM_052707633    | 461 (49.5) | XP_052563593 | <i>Culex pipiens pallens</i>       | Mucin-2 Variant X5         | Whole body                     | Mucin-like                          |
| M_scaff_1411   | 820  | CAKXAJ010025123 | 269 (30.3) | CAH2235359   | <i>Pararge aegeria aegeria</i>     | Jg5928                     |                                | Major outer envelope protein        |
| LOC108904830   | 1520 | XM_018707509    | 576 (32.8) | XP_018563025 | <i>Anoplophora glabripennis</i>    | CWA-3 Variant X1           | Larvae                         | Cell wall integrity Stress response |
| LOC125234523   | 2861 | XM_048140790    | 287 (32.0) | XP_047996747 | <i>Leguminivora glycinivorella</i> | RickA-like Variant X2      | Whole body Larvae              | Arp2/3 complex-activating protein   |
| LOC128271871   | 2177 | XM_053009537    | 548 (59.0) | XP_052865497 | <i>Anopheles cruzii</i>            | Mucin-2 Variant X1         | Tegument                       | Virus-tegument cross-linker         |
| LOC6034358     | 2697 | XM_038253951    | 535 (58.9) | XP_038109879 | <i>Culex quinquefasciatus</i>      | Mucin-2 Variant X1         | Whole body                     | Cell wall protein                   |
| LOC123011164   | 1400 | XM_044408485    | 239 (26.3) | XP_044264420 | <i>Tribolium maedens</i>           | Allergen Tha p1 Variant X2 | Whole body                     | IgE binding protein                 |
| LOC116772069   | 1195 | XM_032664103    | 297 (32.9) | XP_032519994 | <i>Danaus plexipus plexipus</i>    | WAS/WASL                   |                                | Actin skeleton-regulatory complex   |
| LOC129727069   | 2801 | XM_055684473    | 625 (66.2) | XP_055540448 | <i>Wyeomyia smithii</i>            | Mucin-2 Variant X1         | Whole body                     | Mucin-like                          |
| LOC129727069   | 2580 | XM_055684478    | 553 (58.5) | XP_055540453 | <i>Wyeomyia smithii</i>            | Mucin-2 Variant X3         | Whole body                     | Mucin-like                          |
| LOC122855961   | 1243 | XM_044157672    | 128 (14.5) | XP_044013607 | <i>Aphidius gifuensis</i>          | Allergen Tha p1            | Adults                         | IgE binding protein                 |
| LOC129727069   | 2476 | XM_055684480    | 518 (54.9) | XP_055540455 | <i>Wyeomyia smithii</i>            | Mucin-2 Variant X5         | Whole body                     | Mucin-like                          |
| Scaffold_15245 | 372  | CH916367        | 122 (14.0) | GH22042      | <i>Drosophila grimshawi</i>        | Pherokine-3-PC             | Embryo                         | Immunity                            |

|                |      |                 |            |              |                                  |                            |                               |
|----------------|------|-----------------|------------|--------------|----------------------------------|----------------------------|-------------------------------|
|                |      |                 |            |              |                                  |                            | Pupae                         |
| LOC6532152     | 607  | XM_002092892    | 239 (13.6) | XP_002092928 | <i>Drosophila yakuba</i>         | Ebsp-3/PebIII              | Ejaculatory bulb/duct         |
| LOC122859060   | 559  | XM_044162408    | 128 (14.2) | XP_044018343 | <i>Aphidius gifuensis</i>        | Allergen Tha p1            | Adults                        |
| LOC125060609   | 487  | XM_047665563    | 121 (14.2) | XP_047521519 | <i>Pieris napi</i>               | Allergen Tha p1            | IgE binding protein           |
| LOC123307467   | 1891 | XM_044889794    | 457 (50.7) | XP_044745729 | <i>Coccinella septempunctata</i> | Eukaryotic TIF Variant 4   | Translation initiation factor |
| Taxon80765     | 387  | KJ451426        | 128 (14.4) | AHX71992     | <i>Aphis gossypii</i>            | Acid trehalase "CSP3"      | Fat body                      |
| LOC123712760   | 465  | XM_045666008    | 239 (26.3) | XP_045521964 | <i>Pieris brassicae</i>          | Allergen Tha p1            | IgE binding protein           |
| Scaffold00027  | 846  | JANFCV010000027 | 282 (31.4) | KAI5642933   | <i>Phthorimaea operculella</i>   | A10/OS-D domain            | Adults                        |
| Scaffold133    | 393  | JAANIC010002949 | 131 (14.7) | KAG5343447   | <i>Acromyrmex charruanus</i>     | PEB3                       | Whole body                    |
| LOC124166051   | 721  | XM_046543645    | 131 (14.5) | XP_046399601 | <i>Ischnura elegans</i>          | Allergen Tha p1 Variant X2 | IgE binding protein           |
| PHUM594540     | 423  | XM_002432550    | 140 (15.5) | XP_002432595 | <i>Pediculus corporis</i>        | hum. Ebsp-3/PebIII         | Larvae                        |
| LOC123010695   | 476  | XM_044407769    | 124 (14.4) | XP_044263704 | <i>Tribolium maedens</i>         | Allergen Tha p1            | Whole body                    |
| Chromosome 2R  | 339  | CP012524        | 113 (13.1) | ALC42649     | <i>Drosophila busckii</i>        | PebIII                     | Pupae                         |
| Whole body     |      |                 |            |              |                                  |                            | Larva                         |
| Dmel(CG9358)   | 363  | NM_001299941    | 121 (13.4) | NP_001286870 | <i>Drosophila melanogaster</i>   | Pherokine-3 Variant B      | Embryonic gonads              |
| Scaffold_15245 | 366  | CH916367        | 122 (13.8) | EDW02527     | <i>Drosophila grimshawi</i>      | GH22042/OS-D               | Immunity                      |
| LOC107225040   | 553  | XM_015665345    | 127 (14.1) | XP_015520831 | <i>Neodiprion lecontei</i>       | Allergen Tha p1 Variant X1 | Thorax Abdomen                |
| LOC124186858   | 574  | XM_046578909    | 127 (14.2) | XP_046434865 | <i>Neodiprion fabricii</i>       | Allergen Tha p1            | Thorax Abdomen                |
| LOC124166051   | 708  | XM_046543644    | 131 (14.2) | XP_046399600 | <i>Ischnura</i>                  | Allergen Tha p1            | IgE binding protein           |

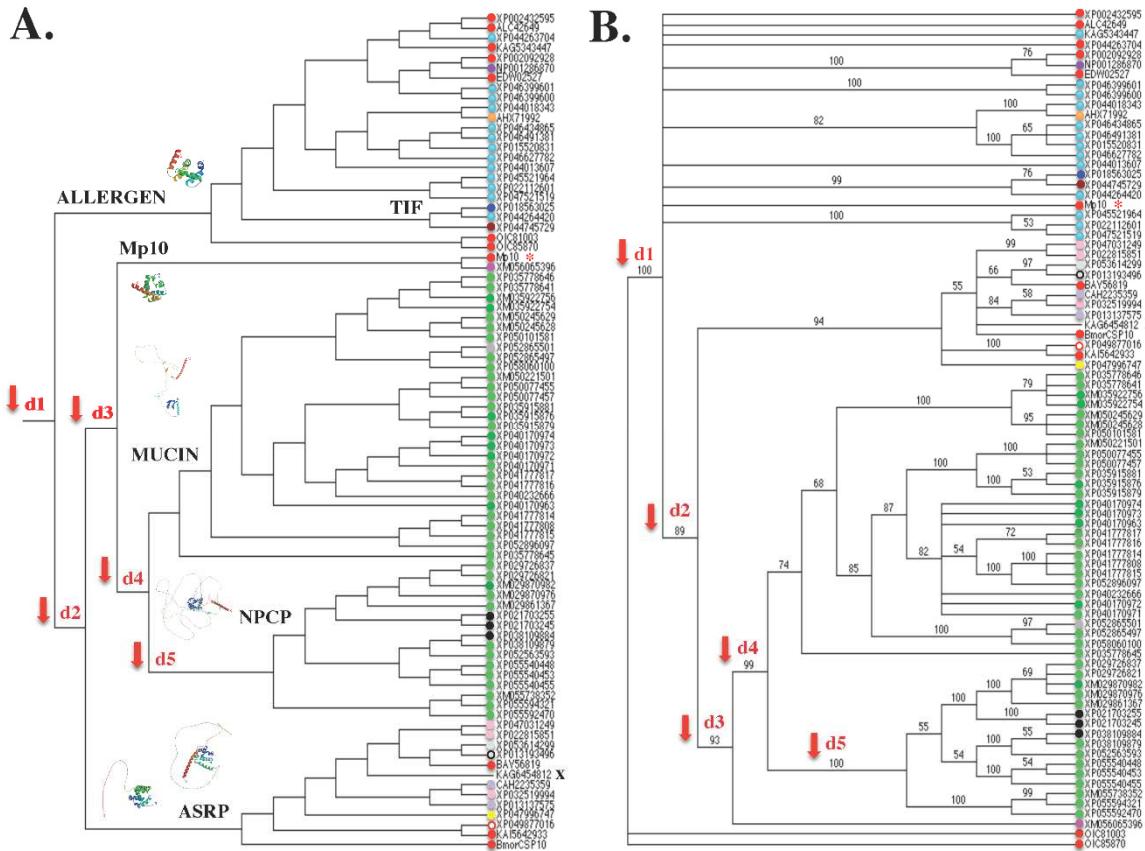
|              |     |              |            |              | <i>elegans</i>                 | Variant X1                    |                   |                     |
|--------------|-----|--------------|------------|--------------|--------------------------------|-------------------------------|-------------------|---------------------|
| LOC124223445 | 658 | XM_046635425 | 127 (14.2) | XP_046491381 | <i>Neodiprion pinetum</i>      | Allergen Tha p1<br>Variant X2 | Thorax<br>Abdomen | IgE binding protein |
| LOC124308780 | 659 | XM_046771826 | 127 (14.2) | XP_046627782 | <i>Neodiprion virginianus</i>  | Allergen Tha p1<br>Variant X2 | Thorax<br>Abdomen | IgE binding protein |
| LOC110991536 | 755 | XM_022256909 | 121 (14.2) | XP_022112601 | <i>Pieris rapae</i>            | Allergen Tha p1               |                   | IgE binding protein |
| Contig_79    | 360 | LYFD01000069 | 120 (13.8) | OIB81003     | <i>Acinetobacter baumannii</i> | A10/OS-D                      | Prokaryotic cell  |                     |
| Contig_92    | 261 | LYGG01000091 | 87 (9.8)   | OIC85870     | <i>Acinetobacter baumannii</i> | A10/OS-D                      | Prokaryotic cell  |                     |

## 5. CSPs' function in regulating gene transcription and nucleotide binding

Insect CSPs and RNA/DNA-binding protein sequences have some links, according to a preliminary phylogenetic analysis in IQ-Tree (Liu et al., 2019). While another orthology group included the coleopteran CSP AAJJ0012J, the DNA-regulatory protein WP\_0449980, and the DNA-binding protein WP\_0162268, the transcriptional/cell division repressor and DNA-regulatory protein (transcriptional regulator; WP\_011671486) branched with BmorCSP16, BmorCSP8, and BmorCSP9 and their bacterial counterparts (Liu et al., 2019). This implied that some CSPs, at least in Lepidoptera and Coleoptera, have evolved to serve tasks related to transcription, DNA/RNA control, and/or nucleotide binding.

This was supported by a phylogenetic examination of BmorCSPs, AAJJ0012J, Mp10, its derivative sequences, Mucin-like sequences, and a variety of RNA-binding proteins (RNA-BPs) in PAUP (\*10Alttivec; Figures 4 & S1-S2). AAJJ0012J molecule clustered with allergen Thp1, and associated RNA-BPs rather than BmorCSPs, according to a neighbour-joining tree (BioNJ study). On the BioNJ tree, CSP4 molecule from the marmalade hoverfly *Episyrphus balteatus* (QIS77191) deviated from the CSPs groups (see G1, Figure S2A). A further "CSP" from *E. balteatus*, EbalDDB\_G0285119X1, linked with BmorCSP10 and showed high similarity to EcorGTPase XP\_055921371 (a rho GTPase-activating protein 17 isoform X3; G2, Figure S2A). *E. balteatus* CSP3 (QIS77910) is important because it joins G2. Mp10 is also important since it does not join the BmorCSP molecules, but instead seems to be much more closely connected to the mucins group (Figure S2A), albeit not clearly forming an orthologous group (Figure S2B). Only the G2 group maintained a relatively high bootstrap value (98%) in the Jackknife analysis (bootstrapping calculation over 1000 repeats), strongly demonstrating the branching of "CSPs" with GTPase enzyme. IpodCAH2042437 was drawn at the bottom of the tree in this analysis using BmorCSP2 and its bacterial (*A. baumannii*) counterparts as the outgroup, but the rest of the sequences combining CSPs, RNA-BPs, Mucins, DNA-regulatory proteins, SAmkC, WAS/WASL-like, CWP, and other transcriptional factors fell on a common branch with a high bootstrap value (82%; Figure S2B). This strongly implied that there were important and predicted relationships among all of these various intracellular proteins.

With an emphasis on aphid 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 when compared to amino acid sequences from the Allergen, Mucin, Transcription Initiation Factor (TIF), Nuclear Pore Complex Protein (NPCP), and Actin skeleton regulatory complex (ASRP), this specific protein sequence (XP\_022173691) constructs a relevant hierarchical clustering UPGMA analysis and a phylogenetic tree backed by a high bootstrap value (Figure 4). It's interesting to note that Mp10 does not cluster on the UPGMA tree with other "CSPs". It belongs to a sizable group that includes TIF and all Mucin taxa, and it exhibits a strong association with Rho GTPase-activating protein (XM\_056065396; Figure 4A). Small GTPases (20–30 kDa GTP-binding proteins) in the Rho family 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, such as events connected to the cytoskeleton and gene transcription. One of the main classes of Rho GTPase regulators, known as RhoGAPs, is present in all eukaryotes (within the cell) and has been demonstrated to control a broad range of cellular processes, including cytoskeletal organization, growth, differentiation, neuronal development, and synaptic functions (Moon and Zheng, 2003). They have nothing to do with smell because they don't trigger transmembrane sensory receptors.

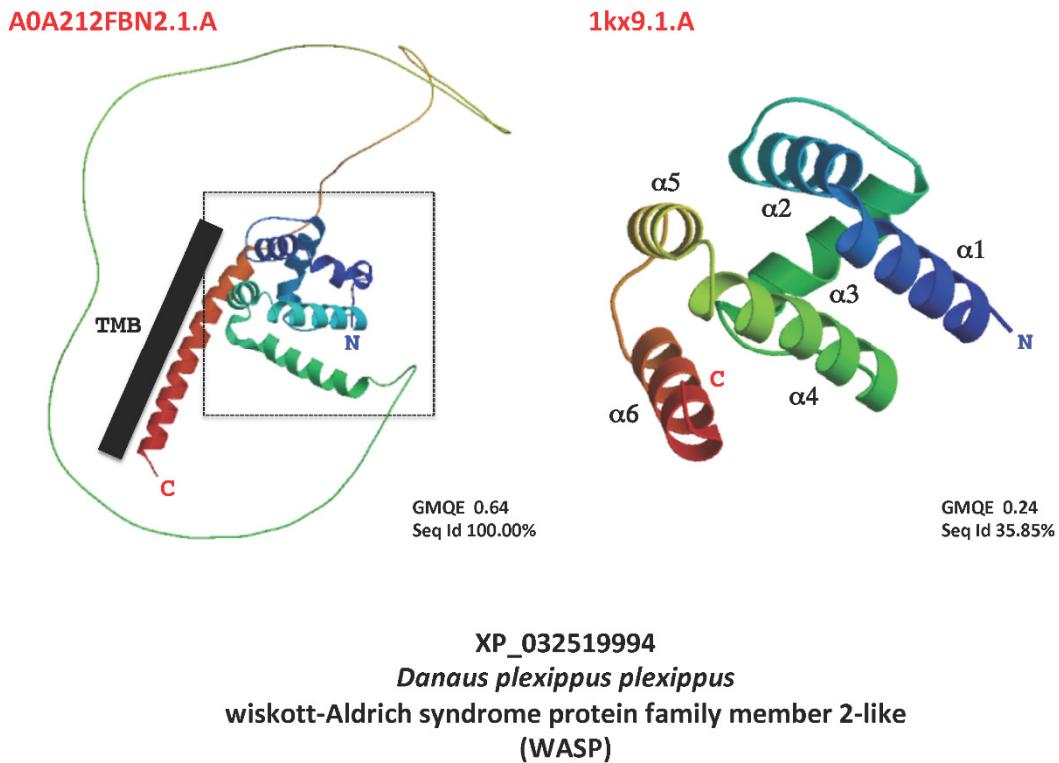


**Figure 4.** Phylogenetic analysis of “CSP” (Ebsp-3/PebIII/A10/OS-D/Pherokine/Mp10) with counterparts in the family of Allergen/IgE binding protein, stress response protein, Rho GTPase-activating protein, transcription initiation factor, Mucin-like, Extensin, Actin skeleton regulatory complex (ASRP), and Nuclear Pore Complex Protein (PAUP\*10alive). **A.** UPGMA analysis of Mp10 and counterparts (Table 1). Agglomerative (bottom-up) hierarchical clustering analysis based on the distance matrix of the analysed taxa that were calculated from a multiple alignment in ClustalW. The red arrows represent gene duplication events (d1-d5) that led to Allergens (IgE-binding protein) and the group of Mp10, Mucins, transcription initiation factors, DAN4, NPCP (Nuclear Pore Complex Protein), and BmorCSP10. The position of Mp10 is indicated by the red asterisk (\*): grouping with XM056065396 (pherokine) and Mucins. **B.** Bootstrap/Jackknife algorithm analysis of Mp10 and counterparts (Table 1) with bacterial *A. baumannii* CSPs (OIC81003 and OIC85870) as outgroup. 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. In red: “CSP” Ebsp-3/PebIII A10/OS-D, light blue: Allergen Tha p1 (IgE binding protein), purple: pherokine-3, orange: “CSP” acid trehalase, dark blue: CWA-3 (cell wall integrity, stress response), brown: TIF (transcriptional initiation factor), pink: Rho GTPase-activating protein, light green: Mucin-like, Extensin-like, translation initiation factor, grey: WAS/WASL (Actin skeleton regulatory complex), black: DAN4/NPCP (Nuclear Pore Protein Complex), salmon: PAN-1 (Actin skeleton regulatory complex), light grey: Formin-1 (Actin microtubule association), light grey in dark circle: WASP-2 (Actin skeleton regulatory complex), light purple: Jg5928 Major outer envelope protein GP Ib-like (Outer surface membrane protein), yellow: RickA-like (Arp2/3 complex activating protein), white in red circle: YLP motif protein 1 (nuclear nucleoside kinase), X: Hypothetical protein (unknown function). The sequences of amino acids of the proteins are utilized in comparative molecular analysis. Figures 5 & S3 contains the molecular structures.

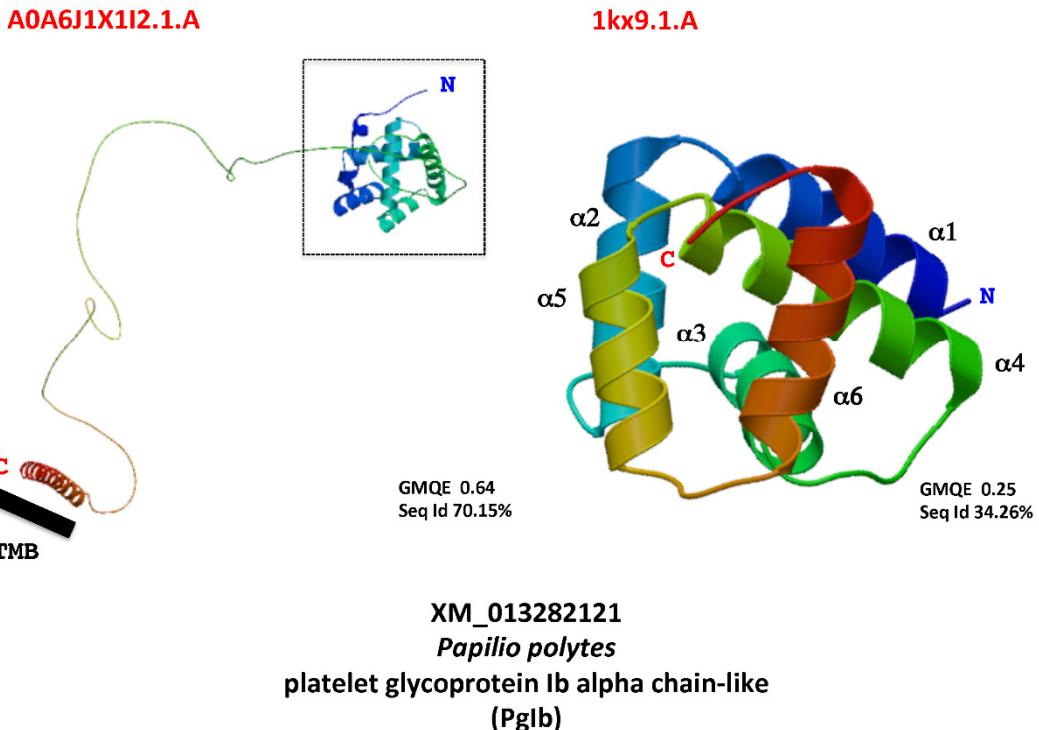
As for the other “CSPs”, they belong to either the ASRP group (see the position of BAY56819, KAI5642933, and BmorCSP10, in the UPGMA analysis on Figure 4A) or the Allergen group (which

appeared to be the most ancestral molecule): XP\_002432595, ALC42649, KAG5343447, XP\_002092928, EDW02527, OIC81003, and OIC85870 (bacterial *A. baumannii* CSPs). IgE binding proteins, pherokines (fly hemolymph proteins), acid trehalase (AHX71992, involved in intracellular trehalose mobilization during postdiauxic growth and severe saline stress in yeasts), CWA-3 (XP\_018563025), and TIF (XP\_044745729) are found in the Allergen group, whereas significant outer envelope proteins, RickA-like (Arp2/3 complex activating protein), and particular nuclear nucleoside kinase are found in the ASRP group (Figure 4A). CWA-3 molecule is a component of the cell wall integrity-signaling pathway, which is regulated by small proteins like G-protein Rho1. In the cell cycle and in response to different types of cell stress, CWA-3 is primarily in charge of controlling gene expression and coordinating periodic modifications to the cell wall (Cronmiller et al., 2019). A nucleoside diphosphate kinase catalyzes the reversible transfer of the terminal phosphate (P) from a donor nucleoside triphosphate (TP) to an acceptor nucleoside diphosphate (DP). Within the actin cytoskeleton of eukaryotic cells, the Arp2/3 complex is a ubiquitous and vital element. In order to build orthogonal networks, it nucleates actin filaments, caps their sharp ends, and cross-links them. As a vast group of proteins linked to Mp10, a protein categorized as a “chemosensory protein” (CSP), all of these groups have functions that are more closely related to the cytoskeleton (actin), transcription of genes, P transfer reaction, P exchange, phosphorylation, mitochondrial energy production, TP-DP conversion, and cell regulation than they are to smell (see Figure 5).

**A.**



B.



**Figure 5.** The modeling of the N'-CSP'prism-long loop-transmembrane domain-C molecular structure. **A.** The modeling of Wiskott-Aldrich Syndrome Protein family member 2-like (WASP) molecular structure. **B.** The modeling of platelet glycoprotein Ib alpha chain-like (PgIb). Locus and Access Number can be found in Table 1. Figure 4 illustrates how WASP and PgIb align with "CSP" molecular sequences. Sequences were aligned with BmorCSP1 in order to identify the signal peptide and cut it off based on N-terminal sequencing by Edman degradation (Picimbon et al., 2000b). 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: A) 1kx9.1 ("Chemosensory Protein A6", X-ray, 1.6 Å, monomer, cabbage moth, *Mamestra brassicae*) and A0A212FBN2.1.A (Wiskott-Aldrich syndrome protein family member 2-like, AlphaFold DB model of A0A212FBN2\_DANPL, LOC116772069 gene, butterfly, *Danaus plexippus plexippus*) for WASP, B) 1kx9.1 and A0A6J1X1I2.1.A (Mucin-2-like, AlphaFold DB model of A0A6J1X1I2\_GALME, LOC113521739 gene, greater wax moth, *Galleria mellonella*) for PgIb. 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 six -helices that make up the CSP prism are numbered 1 through 6. The N-terminal tail of WASP (actin nucleation factor) and PgIb (platelet surface membrane protein) molecules in the intracellular compartment (extracellular exosome) is where the CSP prism is located, as indicated by the location of the black square with dotted lines. The black bar indicates the position of the transmembrane segment (TMB).

The molecules known as "CSP", Allergen, TIF, Mp10, 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" (Ho, 2008) to account for mutation rates). The aforementioned proteins are the outcome of a series of duplication events that produced Allergen, which includes CSP and TIF, prior to Mp10, Mucin, NPCP, and ASRP (Figure 4A). Mp10 (and Rho GTPase-activating protein) emerged later in a sequence of duplication events that produced a wide range of Mucin variants,

especially in mosquito species (*Aedes*, *Anopheles*, *Culex*, *Uranotearia*, and *Wyeomyia*). This is a noteworthy point because it is comparable to the high degree of variation and duplication found in mosquitoes' *CSP* genes (Picimbon, 2020). This is also similar to the high degree of variation in *CSP*-encoding RNA observed in many different insect species, primarily in moths and bees (Xuan et al., 2014; Liu et al., 2020). The gene family enlisting *CSP*, *Mp10*, *Rho GTPase-activating protein*, *Allergen*, *TIF*, *MUCIN*, *NPCP*, and *ASRP* appears to have undergone multiple duplications over its evolutionary history. Some of these duplications are specific to particular taxonomic lineages, such as mosquitoes for the long Mucin precursor proteins needed for growth, development, digestion, oviposition, and infection protection (Wang and Granados, 1997; Dias et al., 2018), while others are more archaic and are shared across lineages (Figure 4A). This is true for the *Allergen* group, which includes taxa from the following species: *Acromyrmex*, *Aphidius*, *Coccinella*, *Drosophila*, *Ischnura*, *Neodiprion*, *Pediculus*, *Phtorimaea*, *Pieris*, and *Tribolium*. This also holds true for the *ASRP* group, which comprises taxa from *Amyelois*, *Anaplophora*, *Bombyx*, *Danaus*, *Heliothis*, *Leguminivora*, *Ostrinia*, *Papilio*, *Pararge*, *Pectinophora*, *Phtorimaea*, and *Spodoptera* (Figure 4A). Given the distance between the *ASRPs/CSPs* and the *Allergen* molecules, the presence of a common root in the UPGMA tree of amino acid sequences indicates that all of these molecules have a distant common origin that is estimated to be 324–440 Mya (the latest Mississippian–Silurian or Devonian; Schachat et al., 2023) ahead of the emergence of the various flying insect species.

With the use of maximum parsimony analysis, the relationships between the *CSP*, *Mp10*, *Rho GTPase-activating protein*, *Allergen*, *TIF*, *MUCIN*, *NPCP*, and *ASRP* proteins were further investigated (Figure 4B). Strict consensus trees were established using MP and the PAUP4.0b10 (Altivec) program, as detailed in Abraham et al. (2005).

A shared characteristic of *CSP*, *Mp10*, *Rho GTPase-activating protein*, *Allergen*, *TIF*, *MUCIN*, *NPCP*, and *ASRP* is their widespread tissue expression. There have been reports of allergens from the abdomen and thorax clones. All body, larvae, and pupae have been found to contain PAN-1, YLP motif, DAN4, *NPCP*, *WASP*, and *Mucins* (types 2 and 5AC; see Table 1). Based on what is known about the tissue distribution of "CSPs", this is strikingly similar. "*Cloned or labeled from the antennae does not imply specificity to the antennae if the distribution of tissue throughout has not been studied*". It has now been established through extensive research on tissue distribution that "CSPs", like the various types of *Mucins* that are examined here, *Allergens*, PAN-1, YLP motif, DAN4, *NPCP*, and *WASP*, occur in all parts of the insect body, including the thorax, abdomen, and larvae and pupae (Liu et al., 2020; Picimbon, 2020; Table 1).

In this case, phylogenetic data and expression analysis combine. From XP\_002432595 (*Pediculus* "CSP") to XP0055592470 (*U. lowii* Mucin-2 type X3, nuclear pore complex protein) and XM\_056065396 (*E. corolla* Rho-GTPase-activating protein), the dipteran, coleopteran, hemipteran, and lepidopteran *CSP*, *Mp10*, *Rho GTPase-activating protein*, *Allergen*, *TIF*, *MUCIN*, *NPCP*, and *ASRP* molecules are closely related to each other; they formed groups with a high bootstrap value (89-100%). *Mucin* and *NPCP* combine to form a group with a high bootstrap value of 99%. They also attach to the 'ASRP' and *Rho-GTPase-activator*, both of which have very high bootstrap values of 93-99%, suggesting that these molecule families are strongly related (Figure 4B). With 94% bootstrap support for molecules involved in actin microtubule association, cytokinesis, actin skeleton regulatory complex, Arp2/3 complex activating protein, and nuclear nucleoside kinase, three *CSPs* (BAV56819, *BmorCSP10*, and KAI5642933) belong to the *ASRP* group (Figure 4B). Antennae, legs, sex pheromone gland, head, wings, gut, fat body, epidermis, and thorax all express *BmorCSP10* (Xuan et al., 2015). This gene's supposed "chemosensory" function is disproved by phylogenetic information, orthology grouping, and molecular expression analysis combined (see Figure 4B and Xuan et al., 2015).

It's interesting to note that bacterial *CSPs* (*A. baumannii* OIC81003 and OIC85870) attracted *Rho GTPase-activating protein* at the base of the tree but not *Mp10* when they were used as an outgroup in the MP analysis of protein sequences. *Mp10* continued to have a closer relationship with the *Allergen* sequence group, in particular with *TIF* (Figure 4B). In contrast to *Mucins*, *NPCPs*, and *ASRPs*, coleopteran, dipteran, hemipteran, hymenopteran, lepidopteran, phthirapteran, and zygopteran (Odonata, blue-tailed damselfly) allergens/*CSPs* from XP\_002432595 to XP\_44264420 and

Mp10 are related, though they do not form a general orthology grouping. The MP tree (Bootstrap Jackknife) in the PAUP4.0b10 (Altivec) program and the UPGMA tree differ greatly in this regard (Figure 4). Allergens/CSPs, including transcription initiation factor, segregated independently in contrast to Mucins, NPCPs, and ASRPs; this separation was consistent with the evolutionary distances between insect species. As previously described for the “CSP” gene family, this indicates a rapid, intense, frequent, and high evolutionary rate for this molecule family, despite its extreme age (Picimbon, 2003, 2019, 2020, 2023; Sánchez-Gracia et al., 2009; Kulmuni et al., 2013).

Our phylogenetic analysis, assuming that the amino acid tree is the “correct” tree, thus reveals that the Mp10 orthologs of the aphid, beetle, dragonfly, fly, louse, mosquito, moth, and sawfly belong to different groups, ranging from Mucins and ASRP to Allergens and TIF. This suggests that before the emergence of *Myzus*, *Acromyrmex*, *Aedes*, *Amyelois*, *Anopheles*, *Anoplopheura*, *Aphidius*, *Aphis*, *Bombyx*, *Coccinella*, *Culex*, *Danaus*, *Drosophila*, *Eupeodes*, *Heliothis*, *Ischnura*, *Leguminovora*, *Neodiprion*, *Ostrinia*, *Palarge*, *Pectinophora*, *Pediculus*, *Pieris*, *Phtorimaea*, *Spodoptera*, *Tribolium*, *Uranotaeria*, and *Wyeomyia* (e.g. the carboniferous period of the Paleozoic era 299-359 Mya), there were multiple local duplications that led to diversification in the Mucin-Allergen-TIF-CSP family (Figure 4). The phylogenetic distances between mosquito genomes were consistent with the grouping of mosquito mucin genes together. The oldest genome belongs to the Anophelinae family. By the Jurassic Period, mosquito genomes had evolved (206-142 Mya). Evolution of the genomes of *Culex* and *Aedes* occurred by the Eocene (60 Mya; Rai and Black, 1999). Unlike *Aedes* and *Culex*, *Anopheles* (which has only six genes placed on two distinct chromosomes) is not truffed with CSPs or allergens. The West Nile virus vector *C. pipiens* and the yellow fever mosquito *A. aegypti* have been shown to retain 29 and 42 CSP genes, respectively (Picimbon, 2020). As a result, Mucin/NPCP/ASRP and Allergen/CSP molecules exhibit distinct evolutionary histories. For insects, Mp10 orthologs appear to have undergone one to five distinct duplication events (see Figure 4). Four consecutive duplication events appear to be the source of mucins in mosquitoes: two early duplications (d1 and d2) produced allergens/CSPs (immune response), TIFs (transcription regulation), and ASRPs (cytoskeleton regulation); two late duplications (d3 and d4) produced Mp10 (innate immunity induction) and mucins (multifunction from cell growth to immunity). CSP amino acid sequences in insects are arranged so that all of the protein molecules in the Allergen, TIF, Mucin, NPCP, and ASRP trees share a common root. Mp10, moth CSPs, *Drosophila* pherokines, and a profusion of genetic elements, immune system proteins, IgE-binding proteins, nuclear complex proteins, cell regulation proteins, cytoskeleton regulation proteins, and actin-related protein complex are all associated. There is no correlation found here for the protein sequences of odor-related or chemosensory receptors. Although a fifth duplication that resulted in DAN4 and NPCP molecules may be exclusive to mosquitoes (see Figure 4), all of these Allergen-CSP 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; Beukers, 2004; Picimbon, 2023).

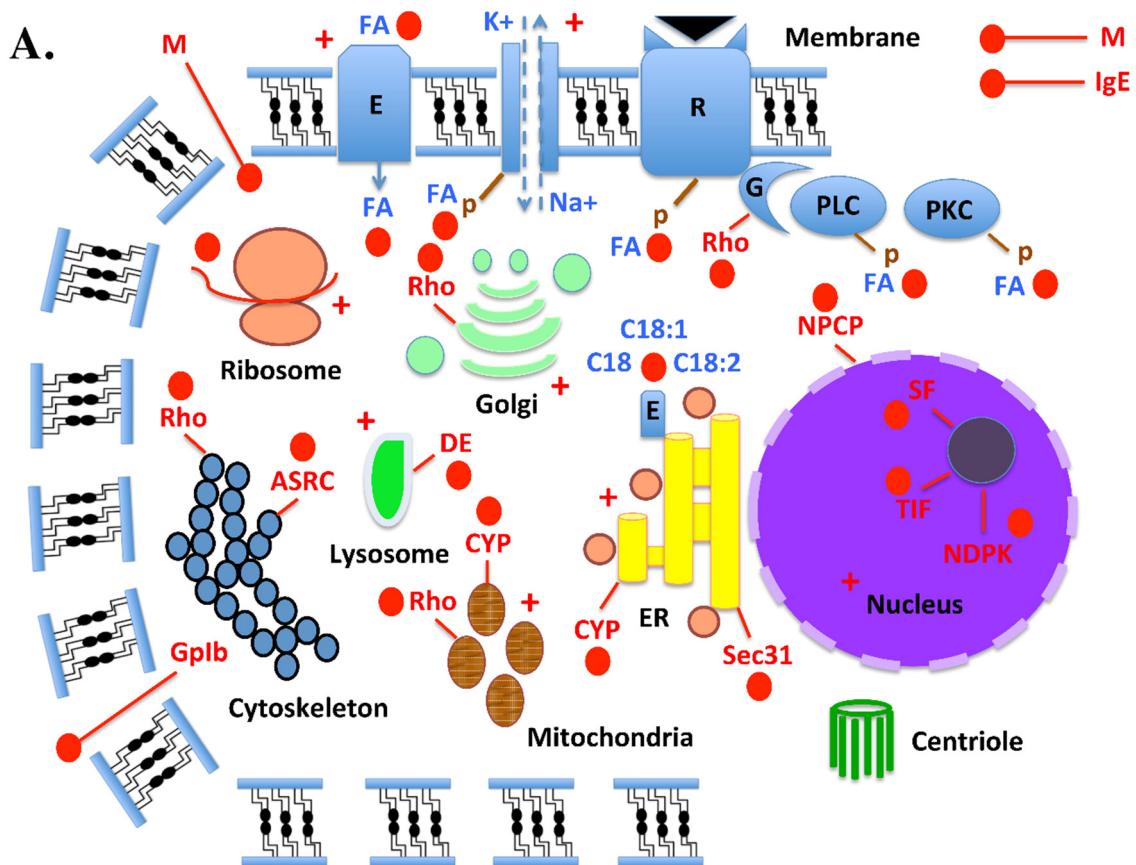
Structural modeling provides an explanation for the strong relationships observed between “CSP”, Mp10, allergen/IgE binding protein (Thap-1), pherokine, acid trehalase, and these larger molecules in several protein families, such as mucin-2-like, CWP, TIF, Rho, NPCP, ACRP, PAN1-like, WAS/WASL, Rick-A, Formin-1, UL36, PgIb, jg5928, SamkC, LRR, WASP, and NNK (see table 1, Figures 4-5 & S2-S3). Using SWISS-MODEL Workspace/ GMQE (Waterhouse et al., 2018), it appears that most of all query molecules fold into a prism of six  $\alpha$ -helices when compared to 1kx9.1 as template reference (X-ray crystal structure; Lartigue et al., 2002). This was especially true for pherokines (Phk-3, hemolymph protein, 9W0X2), immune response protein, acid trehalase, allergen/IgE binding protein (Thap-1), Mp10, small binding proteins known as “BmorCSP1s”, and eukaryotic translation initiation factor 4 gamma-like from *Coccinella septempunctata* (XM\_044889794; Figure S3). All other proteins examined in the SWISS-Model exhibit prism structure, albeit only in particular molecule-specific regions, such as the N-terminus (see BmorCSP10, GMQE 0.68, Sequence Identity 99.38% with Q3LB93.1.A, DB model of Q3LB93\_BOMMO, BABH01021709 gene, silkworm, Figure S3). A “CSP” prism at the N-terminal region and a very long  $\alpha$ -helical stretch (7–13 turns) forming the C-terminal tail characterize the majority of representative molecules of the mucin, CWP, TIF, Rho, NPCP, ACRP, PAN1-like, WAS/WASL, Rick-A, Formin-1, UL36, PgIb, jg5928, SamkC, LRR, WASP, and NNK families (Figure S3). A transmembrane segment at the C-terminal tail is

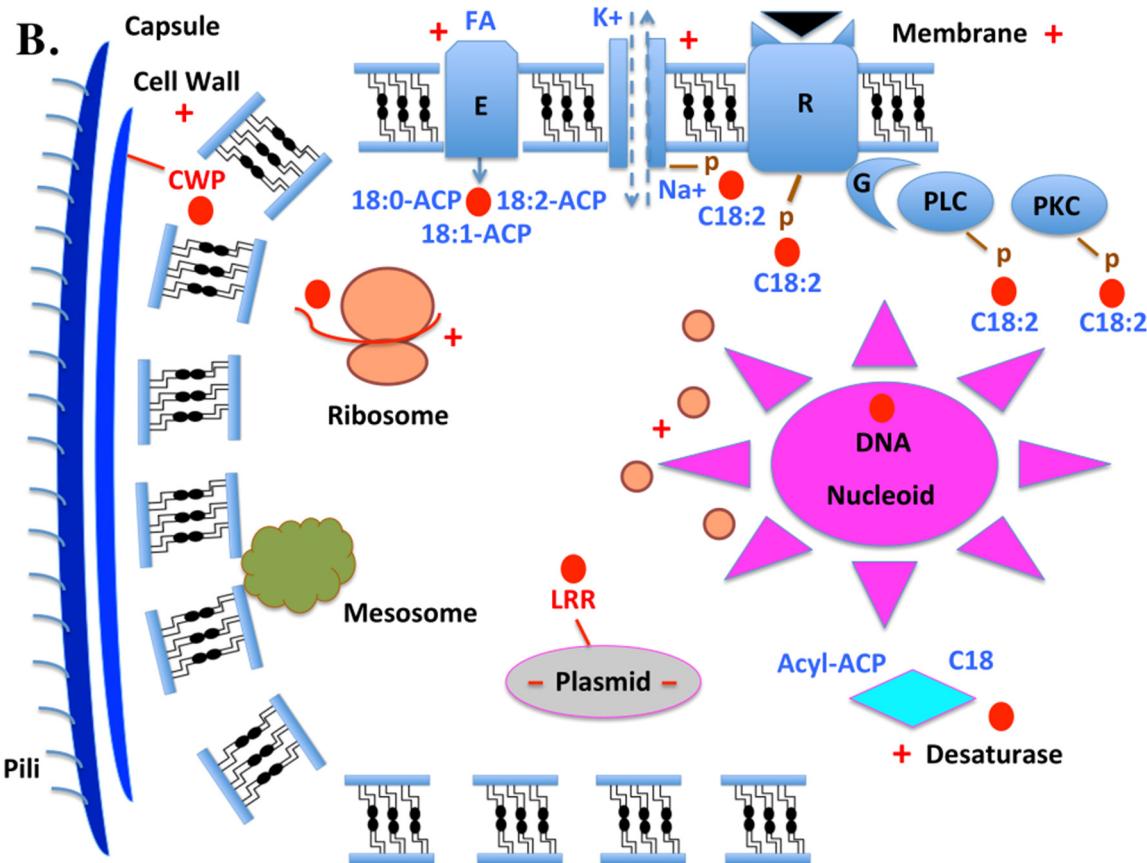
predicted for these structures using Swissmodel (see XP\_032519994 *Danaus plexippus plexippus* wiskott-Aldrich syndrome protein family member 2-like, WASP, Figure 5A; XM\_013282121 *Papilio polytes* platelet glycoprotein Ib alpha chain-like, PgIb, GMQE 0.64, Sequence Identity 70-100%, Figure 5B). The butterfly *D. plexippus* WASP protein's gene locus and molecular structure are known (A0A212FBN2.1.A AlphaFold DB model of A0A212FBN2\_DANPL, LOC11677206), and they do not match the characteristics of a molecule that is classified as "chemosensory" (length: 297 amino acids; mass: 33 kDa; transmembrane domain: 1). The majority of these larger molecules, which are found in the cytoplasm, cell wall, integument, cytoskeleton, microtubule, nuclear pore, and many other intracellular organelles, are distinguished by an incredibly long loop that connects the C-terminal transmembrane  $\alpha$ -helical maillon and the "CSP" structure at the N-terminus (Figures 5 & S3). The AlphaFold DB model of an uncharacterized protein or proteins classified as "Chemosensory Proteins" or members of the "putative insect pheromone-binding family" is used as a template to obtain this (UniProt: A0A0K8TVC4, A0A182NU82, A0A1Q3FP99, A0A1V1WC08, A0A2H4ZB78, A0A345BEP2, A0A6H0D4B0, A0A6I8TVU3, A0A6I8TWU8, A0A835CMW0, D6WUE8, Q6H8Y7, and A7UL73). The presence of the 'CSP prism'-prominent loop-TMB domain in CAH2235359, a Palearctic butterfly molecule, strongly suggests that this type of molecule exists in viral particles, at least in virus and consor of herpes simplex (type 1 and type 2, HSV-1 and HSV-2; Turcotte et al., 2005). This structure evokes the major outer envelope glycoprotein of the herpes virus (jg5928; A0A1V1WC08, GMQE 0.64, Sequence Identity 70.31%, Figure S3). Bacterial species have a similar molecular structure (WP\_096417339, 491 amino acids, 50.6 kDa, Xanthomonadaceae, *Lysobacter capsici*; Liu et al. 2019). As a result, it is more likely an infection (virus or bacterial binding and immune response) than chemosensation.

These molecules' "chemosensory" function is not supported by their location, size, structure, or expression in viruses and microbes.

These striking similarities between "CSPs", allergens, mucins, Rho, NPCP, ASRC, Sec31, GpIb, Leucine-Rich Repeat proteins, and cell wall proteins suggest that "CSP" molecules serve a wide range of diverse functions within the intracellular systems of both eukaryotes and prokaryotes (Figure 6). In eukaryotes: CSP binds to FA (C:18-2) which mediates the phosphorylation of different plasma membrane-bound proton pumps (sodium-potassium exchange pump), enzymes (ATPases, anion transport protein, glyceraldehyde 3-phosphate dehydrogenase, protein kinases, protein acetyl transferases, alkaline phosphatase, phosphodiesterase (PDE), adenylate cyclase (AC), esterase, acetylcholinesterase,  $\beta$ -N-acetylglucosaminidase, fatty acid-binding proteins, fatty acid transporters, CD36, regulators of lipid uptake) and receptors (receptor tyrosine kinase, ligand-gated channels, enzyme-linked cell-surface receptor, and G-protein-coupled receptor) molecules. These activities are necessary for the synthesis of diacylglycerol, which triggers the activation of phospholipase kinase C and the phosphorylation of multiple proteins involved in specific signal transduction pathways. Additionally, these processes are necessary for the production of protein kinase C, which is ubiquitously distributed, phosphorylates serine and threonine residues in a vast array of protein substrates, and regulates numerous cellular processes (Liu et al., 2020). CSP is linked to multifaceted Rho GTPase signaling at the level of endomembranes from cytoskeleton, Golgi, mitochondria and plasma membrane (Michaelson et al., 2001). Desaturase enzyme interacts with CSPs that transport fatty acids, such as linoleic acid (C18:2), and its two precursor molecules (stearic acid C:18 and elaidic acid C:18-1), at the endoplasmic reticulum (ER) level (Liu et al., 2016b, 2020). Insecticide exposure (+) and other stress responses cause CSP molecule to interact with cytochrome P450 (CYP) in the ER and the mitochondrial system. The CSP-CYP interaction is essential for cellular metabolism, homeostasis, and xenobiotic detoxification (Xuan et al., 2015). The lysosome's functions, including the digestion and breakdown of macromolecules (proteins, lipids, carbohydrates, and nucleic acids), repairs of cell membranes, and defense against pathogens like microbes, bacteria, fungi, and viruses, are all mediated by the CSP-Degradative Enzyme (DE). The lysosome releases its enzymes when food is consumed or absorbed by the cell, breaking down complex molecules (such as sugars, proteins, and fatty acids brought by CSPs) into energy that the cell can use to survive (Fujiwara et al., 2017). CSPs participate in the FA and phosphorylation processes, which in turn affect the biosynthesis of stress responses in the nucleus, mitochondria, Golgi, ER, ion pump, lysosomes, and ribosomes (see +, Figure 6A; Hilgemann, 2003; Sasaki and Yoshida, 2015; Albert et al., 2019; Lee and Choi, 2020; Zhang et al., 2021; Wen et al., 2023). The mechanisms of control of lipid metabolism in the cell cycle are mediated

by the stress response, fatty acids, and CSPs at the molecular translational level in the ribosome (Blank et al., 2017a,b). In order to produce transport vesicles from the ER, the CSP-Sec31 protein complex is essential (Kim et al., 2012). Mucin (M), Actin Skeleton Regulatory Complex (ASRC) and Nuclear Pore Complex Protein (NPCP) are associated with CSP molecules, locating them at the level of the plasmic membrane, cytoskeleton, and nuclear membrane for their combined role in protecting cell membranes and controlling membrane interactions, internalizing endosomes (mainly intracellular sorting organelles) during actin-coupled endocytosis, regulating the trafficking of proteins, fatty acids, and lipids among other subcellular compartments of the secretory and endocytic pathway, particularly the Golgi system and lysosomes, and controlling genome functions (Ibarra and Hetzer, 2015; Lin and Hoelz, 2019). Moreover, CSP is the N-terminal region of Gplb, and as such, it resides extracellularly to recognize antigens, activate the immune system, and/or recognize hormones and promote cell growth (see Figure 6A; Kenis et al., 2023). Together, CSP, splicing factor (SF), transcription initiation factor (TIF), and nucleoside diphosphate kinase (NDPK) connect these molecules to nucleotide binding, RNA metabolism, and nucleus-wide control of gene expression (Rosonina and Blencowe, 2002; Kornblith et al., 2004; Georgescauld et al., 2020). As allergens or mucin-like fractions, some soluble fractions of CSP are secreted, bolstering the immune system's resistance against microbial stress, especially in the saliva, gut, and eyes (Shangguan et al., 2018). They mediate substance exchanges (gas, water, etc.) between the environment and internal organs in the eyes and lubricate, which is one function of tears. Extracellular CSP-mucin-like may also be necessary for the adult development of insects, as reported by Huang et al. (2022) and our findings here (Table 1 and Figures 4-6 & S2-S3).





**Figure 6.** The intracellular systems of eukaryotes and prokaryotes use "CSP" molecules for a variety of purposes. **A.** In eukaryotes: CSP binds to fatty acid (FA), which mediates the phosphorylation (p) of different plasma membrane-bound protein molecules. E: Enzymes, K<sup>+</sup> intake/Na<sup>+</sup> outflow: potassium-sodium exchange pump, R: Receptors, G: G-protein, PLC: phospholipase kinase C, PKC: protein kinase C, Rho: Rho GTPase, E on endoplasmic reticulum (ER): Desaturase enzyme (E), which interacts with CSPs that transport fatty acids, such as linoleic acid (C18:2) and its precursors (stearic acid C:18 and elaidic acid C:18-1; see Liu et al., 2016b, 2020). +: 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 CSPs 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 immune system-building soluble allergens, while M stands for mucin-like fractions, secretory molecules that fortify the immune system's defense against microbial stress. **B.** In prokaryotes: CSPs carrying linoleic acid (C18:2) and acyl-carrier-protein (ACP) counterparts (18:0 = stearoyl-ACP, 18:1 = oleoyl-ACP) interact with integral membrane fatty acid (FA) desaturase (E). C18:2-ACP = linoleoyl-acyl-carrier-protein. The phosphorylation (p) of the sodium-potassium pump (K<sup>+</sup>/Na<sup>+</sup>), receptor (R), PLC, and PKC is made possible by the transport of C18:2 (linoleic acid). Acyl-ACP, C18, and Desaturase: microbial soluble intracytoplasmic desaturase system. +: Location of reaction to conditions of biological chemical stress (membrane, cell wall, ribosome, nucleoid, ribosome-nucleoid interaction site, and ACP desaturase systems). The part of the cell's envelope that provides structural stability and a protective reaction to stress is the cell wall protein complex (CWP) attached to CSP. The red bars show the plasmid's CSP genes, which can be subject to horizontal transfer from bacteria to insects and plants. On the plasmid, CSPs are also found as molecules connected to the Leucine-Rich Repeat protein (LRR). The red dots represent every protein and supramolecular complex in every organelle that "CSP" molecules interact with intracellularly in both prokaryotes and eukaryotes.

In prokaryotes, the integral membrane fatty acid desaturase can selectively introduce double bonds or catalyze unusual reactions such as acetylenation, conjugation, hydroxylation, and epoxydation. CSPs carrying linoleic acid and ACP counterparts, such as stearoyl-acyl-carrier-proteins and oleoyl-acyl-carrier-proteins (18:0-ACP and 18:1-ACP) can interact with this trans-membrane enzyme (Nachtschatt et al., 2020; Cerone and Smith, 2022). Like in Eukaryotes, the transport of C18:2 (linoleic acid) is necessary for the phosphorylation of sodium-potassium pump (K<sup>+</sup>/Na<sup>+</sup>), receptor (Sensor histidine kinase, Methyl-accepting chemotaxis protein, KdpD, KdpE, G-protein-coupled receptors and microbial rhodopsin), PLC, and PKC (Titball, 1993; Gestwicki et al., 2000; Dibrova et al., 2015; Shalaeva et al., 2015; Salah Ud-Din and Roujeinikova, 2017; Heinisch and Rodicio, 2018; Ishii and Eguchi, 2021; Monturiol-Gross et al., 2021). Different from the diiron-containing integral membrane desaturases found in plants and other organisms, plastidial acyl-acyl carrier protein (ACP) desaturases in microbes and plants are a soluble class of diiron-containing enzymes (see Figure 6B). Bacteria, fungi, and plants possess soluble intracytoplasmic desaturases, and the interaction of CSP with acyl-lipid ACP molecules mediates both the activity of this enzyme and the activity of transmembrane fatty acid desaturase (Pereira et al., 2003; Sperling et al., 2003; Rodríguez et al., 2015). These enzymes catalyze desaturation of fatty acids esterified at the sn-1 position of the glycerol moiety of glycerolipids at the Δ9, Δ12, and Δ6 positions, respectively, in bacteria (fungi and plants). Oleic acid is formed when the Δ9 desaturase adds the first unsaturated bond to stearic acid. The Δ12 desaturase then further desaturates oleic acid to linoleic acid. CSP molecules are closely linked to the desaturase function in cell and molecular stress response with regard to transporting C:18, C18:1, and C18:2 (Los and Murata, 2002; Rock, 2008). The majority of insect species, including flies and whiteflies, lack these enzymes, so CSP molecules are essential scavenger proteins for linoleic acid (C18:2), which they absorb from food and other nutrients and distribute throughout their entire body (Liu et al., 2016b, 2020). This function is more concerned with providing energy to the cell than it is with binding protein-chemosensory receptor complexes (Figure 6). Engaging in intricate cross-talks, a range of specialized effectors are typically involved in interactions between land plants and other organisms, including insect pollinators, viral pathogens, and bacterial symbionts. These biological interactions involve fatty acids and their lipid derivatives. The targets of biological chemical stress in bacteria are the same as in eukaryotic cells: ribosomes (regulation of the number of ribosomes, regulation of protein synthesis and/or mRNA translation, chaperone burden and protein influx; Zhu and Dai, 2020; Chen-Guang and Gualerzi, 2021; Njenga et al., 2023). Biological stress response also appears to have a major impact on the regulation of Acyl-ACP desaturases. As with plants, bacteria's various monounsaturated fatty acids are generated by these enzymes and are demonstrated to participate actively in many various interactions. Desaturase enzymes, fatty acids, CSPs, and ribosomes are all interconnected, especially in situations of biological chemical stress. The position of the carbon-carbon double bond (C=C) in the acyl chain is intrinsically linked to these protective enzymatic activities (Kazaz et al., 2021). The soluble intracellular enzymes known as ACP desaturases facilitate the introduction of a double bond into saturated fatty acid molecules that are bound to saturated acyl chains that are bound to ACP in higher plants. This process results in the production of cis-monounsaturated fatty acids. There are three different types of soluble acyl-ACP desaturases that have been described: Δ9-acyl-ACP, Δ6-acyl-ACP, and Δ4-acyl-ACP desaturases. The differences between these types of desaturase enzyme are in the position of the double bond and the specificity of their substrate. Nevertheless, this family of enzyme molecules has been reported to exhibit bifunctional activities and/or substrate switching in bacteria, fungi, insects, and plants (Moto et al., 2004; Damude et al., 2006). ACP desaturases are a major factor in determining the composition of monounsaturated fatty acids in glandular cells, sensory neurons, and all other cells in the insect body. They introduce double bonds at specific positions in FAs of defined chain lengths. Mutant enzymes with modified fatty acid chain-length specificities or the ability to insert double bonds into either the Δ6 or Δ9 positions of 16:0- and 18:0-ACP were identified by substituting specific amino acid residues from a Δ9-stearoyl (18:0)-ACP desaturase for those from a Δ6-palmitoyl (16:0)-ACP desaturase enzyme. A Δ6-16:0-ACP desaturase can undergo specific ribosome mutations, like those listed in the CSP molecule family, to become an enzyme molecule that primarily functions as a Δ9-18:0-ACP

desaturase (Xuan et al., 2014, 2016, 2019; Picimbon, 2014a,b; 2016, 2017, 2019; Yue et al., 2023). These results demonstrate the ability to rationally modify acyl-ACP desaturase activities through mutations, which is a first step toward the design of acyl-ACP desaturase-CSP molecular complexes for the production of novel monounsaturated fatty acids in microbes, plants, insects, and potentially even mammalian and human cells (Cahoon et al., 1997; Xuan et al., 2014, 2016, 2019; Picimbon, 2017, 2019). Instead of folding in insect chemosensory systems, the CSP molecule folds in prokaryotes' Cell Wall Protein (CWP) complex, bringing it to the cell's envelope and the primary stress-bearing and shape-maintaining element (Scheffers and Pinho, 2005). Cell wall integrity is essential for maintaining cell shape and structural integrity, and CSP-CWP seems to be involved in this process. The cell wall of gram-positive bacteria acts as an attachment site for proteins that interact with the bacterial environment, even though its main purpose is to provide a strong rigid exoskeleton (turgor) for protection against both mechanical and osmotic lysis (Abanes-De Mello et al., 2002; Schneewind and Missiakas, 2014). Plants also have cell walls that enclose individual cells and serve as a channel for intercellular communication in addition to offering external protection. Cell wall proteins in plants and microorganisms serve as barriers against both biotic and abiotic stresses in addition to giving the cell body shape. Their functions include shielding the cell from chemicals, toxins, and other harsh environmental conditions, as well as preventing the cell from drying out in extremely hot weather (Han et al., 2023). Given that CSPs are located in the N-tail of CWP and may bind to toxins, expose them to enzymes that degrade them, or contribute to the supramolecular agency of the CWP complex, it would be intriguing to investigate these possibilities (see Figures 5 & S2). The plasmid contains CSP genes, which enables their horizontal transfer from microbes to insect and plant organisms (Redondo-Salvo et al. 2020). By transmitting ecologically significant features between strains and species, conjugative plasmids play a pivotal role in horizontal gene transfer, which speeds up bacterial adaptation (Harrison and Brockhurst, 2012). CSPs are also found in bacteria on the plasmid as molecules associated with the Leucine-Rich Repeat protein complex (see Table 1; Figures 4-6B & S3). It is there that they participate in a range of microbial protein-protein interactions that activate the plasmid and boost its pathogenicity and virulence (Ikegami et al., 2004; Hu et al., 2016). Ultimately, the finding that CSP is associated with the N-terminus of transcription initiation factor (TIF) suggests that, similar to eukaryotes, CSPs and DNA-dependent multi-subunit RNA polymerase play a pivotal role in controlling transcription initiation in bacteria (see Table 1, Figures 4-6 & S2-S3). The degree of transcription of a given gene in a bacterium can be influenced by CSP-binding TIF's ability to bind to RNA polymerase and generate a holoenzyme that is capable of transcription or by a multitude of highly diverse factors (Browning and Busby, 2004; Seshasayee et al., 2011). Among these are CSPs.

## 7. Terminology: function or structure?

This molecule family's first member was given the name p10 in honor of the size and molecular weight (in kDa) of a soluble protein found in the regenerated legs of insects. The same protein, called Pam, was found in the fully developed antennae and legs of the adult American cockroach *P. americana* in both sexes (Nomura et al., 1992; Picimbon and Leal, 1999). Similar clones identified in the fruit fly *D. melanogaster* and the migratory locust (*Locusta migratoria*) during a search for olfactory genes were dubbed Olfactory-Sensory type D protein (OS-D or Pheromone Binding Protein A10; McKenna et al., 1994; Pikielny et al., 1994; Picimbon et al., 2000a). Without any concrete proof of specificity towards olfactory sensilla, that was merely an initial attempt to rename this molecule protein gene family in favor of olfactory or chemosensory function. Based on the remarkably high expression of this protein family in sensory structures like the legs and antennae, this nomenclature already addressed chemosensory functions. Likewise, to differentiate themselves from another family of soluble proteins already referred to as odor/pheromone-binding proteins (OBPs/PBPs, 16-17 kDa, six cysteines), later clones of p10, OS-D, or A10 found in the antennae of the tobacco hornworm (the sphingid moth *Manduca sexta*) were renamed and changed to "Sensory Appendage Proteins" (SAPs; Robertson et al., 1999; Ingham et al., 2020). The history of the different names used to characterize this molecule protein family as early as 1999 is compiled in Table 2. Even though p10 was found in leg regeneration tissue rather than antennal sensory structures, it already underwent

three nomenclatures and three changes in 1999–2000, all of which had the same clear obvious meaning (chemosensory function), neglecting the first member of this protein gene family to be reported in the literature. Consequently, the protein gene family was renamed to exclude developmental functions and solely reflect putative chemosensory functions. At the time, it was believed that there was no connection at all between p10 relatives and mucins, cytoskeleton complexes, genetic elements, and intracellular processes (Figure 6).

**Table 2.** The family of "ChemoSensory Proteins" (CSP) goes by several names. It is necessary to adopt a new name or nomenclature without assuming anything about its ambiguous and somewhat perplexing functions (e.g., the small number of *CSPs* in many insect species, their ubiquitous expression, their largely outside expression of the chemosensory paradigm, their general tissue distribution, their ubiquitous presence during developmental patterns, their multi-existence in microbes, their association with fatty acids, their association with multiple different intracellular protein complexes, and their response to stress), a new name or nomenclature is needed.

| Name          | Organism                       | Authors           | Year |
|---------------|--------------------------------|-------------------|------|
| p10           | <i>Periplaneta americana</i>   | Nomura et al.     | 1992 |
| Ebsp-3/PebIII | <i>Drosophila melanogaster</i> | Dyanov et al.     | 1994 |
| A10           | <i>Drosophila melanogaster</i> | Pikielny et al.   | 1994 |
| OS-D          | <i>Drosophila melanogaster</i> | McKenna et al.    | 1994 |
| Pam           | <i>Periplaneta americana</i>   | Picimbon and Leal | 1999 |
| CSP           | <i>Schistocerca gregaria</i>   | Angeli et al.     | 1999 |
| SAP           | <i>Manduca sexta</i>           | Robertson et al.  | 1999 |
| Pherokine     | <i>Drosophila melanogaster</i> | Sabatier et al.   | 2003 |
| Mp10          | <i>Myzus persicae</i>          | Bos et al.        | 2010 |
| LA-BP         | <i>Bemisia tabaci</i>          | Liu et al.        | 2016 |
| Toxin-BP      | <i>Bemisia tabaci</i>          | Liu et al.        | 2016 |
| B-CSP         | <i>Acinetobacter baumannii</i> | Liu et al.        | 2019 |
| Lipid-BP      | <i>Bemisia tabaci</i>          | Liu et al.        | 2020 |
| JHRP          | <i>Aedes aegypti</i>           | Picimbon          | 2020 |
| Mucin module  | <i>Aedes aegypti</i>           | Liu et al.        | 2023 |
| TIF module    | <i>Aedes aegypti</i>           | Liu et al.        | 2023 |
| DNA-BP        | <i>Aedes aegypti</i>           | Liu et al.        | 2023 |

After one (polyclonal) antibody against p10 labeled some chemosensory structures in the adult antennae of the desert locust *S. gregaria*, Angeli et al. changed the name of the p10/OS-D protein family to "ChemoSensory Protein" ("CSP"; Angeli et al., 1999). This marked the fourth change to the p10 nomenclature, which still heavily downplays the developmental role in favor of a chemosensory function. Clones that resemble "CSP" in microbe bacterial species are referred to as "B-CSP" (Liu et al., 2019, 2020; Picimbon, 2023). On the other hand, the fact that bacterial species contain "CSPs" begs the question of what function these might serve in chemical detection. What role, if any, CSP molecules play in chemosensing or olfaction is still unknown. This protein molecule family has been shown to function independently of the chemosensory system since 2003 (Picimbon, 2003; Sabatier et al., 2003). The molecules known as "Pherokines" are those that are highly present in the hemolymph of flies as a result of microbial, bacterial, or viral infection (Sabatier et al., 2003; Einhorn & Imler, 2019; Table 2), introducing the idea of p10 or CSP molecule in the immune response. Eventually, it was proposed to rename these molecules as "Cuticular Sensory Proteins" ("CSPs") to highlight their strong expression in immune barriers, exoskeleton (to which muscles are attached), epidermis, epicuticles, insect major outer surfaces, antennae, legs, and sensory organs (Picimbon and Regnault-Roger, 2008; Einhorn & Imler, 2019). Thus, as information about this molecule protein gene family has grown, new names for it have also appeared, reporting functions far beyond the antennal sensory branch; however, none of them have addressed the possible involvement of "CSP" molecule in connection to actin-related protein complex, RNA/DNA, larger precursor molecules, mucins, nuclear

complex proteins, transcription initiation, nucleoside kinase, and genetic regulation as we have here (Figures 4-6 & S1-S3).

What distinguishes OS-D, A10, SAP, and CSP from one another? Many synonyms (see Table 2) entirely overlook the most remarkable characteristic of this molecule protein family, which is expressed throughout the entire body of the insect. The entire molecule family, "OS-D", "SAP", or "CSP", needs to be renamed in order to account for all family members across the different arthropod and insect species as well as across the different species in the bacterial prokaryotic superkingdom (Liu et al., 2020; Picimbon, 2023). The first "physiological" or "pharmacological" effect that is observed for a neuropeptide is usually used to name it, and this naming convention frequently turns out to be fairly inaccurate. It is probably best to rename all naturally occurring peptides, carrier proteins, "CSPs", and similar molecules in a more methodical and objective manner. The location and wide range of activities listed for this molecule protein family conflict with the different names used to identify SAP or CSP members. It is necessary to come up with a new term or nomenclature without assuming any known functions already in place or limiting it to chemosensory or olfactory structures.

The definition of "SAP" and "CSP" is no longer valid, as there is growing evidence that CSPs express in the gut and metabolic organs and do not play a unique and crucial role in chemosensing. Instead, it now refers to membership in a subset of soluble protein molecules that share a high degree of structural similarity and a particular four-cysteine pattern (see Figure 1), but are all found at high levels in a large variety of organs and tissues (see Figure 2; Xuan et al., 2015; Liu et al., 2020). When referring to the entire molecule protein family, it becomes somewhat inappropriate to use the terms "SAP" and "CSP", which exactly translate to "Sensory Appendage Proteins" and "Chemosensory Proteins" expressed in sensillum, respectively. This term should not be used to refer to all the genes and proteins expressed in a variety of sensory and non-sensory (metabolic) cell tissues and hemolymph, as well as those that have evolutionary ancestry with organisms ranging from bacterial microbes to honeybees (Picimbon, 2023). Perhaps the information needed to correctly name this molecule protein family is now available, heartfelt thanks to the analysis of the genome and Expressed Sequence Tag (EST) databases of bacteria, insects, sea crustaceans, and arthropods in the context of molecular data demonstrating that CSP proteins are not exclusively tuned to olfactory and taste chemosensory organs (Figures 2-6 & S1-S3).

This circumstance is comparable to that of "lipocalins", a term for a superfamily of widely distributed, heterogeneous proteins that transports small hydrophobic molecules such as lipids and steroids (from the Greek *lipos*, which means fat, and Greek *kalyx*, which means cup). Unlike lipocalins, the term "CSP" family describes a group of homogenous, evolutionary well-conserved molecules with a characteristic sequence of four cysteines, tissue profiling (ubiquitously expressed), and rather diverse binding properties (not only to long fatty acids and straight lipid chains, C18:2, but also to cyclic compounds like cinnamonaldehyde; Liu et al., 2016b). In contrast, molecules belonging to the lipocalin family are diverse, heterogeneous, and poorly conserved throughout evolution (Diez-Hermano et al., 2021). Therefore, even though the gut and the fat body—which are believed to be the insect body's main organs for storing lipids and fatty acids as energy—produce the majority of the CSP molecules, it is difficult to distinguish between groups and sub-groups within the CSP family. Through the process of lipolysis, these lipids (C18:2) are made available to other organs for use as fuel in their development, regeneration, or growth, as well as in their response to an infectious agent or insecticide (Xuan et al., 2015; Liu et al., 2014; 2016b, 2020; Picimbon, 2019; Einhorn and Imler, 2019). According to Xuan et al. (2015), the great majority of CSPs are taking part in the moth response to avermectin insecticide. According to Loftus et al. (2005), Verjovski-Almeida et al. (2005), Noriega et al. (2006), Nene et al. (2007), *Plasmodium gallinaceum*, *Brugia malayi*, and Dengue virus infections in *Aedes*, and bacterial infections in *Culex*, CSPs are also highly prevalent. Rather than being related to chemosensing, the abundance of CSP genes in moths and mosquitoes is likely related to toxin, insecticide, viral, or microbial resistance (Picimbon, 2020). Like in all other mosquito species, *Ae. Aegypti*, the dengue fever mosquito, has another use for CSPs. These molecules are found in both adult *Aedes* and larvae, but they are primarily concentrated in the adult's fat bodies, *corpora allata* (sources of JH), and salivary glands (Picimbon, 2020). This suggests that, at least in the

yellow fever mosquito *Ae. aegypti*, “CSPs” are involved in the binding, transport, and biosynthesis of juvenile hormone (JH). The function of CSP molecule in relation to juvenoids 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. According to Kim et al. (2017), this function would be comparable to that of the OBPs in the hemolymphatic transport of JH molecule. JH is known to control the processes of development in insects and crustaceans (Laufer et al., 1987). Therefore, JH-related proteins, or JHRPs would be a better name for these proteins than SAPs or CSPs. There are no sensory appendages on *corpora allata*. Nearly every facet of physiology in both systems is regulated by JH function in both larvae and adults. Caste differentiation, pheromone production, development, male genital morphogenesis, female ovarian egg production, social and mating behaviors, and immune function are all regulated by JH (Jindra et al., 2013). Specifically, it has been shown that JH controls neuronal plasticity linked to brain structure, neurogenesis, and behavioral maturation; this is rather similar to the pattern of CSP expression in crickets, moths and honeybees (Cayre et al., 1994; Liu et al., 2020; Anton and Rossler, 2021). According to Bian et al. (2019), even the prothoracic glands (ecdysone) express CSP, which is incompatible with a function in olfaction. The fact that prokaryotic cells and bacteria share the CSPs and JH relationship is a crucial point in this analysis. Phurelipids are secondary metabolites that bacteria produce. Phurelipids and JH have a similar molecular structure. Not only did JH and other phurelipids hinder insects' immune responses, but they also prevented pupae from developing into adults (Ahmed et al., 2022). From the binding of a wide range of exogenous foreign toxic chemicals to the intracellular transport of endogenous fuel C18:2, linoleic acid, lipids, hormones, and fatty acid precursors, all of these observations strongly imply that “CSP” molecules serve a variety of functions in the immune defense system of microbes, bacteria, crustaceans, insects and arthropods (Liu et al. 2016a,b, 2017, 2020; see Figure 6). This is accurate, unless it can be shown that CSPs are part of the microbe bacterial olfactory hoedonics. Even then, it's not clear how “CSPs” connect to the transcription factors, RNA-BPs, Mucins, nuclear pores, DNA-regulatory proteins, SAmkC, WAS/WASL-like, CWP, actin-linkers, and other transcriptional factors that have been discussed thus far (Figures 4-6 & S1-S3). It is even less clear how CSPs could have evolved and been subject to natural selection in viruses prior to cell-based life (see Figures 4-5B; Moelling and Broeker, 2019).

It is critical to look into the relationship between CSP, Mucin, NPCP, and other cell regulator molecules rather than just concentrating on chemosensing (see Figure 6). The intriguing and significant attachment of CSP proteins to larger molecules, like mucins, has implications for both evolutionary and functional issues. If the exon theory of genes—also referred to as the introns-early theory—is correct, it will depend on how CSPs can connect to mucin and/or what role CSP plays as mucin's N-tail. According to this theory, exons may correspond to the boundaries of autonomous modules or small structural domains that assembled to form protein molecules during their evolutionary process. A CSP module, a block from the amino acid pattern around Cys53–Cys56, is present in both the large mRNA molecules that code for the protein TSSC1 and the pol-like protein (reverse transcriptase, RNase H, and exo/endo phosphatase; ACYPI000612 and ACYPI005051, respectively). Furthermore, transcription factors, mucin variants, GP Ib-like, NPCPs, ASRPs, and an extensive list of nucleotide-binding proteins are just a few of the many large molecules that we assess in our comparative analysis of CSP-RNA binding proteins (Table 1 and Figures 4-6 & S1-S3). The N-terminal module in this instance consists of the entire CSP, which runs counter to the protein's intended function of odor sensing (see Figures 5 & S3).

Long lipid fatty acid chains and CSPs are activated for female sex pheromone synthesis in a variety of insect species, though moths in particular (Ohnishi et al., 2009; Xuan et al., 2014, 2016, 2019; Picimbon, 2017, 2019). Gene expression patterns and intracellular signaling are just two of the many cellular processes in bacteria that are greatly impacted by exogenous long lipid fatty acid chains. DiRusso and Black (2004) state that these transcription patterns primarily affect the proteins needed for fatty acid biosynthesis and breakdown. This puts the “CSP” molecules inside the cell rather than in the extracellular environment of the chemosensory neurons, which is consistent with the current findings: a very ancient gene with a highly conserved structure, that is expressed throughout an

insect's or crustacean's entire body, spreads throughout many developmental stages, and is found in bacteria and possibly viruses (Picimbon, 2023; Figures 1-6). In addition, it binds to a broad range of substrates, lipids, fatty acids, transcription regulator families, DNA and intracellular elements (Liu et al., 2016b; Figures 4-6). Therefore, it is imperative that this protein family be given a new name. [on the basis of their existence in viruses, bacteria and arthropods, their widespread expression profiling, their regulation by xenobiotics, drugs, toxins, viruses, and microbial agents, in many different sensory and non-sensory organs of the insect body, and their ability to bind numerous non-semiochemicals, non-sensory endogenous ligands like long fatty acid lipidic chains and genetic components like DNA or RNA]. The new name should ideally not refer to any particular function because the evidence for a given function may be weak and it may change quickly, even in related groups of proteins. CSP/SAP is an example of a peptide molecule that, in many cases, would be better categorized according to its structure rather than just its function, particularly in cases where the physiological function is still up for debate.

It's probably best if the peptide is just described in the new nomenclature (Figure 1). We could suggest using the general term "*lipoclistins*" (from Greek Λίπος lipos=fat and Greek κλειστό kleistó=closed in) for CSP and OBP molecules ("lipoclistins are proteins that enclose lipid ligands, just like lipocalins do"; citing Rudolf Alexander Steinbrecht, 2003, 2020; Max-Planck Institute for Biological Intelligence, Seewiesen, Germany). The term "*lip-anoiktins*" (perhaps "*lip-aniktins*", as οι in modern Greek is pronounced i; derived from Greek words for "fat" and "open", respectively; open=ἀνοικτός anoiktos) may be a better way to describe the main structural component of the CSP –an open-air structure— according to this system of naming based on words and names' Greek ancestry, following the nomenclature of COVID-19. World Health organization has designated key strains of SARS-CoV-2, the virus that causes COVID-19, with labels that are clear, easy to say, and memorable. From now on, all virus variations will be called by their Greek alphabetic names, like "Alpha", "Beta", "Gamma", "Epsilon", or "Omicron". For CSP protein, the Latin word "*apertus*" might be a simpler, more tasteful term (citing Rudolf Alexander Steinbrecht).

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 bacteria and prokaryotic cells (Liu et al., 2019, 2020; Picimbon, 2023; WP\_149730592). 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 conclusions regarding their physiological task. At the next international conference on very broad nomenclature, this would be a name to discuss. A good starting point for renaming all molecules or molecule families whose names don't accurately reflect their functions would be the renaming of CSPs.

CSPs currently have a completely disorganized nomenclature. Even worse now that we have so much more knowledge than we did twenty years ago when we only considered expression in the antennae. "*Chemosensory*" does not refer to molecules that are expressed widely throughout the insect and linked to intracellular systems (Figure 6).

## 8. Evolutionary Networks and Associated Diseases

Originally, the idea of binding protein matrices for chemo-sensors was developed through research on the potential of "CSPs" as semio-chemical carriers (Lu et al., 2014; Zhu et al., 2019). But if we accept to take into account the entire current status of data gathered on this molecule family, CSP matrices could greatly aid in the development of other types of sensors, such as lipid, RNA mutation, and DNA sensors, as well as sensors of insecticide and bio-environmental pollutants (Liu et al., 2017, 2019, 2020; Kramer et al., 2022). In order to capture, bind, and transport hydrophobic odor molecules across the sensillum in sensillar lymph and control insect behavior, chemosensory proteins (CSPs) are not significant molecular components of the insect olfactory system (Liu et al., 2020). Since CSPs have long been willingly restricted to insects and other arthropods, no biomedical application was anticipated. This was unfortunate since these binding proteins had a high affinity for non-semiochemicals like linoleic acid, as reported by Liu et al. (2016b). This information may pave the way for new applications of CSP research, particularly in biomedicine and disorders linked to lipids. Another example is the study of the moth "pheromone"-binding protein (PBP), which has been shown to bind to specific vitamin compounds with a higher affinity (Guo et al., 2021). This finding is crucial in order to challenge "Reverse Chemical Ecology" (the idea that PBP may be used to discover pheromones; Zhu et al., 2017; Choo et al., 2018; Zaremska et al., 2022) and suggest using PBP in both health and sickness instead (Delrue and Speeckaert, 2023). Vitamin K transport is necessary for strong bones (Kohlmeier et al., 1996). Furthermore, because CSP molecules and other "olfactory binding proteins" can transfer C18-lipids—that is, oleic acid, palmitic acid, and linoleic acid—to degradative enzymes, they can potentially treat COVID-19 (Picimbon, 2023b). Restricting CSPs' (and OBPs') involvement in chemosensing would limit their use to anosmia and aguesia (Picimbon, 2002, 2003; 2023b; Ozaki, 2019). Restricting their involvement in insects and other arthropods, as well as too willingly ignoring their existence in the microbe bacterial kingdom, may result in excluding and ignoring their strong potential impact in infectious diseases and human medicine. Besides, if we can verify that CSPs are a component of glycoproteins in the viral outer surface membrane envelope, the potential in infectious diseases and human medicine will become even more significant.

Several studies show that CSP molecules are found in Eukaryotes, arthropods, insects, winged insects like moths and whiteflies, wingless insects like head and body lice, and Prokaryotes, the most primitive cellular systems lacking lymph and chemosensory dendrites. Many different bacterial strains express CSPs, including very small marine Actinobacteria and Proteobacteria such as *Acinetobacter*, *Aeromonas*, *Escherichia*, *Klebsellia*, *Lysobacter*, *Macrococcus*, *Salinimonas*, and *Vibrio* (Picimbon, 2023a). These strains are known to cause serious infectious diseases in both animals and humans (Bendary et al., 2022). The majority of the bacterial species that harbor CSPs, including *Moraxella*, *Staphylococcus*, *Streptomyces*, *Enterobacter* and *Xanthomonadales*, are also known as critical symbionts of the insect gut, which is crucial for insect control (Rupawate et al., 2023). Among many other physiological systems, the gut microbiota is necessary for the growth, development, detoxification, feeding, immunological response, and environmental adaptation of insect pest hosts (Zhao et al., 2022; Suenami et al., 2023). Combining different bacterial species with CSP molecules could be a big step toward microbial control of insect pests in agriculture and potential human vectors (Hunter, 2019; Vega Rúa and Okech, 2019; Rooney et al., 2019).

Many types of bacteria, such as *E. coli*, *Staphylococcus aureus*, and *A. baumannii*, can cause harm to both humans and invertebrates. Some bug species are potential sources of severe human infection when the pathogen is not environmental in origin. This is the case, for instance, with the multi-resistant *Pseudomonadales* *A. baumannii*, which is known to be a symbiotic species of the head louse and human body louse (*Pediculus*) as well as a bacterial pathogen linked to a variety of diseases (pneumonia, meningitis, urinary tract and bloodstream infections; Liu et al., 2019; Morris et al., 2019; Hammoud et al., 2021; Larkin et al., 2023). Risky bacterial strains including *Bartonella quintana*, *Borrelia recurrentis*, and *Rickettsia prowazekii* cause fever, trench fever, relapsing fever, and epidemic typhus, which are all primarily spread by *Pediculus* lice (Badiaga and Brouqui, 2012; Cheslock and Embers, 2019). Therefore, investigating the potential genes that were transferred from the bacterium to the louse may be crucial to stopping many infectious diseases that affect humans, such as typhus.

Research on CSPs—prospective molecules that can be transferred from human to louse—is necessary to understand infectious diseases spread by louses, investigate bacterial immunity in the body, and investigate epidemic diseases carried by vectors.

According to cutting-edge molecular research on insects, toxicity tolerance to chemical and biological toxins is directly correlated with binding protein molecules like CSPs and OBPs (Sabatier et al., 2003; Liu et al., 2014, 2016, 2020; Xuan et al., 2015; Einhorn and Imler, 2019; Guo et al., 2021). The widespread use of insecticide products like plant-derived pyrethrin neurotoxins (Permethrin) for control has led to a variety of lice developing resistance to many families of insecticide molecules, much like mosquitoes (Le Goff and Giraudo, 2019). This has caused the emergence or more re-emergence of lice and louse-borne infections in many countries around the world (Thakur and Chauhan, 2022). The small evolutionary network of CSP-encoding genes, which are present in the insect immune response primary pathways of *Drosophila*, *Anopheles*, *Apis*, *Nasonia*, and *Pediculus*, does not imply that these molecules are converted into trillions of odorants, but rather to a small repertoire of lipids and/or genetic elements (Liu et al., 2019, 2020; Picimbon, 2020). Although the evolutionary network of *CSPs* is much richer in plant-eating insects and mosquitoes such as *Aedes* and *Culex*, most *CSP* genes are duplicated copies of the same gene that are strictly identical in most cases (Picimbon, 2020). When the number of a gene varies between species due to a phenomenon in which portions of the genome are repeated, it is not chemosensory system growth or development; rather, it is copy number variation or gene dosage (Hastings et al., 2009). It would be interesting to look for the specific gene dosage within the evolutionary network of *CSPs* in bacteria, as gene dosage would represent the number of copies of the gene in a bacterial cell and gene expression would primarily depend on higher and lower dosage. CSP cannot recognize a broad range of ligands in bacteria if the dosage is the same as that of *Drosophila* (4) and *Pediculus* (6). The question of whether CSP can activate or regulate specific genetic elements is raised by the gene dosage. It doesn't require a wide range of CSP molecules to play a control regulatory role in the expression of the genome if it can accomplish both (Figure 3). It would be especially important to investigate CSP molecules in bacterial species such as *A. baumannii*, *M./S. caseolyticus*, *K. griseola*, and *E. coli*, as they may offer an essential gene regulatory mechanism for toxin bactericide resistance in certain dangerous strains that are the main cause of serious infectious diseases. It is likely that early in their evolutionary histories, both microbial and insect host species experienced comparable resistance mechanisms, and it is possible that bacteria transmitted *CSPs* to insects through horizontal transfer (Liu et al., 2016a, 2017, 2020; Liu and Picimbon, 2017). This would explain the similarity between the CSP molecules found in silkworm moths and bacterial symbionts (Xuan et al., 2015; Picimbon, 2023a). CSPs may be involved in the transfer of resistance from one bacterial cell to another, as demonstrated by *E. coli* strains that can spread their resistance to the same antibiotic to nearby bacterial cells (Poirel et al., 2018). This would primarily involve how CSPs interact with promoter sequences and/or other regulatory components of gene expression involved in a cell's or group of cells' response to a hazardous environment, based on the similarity revealed here between CSPs, transcription factors, RNA-BPs, DNA elements, actin complexes, NPCPs, allergens, and mucins (Figures 4-6 & S1-S3).

If CSPs existed before the world of cells and before viruses? If *CSP* genes are present in bacterial cells and all of their progeny, including birds and reptiles, even though it is unknown how insects evolved from bacteria? That is probably the cause of the great diversity of CSPs that we found in crab species that are known to store lipids and to switch between sexual and asexual reproduction based on the activation of specific genes (Ye et al., 2023; Picimbon, 2023a). Furthermore, it is recognized that they trigger the activation of "CSP" genes in reaction to specific environmental stressors, including elevated temperatures, elevated salinities, and chemical toxicity (Picimbon, 2023a). The common pattern of evolution in CSP molecules that we report here in a comparative analysis of "chemosensory proteins", mucins, TIFs, and "DNA-binding proteins" is quite intriguing, despite the fact that the protein structure, in particular the N-terminus, is largely retained from insects to bacteria (Picimbon, 2023a). Although further research is required, our analysis reveals a common function for TIFs, "CSPs", and DNA/RNA binding proteins that must be none other than regulating crucial elements of gene expression.

It is worth noting that the CSP molecule (Mp10, no ligand) induces plant immunological responses when injected into the phloem, in a manner akin to that of microbial species associated with honeydew (Bos et al., 2010; Elzinga et al., 2014; Rodriguez et al., 2014; Wari et al., 2019; Pirtilla et al., 2023). Although the release of phytoalexins, jasmonic acid, salicylic acid, and/or volatile organic compounds (VOCs) from leaves is facilitated by CSP proteins, other salivary effector proteins, and/or honeydew microbes (Cui et al., 2019), it would be interesting to find out if specific CSP structures, the N-terminal tail, and/or the inoculation of specific bacteria, such as *A. baumannii*, *M./S. caseolyticus*, *K. griseola*, and *E. coli*, activate defense genes similar to those reported for plant-virus-vector interactions in human tissue, as well as target transcription factors and/or some protein degradation pathways (Ray and Casteel, 2022). The ability of CSPs to transport linoleic and benzoic acids may mediate the effects of CSP injection in a plant or human body, according to the results in *B. tabaci* whiteflies (Liu et al., 2016b). However, our comparison of CSPs to DNA/RNA-binding proteins (this study) may also suggest that CSPs activate specific genes via their N-terminal tail (Figures 4-6 & S1-S3). Benzoic acid and lipidic acid are the sources of plant hormones belonging to the jasmonate and salicylate classes, respectively. According to Bajguz and Piotrowska-Niczyporuk (2023), "CSP" or "Mp10" injection may help stimulate specific hormone biosynthesis pathways. During metabolism, dihomo- $\gamma$ -linolenic acid (DHA) is produced from linoleic acid (LA), an essential fatty acid for humans, crustaceans, and many insect species. This molecule (DHA) is an essential part of phospholipids found in neuronal membranes and serves as a substrate for the synthesis of prostaglandins, which are believed to be essential for preserving nerve blood flow. LA performs numerous functions both directly and indirectly through its metabolites. Energy production can occur in both sensory and non-sensory tissues. It functions as one of the structural components of the cell membrane. It is thought to be an essential precursor to arachidonic acid, which is vital for various physiological processes, including neurodevelopment. LA maintains the epidermis of the skin resistant to water loss and is also necessary for the health of the human heart. When saturated fat is substituted with LA, randomized clinical trials have shown that it lowers both total and LDL cholesterol (Farvis et al., 2014). There is some proof that LA raises blood pressure and insulin sensitivity (Moloney et al., 2004; den Hartig, 2019). Consequently, "CSP" molecules like Mp10 may be viewed as lipid transporters and human health mediators rather than odor scavengers in insects. Although most organisms, including fish, crabs, lobsters, prawns, shrimp, marine crustaceans, molluscs, and terrestrial isopods, are thought to be incapable of producing LA on their own, it is readily produced by bacteria, protozoa, and plants (Malcicka et al., 2018). For these "animals", odor chemical communication is therefore unlikely to be significantly influenced by LA. This lipid molecule family and dietary linoleic acid serve as metabolic precursors for the synthesis of a wide range of different chemicals, and animals' ability to survive and grow depends largely on their nutrition (Liu et al., 2020). CSPs play a critical role in this process. However, given the striking similarity between CSPs and the N-tail of TIFs, Mucins, and numerous other proteins (Figures S2-S3), it is also possible that "CSP" directly activates genes associated with hormones, enzymes, jasmonates, and salicylates. Further research is required to determine the effects of injecting CSP or Mucin-N-termini through or via the skin of humans, shrimp, or plants on diseases related to lipid metabolism, hormone production, or genetic issues in plants, animals, and humans. Though gene activators have specific activities that are highly dependent on cell or tissue function, developmental stage, or pathological situation, fatty acid lipid transporters are present in almost all intracellular organelles (Ma, 2011; Ko et al., 2017; Cramer, 2019). It's possible that CSP therapy enhances the dispersion of lipid droplets following ingestion, akin to a chylomircron, or during tissue growth or hormone synthesis by promoting the transport of lipids in the blood, lymph, sap, gut, glandular secretion, and body fluids. It is also possible that the use of CSP molecules controls genes related to biosynthesis, which leads to the synthesis of hormones and/or oxidation of lipids. This dual activity—lipid transport and contact with RNA/DNA—may be essential for the functional regulation of prokaryotes and eukaryotic cells (see Figure 6). The fields of clinical care, human health, and plant protection, as well as agriculture, could all benefit greatly from more research on CSP linked to DNA. Many clues point to the presence of a well-established active lipid metabolism in bacteria, plant, or

human cells and tissues (Teng et al., 2017; Lamichhane et al., 2021; Jian et al., 2022; Macabuhay et al., 2022). The hypothesis behind this bacterial lipid hijacking, which is similar to that of a viral particle, has been supported by several studies (van der Meer-Janssen et al., 2010; Toledo and Benach, 2015; Allen and Martinez, 2020; O'Neal et al., 2020). It implies that the pathogen can spread to different tissues and become even more severely infected by breaking down the lipid molecules of the host while enriching its own stocks. Should CSP molecules possess the capacity to simultaneously interact with lipids and impact the expression of metabolic genes, they could play a pivotal role in impeding the process of lipid hijacking and vector-borne illnesses. Lipid molecules are primarily involved in controlling cellular function in bacteria, as they are in many other organisms, such as humans and insects. Furthermore, it is recognized that viruses regulate lipid synthesis, signaling, and metabolism in order to modify their host cells into the ideal conditions for virus replication (Heaton and Randall, 2011). The functions of membrane cells, such as cell division and protection, and selective permeability, which allows all water-soluble nutrients to be exchanged with the environment, depend on fatty acids (Barák and Muchová, 2013; Diomandé et al., 2015; de Carvalho and Caramujo, 2018). As such, the synthesis of fatty acid lipids in bacteria is either a strictly regulated "take" or "make" process. The use of a specific pool of lipids with particular physicochemical characteristics may be required in the event of a major alteration in the environment or in the lifestyle of the microbe. This can be accomplished through the modification of pre-existing fatty acid lipids, the synthesis of new ones, or the efficient extraction of them from their environment. After infection, fatty acid- and "niche-specific" changes in lipid fluxes were observed. The infectious agent is primarily dependent on the host's lipid uptake, with the exception of blood, where the pathogen *A. baumannii* synthesizes its own fatty acids (Adams et al., 2021). Given that CSPs disrupt both lipid transport and the activation of metabolic genes involved in biosynthetic lipid synthesis, it is plausible that they have a substantial effect on the equilibrium between these processes. Modifying CSP molecules may interfere with the lipid pathways in the bacterial cell and impact both cellular and membrane activities, thereby blocking, decreasing, or eliminating the defense mechanisms of a particularly dangerous microbial strain.

## 9. Last reflections and conclusions

The existence of "CSPs" in bacteria and their interactions with fatty acids like LA disprove the overwhelming theory that these proteins have "chemosensory" capabilities. The low gene count, wide tissue distribution, expression at different developmental stages, sensitivity to chemical stress primarily in the gut and fat body, need for adaptation to any change in the cell environment, and excretion with saliva and hemolymph are some other arguments in favor of renaming the molecule family that should be taken into consideration in addition to this. Contrary to the "chemosensory" theory proposed for this molecule, CSP builds the N-terminal tail of some viral envelope glycoproteins (see Figure S3). Functional genomics, genetic engineering, clinical genetic cancer, cell regulation, eukaryotic, prokaryotic, and viral transcriptional mechanisms, medical genetics, medical microbiology, cell regulation, infection, and management of hereditary disorders may all benefit from further research on the role and binding of "CSPs" to lipids and genetic components, such as promoter regions that control gene expression in various systems.

Through the direct regulation of defense gene expression, novel methods to knockdown 'CSP' genes in pathogenic bacteria or insect pest species might potentially be developed. This would change the target species' lipid transport (fuel), lower virulence in drug- or insecticide-resistant strains, and decrease resistance to antibiotics, bactericides, and germicides. While CSP-LA may aid in understanding the connections between lipid metabolism and insulin resistance, CSP-RNA or -DNA may introduce some crucial ideas on the drug resistance of cancer cells. In numerous ways, controlling the activity of bacterial "CSPs"—which are more appropriately thought of as lipid scavengers and gene activators than as chemosensory molecules—looks to be a novel and extremely promising method for observing how lipids, "made or taken", host-virus interaction, RNAs, genes, genome, and chemical medication interact, as well as for lessening the pathogenicity of dangerous bacteria like *A. baumannii*, *E. coli*, *Klebsiella pneumonia*, *Mycobacterium tuberculosis*, *Pseudomonas*,

*Staphylococcus*, and *Streptococcus*, among many others. Changing the physiology of the symbiotic bacterial cell to cut off the parasite's main source of lipid fuel is another possible modern parasite management strategy for *Pediculus* or *Aedes*. Our finding that linoleic acid interacts with *Bemisia* 'CSPs' (Liu et al., 2016b, 2020) and their resemblance to Allergens, Transcription Initiation Factors, Cell Walls Proteins, DNA/RNA-binding proteins, Nuclear Pore Complex Proteins (NPCPs), Actin Skeleton Regulatory Proteins (ASRPs), and Mucins, as reported here (Figures 4-6 & S1-S3), suggests two approaches to treat infectious or vector-borne diseases: either focusing on a specific lipid component of bacterial cells, or eliminating the infectious agent that controls the metabolic genes of numerous insect pests and parasites.

The final conclusion is that too many variables preclude "CSPs" from having a part in chemosensing. In 2023, a chemosensory role may be supported only by the name of the particular molecule protein gene family and all of its synonymous terms. Even after thirty years, clones of this protein have been found in a wide range of organisms, including bacteria, arthropods, crustaceans, and insects, but its actual function is still unclear and the topic of much discussion. Starting from scratch, studies of the CSP molecule in these organisms (and viruses) may be able to solve the puzzle and provide new approaches for controlling transcription and gene expression. These approaches may involve utilizing the novel hypothesis that these proteins are intracellular, linked to lipid transport, long-chain fatty uptake, hormone biosynthesis pathways, actin-linked complexes, nuclear pore protein complex, cell wall, phosphorylation of G-protein, channels, enzymes, and receptors (see Figure 6).

As demonstrated here, CSPs can exist as small modules made up of six  $\alpha$ -helices or as a small module linked to a larger molecule, such as a mucin (which plays a variety of roles in protecting mucosal surfaces), a transcriptional initiation factor (TIF), which increases transcription of a gene or set of genes, or a positive regulator of the genome.

It would be irresponsible to overlook the potential that these new analyses bring to the fields of cell physiology, development, and clinical medicine. CSP molecules attach to linoleic acid to provide energy to all of the bacteria and cells in an insect's body; gel-forming mucins are required for the lubrication and removal of debris, pathogens, and allergens. Based on these two unavoidable facts, we must relaunch p10's plot.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

**Acknowledgements:** 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), and most cited paper of PLoS ONE (Biotype characterization, developmental profiling, insecticide response and binding property of *Bemisia tabaci* chemosensory proteins: role of CSP in insect defense; Liu et al., 2016b, Top 10%). Overseas high-level talent program-Taishan scholar title (#NO.tshw20091015), and Ministry of Science and Technology of China (National Expert, #G2022023033L). We would especially like to thank Prof. Drs. Alejandro Rooney (USDA) and Rudolf Alexander Steinbrecht (MPI) for their insightful and critical conversations regarding the renaming of "chemosensory" proteins (CSPs) and for exchanging opinions on the early draft of the manuscript.

## References

1. Abanes-De Mello, A., Sun, Y.L., Aung, S. and Pogliano K. (2002) A cytoskeleton-like role for the bacterial cell wall during engulfment of the *Bacillus subtilis* forespore. *Genes & Development*, 16, 3253-3264.
2. Abraham, D., Löfstedt, C. and Picimbon, J.-F. (2005) Molecular characterization and evolution of pheromone binding protein genes in *Agrotis* moths. *Insect Biochemistry and Molecular Biology*, 35, 1100-1111.
3. Adams, F.G., Trappetti, C., Waters, J.K., Zang, M., Brazel, E.B., Paton, J.C., Snel, M.F. and Eijkelkamp, B.A. (2021) To make or take: bacterial lipid homeostasis during infection. *mBio*, 12, e00928-21.
4. Albert, B., Kos-Braun, I.C., Henras, A.K., Dez, C., Rueda, M.P., Zhang, X., Gadal, O., Kos, M. and Shore, D. (2019) A ribosome assembly stress response regulates transcription to maintain proteome homeostasis. *eLife*, 8, e45002.

5. Ahmed, S., Hrithik, M.T.H., Roy, M.C., Bode, H. and Kim, Y. (2022) Phurelipids, produced by the entomopathogenic bacteria, *Photorhabdus*, mimic juvenile hormone to suppress insect immunity and immature development. *Journal of Invertebrate Pathology*, 193, 107799.
6. Allen, P.E. and Martinez, J.J. (2020) Modulation of host lipid pathways by pathogenic intracellular bacteria. *Pathogens*, 9, 614.
7. Angeli, S., Ceron, F., Scaloni, A., Monti, M., Monteforti, G., Minnoccia, A., Petacchi, R. and Pelosi, P. (1999) Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from *Schistocerca gregaria*. *European Journal of Biochemistry*, 262, 745-754.
8. Anton, S. and Rossler, W. (2021) Plasticity and modulation of olfactory circuits in insects. *Cell and Tissue Research*, 383, 149-164.
9. Badiaga, S. and Brouqui, P. (2012) Human louse-transmitted infectious diseases. *Clinical Microbiology and Infection*, 18, 332-337.
10. Bajguz, A. and Piotrowska-Niczypruk, A. (2023) Biosynthetic pathways of hormones in plants. *Metabolites*, 13, 884.
11. Barák, I. and Muchová, K. (2013) The role of lipid domains in bacterial cell processes. *International Journal of Molecular Sciences*, 14, 4050-4065.
12. Bendary, M.M., Abdel-Hamid, M.I., Alshareef, W.A., Alshareef, H.M., Mosbah, R.A., Omar, N.N., Al-Sanea, M.M., Alhomrani, M., Alamri, A.S. and Moustafa, W.H. (2022) Comparative analysis of human and animal *E. coli*: serotyping, antimicrobial resistance, and virulence gene profiling. *Antibiotics (Basel)*, 11, 552.
13. Beukes, N. (2004) Biogeochemistry: early options in photosynthesis. *Nature*, 431, 522-523.
14. Bian, H.X., Chen, D.B., Zheng, X.X., Ma, H.F., Li, Y.P., Li, Q., Xia, R.X., Wang, H., Jiang, Y.R., Liu, Y.Q. and Qin, L. (2019) Transcriptomic analysis of the prothoracic gland from two lepidopteran insects, domesticated silkworm *Bombyx mori* and wild silkworm *Antherea pernyi*. *Scientific Reports*, 9, 5313.
15. Blackman, R.L., Takada, H. and Kawakami, K. (1978) Chromosomal rearrangement involved in insecticide resistance of *Myzus persicae*. *Nature*, 271, 450-452.
16. Blank, H.M., Maitra, N. and Polymenis, M. (2017a) Lipid biosynthesis: When the cell cycle meets protein synthesis? *Cell Cycle*, 16, 905-906.
17. 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. and Polymenis, M. (2017b) Translational control of lipogenic enzymes in the cell cycle of synchronous, growing yeast cells. *EMBO Journal*, 36, 487-502.
18. Blobel, G. (2000) Protein targeting (Nobel lecture). *ChemBioChem*, 1, 86-102.
19. Blomquist, G.J., Tittiger, C. and Jurenka, R. (2018) Cuticular hydrocarbons and pheromones of arthropods. *Oils and Lipids: Diversity, Origin, Chemistry and Fate*. (ed. H. Wilkes) Hydrocarbons, Handbook of Hydrocarbon and Lipid Microbiology, pp. 1-32, Springer Cham, Switzerland.
20. Bos, J.I., Prince, D., Pitino, M., Maffei, M.E., Win, J. and Hogenhout, S.A. (2010) A functional genomics approach identifies candidate effectors from the aphid species *Myzus persicae* (green peach aphid). *PLoS Genetics*, 6, e1001210.
21. Browning, D. and Busby, S. (2004) The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, 2, 57-65.
22. Cahoon, E.B., Lindqvist, Y., Schneider, G. and Shanklin, J. (1997) Redesign of soluble fatty acid desaturases from plants for altered substrate specificity and double bond position. *Proceedings of the National Academy of Sciences*, 94, 4872-4877.
23. Campanacci, V., Lartigue, A., Hällberg, B.M., Jones, T.A., Giuici-Orticoni, M.T., Tegoni, M. and Cambillau, C. (2003) Moth chemosensory protein exhibits drastic conformational changes and cooperativity on ligand binding. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 5069-5074.
24. Cayre, M., Strambi, C. and Strambi, A. (1994) Neurogenesis in an adult insect brain and its hormonal control. *Nature* 368, 57-59.
25. Celorio-Mancera, Mde L., Sundmalm, S.M., Vogel, H., Rutishauser, D., Ytterberg, A.J., Zubary, R.A. and Janz, N. (2012) Chemosensory proteins, major salivary factors in caterpillar mandibular glands. *Insect Biochemistry and Molecular Biology*, 42, 796-805.
26. Cerone, M. and Smith, T.K. (2022) Desaturases: Structural and mechanistic insights into the biosynthesis of unsaturated fatty acids. *IUBMB Life*, 74, 1036-1051.
27. Chakraborty, A., Ashraf, M.Z., Modlinger, R., Synek, J., Schlyter, F. and Roy, A. (2020) Unravelling the gut bacteriome of *Ips* (Coleoptera: Curculionidae: Scolytinae): identifying core bacterial assemblage and their ecological relevance. *Scientific Reports*, 10, 18572.
28. Cheng-Guang, H. and Gualerzi, C.O. (2021) The ribosome as a switchboard for bacterial Stress response. *Frontiers in Microbiology*, 11, 619038.
29. Cheslock, M.A. and Embers, M.E. (2019) Human bartonellosis: an underappreciated public health problem. *Tropical Medicine and Infectious Disease*, 4, 69.

30. Choo, Y.M., Xu, P., Hwang, J.K., and Leal, W.S. (2018) Reverse chemical ecology approach for the identification of an oviposition attractant for *Culex quinquefasciatus*. *Proceedings of the National Academy of Sciences of USA*, 115, 714-719.
31. Cramer, P. (2019) Organization and regulation of gene transcription. *Nature*, 573, 45-54.
32. Cronmiller, E., Toor, D., Shao, N.C., Kariyawasam, T., Wang, M.H. and Lee, J.H. (2019) Cell wall integrity signaling regulates cell wall-related gene expression in *Chlamydomonas reinhardtii*. *Scientific Reports*, 9, 12204.
33. Cui, N., Lu, H., Wang, T., Zhang, W., Kang, L. and Cui, F. (2019) Armet, an aphid effector protein, induces pathogen resistance in plants by promoting the accumulation of salicylic acid. *Philosophical Transactions B of the Royal Society B: Biological Sciences*, 374, 20180314.
34. Damude, H.G., Zhang, H., Farrall, L., Ripp, K.G., Tomb, J.F., Hollerbach, D. and Yadav, N.S. (2006) Identification of bifunctional delta12/omega3 fatty acid desaturases for improving the ratio of omega3 to omega6 fatty acids in microbes and plants. *Proceedings of the National Academy of Sciences USA*, 103, 9446-9451.
35. de Carvalho, C.C.C.R. and Caramujo, M.J. (2018) The various roles of fatty acids. *Molecules*, 23, 2583.
36. Delrue, C. and Speeckaert, M.M. (2023) Vitamin D and vitamin D-binding protein in health and disease. *International Journal of Molecular Sciences*, 24, 4642.
37. den Hartig, L.J. (2019) Conjugated linoleic acid effects on cancer, obesity, and atherosclerosis: a review of pre-clinical and human trials with current perspectives. *Nutrients*, 11, 370.
38. Dias, R.O., Cardoso, C., Pimentel, A.C., Damasceno, T.F., Ferreira, C. and Terra, W.R. (2018) The roles of mucus-forming mucins, peritrophins and peritrophins with mucin domains in the insect midgut. *Insect Molecular Biology*, 27, 46-60.
39. Dibrova, D.V., Galperin, M.Y., Koonin, E.V. and Mulkidjanian, A.Y. (2015) Ancient systems of sodium/potassium homeostasis as predecessors of membrane bioenergetics. *Biochemistry (Mosc)*, 80, 495-516.
40. Diez-Hermano, S., Ganfornina, M.D., Skerra, A., Guttiérez, G. and Sanchez, D. (2021) An evolutionary perspective of the lipocalin protein family. *Frontiers in Physiology*, 12, <https://doi.org/10.3389/fphys.2021.718983>.
41. Dillon, R.J., Vennard, C.T. and Charnley, A.K. (2000) Pheromones: exploitation of gut bacteria in the locust. *Nature*, 403, 851.
42. Diomandé, S.E., Nguyen-The, C., Guinebretière, M.H., Broussolle, V. and Brillard, J. (2015) Role of fatty acids in *Bacillus* environmental adaptation. *Frontiers in Microbiology*, 6, <https://doi.org/10.3389/fmicb.2015.00813>.
43. DiRussro, C.C., Black, P.N. (2004) Bacterial long chain fatty acid transport: gateway to a fatty acid-responsive signaling system. *Journal of Biological Chemistry*, 279, 49563-49566.
44. Doudna, J.A. and Batey, R.T. (2004) Structural insights into the signal recognition particle. *Annual Review of Biochemistry*, 73, 539-557.
45. 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. and Garczarek, L. (2023) Differential global distribution of marine picocyanobacteria gene clusters reveals distinct niche-related adaptive strategies. *The ISME journal*, 17, 720-732.
46. Dyanov, H.M., Lyozin, G.I., Dzitoeva, S.G. and Korochkin, L.I. (1994) A cDNA of *Drosophila melanogaster* ejaculatory bulb specific protein III (PEBme III). #AAA87058, DMU08281 (RNA source: imago).
47. Einhorn, E. and Imler, J.L. (2019) Insect immunity; from systemic to chemosensory organs protection. *Olfactory Concepts of Insect Control-Alternative to Insecticides* (ed. J.F. Picimbon), pp. 205-229. Vol. 2 Springer Nature, Switzerland.
48. Elzinga, D.A., De Vos, M. and Jander, G. (2014) Suppression of plant defenses by a *Myzus persicae* (green peach aphid) salivary effector protein. *Molecular Plant Microbe Interactions*, 27, 747-756.
49. Farvis, M.S., Ding, M., Pan, A., Sun, Q., Chiuve, S.E., Steffen, L.M., Willett, W.C. and Hu, F.B. (2014) Dietary linoleic acid and risk of coronary heart disease: a systematic review and meta-analysis of prospective cohort studies. *Circulation*, 130, 1568-1578 (Harvard, School of Public Health).
50. Forêt, S., Wanner, K.W. and Maleszka R. (2007) Chemosensory proteins in the honeybee: Insights from the annotated genome, comparative analysis and expression profiling. *Insect Biochemistry and Molecular Biology*, 37, 19-28.
51. Fujiwara, Y., Wada, K. and Kabuta, T. (2017) Lysosomal degradation of intracellular nucleic acids—multiple autophagic pathways. *Journal of Biochemistry*, 161, 145-154.
52. Georgescu, F., Song, Y. and Dautant, A. (2020) Structure, folding and stability of nucleoside diphosphate kinases. *International Journal of Molecular Sciences*, 21, 6779.
53. Gestwicki, J.E., Lamanna, A.C., Harshey, R.M., McCarter, L.L., Kiessling, L.L. and Adler, J. (2017) Evolutionary conservation of Methyl-Accepting Chemotaxis Protein location in bacteria and Archaea. *Journal of Bacteriology*, 182, 6499-6502.

54. Ghai, R., Mizuno, C.M., Picazo, A., Camacho, A. and Rodriguez-Valera, F. (2013) Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Scientific Reports*, 3, 2471.

55. González-Caballero, N., Valenzuela, J.G., Ribeiro, J.M.C., Cuervo, P. and Brazil, R.P. (2013) Transcriptome exploration of the sex pheromone gland of *Lutzomyia longipalpis* (Diptera: Psychodidae: Phlebotominae). *Parasites and Vectors*, 6, 56.

56. Gopikrishnan, V., Radhakrishnan, M., Krupakar, P., Manigundan, K., Abirami, B., and Reshma, S. (2023) Chapter 36 – Endosymbiotic interactions of actinobacteria with the insects. *Microbial Symbionts: Function and Molecular Interactions on Host*. (ed. D. Dharumadurai), pp. 645-658, Academic Press.

57. Guo, X., Xuan, N., Liu, G.X., Xie, H.Y., Lou, Q.N., Arnaud, P., Offmann, B. and Picimbon, J.F. (2021) An expanded survey of the moth PBP/GOBP clade in *Bombyx mori*: new insight into expression and functional roles. *Frontiers in Physiology*, <https://doi.org/10.3389/fphys.2021.712593>

58. Guo, W., Wang, X., Ma, Z., Xue, L., Han, J., Yu, D., Kang, L. (2011) CSP and Takeout genes modulate the switch between attraction and repulsion during behavioral phase change in the migratory locust. *PLoS Genetics*, 7, e1001291.

59. Hammoud, A., Louni, M., Baldé, M.C., Beavogui, A.H., Gautret, P., Raoult, D., Fenollar, F., Misse, D. and Mediannikov, O. (2021) Molecular characterization and genetic diversity of haplogroup E human lice in Guinea. *Microorganisms*, 9, 257.

60. Han, L. and Luan, Y.S. (2015) Horizontal transfer of small RNAs to and from plants. *Frontiers in Plant Science*, 6, 1113.

61. Han, R., Feng, X.Q., Vollmer, W., Stoodley, P. and Chen, J. (2023) Deciphering the adaption of bacterial cell wall mechanical integrity and turgor to different chemical or mechanical environments. *Journal of Colloid and Interface Science*, 640, 510-520.

62. Harrison, E. and Brockhurst, M.A. (2012) Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends in Microbiology*, 20, 262-267.

63. Hastings, P.J., Lupski, J.R., Rosenberg, S.M. and Ira, G. (2009) Mechanisms of change in gene copy number. *Nature Review Genetics*, 10, 551-564.

64. Heaton, N.S. and Randall, G. (2011) Multifaceted roles for lipids in viral infection. *Trends in Microbiology*, 19, 368-375.

65. Heinisch, J.J. and Rodicio, R. (2018) Protein kinase C in fungi—more than just cell wall integrity. *FEMS Microbiology Reviews*, 42, fux051.

66. Hilgemann, D.W. (2003) From a pump to a pore: How palytoxin opens the gates. *Proceedings of the National Academy of Sciences*, 100, 386-388.

67. Ho, S. (2008) The molecular clock and estimating species divergence. *Nature Education*, 1,168.

68. Hou, S., Guo, L., Xu, B., Dong, H., Zhang, S., Fu, Y., Shi, J., Li, L., Fu, J., Shi, F., Meng, Y. and Jin, Y. (2022) Trans-splicing facilitated by RNA pairing greatly expands sDscam isoform diversity but not hemophilic binding specificity. *Science Advances*, 8, DOI: 10.1126/sciadv.abn9458.

69. Hu, Y., Huang, H., Hui, X., Cheng, X., White, A.P., Zhao, Z. and Wang Y. (2016) Distribution and evolution of *Yersinia* Leucine-Rich Repeat proteins. *Infection and Immunity*, 84, 2243-2254.

70. Huang, Y., Li, L. and Rong, Y.S. (2022) JiangShi (僵尸): a widely distributed Mucin-like protein essential for *Drosophila* development. *G3 (Bethesda)*, 12, jkac126.

71. Hunter, D. (2019) Locusts in the world. *Olfactory Concepts of Insect Control-Alternative to Insecticides*. (ed. J.F. Picimbon), pp. 29-49. Vol. 1, Springer Nature, Switzerland.

72. Hunter, W., Martinez-Torres, D., Rahbe, Y., Sabater-Munoz, B., Stern, D., Tagu, D. and Wincker, P. (2004) An expressed sequence sequence tags database for the pea aphid *Acyrthosiphum pisum* #CN760202.

73. Ibarra, A. and Hetzer, M.W. (2015) Nuclear pore proteins and the control of genome functions. *Genes Development*, 29, 337-349.

74. Ikegami, A., Honma, K., Sharma, A. and Kuramitsu, H.K. (2004) Multiple functions of the leucine-rich repeat protein LrrA of *Treponema denticola*. *Infection and Immunity*, 72, 4619-4627.

75. Ingham, V.A., Anthousi, A., Douris, V., Harding, N.J., Lycett, G., Morris, M., Vontas, J. and Ranson, H. (2020). A sensory appendage protein protects malaria vectors from pyrethroids. *Nature*, 577, 376-380.

76. Ishii, E. and Eguchi, Y. (2021) Diversity in sensing and signaling of bacterial sensor histidine kinases. *Biomolecules*, 11, 1524.

77. Jansen, S., Zídek, L., Löfstedt, C., Picimbon, J.F. and Sklenar, V. (2006) <sup>1</sup> H, <sup>13</sup> C, and <sup>15</sup> N resonance assignment of *Bombyx mori* chemosensory protein 1 (BmorCSP1). *Journal of Biomolecular NMR*, 36: 47.

78. Jansen, S., Chmelik, J., Zídek, L., Padra, P., Novak, P., Zdrahal, Z., Picimbon, J.F., Löfstedt, C. and Sklenar, V. (2007) Structure of *Bombyx mori* Chemosensory Protein 1 in solution. *Archives of Insect Biochemistry and Physiology*, 66, 135-145.

79. Jia, Q., Zeng, H., Zhang, J., Gao, S., Xiao, N., Tang, J., Dong, X., and Xie, W. (2021) The crystal structure of the *Spodoptera litura* Chemosensory Protein CSP8. *Insects*, 12, 602.

80. Jian, Z., Zeng, L., Xu, T., Sun, S., Yan, S., Zhao, S., Su, Z., Ge, C., Zhang, Y., Jia, J. and Dou, T. (2022) The intestinal microbiome associated with lipid metabolism and obesity in humans and animals. *Journal of Applied Microbiology*, 133, 2915-2930.

81. Jin, X., Brandazza, A., Navarrini, A., Ban, L., Zhang, S., Steinbrecht, R.A., Zhang, L. and Pelosi, P. (2005) Expression and immunolocalization of odorant-binding and chemosensory proteins in locusts. *Cellular and Molecular Life Sciences*, 62, 1156-1166.

82. Jindra, M., Palli, S.R. and Riddiford, L.M. (2013) The juvenile hormone signaling pathway in insect development. *Annual Review of Entomology*, 58, 181-204.

83. Kazaz, S., Miray, R. and Baud, S. (2021) Acyl-acyl carrier protein desaturases and plant biotic interactions. *Cells*, 10, 674.

84. Kenis, S., Istiban, M.N., Van Damme, S., Vandewyer, E., Watteyne, J., Schoofs, L. and Beets, I. (2023) Ancestral glycoprotein hormone-receptor pathway controls growth in *C. elegans*. *Frontiers in Endocrinology (Lausanne)*, 14, 1200407.

85. Kim, I.H., Pham, V., Jablonka, W., Goodman, W.G., Ribeiro, J.M.C. and Andersen, J.F. (2017) A mosquito hemolymph odorant-binding protein family member specifically binds juvenile hormone. *Journal of Biological Chemistry*, 292, 15329-15339.

86. Kim, S.D., Pahuja, K.B., Ravazzola, M., Yoon, J., Boyadjiev, S.A., Hammamoto, S., Schekman, R., Orci, L. and Kim, J. (2012) The SEC23-SEC31 interface plays critical role for export of procollagen from the endoplasmic reticulum. *Journal of Biological Chemistry*, 287, 10134-10144.

87. Kirkness, E.F., et al. (2010) Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic life. *Proceedings of the National Academy of Sciences USA*, 107, 12168-12173.

88. Kornblhtt, A.R., de la Mata, M., Fededa, J.P., Muñoz, M.J. and Nogués, G. (2004) Multiple links between transcription and splicing. *RNA*, 10, 1489-98.

89. Ko, J.Y., Oh, S. and Yoo, K.H. (2017) Functional enhancers as master regulators of tissue-specific gene regulation and cancer development. *Molecules and Cells*, 40, 169, 177.

90. Kohlmeier, M., Andreas, S., Jorg, S. and Shearer, M.J. (1996) Transport of vitamin K to bone in humans. *The Journal of Nutrition*, 126, 1192S-1196S.

91. Kramer, J., Kang, R., Grimm, L.M., De Cola, L., Picchetti, P. and Biedermann, F. (2022) Molecular probes, chemosensors, and nanosensors for optical detection of biorelevant molecules and ions in aqueous media and biofluids. *Chemical Reviews*, 122, 3459-3636.

92. Kulmuni, J., Wurm, Y. and Pamilo, P. (2013) Comparative genomics and chemosensory protein genes reveals rapid evolution and positive selection in ant-specific duplicates. *Heredity*, 110, 538-547.

93. Lacroix, B. and Citovsky, V. (2019) Pathways of DNA transfer to plants from *Agrobacterium tumefaciens* and related bacterial species. *Annual Review of Phytopathology*, 57, 231-251.

94. Lamichhane, S., Sen, P., Alves, M.A., Ribeiro, H.C., Raunio Niemi, P., Hyötyläine, T. and Orešič, M. (2021) Linking gut microbiome and lipid metabolism: moving beyond associations. *Metabolites*, 11, 55.

95. Larkin, K., Toloza, A.C., Gabrie, J.A., Rodriguez, C.A., Rueda, M.M., Matamoros, G., Palacio, O., Jamani, S., Fontecha, G. and Sanchez, A.L. (2023) First detection of *Acinetobacter baumannii* in *Pediculus humanus capitidis* from Latin America. *Tropical Medicine and Infectious Disease*, 8, 345.

96. Lartigue, A., Campanacci, V., Roussel, A., Larsson, A.M., Jones, T.A., Tegoni, M. and Cambillau, C. (2002) X-ray structure and ligand binding study of a moth chemosensory protein. *Journal of Biological Chemistry*, 277, 32094-32098.

97. Laufer, H., Borst, D., Baker, F.C., Reuter, C.C., Tsai, L.W., Schooley, D.A., Carrasco, C. and Sinkus, M. (1987) Identification of a juvenile-hormone-like compound in a Crustacean. *Science* 235, 202-205.

98. Lee, G.H. and Choi, K.C. (2020) Adversive effects of pesticides on the functions of immune system. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 235, 108789.

99. Le Goff, G. and Giraudo, M. (2019) Effects of pesticide on the environment and insecticide resistance. *Olfactory Concepts of Insect Control-Alternative to Insecticides*. (ed. J.F. Picimbon), pp. 51-78. Vol. 1, Springer Nature, Switzerland.

100. Lin, D.H. and Hoelz, A. (2019) The structure of the Nuclear Pore Complex (an update). *Annual Review of Biochemistry*, 88, 725-783.

101. Liu, G.X. et al. 2014. Biotype expression and insecticide response of *Bemisia tabaci* chemosensory protein-1. *Archives of Insect Biochemistry and Physiology*, 85, 137-151.

102. Liu, G.X., Ma, H.M., Xie, H.Y., Xuan, N. and Picimbon, J.F. (2016a) Sequence variation of *Bemisia tabaci* Chemosensory protein 2 in cryptic species B and Q: new DNA markers for whitefly recognition. *Gene*, 576, 284-291.

103. Liu, G.X., Ma, H.M., Xie, Y.N., Xuan, N., Guo, X., Fan, Z.X., Rajashekhar, B., Arnaud, P., Offmann, B. and Picimbon, J.F. (2016b) Biotype characterization, developmental profiling, insecticide response and binding property of *Bemisia tabaci* chemosensory proteins: role of CSP in insect defense. *PLoS ONE*, 11, e0154706.

104. Liu, G.X. and Picimbon, J.F. (2017) Bacterial origin of insect chemosensory odor-binding proteins. *Gene & Translational Bioinformatics*, 3, e1548.
105. Liu, G.X., Arnaud, P., Offmann, B. and Picimbon, J.F. (2017) Genotyping and bio-sensing chemosensory proteins in insects. *Sensors*, 17, 1801.
106. Liu, G.X., Yue, S., Rajashekhar, B. and Picimbon, J.F. (2019) Expression of chemosensory protein (CSP) structures in *Pediculus humanis corporis* and *Acinetobacter baumannii*. *SOJ Microbiology and Infectious Diseases*, 7, 1-17.
107. Liu, G.X., Xuan, N., Rajashekhar, B., Arnaud, P., Offmann, B. and Picimbon, J.F. (2020) Comprehensive history of CSP genes: evolution, phylogenetic distribution, and functions. *Genes*, 11, 413.
108. Liu, H., Liu, H., Wang, L., Song, L., Jiang, G., Lu, Q., Yang, T., Peng, H., Cai, R., Zhao, X., Zhao, T. and Wu, H. (2023) Cochlear transcript diversity and its role in auditory functions implied by an otoferlin short isoform. *Nature Communications*, 14, 3085.
109. Liu, Y.L., Guo, H., Huang, L.Q., Pelosi, P. and Wang, C.Z. (2014) Unique function of a chemosensory protein in the proboscis of two *Helicoverpa* species. *Journal of Experimental Biology*, 217, 1821-1826.
110. Loftus, B.J., Utterback, T., Pertea, G., Koo, H., Mori, A., Schneider, J., Lovin, D., de Bruyn, B., Song, Z., Raikhel, A., de Fatima, B.M., Casavant, T., Soares, B. and Severson, D. (2005) *Aedes aegypti* cDNA sequencing. #DV263125, DV263127, DV289920, DV289921, DV297938, DV314711, DV314712, DV316589, DV316619, DV334734, DV334735, DV335057, DV335058, DV344048, DV347318, DV347319, DV365747, DV365766, DV357763, DV368339, DV368340, DV393559, DV400785, DV400787, "...".
111. Los, D.A. and Murata, N. (2002) Chapter 10 – Sensing and responses to low temperature in cyanobacteria. *Cell and Molecular Responses to Stress* (K.B. Storey and J.M. Storey), Vol. 3, pp. 139-153. Elsevier, ScienceDirect, Amsterdam, Netherlands.
112. Lu, Y., Li, H., Zhuang, S., Zhang, D., Zhang, Q., Zhou, J., Dong, S., Liu, Q. and Wang, P. (2014) Olfactory biosensor using odorant-binding proteins from honeybee: ligands of floral odors and pheromones detection by electrochemical impedance. *Sensors and Actuators B: Chemical*, 193, 420-427.
113. Ma, C., Cui, S., Tian, Z., Zhang, Y., Chen, G., Gao, X., Tian, Z., Chen, H., Guo, J. and Zhou, Z. (2019) *OcomCSP12*, a chemosensory protein expressed specifically by ovary, mediates reproduction in *Ophraella communis* (Coleoptera: Chrysomelidae). *Frontiers Physiology*, 10, 1290.
114. Ma, J. (2011) Transcriptional activators and activation mechanisms. *Protein Cells*, 2, 879-888.
115. Macabuhay, A., Arsova, B., Walker, R., Johnson, A., Watt, M. and Roessner, U. (2021) Modulators or facilitators? Roles of lipids in plant root-microbe interactions. *Trends in Plant Science*, 27, P180-190.
116. Malcicka, M., Visser, B. and Ellers, J. (2018) An evolutionary perspective on linoleic acid synthesis in animals. *Evolutionary Biology*, 45, 15-26
117. Maleszka, J., Forêt, S., Saint, R. and Maleszka, R. (2007) RNAi-induced phenotypes suggest a novel role for a chemosensory protein CSP5 in the development of embryonic integument in the honeybee (*Apis mellifera*). *Development Genes and Evolution*, 217, 189-196.
118. Mandrioli, M., Melchiori, G., Panini, M., Chiesa, O., Giordano, R., Mazzoni, E. and Manicardi, G.C. (2019) Analysis of the extent of synteny and conservation in the gene order in aphids: a first glimpse from the *Aphis glycines* genome. *Insect Biochemistry and Molecular Biology*, 113, 103228.
119. Martín-Blázquez, R., Chen, B., Kang, L. and Bakkali, M. (2018) Evolution, expression and association of the chemosensory protein genes with the outbreak phase of the two main pest locusts. *Scientific Reports*, 7, 6653.
120. Matlin, K.S. (2002) The strange case of the signal recognition particle. *Nature Review Molecular Cell Biology*, 3, 538-542.
121. McKenna, M.P., Hekmat-Scafe, D.S., Gaines, P. and Carlson, J.R. (1994) Putative *Drosophila* pheromone-binding-proteins expressed in a subregion of the olfactory system. *Journal of Biological Chemistry*, 269, 16340-16347.
122. Mei, T., Fu, W.B., Li, B., He, Z.B. and Chen, B. (2018) Comparative genomics of chemosensory protein genes (CSPs) in twenty-two species (Diptera: Culicidae): identification, characterization, and evolution. *PLoS ONE*, 13, e0190412.
123. Michaelson, D., Silletti, J., Murphy, G., D'Eustachio, P., Rush, M. and Philips, M.R. (2001) Differential localization of Rho Gtpases in live cells. *Journal of Cell Biology*, 152, 111-126.
124. Moelling, K. and Broecker, F. (2019) Viruses and Evolution – Viruses First? A Personal Perspective. *Frontiers in Microbiology, Sec. Virology*, 10,
125. <https://doi.org/10.3389/fmich.2019.00523>
126. Moon, S.Y. and Zheng, Y. (2003) Rho GTPase-activating proteins in cell regulation. *Trends in Cell Biology*, 13, P13-22.
127. Moloney, F., Yeow, T.P., Mullen, A., Nolan, J.J. and Roche, H.M. (2004) Conjugated linoleic acid supplementation, insulin sensitivity, and lipoprotein metabolism in patients with diabetes mellitus. *American Journal of Clinical Nutrition*, 80, 887-895.

128. Monturiol-Gross, L., Villalta-Romero, F., Flores-Díaz, M. and Alape-Girón, A. (2021) Bacterial phospholipases C with dual activity: phosphatidylcholinesterase and sphingomyelinase. *FEBS OpenBio*, 11, 3262-3275.

129. Morris, F.C., Dexter, C., Kostoulias, X., Uddin, M.I. and Peleg, A.Y. (2019) The mechanisms of disease caused by *Acinetobacter baumannii*. *Frontiers in Microbiology*, 10, 1601.

130. Mosbah, A., Campanacci, V., Lartigue, A., Tegoni, M., Cambillau, C. and Darbon, H. (2003) Solution structure of a chemosensory protein from the moth *Mamestra brassicae*. *Biochemical Journal*, 369, 39-44.

131. Moto, K., Suzuki, M.G., Hull, J.J., Kurata, R., Takahashi, S., Yamamoto, M., Okano, K., Imai, K., Ando, T. and Matsumoto, S. (2004) Involvement of a bifunctional fatty-acyl desaturase in the biosynthesis of the silkworm, *Bombyx mori*, sex pheromone. *Proceedings of the National Academy of Sciences USA*, 101, 8631-8636.

132. Nachtschatt, M., Okada, S., and Speight, R. (2020) Integral membrane fatty acid desaturases: a review of biochemical, structural, and biotechnological advances. *European Journal of Lipid Science and Technology*, 122, 20000181.

133. Nene, V., Worthman, J.R., Lawson, D. et al. (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science*, 22, 1718-1723.

134. Neves, G., Zucker, J., Daly, M. and Chess, A. (2004) Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nature Genetics*, 36, 240-246.

135. Njenga, R., Boele, J., Öztürk, Y. and Koch, H.G. (2023) Coping with stress: How bacteria fine-tune protein synthesis and protein transport. *Journal of Biological Chemistry*, 299, 105163.

136. Nomura, A., Kawasaki, K., Kubo, T. and Natori, S. (1992) Purification and localization of p10, a novel protein that increases in nymphal regenerating legs of *Periplaneta americana* (American cockroach). *International Journal of Developmental Biology*, 36, 391-398.

137. Noriega, F.G., Ribeiro, J.M.C., Koener, J.F., Valenzuela, J.G., Hernandez-Martinez, S., Pham, V.M. and Feyereisen, R. (2006) Comparative genomics of insect juvenile hormone biosynthesis. *Insect Biochemistry and Molecular Biology*, 36, 366-374.

138. Ohnishi, A., Hashimoto, K., Imai, K. and Matsumoto, S. (2009) Functional characterization of the *Bombyx mori* fatty acid transport protein (BmFATP) within the silkworm pheromone gland. *Journal of Biological Chemistry*, 284, 5128-5136.

139. Ollivier, M., Legeai, F. and Rispe, C. (2019) Comparative analysis of the *Acyrthosiphon pisum* genome and expressed sequence tag-based gene sets from other aphid species. *Insect Molecular Biology*, 19, 33-45.

140. O'Neal, A.J., Butler, L.R., Rolandelli, A., Gilk, S.D. and Pedra, J.H.F. (2020) Lipid hijacking: A unifying theme in vector-borne diseases. *eLife*, 9, e61675.

141. Ozaki, K., Utoguchi, A., Yamada, A. and Yoshikawa, H. (2008) Identification and genomic structure of chemosensory proteins (CSP) and odorant binding proteins (OBP) genes expressed in foreleg tarsi of the swallowtail butterfly *Papilio xuthus*. *Insect Biochemistry and Molecular Biology*, 38, 969-76.

142. Ozaki, M., Wada-Katsumata, A., Fujikawa, K., Iwasaki, M., Yokohari, F., Satoji, Y., Nisimura, T. and Yamaoka, R. (2005) Ant nestmate and non-nestmate discrimination by a chemosensory sensillum. *Science*, 309, 311-314.

143. Pedra, J.H. , Brandt, A. , Li, H.M. , Westerman, R. , Romero-Serverson, J. , Pollack, R.J. , Murdock, L.L. and Pittendrigh, B.R. (2003) Transcriptome identification of putative genes involved in protein catabolism and innate immune response in human body louse (Pediculidae: *Pediculus humanus*). *Insect Biochemistry and Molecular Biology*, 33, 1135-1143.

144. Pereira, S.L., Leonard, A.E. and Mukerji P. (2003) Recent advances in the study of fatty acid desaturases from animals and lower eukaryotes. *Prostaglandins, Leukotrienes & Essential Fatty Acids*, 68, 97-106.

145. Perkin, L.C., Friesen, K.S., Flinn, P.W. and Oppert, B. (2015) Venom gland components of the ectoparasitoid wasp, *Anisopteromalus calandrae*. *Journal of Venom Research*, 6, 19-37.

146. Persson, T., Battenberg, K., Demina, I.V., Vigil-Stenman, T., Vanden Heuvel, B., Pujic, P., Facciotti, M.T., Wilbanks, E.G., O'Brien, A., Fournier, P., Cruz Hernandez, M.A., Mendoza Herrera, A., Médigue, C., Normand, P., Pawlowski, K. and Berry, A.M. (2015) *Candidatus Frankia Datiscae Dg1*, the actinobacterial microsymbiont of *Datisca glomerata*, expresses the canonical *nod* genes *nodABC* in symbiosis with its host plant. *PLoS One*, 10, e0127630.

147. Picimbon J.F. (2002) Les périrécepteurs chimiosensoriels des insectes. *Médecine et Science M/S (INSERM Journal)*, 18, 1089-1094.

148. Picimbon, J.F. (2003) Biochemistry and evolution of CSP and OBP proteins. *Insect Pheromone Biochemistry and Molecular Biology, The Biosynthesis and Detection of Pheromones and Plant Volatiles* (eds. G.J. Blomquist and R.G. Vogt), pp. 539-566, Elsevier Academic Press, London, San Diego, UK, USA.

149. Picimbon, J.F. (2005) Synthesis of odorant reception-suppressing agents: odorant binding proteins (OBPs) and Chemosensory Proteins (CSPs) as molecular targets for pest management. *Biopesticides of plant origin* (eds. C. Regnault-Roger, B. Philogène and C. Vincent), pp. 245-266, Intercept Ltd, Hampshire, UK.

150. Picimbon J.F. (2014a) RNA mutations in the moth pheromone gland. *RNA Diseases*, 1, e240.

151. Picimbon J.F. (2014b) RNA mutations: source of life. *Gene Technology* 3, 2.

152. Picimbon J.F. (2016) Mutations in the insect transcriptome. *International Journal of Clinical and Experimental Pathology*, 6, 3.

153. Picimbon, J.F. (2017) A new view of genetic mutations. *Australasian Medical Journal*, 10, 701-715.

154. Picimbon, J.F. (2019) Evolution of protein physical structures in insect chemosensory systems. *Olfactory Concepts of Insect Control-Alternative to Insecticides*. (ed. J.F. Picimbon), pp. 231-263. Vol. 2, Springer Nature, Switzerland.

155. Picimbon, J.F. (2020) Chapter three—bioinformatic, genomic and evolutionary analysis of genes: a case study in Dipteran CSPs. *Methods in Enzymology* 642, 35–79.

156. Picimbon, J.F. (2023a) RNA + peptide editing in chemosensory proteins (CSPs), a new theory for the origin of life on Earth's crust. *Gene*, submitted.

157. Picimbon, J.F. (2023b) Insect OBP, CSPs and NPC2s for control of SARS-CoV-2 and COVID-19: transport/degradation of viral capsid lipids. *Revue Médicale de Bruxelles*, 44, 111-122.

158. Picimbon, J.F. and Leal, W.S. (1999) Olfactory soluble proteins of cockroaches. *Insect Biochemistry and Molecular Biology*, 30, 973-978.

159. Picimbon, J.F. and Regnault-Roger, C. (2008) Composés sémiotiques volatils, phytoprotection et olfaction : cibles moléculaires de la lutte intégrée. *Biopesticides d'Origine Végétale* (eds. C. Regnault-Roger, B. Philogène and C. Vincent), pp. 383-415. Lavoisier Tech & Doc, Paris.

160. Picimbon, J.F., Dietrich, K., Breer, H., Krieger, J. (2000a) Chemosensory proteins of *Locusta migratoria* (Orthoptera: Acrididae). *Insect Biochemistry and Molecular Biology*, 30, 233-241.

161. Picimbon, J.F., Dietrich, K., Angeli, S., Scaloni, A., Krieger, J., Breer, H. And Pelosi, P. (2000b) Purification and molecular cloning of chemosensory proteins from *Bombyx mori*. *Archives of Insect Biochemistry and Physiology*, 44, 120-129.

162. Picimbon, J.F., Dietrich, K., Krieger, J. and Breer, H. (2001) Identity and expression pattern of chemosensory proteins in *Heliothis virescens* (Lepidoptera, Noctuidae). *Insect Biochemistry and Molecular Biology*, 31, 1173-1181.

163. Pikielny, C.W., Hasan, G., Rouyer, F. and Rosbach, M. (1994) Members of a family of *Drosophila* putative odorant-binding proteins are expressed in different subsets of olfactory hairs. *Neuron*, 12, 35-49.

164. Pirtilla, A.M., Brusila, V., Koskimaki, J.J., Wali, P.R., Ruotsalainen, A.L., Mutanen, M. and Markkola, A.M. (2023) Exchange of microbiomes in plant-insect herbivore interactions. *mBio*, 14, <https://doi.org/10.1128/mBio.03210-22>.

165. Pittendrigh, B.R., Clark, J.M., Lee, S.H., Yoon, K.S., Sun, W., Steele, L.D. and Seong, K.M. (2015) Body lice: from the genome project to functional genomics and reverse genetics. *Short views on insect genomics and proteomics, Entomology in focus* (eds: C. Raman, M. Goldsmith & T. Agunbiade), pp. 1-18. Springer Cham, Switzerland.

166. Poirel, L., Madec, J.Y., Lupo, A., Schink, A.K., Kieffer, N., Nordmann, P. and Schwarts, S. (2018) Antimicrobial resistance in *Escherichia coli*. *Microbiology Spectrum-American Society for Microbiology Press*, 6, ARBA-0026-2017.

167. Rai, K.S. and Black, W.C. (1999) Mosquito genomes: structure, organization and evolution. *Advances in Genetics*, 41, 1-32.

168. Ray, S. and Casteel, C.L. (2022) Effector-mediated plant-virus-vector interactions. *Plant Cell*, 34, 1514-1531.

169. Redondo-Salvo, S., Fernández-López, R., Ruiz, R., Vielva, L., de Toro, M., Rocha, E.P.C., Garcillán-Barcia, M.P. and de la Cruz, F. (2020) Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nature Communications*, 11, 3602.

170. Richards, S. , Moran, N. , Morioka, M. , Fukatsu, T. and Nakabachi, A. (2007) *Acyrthosiphon pisum* ESTs. #EX608974.

171. Robertson, H.M., Martos, R., Sears, C.R., Todres, E.Z., Walden, K.K. and Nardi, J.B. (1999) Diversity of odourant binding proteins revealed by an expressed sequence tag project on male *Manduca sexta* moth antennae. *Insect Molecular Biology*, 8, 501-518.

172. Rock, C.O. (2008) Chapter 3 – Fatty acid and phospholipid metabolism in prokaryotes. *Biochemistry of Lipids, Lipoproteins and Membranes Fifth Edition* (eds. D.E. Vance and J.E. Vance), pp. 59-96, Elsevier Science, Amsterdam, Netherlands.

173. Rodriguez, P.A., Stam, R., Warbroek, T. and Bos, J.I. (2014) Mp10 and Mp42 from the aphid species *Myzus persicae* trigger plant defenses in *Nicotiana benthamiana* through different activities. *Molecular Plant Microbe Interactions*, 27, 30-39.

174. Rodríguez, M.F., Sánchez-García, A., Salas, J.J., Garcés, R. and Martínez-Force, E. (2015) Characterization of soluble acyl-ACP desaturases from *Camelina sativa*, *Macadamia tetraphylla* and *Dolichandra unguis-cati*. *Journal of Plant Physiology*, 178, 35-42.

175. Rooney, A.P., Jackson, M.A., Dunlap, C.A., Behle, R.W. and Muturi, E.J. (2019) Discovery and development of microbial biological control agents. *Olfactory Concepts of Insect Control-Alternative to Insecticides*. (ed. J.F. Picimbon), pp. 79-91. Vol. 1, Springer Nature, Switzerland.

176. Rosonina, E. and Blencowe, B.J. (2002) Gene expression: the close coupling of transcription and splicing. *Current Biology*, 12, PR319-R321.

177. Rupawate, P.S., Roylewar, P., Khandagale, K., Gawande, S., Ade, A.B., Jaiswal, D.K. and Borgave, S. (2023) Role of gut symbionts of insect pests: A novel target for insect-pest control. *Frontiers in Microbiology*, 14, 1146390.

178. Sabatier, L., Jouanguy, E., Dostert, C., Zachary, D., Dimarcq, J.L., Bulet, P. and Imler, J.C. (2003) Pherokine-2 and -3: Two *Drosophila* molecules related to pheromone/odor-binding proteins induced by viral and bacterial infections. *European Journal of Biology*, 270, 3398-3407.

179. Salah Ud-Din A.I.M. and Roujeinikova A. (2017) Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. *Cellular and Molecular Life Sciences*, 74, 3293-3303.

180. Sánchez-Gracia, A., Vieira, F.G. and Rozas, J. (2009) Molecular evolution of the major chemosensory gene families in insects. *Heredity*, 103, 208-216.

181. Sasaki, K. and Yoshida, H. (2015) Organelle autoregulation—stress responses in the ER, Golgi, mitochondria and lysosome. *Journal of Biochemistry*, 157, 185-195.

182. Savarit, F., Sureau, G., Cobb, M. and Ferveur, J.F. (1999) Genetic elimination of known pheromones reveals the fundamental chemical bases of mating and isolation in *Drosophila*. *Proceedings of the National Academy of Sciences USA*, 96, 9015-9020.

183. Schachat, S.R., Goldstein, P.Z., Desalle, R., Bobo, D.M., Boyce, K., Payne, J.L. and Labandeira, C.C. (2023) Illusion of flight? Absence, evidence and the age of winged insects. *Biological Journal of the Linnean Society*, 138, 143-168.

184. Scheffers, D.J. and Pinho, M.G. (2005) Bacterial cell wall synthesis: new insights from localization studies. *Microbiology and Molecular Biology Reviews*, 69, 585-607.

185. Schneewind, O. and Missiakas, D. (2014) Sec-secretion and sortase-mediated anchoring of proteins in Gram-positive bacteria. *Biochimica et Biophysica Acta – Molecular Cell Research*, 1843, 1687-1697.

186. Schneider, E.S., Kleineidam, C.J., Leitinger, G. and Römer, H. (2018) Ultrastructure and electrophysiology of thermosensitive sensilla coeloconica in a tropical katydid of the genus *Mecopoda* (Orthoptera, Tettigoniidae). *Arthropod Structure & Development*, 47, 482-497.

187. Schultz, D.J., Suh, M.C. and Ohlrogge, J.B. (2000) Stearyl-acyl carrier protein and unusual acyl-acyl carrier protein desaturase activities are differentially influenced by ferredoxin. *Plant Physiology*, 124, 681-692.

188. Seshasayee, A.S., Sivaraman, K. and Luscombe N.M. (2011) An overview of prokaryotic transcription factors: a summary of function and occurrence in bacterial genomes. *Subcell Biochemistry*, 52, 7-23.

189. Shalaeva, D.N., Galperin, M.Y. and Mulkidjanian, A.Y. (2015) Eukaryotic G protein-coupled receptors as descendants of prokaryotic sodium-translocating rhodopsins. *Biology Direct*, 10, 63.

190. Shangguan, 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. and He, G. (2018) A mucin-like protein of planthopper is required for feeding and induces immunity responses in plants. *Plant Physiology*, 176, 552-565.

191. Shigenobu, S., Richards, S., Cree, A.G., Morioka, M., Fukatsu, T., Kudo, T., Miyagishima, S., Gibbs, R.A., Stern, D.L. and Nakabashi, A. (2010) A full-length cDNA resource for the pea aphid, *Acyrtosiphon pisum*. *Insect Molecular Biology*, 19, 23-31.

192. Siddiqui, J.A., Khan, M.M., Bamisile, B.S., Hafeez, M., Qasim, M., Rasheed, M.T., Rasheed, M.A., Ahmad, S., Shahid, M.I. and Xu, Y. (2022) Role of insect gut microbiota in pesticide degradation: a review. *Frontiers in Microbiology*, 13, <https://doi.org/10.3389/fmicb.2022.870462>.

193. Siebers, M., Brands, M., Wever, V., Duan, Y., Hözl, G. and Dörmann, P. (2016) Lipids in plant-microbe interactions. *Biochimica et Biophysica Acta*, 1861, 1379-1395.

194. Silva-Junior, E.A., Ruzzini, A.C., Paludo, C.R., Nascimento, F.S., Currie, C.R., Clardy, J. and Pupo, M.T. (2018) Pyrazines from bacteria and ants: convergent chemistry within an ecological niche. *Scientific Reports*, 8, 2595.

195. Spealman, P., Burrelli, J. and Gresham, D. (2020) Inverted duplicate DNA sequences increase translocation rates through sequencing nanopores resulting in reduced base calling accuracy. *Nucleic Acids Research*, 48, 4940-4945.

196. Sperling, P., Ternes, P., Zank, T.K. and Heinz, E. (2003) The evolution of desaturases. *Prostaglandins, Leukotrienes & Essential Fatty Acids*, 68, 73-95.

197. Steinbrecht, R.A. (2003) What can we learn from localizing lipoclistins? European Symposium For Insect Taste and Olfaction (ESITO VIII), July 2-7th, Harstad, Norway.

198. Steinbrecht, R.A. (2020) Fine structure immunocytochemistry—An important tool for research on odorant-binding proteins. *Methods in Enzymology*, 642, 259-278.

199. Stern, D.L., Wincker, P. and Tagu, D. (2005) Large-scale gene discovery in the pea aphid *Acyrtosiphon pisum* (Hemiptera). #DY230371.

200. Suenami, S., Koto, A. and Miyazaki, R. (2023) Basic structures of gut bacterial communities in eusocial insects. *Insects*, 14, 444.

201. Taga, M.E. and Bassler, B.L. (2003) Chemical communication among bacteria. *Proceedings of the National Academy of Sciences USA*, 100, 14549-14554.
202. Teng, O., Ang, C.K.E. and Guan, X.L. (2017) Macrophage-bacteria interactions—A lipid-centric relationship. *Frontiers in Immunology*, 8, <https://doi.org/10.3389/fimmu.2017.01836>.
203. Thakur, S. and Chauhan, V. (2022) State of the Globe: Re-emergence of the louse-borne infections. *Journal of Global Infectious Diseases*, 14, 45-46.
204. The International Aphid Genomics Consortium, 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biology*, 8, e1000313.
205. Titball, R.W. (1993) Bacterial phospholipases C. *Microbiology Reviews*, 57, 347-366.
206. Toledo, A. and Benach, J.L. (2015) Hijacking and use of host lipids by intracellular pathogens. *Microbiology Spectrum*, 3,
207. doi:10.1128/microbiolspec.VMBF-0001-2014.
208. Tomaselli, S., Crescenzi, O., Sanfelice, D., Ab, E., Wechsel-berger, R., Angeli, S., Scaloni, A., Boelens, R., Tancredi, T., Pelosi, P. and Picone, D. (2006) Solution structure of a chemosensory protein from the desert locust *Schistocerca gregaria*. *Biochemistry*, 45, 1606-1613.
209. Turcotte, S., Letellier, J. and Lippé, R. (2005) Herpes simplex virus type 1 capsids transit by the trans-Golgi network, where viral glycoproteins accumulate independently of capsid egress. *Journal of Virology*, 79, 8847-8860.
210. van der Meer-Janssen, Y.P.M, van Galen, J., Batenburg, J.J. and Helms, J.B. (2010) Lipids in host-pathogen interactions: Pathogens exploit the complexity of the host cell lipidome. *Progress in Lipid Research*, 49, 1-26.
211. Vega Rúa, A. and Okech' B.A. (2019) The spread of mosquito-borne diseases: a major and global public health problem. *Olfactory Concepts of Insect Control-Alternative to Insecticides*. (ed. J.F. Picimbon), pp.1-27. Vol. 1, Springer Nature, Switzerland.
212. Vela, J., Montiel, E.E., Mora, P., Lorite, P. and Palomeque, T. (2020) Aphids and ants, mutualistic species, share a mariner element with an unusual location on aphid chromosomes. *Genes*, 12, 1966.
213. Verjovski-Almeida, S., Eiglmeier, K., El-Dorry, H., Gomes, S.L., Menck, S.F.M., Nascimento, A.L. and Roth, C.W. (2005) FAPESP and Institut Pasteur/AMSUD Network *Aedes aegypti* cDNA sequencing project. #EG001037.
214. Vogt, R.G. and Riddiford, L.M. (1981) Pheromone binding and inactivation by moth antennae. *Nature*, 293, 161-163.
215. Vogt, R.G., Rogers, M.E., Franco, M.D. and Sun M. (2002) A comparative study of odorant binding protein genes: differential expression of the PBP1-GOBP2 gene cluster in *Manduca sexta* (Lepidoptera) and the organization of OBP genes in *Drosophila melanogaster* (Diptera). *Journal of Experimental Biology*, 205, 719-744.
216. Wada-Katsumata, A., Zurek, L., Nalyanya, G., Roelofs, W.L., Zhang, A. and Schal, C. (2015) Gut bacteria mediate aggregation in the German cockroach. *Proceedings of the National Academy of Sciences USA*, 112, 15678-15683.
217. Wang, F.E., Delannay, C., Goindin, D., Deng, L., Guan, S., Lu, X., Fouque, F., Vega-Rúa, A. and Picimbon, J.F. (2019) Cartography of odor chemicals in the dengue vector mosquito (*Aedes aegypti* L., Diptera/Culicidae). *Scientific Reports*, 9, 8510.
218. Wang, P. and Granados, R.R. (1997) An intestinal mucin is the target substrate for a baculovirus enhancin. *Proceedings of the National Academy of Sciences USA*, 94, 6977-6982.
219. Wanner, K.W., Willis, L.G., Theilmann, D.A., Isman, M.B., Feng, Q. and Plettner, E. (2004) Analysis of the insect os-d-like gene family. *Journal of Chemical Ecology*, 30, 889-911.
220. Wanner, K.W., Isman, M.B., Feng, Q., Plettner, E. and Theilmann, D.A. (2005) Developmental expression patterns of four chemosensory protein genes from the Eastern spruce budworm, *Choristoneura fumiferana*. *Insect Molecular Biology*, 14, 289-300.
221. Wari, D., Kabir, M.A., Mujiono, K., Hojo, Y., Shinya, T., Tani, A., Nakatani, H. and Galis, I. (2019) Honeydew-associated microbes elicit defense responses against brown planthopper in rice. *Journal of Experimental Botany*, 70, 1683-1696.
222. 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. and Schwede, T. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46, W296-W303.
223. Wen, X., Feng, K., Qin, J., Wei, P., Cao, P., Zhang, Y., Yuchi, Z. and He, L. (2023) A detoxification pathway initiated by a nuclear receptor TcHR96h in *Tetranychus cinnabarinus* (Boisduval). *PLoS Genetics*, 19, e1010911.
224. Wicker-Thomas, C., Garrido, D., Bontonou, G., Napal, L., Mazuras, N., Denis, B., Rubin, T., Parvy, J.P. and Montagne, J. (2015) Flexible origin of hydrocarbon/pheromone precursors in *Drosophila melanogaster* (S). *Journal of Lipid Research*, 56, 2094-2102.
225. Wirth, R., Muscholl, A. and Wanner, G. (1996) The role of pheromones in bacterial interactions. *Trends in Microbiology*, 4, 96-103.

226. Xuan, N., Bu, X., Liu, Y.Y., Yang, X., Liu, G.X., Fan, Z.X., Bi, Y.P., Yang, L.Q., Lu, Q.N., Rajashekhar, B., Leppik, G., Kasvandik, S. and Picimbon, J.F. (2014) Molecular evidence of RNA editing in the *Bombyx* chemosensory protein family. *PLoS ONE*, 9, e86932.

227. Xuan, N., Guo, X., Xie, H.Y., Lou, Q.N., Bo, L.X., Liu, G.X. and Picimbon, J.F. (2015) Increased expression of CSP and CYP genes in adult silkworm females exposed to avermectins. *Insect Science*, 22, 203-219.

228. Xuan, N., Rajashekhar, B., Kasvandik, S. and Picimbon, J.F. (2016) Structural components of chemosensory protein mutations in the silkworm moth, *Bombyx mori*. *Agri Gene*, 2, 53-58.

229. Xuan, N., Rajashekhar, B. and Picimbon, J.F. (2019) DNA and RNA-dependent polymerization in editing of *Bombyx* chemosensory protein (CSP) gene family. *Agri Gene*, 12, 100087.

230. Ye, Z., Bishop, T., Wang, Y., Shahriari, R. and Lynch M. (2023) Evolution of sex determination in crustaceans. *Marine Life Science & Technology*, 5, 1-11.

231. Yew, J.Y., Chung, H. (2015) Insect pheromones: An overview of function, form, and discovery. *Progress in Lipid Research*, 59, 88-105.

232. Yue, S., He, Q.X. and Picimbon, J.F. (2023) *Lactobacillus* for ribosome peptide editing cancer. *Clinical Translational Oncology*, 25, 1522-1544.

233. Zhang, C., Shi, Q., Li, T., Cheng, P., Guo, X., Song, X. and Gong, M. (2021) Comparative proteomics reveals mechanisms that underlie insecticide resistance in *Culex pipiens pallens* Coquillett. *PLoS Neglected Tropical Diseases*, 15, e0009237.

234. Zhao, M., Lin, X. and Guo, X. (2022) The role of insect symbiotic bacteria in metabolizing phytochemicals and agrochemicals. *Insects*, 13, 583.

235. Zhao, Z., Gonsior, M., Schmitt-Kopplin, P., Zhan, Y., Zhang, R., Jiao, N. and Chen, F. (2019) Microbial transformation of virus-induced dissolved organic matter from picocyanobacteria: coupling of bacterial diversity and DOM chemodiversity. *The ISME Journal*, 13, 2551-2565.

236. Zaremska, V., Maria Fisher, I., Renzone, G., Arena, S., Scaloni, A., Knoll, W., and Pelosi, P. (2022) Reverse chemical ecology suggests putative primate pheromones. *Molecular Biology and Evolution*, 39, msab338.

237. Zhu, J., Wang, G. and Pelosi, P. (2016a) Plant transcriptomes reveal hidden guests. *Biochemical Biophysical Research Communications*, 474, 497-502.

238. Zhu, J., Iovinella, I., Dani, F.R., Liu, Y.L., Huang, L.Q., Liu, Y., Wang, C.Z., Pelosi, P. and Wang, G. (2016b) Conserved chemosensory proteins in the proboscis and eyes of Lepidoptera. *International Journal of Biological Sciences*, 12, 1394-1404.

239. Zhu, J., Arena, S., Spinelli, S., Liu, D., Zhang, Q., Wei, R., Cambillau, C., Scaloni, A., Wang, G., and Pelosi, P. (2017) Reverse chemical ecology: Olfactory proteins from the giant panda and their interactions with putative pheromones and bamboo volatiles. *Proceedings of the National Academy of Sciences USA*, 114, E9802-E9810.

240. Zhu, J., Iovinella, I., Dani, F.R., Pelosi, P. and Wang, G. (2019) Chemosensory proteins: A versatile binding family. *Olfactory Concepts of Insect Control-Alternative to Insecticides* (ed. J.F. Picimbon), pp. 147-169. Vol. 2, Springer Nature, Switzerland.

241. Zhu, L., Zhang, Y., Zhang, W., Yang, S., Chen, J.Q. and Tian, D. (2009) Patterns of exon-intron architecture variation of genes in eukaryotic genomes. *BMC Genomics*, 10, 47.

242. Zhu, M. and Dai, X. (2020) Bacterial stress defense: the crucial role of ribosome speed. *Cellular and Molecular Life Sciences*, 77, 853-858.

**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.