Symbiosis

ISSN Online: 2372-0956
www.symbiosisonlinepublishing.com

Research Article | SOJ Microbiology & Infectious Diseases | Open Access

Expression of Chemosensory Protein (CSP) Structures in Pediculus humanis corporis and Acinetobacter baumannii

Guoxia Liu[1], Shousong Yue[1], Balaji Rajashekar[2] and Jean-François Picimbon[1,3*]

[1]*Biotechnology Research Center, Shandong Academy of Agricultural Sciences, Jinan, Shandong, China*

[2]*Institute of Computer Science, Tartu University, Tartu, Estonia*

[3]*School of Bioengineering, QILU University of Technology, Jinan, Shandong, China*

Received: 29 April, 2018; Accepted: 8 May, 2019 ; Published: 17 May, 2019

***Corresponding author:** Jean-François Picimbon, Biotechnology Research Center, Shandong Academy of Agricultural Sciences, Distinguished Professor, QLUT, Daxue road, Changqing district, Jinan City, Shandong Province, 250363, PR China, Tel: 00-86-531-89631190; Fax: 00-86-531-89631776; E-mail: jfpicimbon@163.com & jfpicimbon@gmail.com*

Abstract

We parsed the microbial genome database using *Bombyxmori* chemosensory proteins (BmorCSPs) as templates. We extracted eleven bacterial CSPs (B-CSPs) from various microorganisms such as *Kitasatosporagriseola, K. purpeofusca*, K. CB01950, K. MBT66, Escherichia coli, *Macrococcuscaseolyticus* and *Acinetobacterbaumannii,* a known infectious prokaryotic symbiont of various insects, particularly the human body louse. We then parsed the body louse *Pediculushumaniscorporis* genomic database for CSPs. We found six *P. humaniscorporis* (Phum) CSPs all grouped in the same gene cluster. Sequence alignment, structure modeling and phylogenetic analysis of CSP proteins in bacteria and insects reveal duplication, conservation, gene loss, but also diversification and neo-fictionalization that took place at a common stage in this ancestral gene family. Phylogenetic analysis of the amino acid sequences also reveals association of CSPs with other prokaryotic gene families, mainly enzymes and secondary metabolites transporters. Their ability to bind lipids and their proved existence and diversity in infectious bacterial prokaryote systems strongly argue for some important general functions in the cellular metabolism process.

Keywords: *Bombyx; Acinetobacter*; *Escherichia; Kitasatospora; Macrococcus;* human body louse; binding protein;

Introduction

Interactions of all various organisms with environment largely drive the process of evolution and/or co-evolution for adaptation [1]. "Life originates in simple forms and develops with time into more and more complex systems. Species from reptiles to birds, plants to insects and mammals to human are not fixed all time but evolve as a result of natural selection. All life is related and has descended from a common ancestor: the birds and the bananas, the fishes and the flowers – all related" (Darwin's theory of evolution) [2]. Different species of living organisms can ultimately be traced to a single common ancestor, the Last Universal Common Ancestor (LUCA), which is probably the most ancient colony of bacteria from which all organisms now living on earth has a common descent. Modern genetic such as high throughput gene sequencing that reveals genome, gene structure and protein expression of many various species gives full-support to Lamarck's and Darwin's theories of evolution: a species is not fixed since creation [3]. The analysis of a particular gene family, namely *chemosensory proteins* or CSPs, strongly supports the inheritance of the hologenome, the genomes of all an organism's symbiotic microbes mixed with its own genome, which contributes to reject the theory of unchangeability in biological life.

Chemosensory Proteins (CSPs) traditionally refer to small soluble binding proteins that are believed to mediate olfactory recognition at the periphery of sensory receptors, similarly to odor-binding proteins or obps [4,5,6,7]. CSPs are made of six (-seven) ∝-helical chains of about 110amino acids, including four cytokines that build two small loops, two adjacent disulfide bridges, in the globular "prism-like" functional structure of the protein [8,9,10,11]. The CSP structure is not unchangeable. CSPs are characterized by RNA editing and/orpost-translational modifications as reported in the silkworm moth, *Bombyxmori* [12,13,14,15,16,17]. In addition, they are capable of breathing or specific conformational changes, which may represent another key feature of this ancestral primitive multifunctional soluble binding protein family [18].

CSP expressing secretions and tissues are notablypheromone gland secretion, antennal branch, mandibles, salivae, venoms, cephalic capsula, eyes, proboscis, thorax and abdomen, head, epidermis, fat body, gut, wings and legs, among many others [19,20,21,22,23,24,25,26]. Such a broad pattern in gene expression over such a wide range of sensory and non-sensory fluids or tissues is in strong agreement with a very general basic function for this gene family.

A role of CSPs in insecticide resistance and xenobiotic degradation has been clearly brought up by Xuan et al. (2015),

 ***Corresponding author email:** jfpicimbon@163.com & jfpicimbon@gmail.com*

who showed drastic up-regulation of CSP genes in many various tissues over exposure to abamectin insecticide molecule [27]. The role of CSPs and/or other binding proteins in lipid transport, xenobiotic degradation and insecticide resistance has been brought up further by Liu et al. (2016) in whiteflies [28]. While Xuan et al. showed up-regulation of all CSP genes in response to insecticide exposure in a tissue-dependent manner, Liu et al. showed insecticide-mediated up regulation and interaction of the protein with C18-lipid, demonstrating a metabolic role of CSP in insect defense rather than olfactory functions [27,28]. The results of CSP protein bound tightly to fatty acid lipid chains are consistent with up-regulation of CSPs in response to microbial/viral infection as found in flies [29]. In addition, a function of CSPs in lipid transport is consistent with a crucial role not only in general and innate immunity [27,28,29], but also in development, pheromonogenesis and behavioral phase change transition as described in honeybees, moths, white flies and solitarious/gregarious locusts [12,13,14,15,28,30,31,32,33].

Identification of genomic sequences encoding CSPs in the whitefly B. tabaci indicated horizontal transfer of genetic materials through end symbiotic bacteria between insects and plants [28,29,30,31,32,33,34]. Here, we report about the finding of CSP expression in bacterial species such as *coccobacillus Acinetobacter baumannii, Micrococcus/Staphylococcuscaseolyticus*, the filamentous actinomycete *Kitasatosporagriseola*, an Actinobacteria genus in the family Streptomycetaceae, and *Escherichia coli* (E. coli) which are known as common bacteria from the digestive tract, main prokaryotic secondary metabolites, opportunistic multi-drug resistant pathogens, high positive cytochrome c oxidase reactions, and symbionts of multiple insect species. The bacterial species *A. baumannii, E. coli* and *M. caseolyticus* CSP sequences are twins or identical twins to BmorCSPs (BmorCSP2, BmorCSP4, BmorCSP6 and BmorCSP15), while *Kitasatospora* CSPs are more distantly related to *Bmor CSP* and orthologous genes in Coleopteran and Hymenopteran species. Because these bacterial pathogens, *A. baumannii, E. coli, Kitasatasporaand M. caseolyticus,* are mediators of infectious diseases in human and can be carried by insects exclusively feeding on human blood, i.e. lice, we also show the genomic repertoire of CSPs in *Pediculushumaniscorporis (PhumCSPs)*. The six PhumCSPs build six different orthology groups. PhumCSPs and the CSP family identified in bacteria (B-CSPs) show the same typical folding, a prism with pairs of antiparallel alpha-helices, which certainly represent a protein structure that existed much before the split of the countless types of bacterial strains. In addition, we show protein structure models only differing in alpha-helical motifs in bacterial CSPs, strongly suggesting that the CSP structure had multifunction and/or acquired new function at a far remote time, probably during the early stages of prokaryote evolution (at least 3.8 billion years ago).

Materials and Methods

Identification of bacterial CSPs

Microbial Genomes at the National Center for Biotechnology Information were used to identify CSP counterparts in the bacterial super kingdom. We looked for bacterial counterparts for each *B. moriCSP* amino acid sequence. *B. moriCSPs* are known to be expressed throughout many various tissues in the insect body [21,22,23,24,25,26,27]. BmorCSP expression is known to be drastically up-regulated by insecticide exposure [27]. BmorCSP expression is also known to be regulated by specific post-transcriptional and/or post-translational events [12,13,14,15,16,17]. The 3D structure of *B. mori* chemosensory protein-1 (BmorCSP1) in solution is known [9,10]. We therefore used BmorCSP1-BmorCSP20 protein query sequence to scan the microbial genome database. Eleven clones (hit rate: 23%-100%) were extracted from bacterial species table 1. They were identified as CSP clones on the basis of molecular weight and the number of cysteines at conserved position in the primary structure table 1; figure 1.

Table 1: Repertoire of CSP clones in bacteria (NCBI Microbial database; Nov 11th 2018) *: Transmembrane protein (CSP precursor)

Locus	Nb aa	Nb Cys	MW (kDa)	Organism	Source	Percent identity (BmorCSP)
WP_043907137	142	4	14.7	*Streptomycetaceae*	*Kitasatospora griseola*	23% (BmorCSP10)
WP_071212566	120	4	13.7	*Moraxellaceae*	*Acinetobacter baumannii*	99% (BmorCSP2)
WP_071222707	87	4	9.7	*Moraxellaceae*	*Acinetobacter baumannii*	100% (BmorCSP2)
WP_073810176 WP_083646628	142	4	15.1	*Streptomycetaceae*	*Kitasatospora sp. CB01950*	22% (BmorCSP10)
WP_078880044	142	4	16.0	Streptomycetaceae	*Kitasatospora purpeofusca*	46% (BmorCSP14)
WP_082558797	142	4	16.0	*Streptomycetaceae*	*Kitasatospora sp. MBT66*	46% (BmorCSP14)
WP_089438515	131	4	14.9	*Enterobacteriaceae*	*Escherichia coli*	100% (BmorCSP6)
WP_096417339*	491	10	50.6	*Xanthomonadaceae*	*Lysobacter capsici*	72% (BmorCSP5)
WP_120787151	127	4	14.5	*Staphylococcaceae*	*Micrococcus caseolyticus*	72% (BmorCSP1)
WP-120787152	131	4	14.9	*Staphylococcaceae*	*Micrococcus caseolyticus*	61% (BmorCSP2)
WP_120787167	150	4	17.2	*Staphylococcaceae*	*Micrococcus caseolyticus*	49% (BmorCSP4)
WP_120787175	122	4	13.9	*Staphylococcaceae*	*Micrococcus caseolyticus*	72% (BmorCSP15)

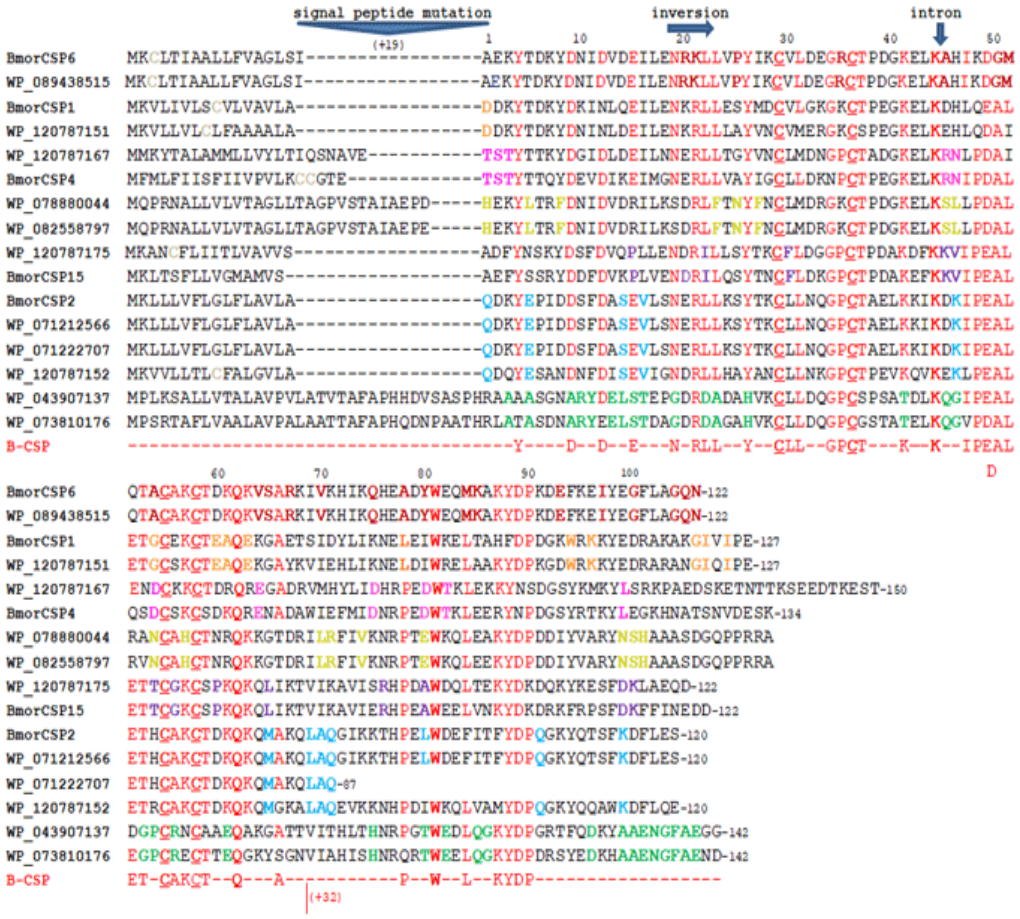

Figure 1: Identification of bacterial CSPs (B-CSPs). The sequences are from *Acinetobacter baumannii* (WP_071212566/WP_071222707), *Escherichia coli* (WP_089438515), *Kitasatosporagriseola* (WP_043907137), *K. purpeofusca* (WP_078880044), K. sp. CB01950 (WP_073810176), K. sp. MBT66 (WP_082558797) and *Macrococcuscaseolyticus* (WP_120787151, WP_120787152, WP_120787167 and WP_120787175; Table 1). Highly conserved amino acids are shown in red. Strictly conserved amino acid residues are shown in red and bold. The four Cys residues characteristic of CSPs are underlined. +Cys residues in signal peptide are shown in bold. Subtype-specific amino acid residues or motifs are shown by different colors. Brown: BmorCSP6/WP_089438515, orange: BmorCSP1/WP_120787151, pink: BmorCSP4/WP_120787167, kaki: WP_078880044/WP_082558797, purple: BmorCSP15/WP_120787175, blue: BmorCSP2/WP_071212566/WP_071222707/WP_120787152, green: WP_043907137/WP_073810176. The typical primary sequence of B-CSPs is shown below the sequence alignment. BmorCSP1, BmorCSP2, BmorCSP4, BmorCSP6 and BmorCSP15 are CSP sequences from the silkworm moth, B. mori [9-10, 12-13, 18, 26]. The arrows in blue indicate signal peptide mutation, inversion in NRKL/NKRL motif and intron insertion site after Lysine at position 44, respectively.

Using *BmorCSP*1 as a template to screen the bacterial genome database extracted Staphylococcaceae *Macrococcuscaseolyticus* WP_120787151 (Evalue: 2e-54, 72% identity) and WP_120787152 (Evalue: 2e-27, 42% identity), and WP_089438515 in Enterobacteriaceae Escherichia coli (Evalue: 2e-27, 42% identity). Using BmorCSP2 as a template extracted WP_071212566 (Evalue: 1e-81, 99% identity) and WP_071222707 (Evalue: 8e-55, 100% identity) from *Acinetobacter calcoaceticus/baumannii complex* (Moraxellaceae). *A. baumannii* WP_071212566 corresponded to a truncated form of WP_071222707/BmorCSP2. Using BmorCSP2 as a template also extracted WP_043907137 (Evalue: 4e-08, 30% identity) from *Kitasatosporagriseola,* WP_073810176/WP_083646628/OKJ13717 (Evalue: 2e-09, 32% identity) from *K. sp*. CB01950, and two new clones from M. caseolyticus, WP_120787175 (Evalue: 2e-28, 46% identity) and WP_120787167 (Evalue: 1e-20, 36% identity) in addition of McasWP_120787152 (Evalue: 6e-51, 61% identity) and *E. coli* WP_089438515 (Evalue: 2e-20, 38% identity). Using *BmorCSP3* did not show other clones, but using BmorCSP4 extracted two additional CSP clones from bacterial strains, Streptomyceteae *Kitasatosporasp*. MBT66 WP_082558797 (Evalue: 1e-26, 40% identity) and WP_078880044 from *K. purpeofusca* (Evalue 4e-26, 40% identity). Using *BmorCSP5* (truncated gene) extracted a CSP clone with a usual size (50 kDa) and an usual number of cysteines (10) from Xanthomonadaceae *Lysobactercapsici* (WP_096417339; Evalue: 9e-10, 52% identity). WP_096417339 is a hypothetical Tran's membrane protein. It contains a protein structure related to E. coliTolB protein and a portion of CSP protein (∝1-∝3) terminated by a RIQPGALADN motif.

Using *BmorCSP*6 extracted WP_089438515 from *E. coli* with an E-value score of 2e-85 and 100% identity match. Using *BmorCSP15* extracted WP_120787175 from *M. caseolyticus* with an E-value score of 2e-61 and 74% identity match. Using other BmorCSPs, even *BmorCSP*10, a CSP clone from *B. mori* with an unusual size and an unusual number of cysteines (6) and prolines (21), did not extract any supplementary sequences possibly corresponding to CSP (four cysteines, MW: about 10-15 kDa).Using BmorCSP17 extracted a sequence of 3409 amino acids (about 363 kDa) corresponding to a hypothetical precursor protein for large tegument protein UL36 and retinal protein (pfam15449; WP_121154966) from the actinobacteria *Micromonosporapisi*. This sequence did not have the structure (Cysteine profile) characteristic of CSPs, but shared some typical motifs (RLLLG—Y and KYEA) with CSPs. Using BmorCSP consensus sequences or consensus sequences from CSP genes of other insect species did not extract any additional CSPs from bacteria in the general microbial protein blastome. We also blasted the sequences obtained from *Baumannii, Kitasatospora* and *Macrococcusfor* an iterative search in the microbial protein database. Only twelve bacterial CSP amino acid sequences and the odd WP_096417339 sequence (Tran's membrane protein and/or CSP precursor) could be extracted so far tables 1 & 2.

Table 2: Identity matrix (mean percent identity values) between every pair of B-CSPs *Transmembrane protein (CSP precursor)

Locus	07137	12566	22707	10176	80044	58797	38515	87152	87167	87151	87175	17339*
WP_043907137		25	26	68	27	27	20	25	22	25	30	25
WP_071212566	25		100	25	29	29	35	60	33	35	45	28
WP_071222707	26	100		26	31	31	39	64	39	39	48	28
WP_073810176	68	25	26		26	26	19	24	25	28	26	19
WP_078880044	27	29	31	26		98	36	31	37	39	30	42
WP_082558797	27	29	31	26	98		35	31	37	38	31	42
WP_089438515	20	35	39	19	36	35		31	32	42	38	37
WP-120787152	25	60	64	24	31	31	31		27	39	42	27
WP_120787167	22	33	39	25	37	37	32	27		43	32	32
WP_120787151	25	35	39	28	39	38	42	39	43		32	31
WP_120787175	30	45	48	26	30	31	38	42	32	32		39
WP_096417339*	25	28	28	19	42	42	37	27	39	31	32	

Identification of louse CSPs

Fly Base and Broad Institute database at http://www.broadinstitute.org (*Aedes, Culex* and *Pediculus*) were used to identify CSP sequences from Phthiraptera *Pediculushumanuscorporis* (Human body louse) and retrieve genomic sequences by blasting the nucleotide sequence of each gene against the assembly of scaffolds and contigs [35,36]. Not only BmorCSPs but also CSP consensus sequences from different insect species such as *Acyrtosiphonpisum, Aedesaegypti, Anopheles gambiae, Apismellifera, Bemisiatabaci, Culexpipiens, Drosophila melanogaster, Harpegnathossaltator/Camponotus floridanus, Nasoniavitripennis and Triboliumcastaneum* were used as queries to search for *Pediculushumanuscorporis* chemosensory proteins (PhumCSPs). Consensus cobbler sequences (blocks or motifs) were built using block maker at http://bioinformatics.weizmann.ac.il/blocks. Similarly to our search for bacterial CSPs, PhumCSPs were identified on the basis of molecular weight and the number of cysteines at conserved position in the primary structure table 3 & figure 4. Exon and intron boundaries were determined by plotting each CSP RNA sequence against the scaffold or contig open reading frame in the search launcher nucleotide algorithm as described for BmorCSPs [27]. Access numbers and scaffolds containing the complete repertoire of PhumCSP genomic DNA sequences (six genes) are reported in Table 3. The degree of identity across PhumCSPs is reported in table 4.

Phylogenetic analysis of bacterial and insect CSPs

The protein sequences were selected from our present study about PhumCSPs and bacterial CSPs (B-CSPs), which have been identified in Enterobacteriaceae, Moraxellaceae, Staphylococcaceae, Streptomycetaceae and Xanthomonadaceae.

Other CSP protein sequences were selected among those identified in model insect species such as *A. mellifera* (Amel), *B. mori* (Bmor), *B. tabaci* (Btab), *D. melanogaster* (Dmel), *H. saltator / C. floridanus* (EFN), *N. vitripennis* (NV) and *T. castaneum* (Tcas), for which gene structure and/or genome organization were known [7,12,27,28,34,37,38,39,40,41,42,43,44]. We add edarthropod (crustacean) CSP sequences from the brine shrimp Artemia franciscana (AfraCSP; ABY62736, ABY62738) and the water flea zooplankton Daphnia pulex (DpulCSP1, DpulCSP2; ABH88166, ABH88167). We provided protein structure data analysis using software tools for AfraCSP, DpulCSP1 and DpulCSP2 in our study (see below). AfraCSP, DpulCSP1 and DpulCSP2 melted with BtabCSP1, a protein known to transport fatty acid lipids such as linoleic acid [28].

Table 3: Repertoire of CSP genes identified by in silico analysis of *Pediculushumaniscorporis* genome and EST database (Broad Institute database http://www.broadinstitute.org (*Aedes*, *Culex* and *Pediculus*), VectorBase(Bioinformatics Resource for Invertebrate Vectors of Human Pathogens, https://www.vectorbase.org). Genomic sequence is in italics. *: Scaffold number, **: Locus.

Locus	Genome size (Mb)	Gene size (bp)	Intron size (bp)	Accession number
	110			
PHUM594410		716	187 (int1)/100 (int2)	*AAZ001007241**, *1103172108314, DS235878**,*NW_002987883,* XM_002432547, XP_002432592, EEB19854,
PHUM594420		687	288	*AAZ001007241**, *1103172108314, DS235878**, *NW_002987883,* XM_002432548, XP_002432593, EEB19855
PHUM594430		667	277	*AAZ001007241**, *1103172108314, DS235878**, *NW_002987883,*XM_002432549, AJ973467, XP_002432594, EEB19856, CAJ01514
PHUM594540		631	208	*AAZ001007242**, *1103172108314, DS235878**, *NW_002987883,*XM_002432550, XP_002432595, EEB19857
PHUM594550		473	89	*AAZ001007242**, *1103172108314, DS235878**, *NW_002987883,*XM_002432551, XP_002432596,EEB19858
PHUM594660		491	143	*AAZ001007243**, *1103172108314, DS235878**, *NW_002987883,*XM_002432552, XP_002432597, EEB19859

Table 4: Identity matrix (mean percent identity values) between every pair of PhumCSPs.

Locus	PHUM594410	PHUM594420	PHUM594430	PHUM594540	PHUM594550	PHUM594660
PHUM594410		39	42	39	30	37
PHUM594420	39		48	80	31	37
PHUM594430	42	48		45	34	47
PHUM594540	39	80	45		29	30
PHUM594550	30	31	34	29		37
PHUM594660	37	33	47	30	37	

The protein sequences were aligned using MUSCLE (www.drive5.com/muscle) [45]. The phylogenetic trees were constructed on the protein sequence alignment using IQ-TREE (version 1.5.6) [46]. The parameters used to construct the phylogenetic tree were ultrafast bootstrap (UF Boot, using the –bb option of 1000 replicates), and a standard substitution model (-st AA –m TEST), and alrt 1000 –NT AUTO was given to generate the tree. Trees generated from IQ-TREE were visualized using Fig Tree software [47]. For better visualization, trees were modified under the "Trees" section, by applying the increasing order nodes, transform branches to proportional and bootstrap support of greater than 50 were displayed on the branches. The branch-support values were obtained from the output ".tree file".

Construction of 3D structure models

To compare the CSP structure between bacteria and insects, we used modeling implemented in ahttp://swissmodel.expasy.org/, as a preliminary characterization of functional features that bridge the gap between prokaryotic and eukaryotic systems. X-Ray structure of chemosensory protein A6 in solution (1kx9.1.A) [8] was used as template to predict the structure of each CSP identified in Entero bacteriaceae, Moraxellaceae, Staphylococcaceae, Strepto mycetaceae and Phthiraptera, respectively.

MbraA6-1kx9.1A (monomer, 1.6Å) modeled a compact structure made of six ∝-helices (∝1-∝6) [8,9,10,11]. Amino acid 1 is the first residue of the mature protein after removing signal peptide as shown by Edman sequencing at the N-terminus in cockroach and moth CSPs [5,21].We used Crustal to produce an alternative alignment of CSP sequences and then used it as a point of comparison to determine amino acid 1 in bacterial and louse CSPs, respectively. The protein without ligand (apoprotein) was used as a model because specific cognate ligands of CSPs in bacteria and lice remain to be found.

Results and Discussion

We searched for "chemosensory" proteins (CSPs) in bacteria. We found a group of genes that share important characteristics, four cysteine codons and similar sequences of RNA building blocks (nucleotides), providing instructions for making products (small soluble alpha-helical proteins) that have a similar structure or function than those identified in insects and arthropods.

Using *Bombyxmorichemosensory* protein (BmorCSP) sequence as a template to screen the protein library of the bacterial genome database (NCBI) extracted twelve mRNA clones that shared 22-46% up to 72-100% identity to *BmorCSP* clones (see table 1). In summary, in our search for microbial sequences related to *Bombyx CSP*s, we extracted one clone (WP_089438515) related to *BmorCSP*6 in the gut bacteria *Escherichia coli,* and four clones (WP_120787151, WP_120787152, WP_120787167 and WP_120787175) related to BmorCSP1, BmorCSP2, BmorCSP4 and BmorCSP15, respectively, in *Macrococcuscaseolyticus,* another bacterial genus inhabiting the insect gut. Additionally, we found four clones (WP_043907137, WP_073810176, WP_078880044 and WP_082558797) related to BmorCSP14 and BmorCSP10 in colonic Kitasatospora bacteria, and two clones related to BmorCSP2 (WP_071212566 and WP_071222707) in *Acinetobacter Baumannii*, a known infectious agent for human health, vehiculated by many biting insect species such as the human body louse, *Pediculushumaniscorporis* (table 1; figure 1).Therefore, the csps that we are describing here are members of a vast protein gene family that exists not only in insects and arthropods, but also in multiple species from the prokaryote super kingdom.

Analyzing the mRNA database from the microbes, eleven different sequences significantly related to CSPs were identified and called B-CSPs to emphasize on the fact that these clones are not from insects or arthropods, but have a common bacterial origin (table 1; figure 1).The B-CSPs share between 22-98% identity (table 2). The four CSPs found in C. caseolyticusdisplay about 27-43% sequence identity with each other, which is an illustrative example for understanding the specific forms of high molecular diversity in CSPs from prokaryotes (table 2; figure 1).

B-CSPs are proteins of about 13.9-17.2 kDa, similarly to moth CSPs and other CSP proteins from other insect species [5-30].The truncated form of WP_071222707 or BmorCSP2 leads to a protein of about 9.7 kDa. This could correspond to a functional protein because the protein is truncated after the intron insertion of CSP (figure 1). Truncated CSP genes are known to be unexpressed in moths [27], but mutations leading to stop codon mutation and truncated mRNAs have been described, not only in bacteria but also in insects and even in human [12,13,14,15,16,17,48,49]. More interestingly, B-CSPs can be divided in seven subtypes, which can be distinguishable from each other on the basis of amino acid sequences (figure 1). All B-CSPs display four Cysteine residues (Cys29, Cys36, Cys55 and Cys58), Glutamine 62 and Tryptophan 81 at key positions that are conserved across the whole CSP gene family (figure 1) [7]. Similarly to BmorCSPs, amino acid motifs such as N--RLL—Y (26), GPCT (37), I/LPD/EALET (53), CAKCT (59), KQK—A (66) and KYDP (90) are well-conserved between B-CSPs (figure 1). The main differences between the seven B-CSP subtypes are found in the N-and C-terminal regions, and in the central region corresponding to the amino acids at position [54-70], after the IPE/DAL motif. Other key replacements are found at position [46,47,75,78,80] which may underlie subtype-specific functions (figure 1).

B-CSPs were called B-AbauCSP2, (-32aa) B-AbauCSP2, B-KcbCSP1A, B-KgriCSP1A, B-KmbtCSP14, B-KpurCSP14, B-McasCSP1, B-McasCSP2, B-McasCSP4, B-McasCSP15 and B-EcolCSP6 on the basis of their sequence identity to BmorCSP1, BmorCSP2, BmorCSP14, BmorCSP15 and BmorCSP6, respectively (see figure 1) [9,10,12,13,14,15,16,17,21,27]. The homology modeling structures of B-CSPs was built using M. brassicae CSP MbraA6 crystal structure (1kx9.1A) [8]. First, we compared the NMR structure of BmorCSP1 and the structure predicted for BmorCSP1 using 1kx9.1A as a reference model [8,9,10]. For BmorCSP1, the model and NMR structures only differed in the C-terminal region; the protein folding was the same, which validated our approach of using SWISS-MODEL as a very preliminary analysis of B-CSP structures (figure 2). In this analysis, the two CSP sequences from the arthropod crustacean daphniid water flea Daphnia pulex(ABH88166, ABH88167; DpulCSP1, DpulCSP2)and the CSP from the extremophile brine shrimp Arthemia franciscana (AfraCSP) folded in a compact structure made of five alpha(∝)-helices. The C-terminal tail (∝5) paired with ∝2, while N-terminal tail (∝1) paired with ∝4, leaving ∝3 at the basis of the triangle CSP structure (figure 2). Lepidopteran moth CSPs such as BmorCSP1 had a rather different structure. Soluble MbraCSPA6 and BmorCSP1 protein structures display a compact structure made of six alpha (∝)-helices connected by small ∝-∝ loops, as found for the structure of the locust CSPsg4 protein [11]. In this typical CSP structure characterized by six ∝-helices, ∝1/∝4 and ∝2/∝5/ ∝1/∝6 align to form the m□□leaving ∝3 at the basis of the triangle CSP structure, similarly to water flea and shrimp CSPs.

Interestingly, in our survey of B-CSP structure models, the same folding was observed for all B-CSPs, suggesting that this -helices into a compact flexibleafolding of six anti-parallel triangle or prism is rather ancient and expressed not only in eukaryotic organisms but also in prokaryotic cells. However, some differences were found between the various B-CSP structure models. B-McasCSP2 displayed a long free N-terminal tail. The was much shorter than in other B-CSPs)1ahelix (-aN-terminal which ,5a dna 2a(figure 2). B-McasCSP2 had also shortened are two parallel helices on the B-CSP structure (figure 2). Other heT .1adifferences were found in the number of maillons in griCSP1A, but not inK-B in 1amaillon was removed in athird the model structure of B-KcbCSP1A (figure 2). The first maillon CSP4 (figure 2). The C-terminussa cM-B in 5awas removed in to form a loop in B-KurCSP14 and B-McasCSP1, 6aexpanded in similarly to BmorCSP1; all other B-CSP models were lacking this C-terminal loop (figure 2). (-32aa)B-AbauCSP2 built a small prism -helices inawith N- and C-terminal tails folded in two antiparallel 3 at the bottom of the structure like in alla2, leaving athe front of

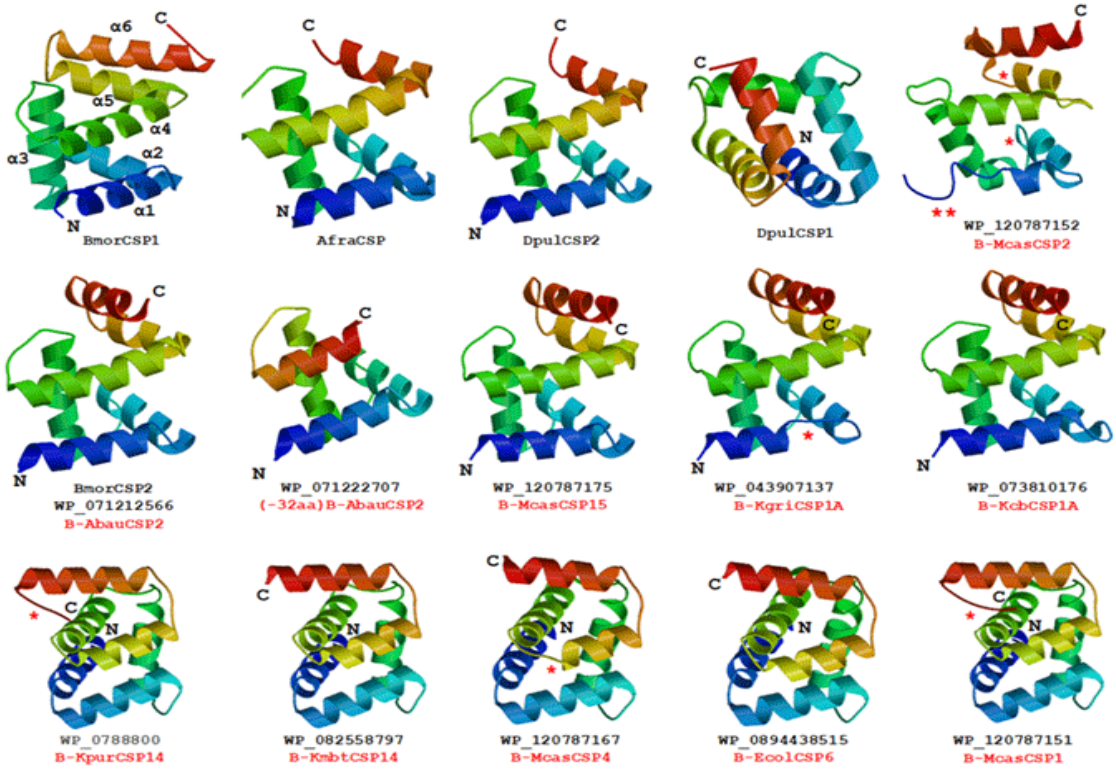

Figure 2: Predicted tertiary structure of B-CSPs. Homology modelling using *M. brassicae* CSPA6 (1kx9.1.A; X-ray, 1.6 Å) as template (http://swiss-model.expasy.org) [8]. Afra: *Arthemia franciscana,* Bmor: *Bombyxmori,* Dpul: *Daphnia pulex* (Arthropod); Abau: *Acinetobacter baumannii*, Ecol: *Escherichia coli*, Kgri: Kitasatosporagriseola, Kpur: *K. purpeofusca*, Kcb: *K. sp.* CB01950, Kmbt: K. sp. MBT66, Mcas: M*acrococcuscaseolyticus* (Bacteria). The asterisk in red shows mutation sites in protein structure.

andgniliforp-aother CSPs (figure 2). These differences in C-loop probably indicate different functions for B-CSPs.

These results demonstrate the existence of CSP genes not only in Eukaryotes, in arthropods and insects, in winged insects such as moths and in wingless insects such as head and body lice, but also in Prokaryotes, microbial systems, and their expression in many various bacterial strains from E. coli to *A. baumannii,* therefore across many different types or genera of bacteria, known for severe infectious diseases in human. Interestingly, most of all these bacterial species (*Moraxella, Staphylococcus, Streptomyces, Enterobacter* and *Xanthomonadales*) that house CSPs are also known as symbionts of the insect gut [50,51,52,53,54]. The gut microbiota plays crucial roles in many various physiological systems such as the growth, development, detoxification, nutrition, immune response and environmental adaptation to insect hosts [55]. Many bacterial species such as *E. coli, Staphyloccocus* and *A. baumannii* are known to affect invertebrates and human alike. Furthermore, many insect species are sources of potential human pathogens if not obtained from the environment. This is the case, for instance, of *A. baumannii,* which is known not only as a bacterial pathogen with a long range of various diseases (pneumonia, meningitis, urinary tract and bloodstream infections) but also as a symbiotic species of the head louse and human body louse, Pediculushumaniscapitis and *Pediculushumaniscorporis* (*humanus*) [56,57].

Body louse and head louse are known to infest humans. They are common species of sucking louse in the family Pediculidae (Anoplura, Phthiraptera) that is found where human lives. Like ticks and mites, the head louse is an obligate ectoparasite of humans.

It spends its entire life on the human scalp, not to clean the hairs (cleaning symbiosis) but to feed off the blood from our brain vessels. In addition to bite our scalp from one hair shaft to another hair shaft, they will lay eggs at the bottom of our hairs (called nits). These nits will start a new life cycle and later jump onto new comers. While they feed on us for their health and growth, they will propagate many various epidemic diseases among people, especially young children who will suffer open wounds, scratches, irritations or sores on the head from the constant sucking of blood from the head lice. Head lice infestation usually leads to *Pediculosiscapitis* and *Phtiriasis* which has been a problem for the human population for thousands years [58].

Besides head louse, there are many other different types of lice that are found on the body, in the armpits or on the pubic area. While they live on different places on the human body, they all live on or in the skin of the human body, and they all feed on

the human blood and soon can infest the whole body, spreading not only scratches or wounds that will be nests for infectious bacteria, but also more serious pathologies.

Lice are the primary vectors for the transmission of diseases such as epidemic thyphus, trench fever and relapsing fever, which are caused by dangerous strains of bacteria such as *Rickettsia prowazekii, Borreliarecurrentis* and *Bartonella Quintana* [59]. Therefore, the analysis of potential genes transferred from the bacterium to the louse might be extremely important in the prevention of numerous infectious diseases in human.

We checked for the occurrence of horizontal gene transfer of CSPs from bacteria to lice by analyzing the repertoire of CSPs in the *Pediculushumaniscorporis* genome. We found only six CSP genes in the louse genome at six specific loci, namely PHUM594410, PHUM594420, PHUM594430, PHUM594540, PHUM594550 and PHUM594660 (figure 3 and tables 3 & 4). Such a low number of CSPs was also found in ants, flies, bees, wasps and anopheline mosquitoes (n= 4-8) [7,37-44].

They are all placed in the same scaffold, which occupies the same chromosomal region of the louse (DS253878; table 3). Out of the six fully identified *Pediculushumaniscorporis* CSP genes (*PhumCS*Ps), one is found to have two introns (PHUM594410) and the other five genes are found to have a single intron always located at the same position, after Lysine residue at position 44, indicating common ancestry for all these genes (figure 3). In contrast to BmorCSPs, the intron in PhumCSPs does not very much in size. It does not exceed 288 bps (figure 3 & table 3). Intron 1 in PHUM594410 is inserted in the signal peptide similarly to AAJJ1196A, BmorCSP19 and GB19453 [27,38,39,43,44]. These 2 introns/3 exons genes must be ancestrally related to a CSP gene that occurred in the last common ancestor of moths, beetles, bees and lice, as ancient as more than 300 Mya [60]. Similarly to moth CSPs, PhumCSP genes sit close to each other in the same genomic region separated by about 1500-8500 bps (figure 3). So the cluster is not a recent expansion of CSPs, but the CSPs have remained in this cluster throughout their very long history.

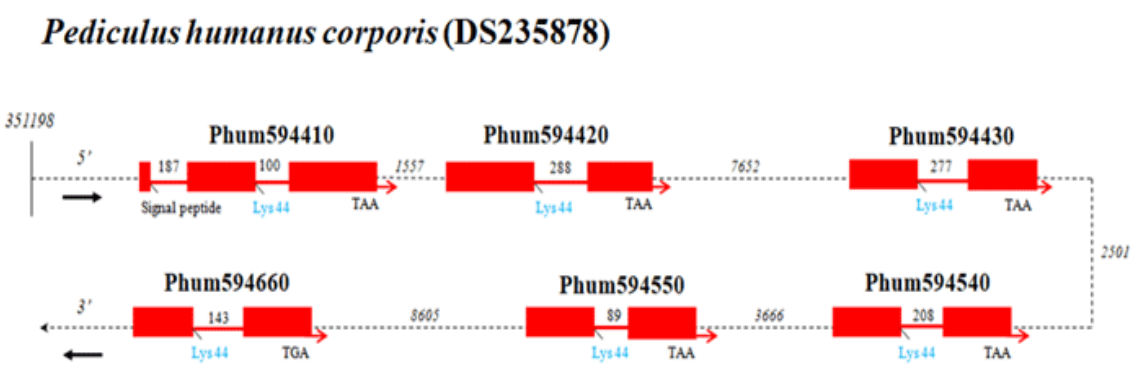

Figure 3: Genomic organization of *Pediculus humaniscorporis* chemosensory protein (PhumCSP) genes on chromosomal region DS253878.PhumCSP genes are shown in red. Exons are shown as red boxes and introns as bold red plain lines. Dotted lines in black are intergenic intron regions. The numbers in italic above the line give the distance between *PhumCSP*s. The numbers above the intron give the intron size. The amino acid residue in blue shows conserved intron insertion site, after lysine at conserved position (amino acid 44). The arrow indicates the orientation of the gene (5'-3': Phum594410, Phum594420 and Phum594430; 3'-5': Phum594540, Phum594550, Phum594660). Stop codons are indicated, showing that PhumCSPs have either TGA (Phum594660) or TAA stop codons.

Similarly to Bmor- and B-CSPs, PhumCSPs show about 29-80% sequence identity with each other (table 4, figure 4). Conservation of specific amino acid residues shows duplication, pair wise evolution, expression and/or function of PhumCSP genes: PHUM594420/PHUM594540, PHUM594410/PHUM594430 and PHUM594550/PHUM594660. However, all PhumCSPs display Threonines T6 and T37, Aspartate 14, Leucine 17, Lysine 44 (intron insertion site), Proline 48, Lysine 77, 21-RLL-23and 87-KYDP-90 motifs, in addition of four Cysteine residues (Cys29, Cys36, Cys55 and Cys58), Glutamine 62 and Tryptophan 81 at key positions that are conserved across the whole CSP gene family [7], including the bacterial CSPs reported in our study (see figures 1 & 4). The main differences between the PhumCSP subtypes are found in the N-and C-terminal regions, and in the central region corresponding to the amino acids at position 54-70, after the LPV/E/DAL motif, as found for B-CSPs (figures 1 & 4). This corresponds to ∝ -helix ∝4 on the structure model of Phum CSP (figure 4).

Like B-CSPs, PHUMCSPs are predicted to fold into a prism characterized by six ∝-helices, with ∝1/∝4 and ∝2/∝5/∝1/∝6 as square sides ∝ and ∝3 as a base (figures 2 & 4). Interestingly, intron is inserted in ∝3, [illegible] right [illegible] protein, suggesting that the prism was built by association of two exons, each of them coding for small alpha-helical subunits (figure 4). This joint seems to have happened much before the time when insect species became diverse and highly specialized, and much before any possible insect-bacteria associations. The degree of identity between insect and bacterial CSPs suggests that it occurred a long time ago, most likely some Bya. Insect species were even not born

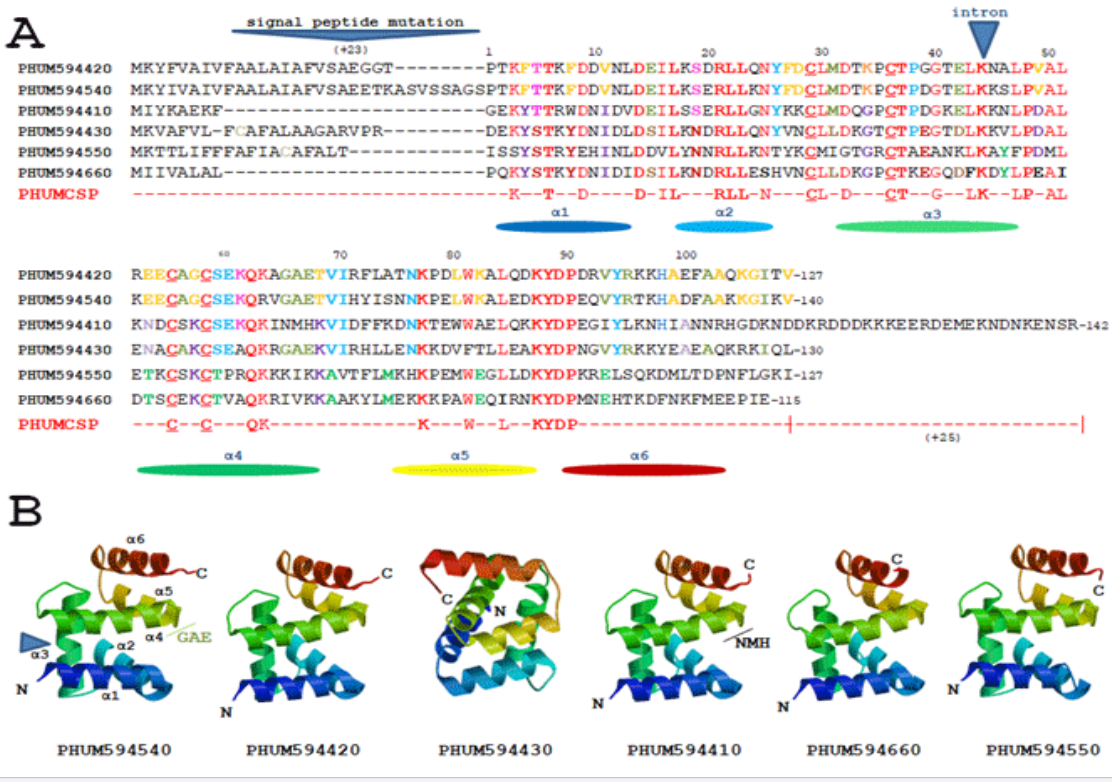

Figure 4: Identification of Pediculus humaniscorporis CSPs (PhumCSPs). (A) Sequence alignment of the six *PhumCSPs* extracted from *Pediculus* genome: *PHUM594410, PHUM594420, PHUM594430, PHUM594540, PHUM594550* and *PHUM594660.* Sequences are from in silico analysis of Fly Base and Vector Base (Table 3). Highly conserved amino acids are shown in red. Strictly conserved amino acid residues are shown in red and bold. The four Cys residues characteristic of CSPs are underlined. +Cys residues in signal peptide are shown in grey. Subtype-specific amino acid residues or motifs are shown by different colors. Switch to conserved amino acids at position 43, 49, 51 and 84 are shown in black bold. The arrows in blue indicate signal peptide mutation and intron insertion site after Lysine at position 44, respectively. The circles below the alignment show the position of six ∝-helices (∝1-∝6) from model structure. (B) Predicted tertiary structure of PhumCSPs. Homology modelling using MbraCSPA6 (1kx9.1.A; X-ray, 1.6 Å) as template (http://swissmodel.expasy.org) [8]. The blue arrow shows the position of intron (inserted in ∝3 at the bottom of the CSP prism). GAE and NMH correspond to subtype-specific motifs (mutation sites) at the extremity of ∝4-helix.

or diversified by this time, with sequence evidence reflecting the common structure of CSP, the presence of multiple subtypes and conserved intron insertion sites as early as the rapid emergence and dissemination of various eukaryote and prokaryote cells [61].

We ran a phylogenetic analysis using IQ-TREE to check for the distribution of B-CSPs among PHUMCSPs and CSP sequences from other insect and arthropod species. We found that B-CSPs and PHUMCSPs segregated together at least in two major groups (CSP1 and CSP2), containing also DmelOS-D and BmorCSP15, respectively (figure 5). *PHUM594420* and *PHUM594540* were clearly two gene duplicates related to Coleopteran *AAJJ0283A* [43]. *PHUM594420* and *PHUM594540* fell closely related to BtabCSP3, a gene from *B. tabaci* involved in metabolism of toxic plant oil chemicals [28], while *PHUM594430* fell closer to AmelASP3c, a gene supposedly involved in pheromone recognition in honeybees [62]. *PHUM594550* grouped with hymenopteran CSPs orthologous to *AmelGB17875* (figure 5). There is no cluster of lice-specific duplications, so PHUMCSPs are not involved in speciation. However, the amino acid tree shows that PHUMCSPs might originate from five serial gene duplication events (D1-D5) from *PHUM594660* to *PHUM594420/PHUM594540,* each of them corresponding to some specific amino acid replacements in the composition of CSP (figures 3 & 4). Very importantly, *PHUM594660* groups not only with eight BmorCSPs (BmorCSP3, BmorCSP11, BmorCSP12, BmorCSP13, BmorCSP15, BmorCSP17, BmorCSP18 and BmorCSP20) and ant/wasp CSP gene duplicates, but also with the microbial CSP gene *WP_120787175* from the bacterial firmicute bacillus bacillale Staphylococcaceae, *M. caseolyticus*. This suggests that CSP gene has been duplicated already in the very far common ancestor of bacteria and later on passed on to insects. Also very importantly, bacteria and insects share many other CSP genes such as *WP_120787167, WP_078880044, WP_082558797, PHUM594410* and *PHUM594420/PHUM594540*, which were duplicated or triplicated in lice and beetles for a function that remains to be found. Beetles, moths and bacteria share the same CSP genes such as *AAJJ0283B*, *BmorCSP14, WP_073810176/WP_083646628 and WP_043907137* (93% bootstrap). Similarly, bacteria and moths share the same genes WP_089438515 and *WP_120787151*

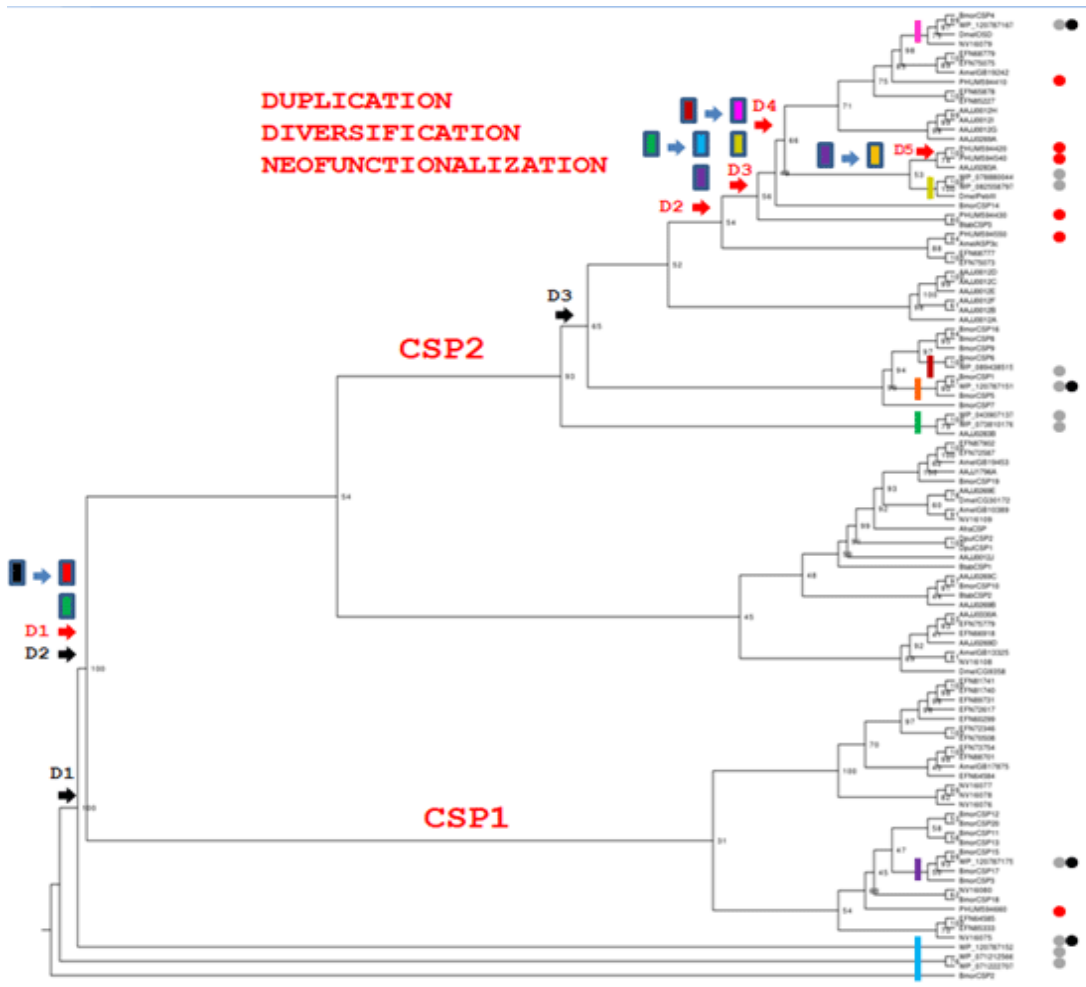

Figure 5: Phylogenetic analysis of bacterial and insect CSP amino acid sequences. Bacterial CSP sequences are those from Figure 1. WP_043: Kitasatosporagriseola (streptomycete), WP_071: Acinetobacter baumannii (cocobacillus), WP_073: K. sp. CB01950 (streptomycete), WP_078: K. purpeofusca (streptomycete), WP_082: K. sp. MBT66 (streptomycete), WP_089: Escherichia coli (proteobacter), WP_120: Micrococcus (staphylococcus). PhumCSPs are those from Figure 4. Other insect CSP sequences are from Picimbon [7], Picimbon et al.[21], Honeybee Genome Consortium [38]Forêt et al. [39], Xuan et al.[27], Liu et al. [28, 34,43-44], Bonasio et al. [40], Kulmuni et al.[41] and Zhao et al. [42].AAJJ: Tribolium castaneum (beetle), Amel: Apismellifera (bee), Bmor: Bombyxmori (moth), Btab: Bemisiatabaci (whitefly), Dmel: Drosophila melanogaster (fly), EFN: Harpegnathossaltator/ Camponotus floridanus (ant), NV: Nasoniavitripennis (wasp),Phum: Pediculushumaniscorporis (louse). Arthropod (crustacean) CSP sequences are from the brine shrimp Artemia franciscana (Afra CSP; ABY62736, ABY62738) and the water flea zooplankton Daphnia pulex: (DpulCSP1, DpulCSP2; ABH88166, ABH88167). Bootstrap values >50% after 1000 replicates are shown in support of the branching in IQ-TREE. BmorCSP2 (100% identity to bacterial counterparts) was used as out group. Grey dots indicate the position of B-CSPs. Red dots indicate the position of PhumCSPs. Black dots show the position of the four CSP subtypes identified in Macrococcus (staphylococcus). Black arrows indicate inferred gene-duplication events in bacteria (D1-D3). Red arrows indicate inferred gene-duplication events in insects (D1-D5). Vertical solid lines of different colors indicate different B-CSP subtypes (see Figure 1). Vertical solid squares of different colors indicate amino acid changes for specific motif transition mutation in PhumCSPs (see Figure 4). CSP1 group is a gene copy made at the time of a very early duplication (D1). Transition mutation rate is much higher in the second group of CSPs, CSP2 (D2-D5).

(BmorCSP1 and BmorCSP6), which were subsequently subjected to further duplication in the silk worm perhaps as an adaptive response to herbivory and plant chemical detoxification (figure 5) [27,28].

In CSP1, "Tegument-Retinal protein" WP_121154966 from the bacterial actinobacteria, micromonosporales, Micromonosporaceae, M. pisi grouped with BmorCSP17 (93% bootstrap; figure 5). "Tran membrane protein" WP_0964117339 did not group with BmorCSP5 and fell at the bottom of CSP2 group. It may suggest that one group of CSPs diverged from a prominent precursor tegument-retinal protein, while another group of CSPs (CSP2) diverged from a Toll-like precursor molecule.

Most importantly, our phylogenetic analysis suggests that serial duplication events occur not only in insects but also in bacteria. Profiling of gene duplication patterns of *pediculus* and *macrococcus* CSPs suggest that the very ancient gene duplication has shaped a CSP1 group that may be important for cell differentiation. The formation of the prokaryotic cell and later the development of multi cellular organisms are known to be accompanied by gene expression changes in differentiated cells. Thus, CSP1 may be one of the most ancient form of CSP and may represent a gene set that contributed to the morphological diversity across the bacterial and arthropod super kingdoms and/or contributed to protect both bacteria and arthropods from host plant poisoning [63,64,65]. The function of PHUM594660 and WP_120787175 (BmorCSP15) in the CSP protein family has been highly conserved after duplication events and over evolutionary time.

Profiling of gene duplication patterns of pediculus and macrococcus CSPs also suggest that further tandem duplications occurred in both prokaryotes and eukaryotes, leading to diversification and neo fictionalization of the CSP gene family at a common stage of evolution(CSP2; figure 5).Our phylogenetic analysis of the PHUMCSP family indicates that multiple duplication events have taken place after PHUM594550, at least three (D3, D4 and D5) of which led to the formation of four proteins with novel functions (figure 5). Comparative analysis of amino acid sequences from PhumCSPs show different steps of sequence mutations. Following duplication 1, CSP acquired specific elements such as N26, L44, V50, L52 and L85. Mutations were also observed in the signal peptide, C-tail and stop codon (figure 6A). Duplication 2 was accompanied by a few subtle amino acid replacements, but more drastic change occurred in the primary structure of CSP after duplication 3. Over duplication 3, some key residues changed in ∝1, ∝2 and ∝3, respectively. Signal peptide lost Cys residue and C-terminus acquired a long prominent tail (+25 amino acids) (figures 4 - 6A). Thus, CSPs probably acquired diversification and novel function(s) at that time. Interestingly, neofunctionalization and diversification of CSP gene continued over duplication 4, where mutations were observed mainly in the parallel ∝-helices ∝1 and ∝4.C-terminus (∝6) changed to A--AA-KGI-V motif (figures 4-6A). Following duplication 5, the motif corresponding to ∝4 and ∝5 loop was entirely replaced, as found for duplications 3 and 4 (figures 4-6A). These observations strongly argue for gene duplication and neofunctionalization of the CSP gene family over three steps in the louse genome.

In addition, focusing on PHUMCSPs suggests at least five independent duplication events followed by partial deletions or loss of one copy (figures 4-6A). Only one CSP gene is present in each orthology group, except for *PHUM594540* and *PHUM594420* (figure 5). The topology of the tree and sequence alignment shows that the two copies of the genes were preserved at each duplication event, but only one copy of the gene was subsequently subjected to duplication, the other gene remaining with the old function (figures 4-6). However, the repertoire of CSP sequences expressed in Pediculus cannot be explained by five serial duplications in six genes. Sequence alignment shows conservation of specific residues between *PHUM594430* and *PHUM594420*/*PHUM59440* (in green on figure 4), and a switch from GAE to NMH motif at the extremity of ∝4 in *PHUM594410.*This can be explained only by a situation where *PHUM594430* led to two copies, *PHUM594430* with the original GAE motif, and *PHUM594410* characterized by GAE substitution by NMH motif and expansion of C-tail (figure 6B). These two gene copies, *PHUM594430*and *PHUM594410*, were subsequently subjected to duplication. Duplication of *PHUM594430* led to *PHUM594540* and *PHUM594420*, while the post-duplication scenario was different for *PHUM594410*. In the louse genome there is only one copy of *PHUM594410* left, and somewhere along evolution, at least Pediculus has lost a duplicated copy of *PHUM594410.* This gene was not subjected to a further duplication event, so degeneration and gene loss was associated with *PHUM594410,* while PHUM594430associated with gene amplification, gene duplication and subsequent modification for neofunctionalization of the CSP gene family (figure 6B).PhumCSPs fall in the same orthology group than B-CSPs with significantly high boostrap value (53%; figure 5). Therefore, these genetic events that led to CSP diversification might have happened much before the diversification of insect species and the emergence of lice, i.e. much before the split of Eukaryotes and Prokaryote species.

Recent studies indicate possible links between insect CSPs and resistance to biological and chemical toxins [27,28,29]. Outbreaks of louse-borne diseases are usually frequent by-products of poverty, famine, poor hygiene, poor living conditions, droughts, earthquakes, floats, hurricanes, tornadoes, tsunamis and various other natural disasters, before the advent of insecticides. Largely due to the widespread use of insecticide products such as pyrethroid permethrins for control, a variety of lice has developed resistance to many families of insecticides or pediculicides, resulting in the emergence ore re-emergence of lice and lice-diseases in many areas, regions, provinces or states of the world [66-70]. The repertoire of (six) CSP genes identified in the body louse *P. humaniscorporis* may explain the resistance of the louse to permethrin or pediculicide, following the results from Xuan et al. (2015) and Liu et al. (2016) in moths and whiteflies, respectively, about the role of CSPs in insecticide resistance [27-28]. Similarly, the discovery of CSP proteins from *A. baumannii, M. /S. caseolyticus, K. griseola* and *E. coli* is important because it may reveal a key mechanism of bactericide resistance in some particular noxious strains mediators of severe infectious diseases

[71,72] Perhaps they went through a mechanism of toxic chemical resistance at an early stage of evolution. CSPs may be involved in the transmission of resistance from one cell to another as found in strains of *E. coli* that are resistant to one antibiotic and are proved to protect against the same antibiotic the bacterial cells growing nearby [73].

As a first step unraveling the role of CSPs in bacteria, we performed a phylogenetic analysis of B-CSPs not only with insect CSP sequences but also with related genes of various functions. These genes were extracted from microbial database by iterative blast search using CSP as a template. They do not share the configuration of disulfide bridges characteristic of CSPs, but

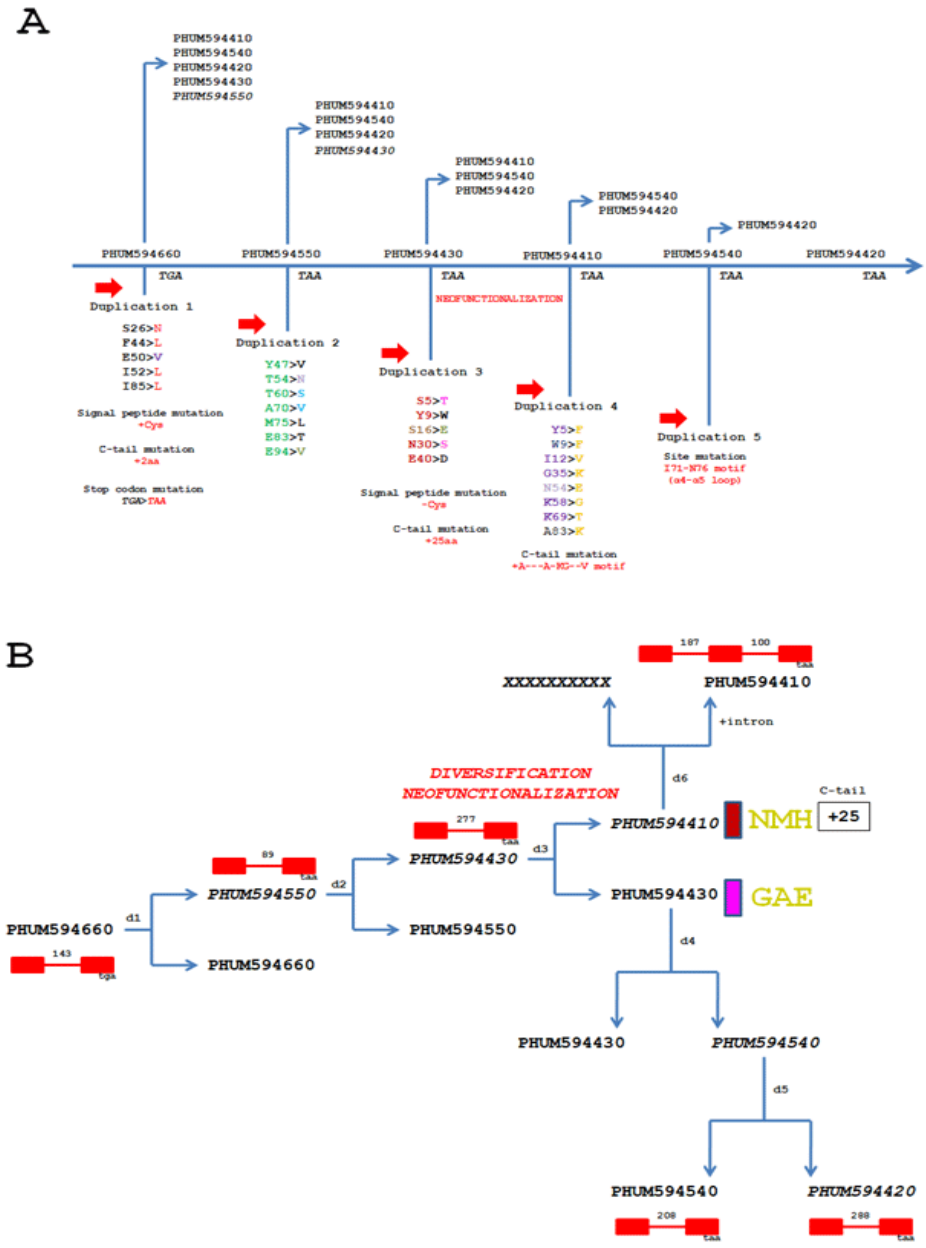

Figure 6: Duplication and transition mutation in PhumCSPs. (A) Point mutations and motif transitions resulting in neo functionalization in PhumCSP after duplication 3. (B) Inferring evolutionary scenarios in the duplication, loss and diversification of PhumCSP. PHUM594660 corresponds to the original CSP that became a major source of evolutionary innovation because it led to all other PhumCSPs through six gene duplication events (d1-d6). D3 led to the separation of two different subtypes characterized by NMH and GAE, respectively. The GAE-gene copies were preserved and diversified (d3, d4 and d5). The NMH-gene copy was lost following duplication (d6). The NMH-gene was subjected to specific mutation (intron gain), which brought a gene that can be clearly distinguished from the original copy, perhaps in response to major evolutionary transitions in eukaryotic and/or prokaryotic cells.

they share some amino acid motifs with the CSP gene family [74]. Perhaps most interestingly, the CSPs grouped specifically with genes related to various functions such as DNA-binding proteins, secondary metabolites transporters, pathway signal proteins, phosphatases, amino acid syntheses and transcriptional regulators, which may be some examples of protein partners by sub/neofunctionalization of the CSP gene family (figure 7). The tree is showing a very high bootstrap value for transmembrane protein WP_096417339 and a truncated gene of *B. mori*, BmorCSP5 (94%) in a large clade of BmorCSPs together with B-CSPs (51-99% bootstrap value) (figure 7). This may suggest that the original functions of CSPs may be linked to a transmembrane

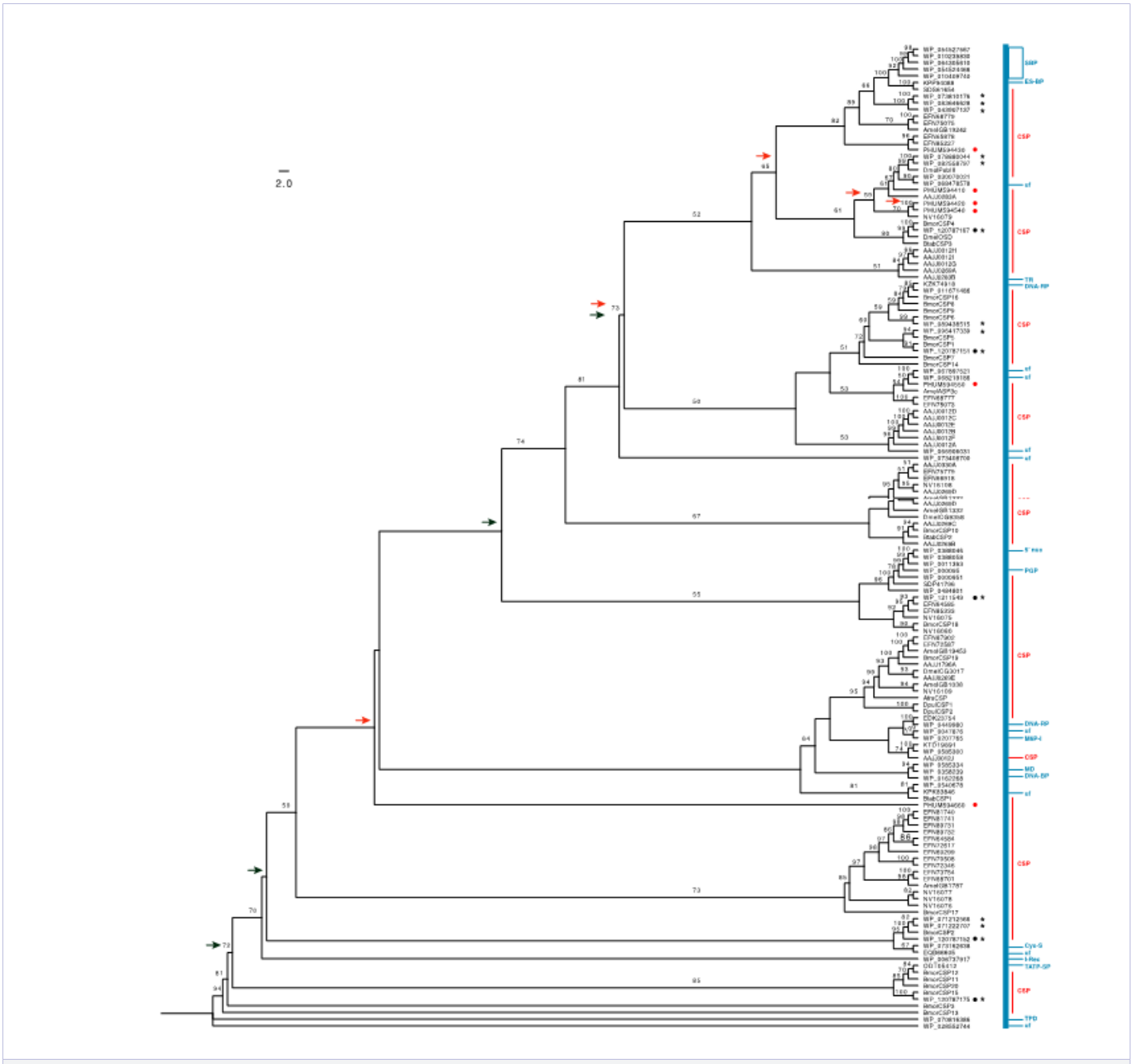

Figure 7: Phylogenetic analysis of CSPs and CSP-related gene families in bacteria. CSPs are those from figure 5. CSP-related genes in bacteria are from Liu and Picimbon [64]. Bootstrap values >50% after 1000 replicates are shown in support of the branching in IQ-TREE. 5'nuc: 5' nuclease, I-Rec: Type I deoxyribonuclease (DNase I), Cys-S: Cysteine synthase, DNA-BP: DNA-binding protein, DNA-RP: DNA-regulatory protein (transcriptional regulator), ES-BP: Extracellular solute-binding protein, FAD-BP: Oxidoreductase FAD-binding protein, M6P-I: Mannose-6-phosphate isomerase, MD: Malonate decarboxylase, PGP: Phosphoglycolate phosphatase, SBP: Substrate-binding protein, TATP: Twin-arginine translocation pathway signal protein, TPD: Thiamine-phosphate diphosphorylase, TR: Transcriptional/cell division repressor, uf: unknown function. Red dots indicate the position of PhumCSPs. Black dots indicate the position of Macro coccus (staphylococcus) CSPs. Black arrows indicate inferred gene-duplication events in bacteria. Red arrows indicate inferred gene-duplication events in insects. The asterisks in black show the position of all B-CSPs reported in our study.TPD (Thiamine-phosphate diphosphorylase) falls at the bottom of the tree, attracting BmorCSP3, BmorCSP13 and a group of BmorCSPs (BmorCSP11, BmorCSP12, BmorCSP15, BmorCSP20) with B-CSP WP_120787175 and twin-arginine translocation pathway signal protein (ODT05412) (85%).

protein similar to a toll-like receptor such as WP_096417339 and later on evolved through new functions through loss or truncation of the transmembrane component. The tree is also showing a very high bootstrap value for soluble BmorCSP1 and WP_120787151 (91%), and for soluble BmorCSP6 and WP_089438515 (99%), strongly suggesting that the coming of soluble CSP proteins from membrane has happened not in Lepidoptera, Phthiraptera or Mallophaga, not in insects or some other arthropod species, but in the bacterial prokaryote cell.

The CSP (OS-D or p10) multi gene protein family is gigantic, perhaps an amazing summary of the cell or organism evolution. From their findings in *Periplanetare* generating legs (p10) and *Drosophila* olfactory sensilla (OS-D) to their high diversity in a bacterial strain such as *A. baumannii* or *M. caseolyticus,* several questions arise. How many organisms express them? Where are they found and in which conditions are they expressed? How could this protein cause a cell to respond to xenobiotic chemical, plant metabolite or infectious agent? And, most pressing of all, what are their functions? A gene family that is a group of genes that share important characteristics, similar sequences of DNA building blocks (nucleotides), providing instructions for making products (proteins) that have a similar structure, motif or function, not only in insects and arthropods, but also in microbes and bacteria, certainly retains crucial importance for functional evolution in cells eventually compatible with the beginning of life.

We do not know if insect is evolved from bacteria or if LUCA and all his descendants from birds to reptiles have CSPs. However, the common pattern of evolution in CSPs we report here in a comparative analysis of lice and some infectious bacterial agents is very intriguing in the fact that the protein structure is conserved from Acinetobacter to louse. Lice have been parasitizing human and mammals since Prehistory and the origin of human lice and their role in vector pathogenic bacteria that cause human infectious disease may date as far as >6 Mya [75]. Ape body or head, louse and the infectious bacterium they carry probably show an intrinsic association in a faithful long and complex evolutionary history that can be traced back to [20-25] Mya. Bacteria provide essential vitamins to the lice, and the lice pass on diseases to humans. Symbiotic bacterial evolution is driven by highly specific associations with the louse, and bacterium/louse evolution is tied to mammal/human host [76]. We speculate that CSPs had a key role to play in the evolutionary history of bacteria residing in lice and later in the infectious process of lice parasitizing the mammalian/human body or head.

Considering more the association of this group of gene with human diseases, a related point to consider is that the CSP structure triggers plant immune responses when injected into the phloem, similarly to honeydew-associated microbes [77,78,79]. While CSP and/or honeydew microbes help release phytoalexins, jasmonic acid, salicylic acid and volatile organic compounds (VOCs) from the leaves, it would be interesting to see if the injection of CSP structures and/or the inoculation of specific bacteria such as *A. baumannii, M. / S. caseolyticus, K. griseola* and *E. coli* result in the activation of similar defense genes in the human tissue. The effects of CSPs in plants may be mediated by their ability to transport lipid and benzoic types of compounds as shown in whiteflies [28]. Lipid acid and benzoic acid are precursors of plant hormones in the jasmonate and salycilate class, respectively. Therefore, it is likely that CSPs can help stimulate specific hormone pathways by stimulating the transport of small precursor molecules. Whether this could be beneficial to treat some diseases related to lipid metabolism in human needs to be investigated further about the effects of injection of insect/bacterial CSPs through or via the human or mammalian skin. Lipid transporters are found in almost every type of intracellular organelle, but individual transporters have multiple functions, which strongly depend on the cell or tissue function [80]. It could be that a treatment with CSPs helps transport of lipids in the blood, lymph and body fluids, facilitating the dispersal of lipid droplets and chylomicrons after digestion. Combining CSPs to natural bacterial bio-products that target the gut flora may be an extremely beneficial approach for the treatment of obesity and other metabolic diseases [81]. There are numerous indications of an active lipid metabolism that is set out between bacteria and human cells or tissues. It has been proposed that this hijacking of lipids by the bacteria may be a way for the infectious agent to degrade host lipids while enriching its own stocks to survive, adapt and propagate to various tissues developing even more severe infection [82]. Thus, inhibiting B-CSPs may stop bacteria from hijacking human lipids. Then, an agent or a molecule capable to block B-CSPs in a specific manner may be able to influence the speed of recovery from an infectious disease. Lipid biological functions in bacteria are fuel for the regulation of cellular function, but mainly concern the membrane cell processes, i.e. selective permeability for exchange of essential water-soluble nutrients with the environment, cell division and cell protection [83]. So, altering the B-CSPs may interfere with the lipid pathways in the bacterial cell, affecting not only the membrane functions but also the cellular processes, which could be useful to counteract, exhaust or kill the defence mechanisms of particularly harmful microbial strains.

In addition, investigations about the function and binding properties of B-CSPs may open other new fields of research in medical microbiology. We may be able to increase the major beneficial activities of bacteria by cloning a specific CSP or CSP variant in a selection of microorganisms particularly useful or effective for human health protection [12,13,14,81,82,83]. We may also be able to develop new methods to knockdown specific CSP genes in toxic bacterial microorganisms, resulting not only in the alteration of lipid transport in the target bacterial cell, but also in the reduction of antibiotics/bactericide/germicide resistance and virulence in drug resistant strains. This area of research may bring some important concepts in cancer cell drug resistance [84]. It may also help understand the links between lipid metabolism and insulin resistance [85]. Therefore, controlling the function of bacterial CSPs appears to be in many ways a new very promising approach for controlling the pathogen city of harmful bacteria such as *A. baumannii, E. coli, Staphylococcus* and *Streptococcus,* among many others, as well as for seeing the interplay between lipids and drugs. A modern strategy of louse pest control may also consist in altering the physiology of the symbiotic bacterial cell, disrupting the main vitamin or fuel supply of the parasite. Our finding of the existence of the protein gene family, CSPs, in so numerous bacteria genera that belong to infectious pathogens and/or insect-bacteria symbiosis is crucial to treat human diseases by two means, i.e. targeting a component of bacterial cells and/or eliminating the infectious agent necessary for the growth of many various human body parasites.

Acknowledgements

Natural Sciences Foundation of Shandong Province

(ZR2011CM046) and Shandong Province Overseas High-Level Talents Program (Taishan scholar, No. tshw20091015)

References

1. Lamarck JB. PhilosophieZoologique, GermerBailliere. Paris. 1830.
2. Darwin C. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, Nature. 1859;5:502.
3. Owen R. Review of Darwin's Origin of Species. Edinburg Rev. 1860;3:487-532.
4. Vogt RG, Riddiford LM. Pheromone binding and inactivation by moth antennae. Nature. 1981;293:161-163.
5. Picimbon JF, Leal WS. Olfactory soluble proteins of cockroaches. Insect BiochemMolBiol. 1999;29(11):973-978.
6. AngeliS, CeronF, ScaloniA, Monti M, Monteforti G, Minnocci A, et al. Purification, structural characterization, cloning and immunocytochemical localization of chemoreception proteins from *Schistocercagregaria.* Eur J Biochem. 1999;262(3):745-754.
7. Picimbon JF. Biochemistry and evolution of CSP and OBP proteins. In: Blomquist GJ, Vogt RG, editors. Insect Pheromone Biochemistry and Molecular Biology, the Biosynthesis and Detection of Pheromones and Plant Volatiles. Elsevier Academic Press, London, San Diego. 2003;539-566.
8. Lartigue A, Campanacci V, Roussel A, Larsson AM, Jones TA, Tegoni M, et al. X-ray structure and ligand binding study of a moth chemosensory protein. J Biol Chem. 2002;277:32094-32098.
9. Jansen S, Zídek L, Löfstedt C, Picimbon JF, Sklenar V. 1H, 13C, and 15N resonance assignment of *Bombyx mori* chemosensory protein 1 (BmorCSP1). J Biomol NMR. 2006;36:47.
10. Jansen S, Chmelik J, Zídek L, Padrta P, Novak P, Zdrahal Z, et al. Structure of *Bombyxmori* Chemosensory Protein 1 in solution. Arch Insect Biochem Physiol. 2007;66(3):135-145.
11. Tomaselli S, Crescenzi O, Sanfelice D, Ab E, Wechselberger R, Angeli S, et al. Solution structureof a chemosensory protein from the desert locust *Schistocerca gregaria.* Biochemistry. 2006;45:1606-1613.
12. Xuan N, Bu X, Liu YY, Yang X, Liu GX, Fan ZX, et al. Molecular evidence of RNA editing in *Bombyx* chemosensory protein family. PLoS ONE. 2014.
13. Xuan N, Rajashekar B, Kasvandik S, Picimbon JF. Structural components of chemosensory protein mutations in the silkworm moth, *Bombyx mori*. Agri Gene. 2016;2:53-58.
14. Xuan N, Rajashekar B, Picimbon JF. DNA and RNA-dependentpolymerization in editingof*Bombyx*chemosensory protein (CSP) gene family. Agri Gene. 2019;12:100087.
15. Picimbon JF. Mutations in the insect transcriptome. J Clin Exp Pathol. 2016;6:3.
16. Picimbon JF. A new viewofgenetic mutations. Australas Med J. 2017;10:701-715.
17. Picimbon JF. Evolution of protein physicalstructures in insectchemosensory systems. In: Picimbon JF, editor. Olfactory Concepts of Insect Control-Alternative to Insecticides. Springer Nature, Switzerland. 2019.
18. Campanacci V, Lartigue A, Hällberg BM, Jones TA, Giuici-Orticoni MT and Tegoni M, et al. Moth chemosensory protein exhibits drastic conformational changes and cooperativity on ligand binding. ProcNatlAcad Sci. 2003;100(9):5069-5074.
19. NomuraA, Kawasaki K, Kubo T, Natori S. Purification and localization of p10, a novel protein that increases in nymphal regenerating legs of *Periplanetaamericana* (American cockroach). Int J Dev Biol. 1992;36(3):391-398.
20. Picimbon JF, Dietrich K, Breer H, Krieger J. Chemosensory proteins of *Locustamigratoria* (Orthoptera :Acrididae). Insect Biochem Mol Biol. 2000;30(3):233-241.
21. Picimbon JF, DietrichK, AngeliS, Scaloni A, KriegerJ and Breer H et al. Purification and molecular cloning of chemosensory proteins from *Bombyx mori*. ArchInsectBiochemPhysiol. 2000;44(3):120-129.
22. Celorio-ManceraMdP, Sundmalm SM, Vogel H, Rutishauser D, Ytterberg AJ, Zubarv RA et al. Chemosensory proteins, major salivary factors in caterpillar mandibular glands. Insect Biochem Mol Biol. 2012;42:796-805.
23. González-Caballero N, Valenzuela JG, Ribeiro JMC, Cuervo P, Brazil RP. Transcriptome exploration of the sex pheromone gland of *Lutzomyialongipalpis* (Diptera: Psychodidae: Phlebotominae). Parasit Vect. 2013;6:56.
24. Liu YL, Guo H, Huang LQ, Pelosi P, Wang CZ. Unique function of a chemosensory protein in the proboscis of two *Helicoverpa* species. J Exp Biol. 2014;217:1821-1826.
25. Perkin LC, Friesen KS, Flinn PW, Oppert B. Venom gland components of the ectoparasitoid wasp, *Anisopteromalus calandrae.* J. Venom Res. 2015;6:19-37.
26. Zhu J, Iovinella I, Dani FR, Liu YL, Huang LQ and Liu Y, et al. Conserved chemosensory proteins in the proboscis and eyes of Lepidoptera. Int J Biol Sci. 2016;12:1394-1404.
27. Xuan N, Guo X, Xie HY, Lou QN, Bo LX and Liu GX, et al. Increased expression of CSP and CYP genes in adult silkworm females exposed to avermectins. Insect Sci. 2015;22(2):203-219.
28. Liu GX, Ma HM, Xie YN, Xuan N, Xia G and Fan ZX, et al. Biotype characterization, developmental profiling, insecticide response and binding property of *Bemisiatabaci* chemosensory proteins: role of CSP in insect defense. PLoS ONE. 2016;11:e0154706.
29. Sabatier L, Jouanguy E, Dostert C, Zachary D, Dimarcq JL and Bulet P, et al. Pherokine-2 and -3: Two *Drosophila* molecules related to pheromone/odor-binding proteins induced by viral and bacterial infections. Eur J Biol. 2003;270:3398-3407.
30. Picimbon JF, DietrichK, Krieger J, Breer H. Identity and expression pattern of chemosensory proteins in *Heliothis virescens* (Lepidoptera, Noctuidae). Insect Biochem Mol Biol. 2001;31(12):1173-1181.

31. Wanner KW, Isman MB, Feng Q, Plettner E, Theilmann DA. Developmental expression patterns of four chemosensory protein genes from the Eastern spruce budworm, *Choristoneurafumiferana*. Insect Mol Biol. 2005;14(3):289–300.

32. Maleszka J, Forêt S, Saint R, Maleszka R. RNAi-induced phenotypes suggest a novel role for a chemosensory protein CSP5 in the development of embryonic integument in the honeybee (*Apismellifera*). Dev Genes Evol. 2007;217(3):189-196.

33. Guo W, Wang XH, Ma ZY, Xue L, Han JY, et al. CSP and Takeout genes modulate the switch between attraction and repulsion during behavioral phase change in the migratory locust. PLoS Genet. 2011;7:e1001291.

34. Liu GX, Ma HM, Xie HY, Xuan N, Picimbon JF. Sequence variation of *Bemisia tabaci* Chemosensory protein 2 in cryptic species B and Q: new DNA markers for whitefly recognition. Gene. 2016;576:284-291.

35. Pedra JH, Brandt A, Li HM, Westerman R, Romero-Serverson J and Pollack RJ, et al. Transcriptome identification of putative genes involved in protein catabolism and innate immune response in human body louse (Pediculicidae: *Pediculushumanus*). Insect Biochem Mol Biol. 2003;33(11):1135-1143.

36. Kirkness EF, Haas BJ, Sun W, Braig HR, Perotti MA and Clark JM, et al. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc Natl Acad Sci. 2010;107(27):12168-12173.

37. Wanner KW, Willis LG, Theilmann DA, Isman MB, Feng Q and Plettner E. Analysis of the insect os-d-like gene family. J Chem Ecol. 2004;30(5):889-911.

38. Honeybee Genome Sequencing Consortium. Insights into social insects from the genome of the honeybee *Apismellifera*. Nature 2006;443:931-949.

39. Forêt S, Wanner K, Maleszka R. Chemosensory proteins in the honeybee: insights from the annotated genome, comparative analysis and expressional profiling. Insect Biochem Mol Biol. 2007;37(1):19-28.

40. Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, Qin N, et al. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Science. 2010;329:1068-1071.

41. Kulmuni J, Wurm Y, Pamilo P. Comparative genomics of chemosensory protein genes reveals rapid evolution and positive selection in ant-specific duplicates. Heredity. 2013;110(6):538-547.

42. Zhao Y, Wang F, Zhang X, Zhang S, Guo S and Zhu G, et al. Transcriptome and expression patterns of chemosensory genes in antennae of the parasitoid wasp *Chouioiacunea*. PLoS ONE. 2016;11:e0148159.

43. Liu GX, Arnaud P, Offmann B, Picimbon JF. Genotyping and bio-sensing chemosensory proteins in insects. Sensors. 2017;17(8): 1801.

44. Liu GX, Xuan N, Rajashekar B, Arnaud P, Offmann B, Picimbon JF. Genetic plasticity of insect chemosensory proteins. Front Physiol. submitted.

45. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792-1797.

46. Nguyen LT., Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268-274.

47. Fig Tree, a graphical viewer of phylogenetic trees.

48. Wong TY, Fernandes S, Sankhon N, Leong PP, Kuo J, Liu JK. Role of premature stop codons in bacterial evolution. J Bacteriol. 2008;190:6718-6725.

49. Fiszer D, Rozwadowska N, Rychlewski L, Kosicki W, Kurpisz M. Identification of IL-18RAP mRNA truncated splice variants in human testis and the other human tissues. Cytokine. 2007;39(3):178-183.

50. Yun JH, Roh SW, Whon TW, Jung MJ, Kim MS, Park DS, et al. Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. Appl Environ Microbiol. 2014;80:5254-5264.

51. Book AJ, Lewin GR, McDonald BR, Takasuka TE, Doering DT, Adams AS, et al. Cellulolytic Streptomyces strains associated with herbivorous insects share a phylogenetically linked capacity to degrade lignocellulose. Appl Environ Microbiol.2014;80:4692-4701.

52. Gonzales-Escobar JL, Grajales-Lagunes A, Smolinski A, Chagolla-Lopez A, De Leon-Rodriguez A, Barba de la Rosa AP. Microbiota of edible *Liometopumapiculatum* ant larvae reveals potential functions related to their nutritional value. Food Res Int. 2018;109:497-505. doi: 10.1016/j.foodres.2018.04.049

53. Salzman S, Whitaker M, Pierce NE. Cycad-feeding insects share a core gut microbiome. Biol J Linn Soc. 2018; 123:728-738.

54. Paniagua Voirol LR, Frago E, Kaltenpoth M, Hilker M, Fatouros NE. Bacterial symbionts in Lepidoptera: their diversity, transmission, and impact on the host. Front Microbiol. 2018;9:556. doi: 10.3389/fmicb.2018.00556

55. Engel P, Moran N. The gut microbiota of insects-diversity in structure and function. FEMS Microbiol Rev. 2013;37(5):699-735. doi: 10.1111/1574-6976.12025

56. La Scola B, Raoult D. *Acinetobacterbaumannii* in human body louse. Emerg Infect Dis. 2004;10:1671-1673.

57. Sunantaraporn S, Sanprasert V, Pengsakul T, Phumee A, Boonserm R, Tawatsin A, et al. Molecular survey of the head louse *Pediculushumaniscapitis* in Thailand and its potential role for transmitting Acinetobacter spp. ParasitVect. 2015;8:127. doi: 10.1186/s13071-015-0742-4

58. Araujo A, Ferreira LF, Guidon N, Maues Da Serra Freire N, Reinhard KJ, Dittmar K. Ten thousand years of head lice infection. Parasitol Today. 2000;16(7):269.

59. Badiaga S, Brouqui P. Human louse-transmitted infectious diseases. ClinMicrobiol Infect. 2012;18(4):332-337. doi: 10.1111/j.1469-0691.2012.03778.x

60. Koonin EV. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus intron late debate? Biol Direct. 2006;1:22. doi: 10.1186/1745-6150-1-22

61. Cooper GM. The origin and evolution of cells. In: The cell: a molecular approach, 2nd edition, Sunderland (MA), Sinauer Associates, 2000;689.

62. Briand L, Swasdipan N, Nespoulos C, Bezirard V, Blon F, Huet JC, et al. Characterization of a chemosensory protein (ASP3c) from honeybee (*Apismellifera* L.) as a brood pheromone carrier. Eur J Biochem. 2002;269:4586-4596. doi:10.1046/j.1432-1033.2002.03156.x

63. Sprusansky O, Zhou L, Jordan S, White J, Westpheling J. Identification of three new genes involved in morphogenesis and antibiotic production

in Streptomyces coleicolor. J Bacteriol. 2003;185(20):6147-6157. doi: 10.1128/JB.185.20.6147-6157.2003

64. Vollmer J, Schiefer A, Schneider T, Julicher K, Johnston KL, Taylor MJ, et al. Requirement of lipid II biosynthesis for cell division in cell wall-less *Wolbachia*, endobacteria of arthropods and filial nematodes. Int J Med Microbiol. 2013;303:140-149. doi: 10.1016/j.ijmm.2013.01.002
65. Wybouw N, Dermauw W, Tirry L, Stevens C, Grbic M, Feyereisen R et al. A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. Elife. 2014;3:e02365. doi: 10.7554/eLife.02365
66. Downs AMR, Stafford KA, Coles GC, Downs AMR, Stafford KA, Coles GC. Head lice: prevalence in schoolchildren and insecticide resistance. Trends Parasitol. 1999;15(1):P1-4.
67. Downs AMR, Stafford KA, Hunt LP, Ravenscroft JC, Coles GC. Widespread insecticide resistance in head lice to the over-the-counter pediculocides in England, and the emergence of carbaryl resistance. B J Dermatol. 2002;146:88-93.
68. Thomas DRh, McCarroll L, Roberts R, Karunaratne P, Roberts C, Casey D, et al. Surveillance of insecticide resistance in headlice using biochemical and molecular methods. Arch Dis Child. 2006;91:777-778. doi: 10.1136/adc.2005.091280
69. Durand R, Bouvresse S, Berdjane Z, Izri A, Chosidow O, Clark JM. Insecticide resistance in head lice: clinical, parasitological and genetic aspects. Clin Micro biol Infect. 2012;18(4):338-344. doi: 10.1111/j.1469-0691.2012.03806.x
70. Drali R, Benkouiten S, Badiaga S, Bitam I, Rolain JM, Brouqui P. Detection of a knockdown resistance mutation associated with permethrin resistance in the body louse *Pediculushumaniscorporis* by use of melting curve analysis genotyping. J ClinMicrobiol. 2012;50(7):2229-2233. doi: 10.1128/JCM.00808-12
71. Cloete TE. Resistance mechanism of bacteria to antimicrobial compounds. Int Bio deter Bio degr. 2003;51(4):277-282.
72. Wyrsch ER, Chowdhury R, Chapman TA, Charles IG, Hammond JM, Djordjevic SP. Genomic microbial epidemiology is needed to comprehend the global problem of antibiotic resistance and to improve pathogen diagnosis. Front Microbiol. 2016:7:843. doi: 10.3389/fmicb.2016.00843
73. Yurtsev EA, Conwill A, Gore J. Oscillatory dynamics in a bacterial cross-protection mutualism. Proc Natl Acad Sci USA. 2016;113(22):6236-6241. doi: 10.1073/pnas.1523317113
74. Liu GX, Picimbon JF. Bacterial origin of insect chemosensory odor-binding proteins. Gene TranslBioinform. 2017; 3: e1548.
75. Boutellis A, Abi-Rached L, Raoult D. The origin and distribution of human lice in the world. Infect Genet Evol. 2014;23:209-217. doi: 10.1016/j.meegid.2014.01.017
76. Boyd BM, Allen JM, Nguyen NP, Vachaspati P, Quickcall Z, Warnow T, et al. Primates, lice and bacteria: speciation and genome evolution in the symbionts of hominid lice. MolBiolEvol. 2017;34(7):1743-1757. doi: 10.1093/molbev/msx117
77. Bos JI, Prince D, Pitino M, Maffei ME, Win J, Hogenhout SA. A functional genomics approach identified candidate effectors from the aphid species *Myzuspersicae* (green peach aphid). PLoS Genet. 2010;6(11):e1001216. doi: 10.1371/journal.pgen.1001216
78. Rodriguez PA, Stam R, Warbroek T, Bos JI. Mp10 and Mp42 from the aphid species *Myzuspersicae* trigger plant defenses in *Nicotianabenthemiana* through different activities. Mol Plant Microbe Interact. 2014;27(1):30-39. doi: 10.1094/MPMI-05-13-0156-R
79. Wari D, Kabir MA, Mujiono K, Hojo Y, Shinya T, Tani A, et al. Honeydew-associated microbes elicit defense responses against brown planthopper in rice. J Exp Bot. 2019;70(5):1683-1696. doi: 10.1093/jxb/erz041
80. Tarling EJ, de AguiarVallim TQ, Edwards PA. Role of ABC transporters in lipid transport and human disease. Trends EndocrinolMetabol. 2013;24(7):342-350.
81. Yue S, Chinnapandi B, Ge H, Zou X, Chen X, Wang C, et al. A lactobacillus cocktail changes gut flora and reduces cholesterolemia and weight gain in hyperlipidemia mice. SOJ Microbiol Infect Dis. 2014;2(2):1-14.
82. Kumar NG, ContaiferJr D, Baker PRS, Ekroos, Jefferson KK, Wijesinghe DS. Untargeted lipidomic analysis to broadly characterize the effects of pathogenic and non-pathogenic staphylococci on mammalian lipids. PLoS ONE. 2018;13(10):e0206606.
83. Barak I, Muchova K. The role of lipid domains in bacterial cell processes. Int J Mol Sci. 2013;14(2):4050-4065. doi: 10.3390/ijms14024050
84. Iwamoto H, Abe M, Yang Y, Cui D, Seki T, Nakamura M, et al. Cancer lipid metabolism confers antiangiogenic drug resistance. Cell Metabol. 2018;28(1):104-117.e5. doi: 10.1016/j.cmet.2018.05.005
85. Savage DB, Petersen KF, Shulman GI. Disordered lipid metabolism and the pathogenesis of insulin resistance. Physiol Rev. 2007;87(2):507-520.