2019

Genome-wide annotation and comparative analysis of cuticular protein genes in the noctuid pest *Spodoptera litura*

Jianqiu Liu

Shenglong Li

See next page for additional authors

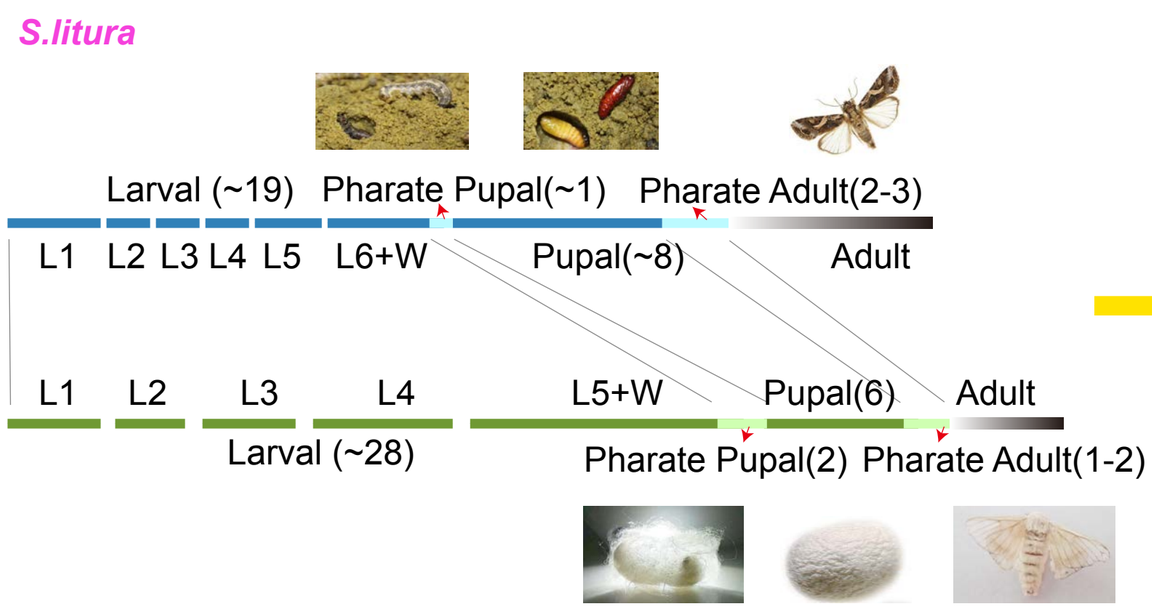
Citation/Publisher Attribution
Liu, J., Li, S., Li, W., Peng, L., Chen, Z., Xiao, Y., Guo, H.,...Mita, K. (2019). Genome-wide annotation and comparative analysis of cuticular protein genes in the noctuid pest *Spodoptera litura*. *Insect Biochemistry and Molecular Biology*, *110*, 90-97. doi: 10.1016/j.ibmb.2019.04.012
Available at: https://doi.org/10.1016/j.ibmb.2019.04.012

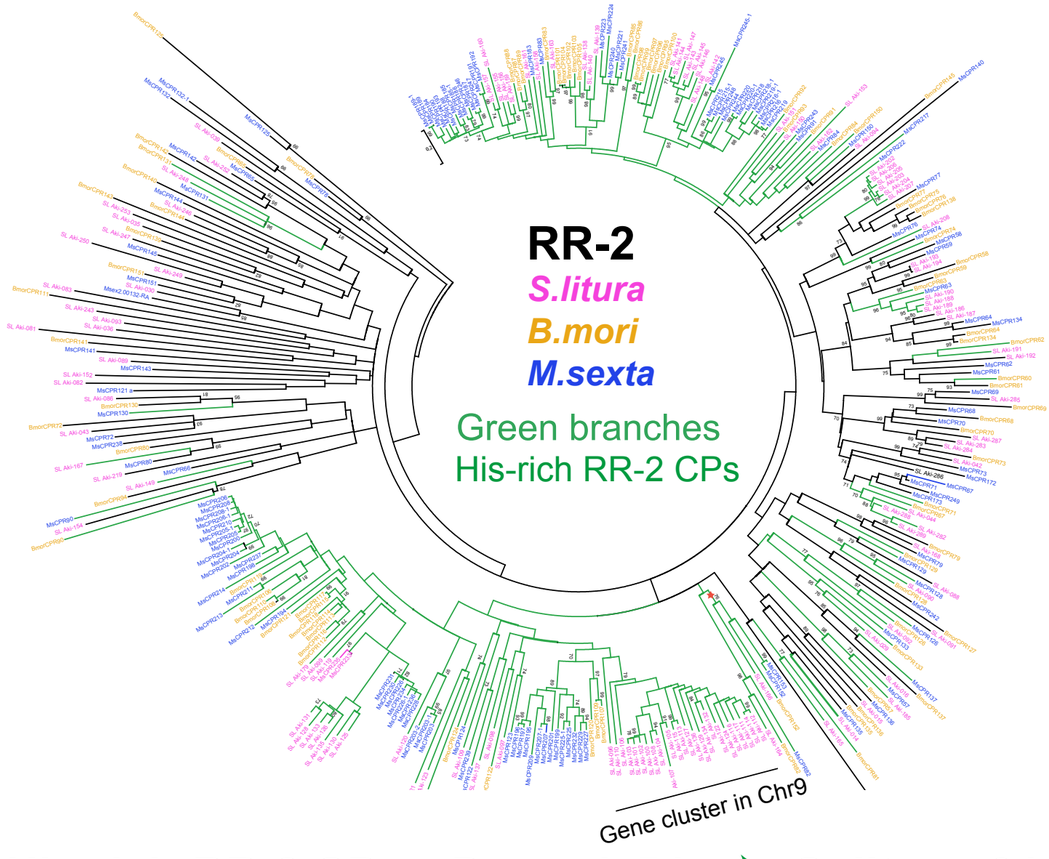
Highlights

- In total, 287 cuticular protein genes are annotated in *Spodoptera litura* genome.
- Most RR-1 and His-rich RR-2 CPs form a major cluster in *S. litura* chromosome 9
- Histidine-rich RR-2 CPs are greatly expanded in *S. litura* compared to *Bombyx mori*.
- RR-1 CP genes are mostly conserved among *S. litura*, *B. mori*, and *Manduca sexta*.
- His-rich RR-2 CP genes form species-specific clades in phylogenetic analysis.

Graphical Abstract

Major histidine-rich CP expansion in *S. litura*





His-rich RR-2 CPs: *B. mori*: 54 → *S. litura*: 91

In press, Insect Biochemistry and Molecular Biology, April 13, 2019

Genome-wide annotation and comparative analysis of cuticular protein genes in the noctuid pest *Spodoptera litura*

Jianqiu Liu[1,2], Shenglong Li[1,2], Wanshun Li[1,2], Li Peng[1], Zhiwei Chen[1,2] Yingdan Xiao[1,2], Huizhen Guo[1,2], Jiwei Zhang[1,2], Tingcai Cheng[1,2], Marian R. Goldsmith[1, 2, 3], Kallare P. Arunkumar[1, 2, 4], Qingyou Xia[1,2], Kazuei Mita[1,2]

[1]State Key Laboratory of Silkworm Genome Biology, Biological Science Research Center, Southwest University, Chongqing 400716, China.

[2]Chongqing Key Laboratory of Sericultural Science, Chongqing Engineering and Technology Research Center for Novel Silk Materials, Southwest University, Chongqing 400716, China.

[3]University of Rhode Island, Kingston 02881, USA

[4]Central Muga Eri Research and Training Institute, (CMER&TI), Central Silk Board, Lahdoigarh, Jorhat 785700, India

Corresponding author

State Key Laboratory of Silkworm Genome Biology, Biological Science Research Center, Southwest University, Chongqing 400715, China. E-mail: mitakazuei@gmail.com

Keywords:

Cuticular protein

Comparative analysis

Histidine-rich

RNA-Seq

Spodoptera litura

Abstract

Insect cuticle is considered an adaptable and versatile building material with roles in the construction and function of exoskeleton. Its physical properties are varied, as the biological requirements differ among diverse structures and change during the life cycle of the insect. Although the bulk of cuticle consists basically of cuticular proteins (CPs) associated with chitin, the degree of cuticular sclerotization is an important factor in determining its physical properties. *Spodoptera litura*, the tobacco cutworm, is an important agricultural pest in Asia. Compared to the domestic silkworm, *Bombyx mori*, another lepidopteran whose CP genes have been well annotated, *S. litura* has a shorter life cycle, hides in soil during daytime beginning in the 5th instar and is exposed to soil in the pupal stage without the protection of a cocoon. In order to understand how the *CP* genes may have been adapted to support the characteristic life style of *S. litura*, we searched its genome and found 287 putative cuticular proteins that can be classified into 9 CP families (CPR with three groups (RR-1, RR-2, RR-3), CPAP1, CPAP3, CPF, CPFL, CPT, CPG, CPCFC and CPLCA), and a collection of unclassified CPs named CPH. There were also 112 cuticular proteins enriched in Histidine residues with content varying from 6% to 30%, comprising many more His-rich cuticular

proteins than *B. mori*. A phylogenetic analysis between *S. litura, M. sexta* and *B. mori* uncovered large expansions of *RR-1* and *RR-2* CPs, forming large gene clusters in different regions of *S. litura* chromosome 9. We used RNA-seq analysis to document the expression profiles of *CPs* in different developmental stages and tissues of *S. litura*. The comparative genomic analysis of CPs between *S. litura* and *B. mori* integrated with the unique behavior and life cycle of the two species offers new insights into their contrasting ecological adaptations.

1. Introduction

Insect cuticle must provide an effective barrier from the natural environment. Consequently, its physical properties, such as thickness, stiffness, strength, elasticity and color, show large variations at different metamorphic stages and in different anatomical regions. Cuticles have a common fundamental structure, consisting of a procuticle composed of a filamentous chitin structure within a protein matrix covered by an epicuticle consisting of lipids and protein above which there is a dense cuticulin lamina (Locke, 2001). The variation in physical properties of cuticle is partly due to different degrees of cross-linking and hardening occurring during the process of sclerotization, whereby phenolic material is incorporated into the CPs and/or other cross-links are formed. Additionally, numerous CPs are identified in all insect species studied, and their number and features are quite different among diverse species (Willis et al., 2012).

Information about the CPs of various insects and their underlying genes has been obtained at the transcriptome and protein levels in the past few decades (Andersen, 2000; Baton et al., 2009; Dittmer et al., 2015; Dong et al., 2016; Futahashi. et al., 2008; Gu and Willis, 2003; He et al., 2007; Pan et al., 2018). With the improvement of sequencing technology, more and more genomic information for insect CPs has become available, with many thousands of CP coding genes accumulated in sequence databases. Many CPs have been identified by their conserved protein sequence motifs. Andersen et al. (1995), Willis (2010) and Willis et al. (2012) have defined several CP families. The most abundant family of CPs contains the Rebers and Riddiford Consensus (R&R Consensus), which in an extended form has been shown to bind chitin (Dong et al., 2016; Rebers and Willis, 2001; Tang et al., 2010; Togawa et al., 2007; Togawa et al., 2004). Three distinct forms of this consensus have been classified as RR-1, RR-2 and RR-3 (Andersen, 1998; Andersen, 2000). The other families with conserved motifs are CPs with a 44 amino acid motif (CPF), CPF-like in the conserved C-terminal region (CPFL), the Tweedle motif (CPT), alanine-rich CPs of low complexity (CPLCA), CPs of low complexity with two invariant glycine residues in the conserved domain (CPLCG), CPs of low complexity with an invariant tryptophan in the conserved domain (CPLCW), proline-rich CPs of low complexity (CPLCP), CPs with well-conserved cysteine residues (CPCFC), a glycine-rich CP (CPG), and analogs to peritrophins (CPAP1 and CPAP3).

Spodoptera litura (Lepidoptera, Noctuidae) is an important agricultural pest distributed in the tropical and subtropical areas of Asia. Compared with the domestic silkworm, *Bombyx mori*, *S. litura* has a shorter life cycle, although it has one more larval instar (Fig. 1). In *S. litura*, instars L2-L5 are shorter than L1, whereas in *B. mori,* larval stages following L1 take increasingly longer times. Moreover, from the 5th instar *S. litura* hides in the soil during the daytime, comes out from the soil in the evening to eat crops throughout the night and then goes back into the soil with daybreak. Further, instead of residing in a protective cocoon from pupation until adult eclosion like *B. mori*, *S. litura* stays in the soil from the wandering stage until the moth emerges. This exposure to soil requires CPs to form a more protective cuticle for protection against abrasion and fungal and bacterial infection compared with *B. mori*. It is also informative to compare *S. litura* with *M. sexta*, a close relative of *B. mori* that stays underground during pupal and pharate adult stages.

In this study, we annotated *CP* genes based on the recently published complete genome sequence of *S. litura* (Cheng et al., 2017). Combined with transcriptome analysis, we then estimated what different kinds of CPs contribute to different types of cuticles among various tissues and in different metamorphic stages. In addition, comparative genomics and phylogenetic analysis among Lepidoptera provided information on how *CP* genes evolved to adapt to the different ecological niches of each of these three species.

2. Materials and Methods

2.1 Annotation of cuticular protein genes

To predict putative *CP* genes of *S. litura,* reported sets of lepidopteran CPs were collected from KAIKObase (http://sgp.dna.affrc.go.jp/KAIKObase/) and the NCBI Reference Sequence database (https://www.ncbi.nlm.nih.gov/genbank/). *CP* genes were predicted for the *S. litura* genome assembly (Cheng *et al*., 2017) using TBLASTN (E-value $<10^{-5}$) and BLASTP in the non- redundant GenBank database. Predicted *CP* genes were further examined by HMMER3 search (cutoff E-value = 0.001) using the Pfam database to confirm conserved domains and subsequently classified into 9 families based on conserved motifs with the help of an online tool (http://bioinformatics.biol.uoa.gr/CutProtFam-Pred/) (Ioannidou et al., 2014). The annotation sequences were deposited at GenBank under BioProject accession PRJNA344815.

2.2 Phylogenetic tree construction

A total of 199 RR-1 and 361 RR-2 CDS sequences from *S. litura*, *B. mori,* and *M. sexta* were aligned using ClustalW in MEGA6 (Tamura et al., 2013). The tree was constructed using the Maximum Likelihood method based on the Jukes-Cantor model

(Jukes and cantor, 1969). A bootstrap consensus tree was inferred from 1000 replicates (Felsenstein, 1985). Adobe Illustrator CS6 was used for editing and drawing the trees.

For the CPFL family, 17 sequences from *S. litura*, *B. mori,* and *M. sexta* were used for phylogenetic tree construction. The sequences were aligned using the program ClustalW in MEGA6 (Tamura et al., 2013). Trees were constructed using the Neighbor-Joining method (Nei, 1987), and gaps were treated by pairwise deletion.

2.3 Insect rearing, RNA library preparation, and RNA-Seq

S. litura (Ishihara inbred strain) was reared on artificial diet at 25°C under a 12h light/ 12h dark cycle as described previously (Cheng et al., 2017), and RNA-Seq libraries were prepared from different developmental stages. For 1st and 2nd instar larvae (L1D2 and L2D2), RNA was extracted from the whole body of second-day larvae. For 3rd to 5th (L3D2, L4D2 and L5D2) instar samples, RNA was derived from epidermis of second-day larvae. For the 6th instar and wandering stages (L6D2 and W), epidermal RNA samples were collected from second day larvae in the daytime sleeping (D) phase and nighttime feeding phase (N). Although other tissues were removed carefully with tweezers under a ZEISS Stemi 2000 microscope (ZEISS, Germany), small amounts of muscle, fat body or trachea might have contaminated the libraries. RNA samples from epidermis and wing were extracted separately from 2nd (P-2), 6th (P-6), 9th and 12th-day pupae (P-9 and P-12: pharate adults). RNA isolation, library construction, sequencing and analysis of transcripts were carried out as described in Cheng et al. (2017). The

RNA-seq data are deposited in the Sequence Read Archive database under SRA accession: PRJNA498147 (https://www.ncbi.nlm.nih.gov/sra/PRJNA498147). The Log_2 (FPKM+0.01) (fragments per kilobase per million fragments mapped) value was used for making the heat maps.

3. Results and Discussion

3.1 Annotation of cuticular proteins

As shown in Table 1, 287 putative CP coding genes (Supplementary Table 1) were predicted and classified into eight CP families: CPR subdivided into three groups, (RR-1, RR-2 and RR-3), Tweedle, CPF, CPFL, CPLCA, CPCFC, CPAP1, CPAP3 and Glycine-rich (CPG). In addition, a collection of unclassified CPs was named CPH following Futahashi et al. (2008).

Among the 3 R&R consensus groups used to subdivide the CPR family in *S. litura,* 63 RR-1 protein genes and 129 RR-2 protein genes were identified with CutProtFam-Pred, and one RR-3 protein gene was identified by manual annotation based on its similarity to RR-3 proteins in *B. mori*.

CPAP1 and CPAP3 families, which contain one or three peritrophin A-type chitin-binding domains, are expressed only in cuticle-forming tissues (Jasrapuria et al., 2010). Thirteen *CPAP1s* and 9 *CPAP3s* were annotated in the *S. litura* genome (Table 1) (Tetreau et al., 2015).

158 The CPLCA family is defined by the presence of the retinin domain (pfam04527)
159 and richness in alanine residues, varying from 13-26% (Cornman and Willis, 2009).
160 Four *CPLCA* genes (*Sl_Aki-270, 271, 272* and *273*) which formed a gene cluster in *S.*
161 *litura* chromosome 28 (Chr28) were annotated. Two CPs, named *BmorCPH6* and
162 *BmorCPH7* by Futahashi *et al.* (2008) in *B. mori*, were related to CPLCA by homology
163 search. This is the first time CPLCAs, originally described in *Anopheles gambiae*
164 (Cornman and Willis, 2009; Willis et al., 2012), were found in Lepidoptera.

165 CPF has a conserved region with 44 amino acids (Togawa et al., 2007). One putative
166 *CP* gene belonging to the CPF family in the *S. litura* genome (SWUSl0111680) was
167 identified which we named *Sl_Aki-CPF*. Seven *CPFL* genes, which have a conserved
168 C-terminal region similar to CPF (Togawa et al., 2007), were found. All of the *CPFL*
169 genes together with two other *CPG* genes (*Sl_Aki-235* and *242*) formed a cluster on
170 Chr25. This genomic structure was similar to *B. mori*, in which three CPFL genes
171 (*BmorCPFL2*, *BmorCPFL3* and *BmorCPFL4*) and two CPG genes (*BmorCPG23* and
172 *BmorCPG42*) form a gene cluster on Chr23. In addition, *Sl_Aki-235* and *Sl_Aki-242*
173 were orthologous to *BmorCPG23* and *BmorCPG42,* respectively.
174 We identified five genes with a Tweedle motif, among which *Sl_Aki-292, 293* and
175 *294* formed a cluster on ChrZ; the other two *CPT* genes were not linked to other *CPs*.

176 The CPG family contains GG repeats but does not have any definite motif described
177 in other CP families (Futahashi. et al., 2008; Willis et al., 2012). However, although

178 some glycine-rich cuticular proteins also had an R&R Consensus or Tweedle motif
179 (i.e., *Sl_Aki-008* (RR-1), *Sl_Aki-185* (RR-2), and *Sl_Aki-34* (CPT)), we left them in
180 their well defined families. In *S. litura*, 28 putative *CP* genes were classified as CPGs.

181 Futahashi et al. (2008) classified a group of 34 proteins as CPH which stands for
182 cuticular protein, hypothetical, among which we assigned BmorCPH6 and 7 to CPLCA
183 and BmorCPH1 to CPCFC. They all have signal peptides and some sequence similarity
184 with known *CP* genes or with the AAP (A/V) motif often found in CPs (Futahashi. et
185 al., 2008; Magkrioti et al., 2004). Twenty-six putative *CP* genes which showed
186 sequence similarity with 31 already known CPH proteins in *B. mori* were classified as
187 *CPH* in *S. litura*. Most of them had signal peptides*;* additionally, the AAP (A/V) motif
188 was found in 20 CPH proteins. Only 6 of these had sequence similarity with already
189 known CPH proteins in *B. mori*.

190 **3.2 Characterization of histidine–rich CPs**

191 The amino acid composition of the 287 CPs revealed 112 His-rich CPs in which
192 histidine residues ranged from 6.00–30.14% (Fig. 2). Interestingly, most His-rich CPs
193 belonged to the RR-2 family, which is common to both *B. mori* and *S. litura*. *B. mori*
194 has one extremely high His-rich CP, BmorCPR152, of 45%, while the orthologous CP
195 with the highest His content in *S. litura* is Sl_Aki-166 with 30.14%. Figure 3 shows a
196 large expansion of His-rich *CPs* of the RR-2 family in the *S. litura* genome. As
197 Andersen and Roepstorff (2007) stated, histidine and several other amino acid residues

198 in CPs could be involved in cuticular adduct formation based on biochemical analysis
199 of the cuticular hydrolysates from different cuticular types of *M. sexta*, desert locust
200 *Schistocerca gregaria,* and yellow mealworm, *Tenebrio molitor* (Andersen, 2007;
201 Andersen and Roepstorff, 2007; James L.Kerwin, 1999; Kerwin, 1999; SO Andersen
202 1997). EM immunolocalization utilized by Vannini and Willis (2017) support the
203 hypothesis about the deployment of RR-1 and RR-2 localization first proposed by
204 Andersen (1998). Namely, the RR-1 CPs are mostly found in soft cuticle like
205 inter-segmental membranes, whereas RR-2s are restricted to hard cuticles in *A.*
206 *gambiae* (Vannini and Willis, 2017). Histidino-β-dopamine is the dominating adduct in
207 hard cuticles like those of adult beetle cuticle, lepidopteran pupae and dipteran puparia
208 (Andersen, 2008). Consistent with these reports is that His-rich CPs were mostly found
209 in RR-2, but not RR-1 CPs of *S. litura*.

210 **3.3 Major clusters of *CPR* genes in *S. litura***

211 Compared to *B. mori* which has a cluster of CPR genes on chromosome 22, major
212 expansions of *S. litura CPR* genes derived from this family were on *S. litura* Chr9.
213 Figure 3 shows that 34 RR-1 (*Sl-Aki-48 - 80*) genes and 82 RR-2 (*Sl-Aki-86, 88 -94,*
214 *96-169*) genes were present as two large clusters located on different regions of Chr9.
215 Intriguingly, all of the RR-2 *CP* members belonged to the cluster encoding His-rich
216 CPs, whereas none of the 34 RR-1 *CPs* in the large cluster on Chr9 were His-rich.
217 Chr22 in *B. mori* also has the orthologs of these RR-1 genes, but none of them were
218 found to be His-rich (Fig. 3). Although much smaller than *S. litura*, there was a

219 separate His-rich RR-2 *CP* cluster (*BmorCPR79-129*) on Chr22 in *B. mori.* Thirteen
220 RR-1 (*Sl-Aki-001 - 013*) genes were also localized on *S. litura* Chr1 as a cluster
221 (Supplementary Fig. 1). Their orthologs in *B. mori* also formed a cluster on Chr9.

222 **3.4 Phylogenetic analysis of *S. litura*, *B. mori* and *M. sexta* CPs**

223 To compare the *S. litura* CPs with *B. mori* and *M. sexta* CPs, phylogenetic trees of
224 RR-1 and RR-2 CPs were constructed separately (Fig. 4). A single species clade was
225 very rare in the RR-1 tree. Although a small number of RR-1 *CP* genes were expanded
226 in comparison with *B. mori*, more than half of them showed one-to-one correspondence
227 among the three lepidopteran species. Sl_Aki-2, 3, 4, 5 and 6 and MsCPR13, 174, 175,
228 176 and 177 formed separate clades, which corresponded to *BmorCPR13* (black star in
229 Fig. 4A). Another small species-specific RR-1 CP clade was observed (thick bar in Fig.
230 4A).

231 In sharp contrast to the situation with RR-1s, more than half of the RR-2 CPs formed
232 species-specific clades (Fig. 4B), indicating species-specific expansions by gene
233 duplication events. Intriguingly, all CP members of *S. litura*, *M. sexta* and *B. mori*
234 belonging to large species specific clades (green branch in Fig. 4B) were His-rich.
235 Twenty-five RR-2 CPs (Sl_Aki-100-108, Sl_Aki-110-117, Sl_Aki-122, 124, 127, 129,
236 132, 134 136) formed the biggest *S. litura* specific clade (Blue star in Fig.4B), all of
237 which belonged to the *RR-2 CP* gene cluster in chr9. BmorCPR152, MsCPR152,
238 MsCPR153 and Sl_Aki-166, which had the highest His-residue content in each species,

made a single clade (red star in Fig.4B), indicating that this highest His CP is conserved and may play a common role in some specific structure among the three Lepidoptera. Five of the CPFLs (Sl_Aki-236, 237, 238, 239, 240) formed a clade in *S. litura* (Supplementary Fig. 2).

3.5 Transcript abundance of CPs

We conducted RNA-Seq analysis to study the transcript distribution of *CP* genes in various developmental stages and tissues. In total, transcriptional evidence was obtained for 283 of the 287 annotated *CPs* (see Supplementary Table 2 for numbers of genes from each CP group expressed per library). Transcripts were found for a maximum of 233 *CP* genes in the 2nd instar larvae and a minimum of 94 *CP* genes in the epidermis of the 6th day pupa (Fig. 5A). The pattern of transcript levels of different *CP* groups (Fig. 5B) showed dynamic changes in epidermis from larval to pupal stages (Supplementary Table 3 for the total FPKM value for CP groups in each library). The transcript pattern of *RR-1* CPs contrasted extremely compared to *RR-2* CPs (Fig. 5B). Most *CP* transcripts in larval epidermis were derived from *RR-1 CP* genes whereas RR-2 transcripts predominated in pharate adults. The difference in transcript pattern between *RR-1* and *RR-2* genes was consistent with published reports that the *RR-1* transcripts are much more abundant in soft and flexible cuticles typical of larvae than in hard cuticles, whereas *RR-2* transcripts are associated with hard structures typical of

258 adults (Dittmer et al., 2015; Ettershank, 1964 ; Futahashi. et al., 2008; Vannini and
259 Willis, 2017).

260 The transcripts of *CPG* and *CPH* amount to a large percentage in pupal and pharate
261 adult stages, despite their small numbers of genes compared with the CPR family
262 (CPG 16-22; CPH 10-21; CPR 38-151) (Fig. 5A,B). In *B. mori*, transcripts from *CPG*
263 genes are also reported to be present in some hard cuticles such as pheromone gland,
264 compound eye and maxillary galea (Futahashi. et al., 2008). Most CPHs contain AAP
265 (A/V) repeats which might cause a protein to fold into a more or less regular helix,
266 leading to an elastin-like structure which is easily and reversibly deformed by external
267 forces (Andersen, 1995). Transcripts of *CPH* genes were abundant in pupal epidermis
268 and wing, especially in early pupal epidermis (Fig. 5B). Further study is needed to
269 determine the function of the CPGs and CPHs in cuticle formation.

270 Wolfgang and Riddiford (1986) reported changes of CP synthesis correlated with
271 changes of lamellar structure in *M. sexta* cuticle during the final larval instar when they
272 dig into soil in preparation for pupation. Our RNA-Seq analysis of *S. litura* suggested
273 that epidermal cuticular layers of 6th larval instar would be mainly composed of RR-1s
274 as well as CPAPs, whereas, based on transcript abundance, CPH and CPGs together
275 with CPAPs would contribute to the cuticular layers of the wandering stage (Fig. 5A;
276 Fig. 6). Especially, CPHs encoded by *Sl_Aki-260* (Supplementary Fig. 4) and *Sl_Aki-*
277 *261* (Supplementary Fig. 5) were extremely highly expressed in the wandering stage at
278 night compared to the daytime. Nevertheless, the ratios of expressed gene numbers

among *CP* gene families did not change so much in epidermis between 6th larval instar and wandering stage (Fig. 5B).

The heat map of transcripts for each *CP* gene *in S. litura* showed several characteristic patterns (Fig. 6). The transcripts from *RR-1* and *CPG* genes were continuously and highly abundant in epidermis of larval stages (Fig. 6A). However, transcripts from some *RR-1* genes (Fig. 6B) that were abundant in larval epidermis were also found in the early pupal stage (P-2) or pharate adult stage (P-12). This is similar to *An. gambiae*, where Willis (2010) reported that both RR-1 and RR-2 transcripts are present in pharate adults and post-eclosion, but many fewer are RR-1 compared with RR-2.

The secondary structure predicted by online software Phyre^2 (Kelley et al., 2015) (http://www.sbg.bio.ic.ac.uk/) for RR-1 CPs of pattern B (Fig. 6B) suggested that most of them shared a common structure homologous to the Polo-Box domain (Park et al., 2010), which comprises a six-stranded antiparallel β-sheet shielded by one α-helix. However, we could not find such a common domain in RR-1 CPs of pattern A (Fig. 6A). As Vannini and Willis (2017) reported, the location of RR-1s and RR-2s depends on properties of individual proteins in *An. gambiae*. It will be interesting to learn how RR-1s are involved in specific protein structures of adult cuticle.

The *CPH* genes (*Sl_Aki-262, 265, 266, 267*and *268*) which formed a cluster on Chr28 and a small cluster of *RR-2* genes (*Sl_Aki-186, 187, 188, 189* and *190*) in Chr15

299 (Supplementary table. 1) had transcripts mainly in late larval and early pupal stages
300 (Fig. 6C).

301 *RR-2* genes, which are the main members of "pattern D" of the heat-map (Fig. 6D),
302 were highly expressed in the pharate adult wing and epidermis (P-9 and P-12). These
303 genes were also well conserved among the three moth species with one-to-one
304 correspondence in the phylogenetic trees (Fig. 4B). *Sl_Aki-166*, encoding the highest
305 His-rich CP, had abundant transcripts in the pharate adult stage. This expression
306 pattern was similar to its *B. mori* ortholog, *BmorCPR152*, which encodes the highest
307 His-rich CP (Suetsugu et al., 2013). These *CP* genes, which were mainly expressed in
308 the pharate adult stage, may contribute to scales or other specific structures in the adult.
309 Not only *RR-2* transcripts, but also transcripts from other *CP* families such as *RR-1,*
310 *CPG, CPFL, CPAP* and *CPH*, were also observed in high abundance in pupal wing and
311 epidermis (Fig. 6C, D). The finding that two *RR-1* genes (*Sl_Aki-182* and *210*) were
312 expressed highly only in the pupal stage, but not in the larval stage (Fig. 6D), is
313 interesting since most *RR-1* genes had high transcript levels in larval epidermis. The
314 contrasting expression patterns of the genes encoding these two RR-1 proteins suggests
315 that their function merits further study.

316 Transcripts from a few *CPAP*, *RR-1* and *CPH* genes were observed in abundance in
317 epidermis throughout the development and wing of pharate adult (Fig. 6E). The
318 transcripts of three *CPAP* genes (*Sl_Aki*-20, 22 and 23) showed a steady high
319 abundance in epidermis in all stages of pupal wing. These three genes are orthologs of

CPAP3-A1, CPAP3-B and CPA3-C in both *M. sexta* and *T. castaneum*. Their biological importance is indicated by reports that RNAi knockdown of *TcCPAP3-A1* causes adult lethality, down-regulation of the *TcCPAP3-B* causes a walking defect and RNAi treatment for *TcCPAP3-C* leads to molting arrest at the pharate adult stage (Jasrapuria et al., 2012; Petkau et al., 2012).

His-rich *RR-2 genes* (*Sl_Aki-100-108, Sl_Aki-110-117, Sl_Aki-122, 124, 127, 129, 132, 134 136*), which form the large species-specific clade in the phylogenetic tree (Blue star in Fig.4B), unexpectedly showed few transcripts in epidermis or other samples (Supplementary Fig. 3). However, since the transcripts were only examined in limited tissues in this study, we estimate these species-specific *His-rich CP* genes may be expressed in some hard cuticle structures such as the cornea of the compound eye, maxilla, or antenna.

Another unexpected finding was that transcripts from *CPG, CPAP, CPT*, *CPH*, *RR-1*and *RR-2* were present in relatively high abundance at the 2nd instar (Supplementary Fig. 4), which was quite different from those of other larval stages. Other CPs that were abundant at later stages also had high levels of transcripts in the 2nd instar (Supplementary Fig. 5). Among several possibilities to explain this peculiar expression pattern across all the samples is the large/major morphological change that occurs in larvae during the transition from 1st to 2nd instar. Further study of the ultrastructure and physical properties of larval cuticle at these stages may help to explain this observation. Another likely possibility is related to the timing of sample

collection. If the sampling was performed at the pharate third instar stage, CP expression would be higher than in mid-instar larvae. It is also of interest to check the transcripts of *CP* genes in the pharate stage. Levels of other *CP* transcripts are shown in Supplementary Figure 5.

4. Conclusion

S. litura, among the most economically important global agricultural pests, is characterized by a short life cycle, direct contact with soil during both late larval and pupal/pharate adult stages, long distance migration and quick adaptation to diverse ecological niches. RNA-Seq analysis suggests that differential expression of various CP groups to produce cuticle layers with different physical properties at different stages. This study aimed to illustrate how the *CP* genes have been adapted to the characteristic life style of this pest by analyses of their genome organization, phylogenetics and transcriptomics. Similar to *M. sexta* (Dittmer et al., 2015), the RR-2 group has expanded more than other CP groups, largely by gene duplication events after speciation. Additionally, we found that 91 of 129 RR-2 CPs are His-rich, amounting to twice the number and a greater fraction than in *B. mori* (from 58% in *B. mori* to 71% in *S. litura*). These His-rich proteins are likely involved in cuticular sclerotization since a His residue, which contains a nucleophilic imidazole group, can react with ortho-quinone or dehydrobenzodioxine (Andersen, 2010, 2012). Although more evidence is needed, we speculate that the high content of His-residues and the

large expansion of *His-rich RR-2 CP* genes that occur in *S. litura* could contribute to construct tougher cuticles that could protect against fungal and bacterial attacks and abrasion throughout their larval and pupal stages. Comparison with *M. sexta* CPs also strengthened this idea, since both species which stay under soil during pupal/pharate adult stages without cocoons share a large expansion of *His-rich RR-2* CP genes compared with *B. mori.*

The present work clearly showed how this pest established its unique life cycle through the expansion of *His-rich* CP genes, which provides insights into the possible regulation of CPs for pest control and their properties as new biomaterials.

Acknowledgements

We are very grateful to Dr. Judith H. Willis for many helpful discussions and comments for this MS. We also thank a reviewer and Prof. H. Kishino (University Tokyo) for their valuable comments and discussion on phylogenetic analysis. This research was supported by a grant from the One Thousand Foreign Experts Recruitment Program of the Chinese Government (No. WO20125500074).

References

Andersen, S.O., Roepstorff, P., 1997. Sequence studies of proteins from larval and pupal cuticle of the yellow meal worm, *Tenebrio molitor*. Insect Biochem. Mol. Biol. 27, 121-131.

380 Andersen, S.O., 1998. Amino acid sequence studies on endocuticular proteins form the
381 desert locust *Schistocerca gregaria*. Insect Biochem. Mol. Biol. 28, 421.
382 Andersen, S.O., Hojrup, P., Roepstorff P., 1995. Insect Cuticular Proteins. Insect
383 Biochem. Mol. Biol. 25, 153-176.
384 Andersen, S.O., 2000. Studies on proteins in post-ecdysial nymphal cuticle of locust,
385 *Locusta migratoria*, and cockroach, *Blaberus craniifer*. Insect Biochem. Mol. Biol. 30,
386 569-577.
387 Andersen, S.O., 2007. Involvement of tyrosine residues, N-terminal amino acids, and
388 beta-alanine in insect cuticular sclerotization. Insect Biochem. Mol. Biol 37, 969-974.
389 Andersen, S.O., 2008. Quantitative determination of catecholic degradation products
390 from insect sclerotized cuticles. Insect Biochem. Mol. Biol. 38, 877-882.
391 Andersen, S.O., 2010. Insect cuticular sclerotization: a review. Insect Biochem. Mol.
392 Biol. 40, 166-178.
393 Andersen, S.O., 2012. Cuticular Sclerotization and Tanning. In: Gilbert, L.I., Iatrou, K.,
394 Gill, S. (Eds.), Comprehensive Molecular Insect Science, vol. 4. Elsevier Press, Oxford,
395 UK, pp. 145-170.
396 Andersen, S.O., Roepstorff, P., 2007. Aspects of cuticular sclerotization in the locust,
397 *Scistocerca gregaria*, and the beetle, *Tenebrio molitor*. Insect Biochem. Mol. Biol. 37,
398 223-234.
399 Baton, L.A., Robertson, A., Warr, E., Strand, M.R., Dimopoulos, G., 2009.
400 Genome-wide transcriptomic profiling of *Anopheles gambiae* hemocytes reveals

401 pathogen-specific signatures upon bacterial challenge and Plasmodium berghei
402 infection. BMC genomics. 10, 257.
403 Cheng, T., Wu, J., Wu, Y., Chilukuri, R.V., Huang, L., Yamamoto, K., Feng, L., Li,
404 W., Chen, Z., Guo, H., Liu, J., Li, S., Wang, X., Peng, L., Liu, D., Guo, Y., Fu, B., Li,
405 Z., Liu, C., Chen, Y., Tomar, A., Hilliou, F., Montagne, N., Jacquin-Joly, E.,
406 d'Alencon, E., Seth, R.K., Bhatnagar, R.K., Jouraku, A., Shiotsuki, T., Kadono-Okuda,
407 K., Promboon, A., Smagghe, G., Arunkumar, K.P., Kishino, H., Goldsmith, M.R.,
408 Feng, Q., Xia, Q., Mita, K., 2017. Genomic adaptation to polyphagy and insecticides in
409 a major East Asian noctuid pest. Nat Eco Evol. 1, 1747-1756.
410 Cornman, R.S., Willis, J.H., 2009. Annotation and analysis of low-complexity protein
411 families of *Anopheles gambiae* that are associated with cuticle. Insect Mol Biol. 18,
412 607-622.
413 Dittmer, N.T., Tetreau, G., Cao, X., Jiang, H., Wang, P., Kanost, M.R., 2015.
414 Annotation and expression analysis of cuticular proteins from the tobacco hornworm,
415 *Manduca sexta*. Insect Biochem. Mol. Biol. 62, 100-113.
416 Dong, Z., Zhang, W., Zhang, Y., Zhang, X., Zhao, P., Xia, Q., 2016. Identification and
417 Characterization of Novel Chitin-Binding Proteins from the Larval Cuticle of
418 Silkworm, *Bombyx mori*. J. Proteome Res. 15, 1435-1445.
419 Felsenstein J, 1985. Confidence limits on phylogenies: An approach using the
420 bootstrap. Evolution 39, 783-791.

421 Futahashi., R., Okamoto., S., Kawasaki., H., Zhong, Y.-S., Iwanaga., M., Mita, K.,
422 Fujiwara, H., 2008. Genome-wide identification of cuticular protein genes in the
423 silkworm, *Bombyx mori*. Insect Biochem. Mol. Biol. 38, 1138-1146.
424 Gu, S., Willis, J.H., 2003. Distribution of cuticular protein mRNAs in silk moth
425 integument and imaginal discs. Insect Biochem. Mol. Biol. 33, 1177-1188.
426 He, N., Botelho, J.M., McNall, R.J., Belozerov, V., Dunn, W.A., Mize, T., Orlando, R.,
427 Willis, J.H., 2007. Proteomic analysis of cast cuticles from *Anopheles gambiae* by
428 tandem mass spectrometry. Insect Biochem. Mol. Biol. 37, 135-146.
429 Ioannidou, Z.S., Theodoropoulou, M.C., Papandreou, N.C., Willis, J.H., Hamodrakas,
430 S.J., 2014. CutProtFam-Pred: detection and classification of putative structural
431 cuticular proteins from sequence alone, based on profile hidden Markov models. Insect
432 Biochem. Mol. Biol. 52, 51-59.
433 Kerwin, J.L., Turecek, F., Xu, R., Kramer, K.J., Hopkins, T.L., Gatlin, C.L., John
434 R.Yates, J.R.,III, 1999. Mass spectrometric analysis of catechol-histidine adducts from
435 insect cuticle. Anal Biochem. 268, 229-37.
436 Jasrapuria, S., Arakane, Y., Osman, G., Kramer, K.J., Beeman, R.W., Muthukrishnan,
437 S., 2010. Genes encoding proteins with peritrophin A-type chitin-binding domains in
438 *Tribolium castaneum* are grouped into three distinct families based on phylogeny,
439 expression and function. Insect Biochem. Mol. Biol. 40, 214-227.

Jasrapuria, S., Specht, C.A., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., 2012. Gene families of cuticular proteins analogous to peritrophins (CPAPs) in *Tribolium castaneum* have diverse functions. PloS one. 7, e49844.

Kerwin, J.L., Turecek, F., Xu, R., Kramer, K. J., Hopkins, T. L., Gatlin, C. L.,Yates, J.R.,, 1999. Mass spectrometric analysis of catechol–histidine adducts from insect cuticle. Anal Biochem. 268, 229-237.

Locke, M., 2001. The Wigglesworth Lecture: Insects for studying fundamental problems in biology. J Insect Physiol 47, 495–507.

Magkrioti, C.K., Spyropoulos, I.C., Iconomidou, V.A., Willis, J.H., Hamodrakas, S.J., 2004. cuticleDB: a relational database of Arthropod cuticular proteins. BMC bioinformatics. 5, 138.

Nei, N.S.M., 1987. The neighbor-joining method- a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-25.

Pan, P.L., Ye, Y.X., Lou, Y.H., Lu, J.B., Cheng, C., Shen, Y., Moussian, B., Zhang, C.X., 2018. A comprehensive omics analysis and functional survey of cuticular proteins in the brown planthopper. Pro. Natl. Acad. Sci. USA 115, 5175-5180.

Park, J.E., Soung, N.K., Johmura, Y., Kang, Y.H., Liao, C., Lee, K.H., Park, C.H., Nicklaus, M.C., Lee, K.S., 2010. Polo-box domain: a versatile mediator of polo-like kinase function. Cell Mol Life Sci. 67, 1957-1970.

459 Petkau, G., Wingen, C., Jussen, L.C.A., Radtke, T., Behr, M., 2012. Obstructor-A Is
460 Required for Epithelial Extracellular Matrix Dynamics, Exoskeleton Function, and
461 Tubulogenesis. J Biol Chem. 287: 21396–21405.
462 Rebers, J.E, Willis, J.H., 2001. A conserved domain in arthropod cuticular proteins
463 binds chitin. Insect Biochem. Mol. Biol. 31, 1083-1093.
464 Andersen, S.O., Rafn, K., Roepstorff, P., 1997. Sequence studies of proteins from
465 larval and pupal cuticle of the yellow meal worm, *Tenebrio molitor*. Insect Biochem.
466 Mol. Biol. 27, 121-131.
467 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and
468 post-analysis of large phylogenies. Bioinformatics. 30, 1312-1313.
469 Suetsugu, Y., Futahashi, R., Kanamori, H., Kadono-Okuda, K., Sasanuma, S.,
470 Narukawa, J., Ajimura, M., Jouraku, A., Namiki, N., Shimomura, M., Sezutsu, H.,
471 Osanai-Futahashi, M., Suzuki, M.G., Daimon, T., Shinoda, T., Taniai, K., Asaoka, K.,
472 Niwa, R., Kawaoka, S., Katsuma, S., Tamura, T., Noda, H., Kasahara, M., Sugano, S.,
473 Suzuki, Y., Fujiwara, H., Kataoka, H., Arunkumar, K.P., Tomar, A., Nagaraju, J.,
474 Goldsmith, M.R., Feng, Q., Xia, Q., Yamamoto, K., Shimada, T., Mita, K., 2013.
475 Large scale full-length cDNA sequencing reveals a unique genomic landscape in a
476 lepidopteran model insect, *Bombyx mori*. G3. 3, 1481-1492.
477 Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6:
478 Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30, 2725-2729.

Tang, L., Liang, J., Zhan, Z., Xiang, Z., He, N., 2010. Identification of the chitin-binding proteins from the larval proteins of silkworm, *Bombyx mori*. Insect Biochem. Mol. Biol. 40, 228-234.

Tetreau, G., Dittmer, N.T., Cao, X., Agrawal, S., Chen, Y.R., Muthukrishnan, S., Haobo, J., Blissard, G.W., Kanost, M.R., Wang, P., 2015. Analysis of chitin-binding proteins from *Manduca sexta* provides new insights into evolution of peritrophin A-type chitin-binding domains in insects. Insect Biochem. Mol. Biol. 62, 127-141.

Togawa, T., Augustine Dunn, W., Emmons, A.C., Willis, J.H., 2007. CPF and CPFL, two related gene families encoding cuticular proteins of *Anopheles gambiae* and other insects. Insect Biochem. Mol. Biol. 37, 675-688.

Togawa, T., Nakato, H., Izumi, S., 2004. Analysis of the chitin recognition mechanism of cuticle proteins from the soft cuticle of the silkworm, *Bombyx mori*. Insect Biochem. Mol. Biol. 34, 1059-1067.

Vannini, L., Willis, J.H., 2017. Localization of RR-1 and RR-2 cuticular proteins within the cuticle of *Anopheles gambiae*. Arthropod Struct Devl. 46, 13-29.

Willis, J.H., Papandreou, N.C., Iconomidou, V.A., Hamodrakas, S.J., 2012. Cuticular proteins. In: Gilbert, L.I. (Ed.), Insect Molecular Biology and Biochemistry. Academic Press, London, Waltham & San Diego, pp. 134e166.

Wolfgang, W.J., Riddiford, L.M., 1986. Larval cuticular morphogenesis in the tobacco hornworm: *Manduca sexta*, and its hormonal regulation. Devl. Biol. 113, 305-316.

499 Yang, Z., 1994 . Maximum likelihood phylogenetic estimation from DNA sequences
500 with variable rates over sites: approximate methods. J Mol Evol. 39, 306-314.
501

502 **Figure Captions**

503 Figure 1. Differences in behavior and life cycle between *S. litura* and *B. mori.* L1-L6
504 refer to larval instars 1-6; W is the wandering stage; numbers in parentheses represent
505 days at a particular stage.

506 Figure 2. The number of His-rich CPs in each CP group. His-rich RR-2 CPs in *S. litura*,
507 91; *B. mori*, 54; His-rich RR-1 CPs in *S. litura*, 0; *B. mori*: 2; His-rich CPGs in *S. litura*,
508 11; *B. mori* 9; His-rich CPLCA in *S. litura*, 3; *B. mori*, 1; His-rich CPAP in *S. litura*, 3;
509 *B. mori*, 1; His-rich CPH in *S. litura*, 4; *B. mori*, 2. Blue, *B. mori* CPs; red, *S. litura*
510 CPs.

511 Figure 3. Chromosomal location of the largest cluster of CPR genes in *S. litura*
512 compared with *B. mori.* RR-1 genes (red) and RR-2 genes (blue) are located primarily
513 in separate clusters on chromosome 9. Arrows indicate direction of transcription.

514 Figure 4. Phylogenetic trees of annotated RR-1 and RR-2 proteins among *S. litura, B.*
515 *mori* and *M. sexta*. (A) RR-1 protein genes. Black star, BmorCPR13; thick bar,
516 species-specific clade. (B) RR-2 protein genes. blue, *M. sexta* CPs; yellow, *B. mori*
517 CPs; magenta, *S. litura* CPs; blue star, 25 RR-2 CPs forming the biggest *S. litura*
518 specific clade; red star, highest His-content RR-2 CP among the three species, forming

519 a clade. Green branch, His-rich CP genes. Bootstrap values of 70 or higher are shown
520 in the branches.
521 Figure 5. Transcript distribution of *CP* genes in RNA-Seq libraries. (A) Total FPKM
522 values of transcripts for CP genes in each family. (B) Numbers of CP genes with
523 transcripts in each family. EP, Epidermis; D, daytime; N, nighttime; Wg, wing.
524 Figure 6. Heatmap of transcripts of *S. litura CP* genes grouped into five distinct
525 patterns. (A) *CP* transcripts mainly found in the larval epidermis. (B) *CP* transcripts
526 expressed highly in larval and pharate adult stages. (C) *CP* transcripts expressed highly
527 in epidermis of late larval to early pupal stages. (D) *CP* transcripts found mainly in the
528 pharate adult stage. (E) Transcripts of *CP* genes highly abundant through the larval to
529 pupal stage.

Figure 1



Larval (~19) Pharate Pupal(~1) Pharate Adult(2-3)

L1 L2 L3 L4 L5 L6+W Pupal(~8) Adult

S.litura

L1 L2 L3 L4 L5+W Pupal(6) Adult

B.mori

Larval (~28) Pharate Pupal(2) Pharate Adult(1-2)



Figure 2

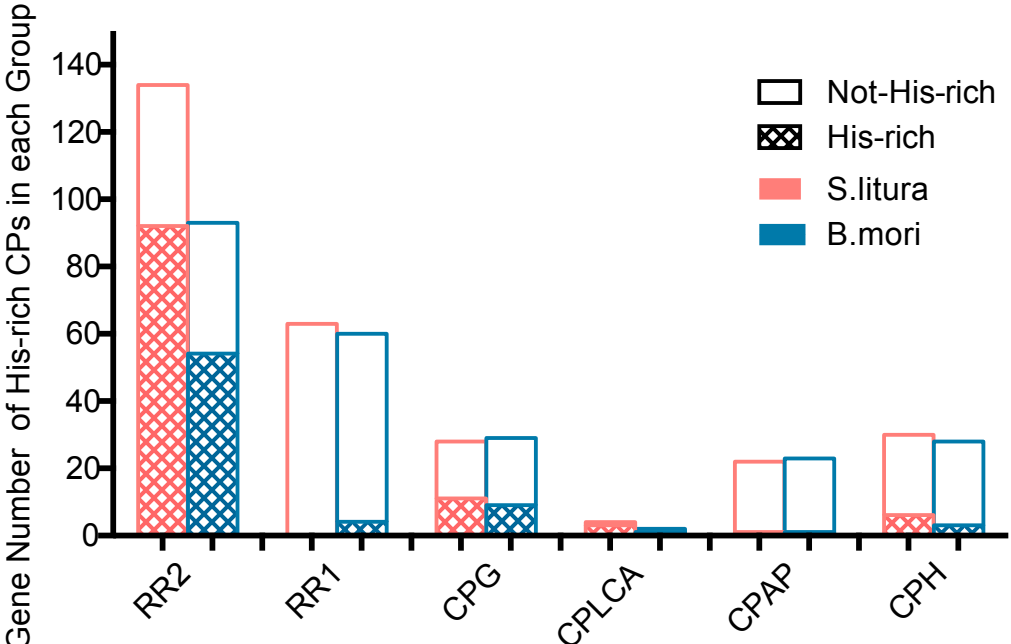

Gene Number of His-rich CPs in each Group
140
120
100
80
60
40
20
0
Not-His-rich
His-rich
S.litura
B.mori
RR2
RR1
CPG
CPLCA
CPAP
CPH

Figure 3

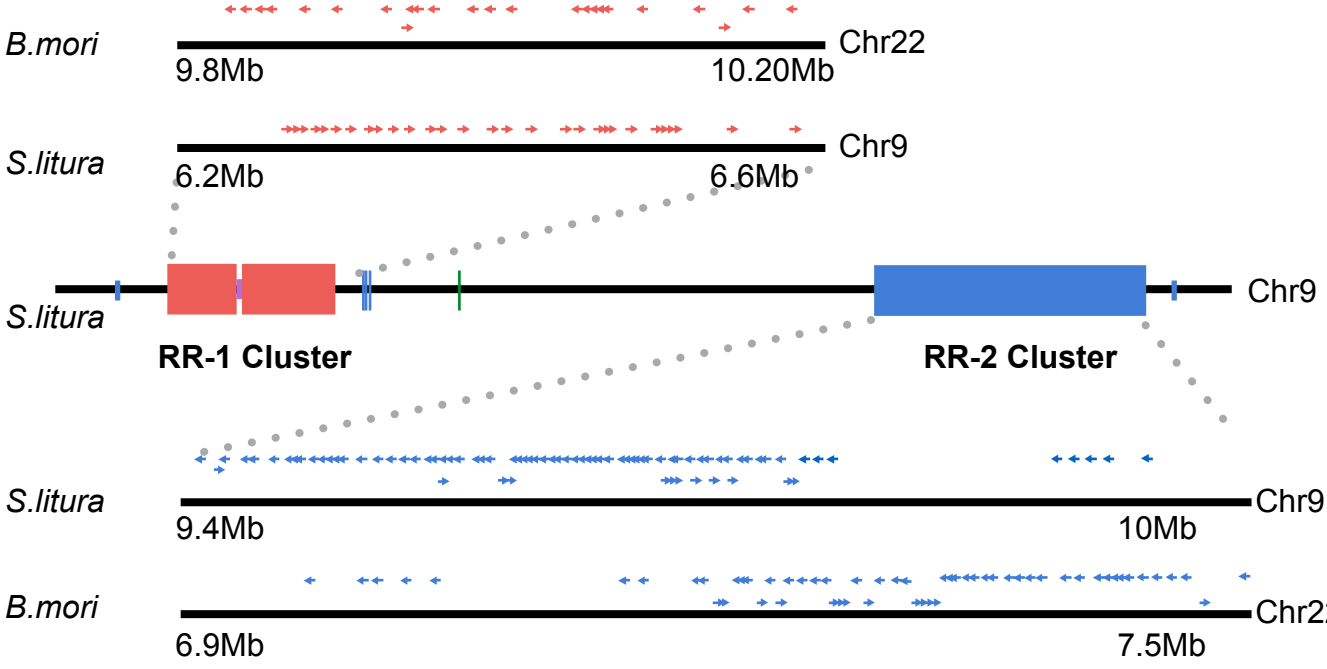

B.mori
Chr22
9.8Mb
10.20Mb
S.litura
Chr9
6.2Mb
6.6Mb
S.litura
Chr9
RR-1 Cluster
RR-2 Cluster
S.litura
Chr9
9.4Mb
10Mb
B.mori
Chr22
6.9Mb
7.5Mb

Figure 4

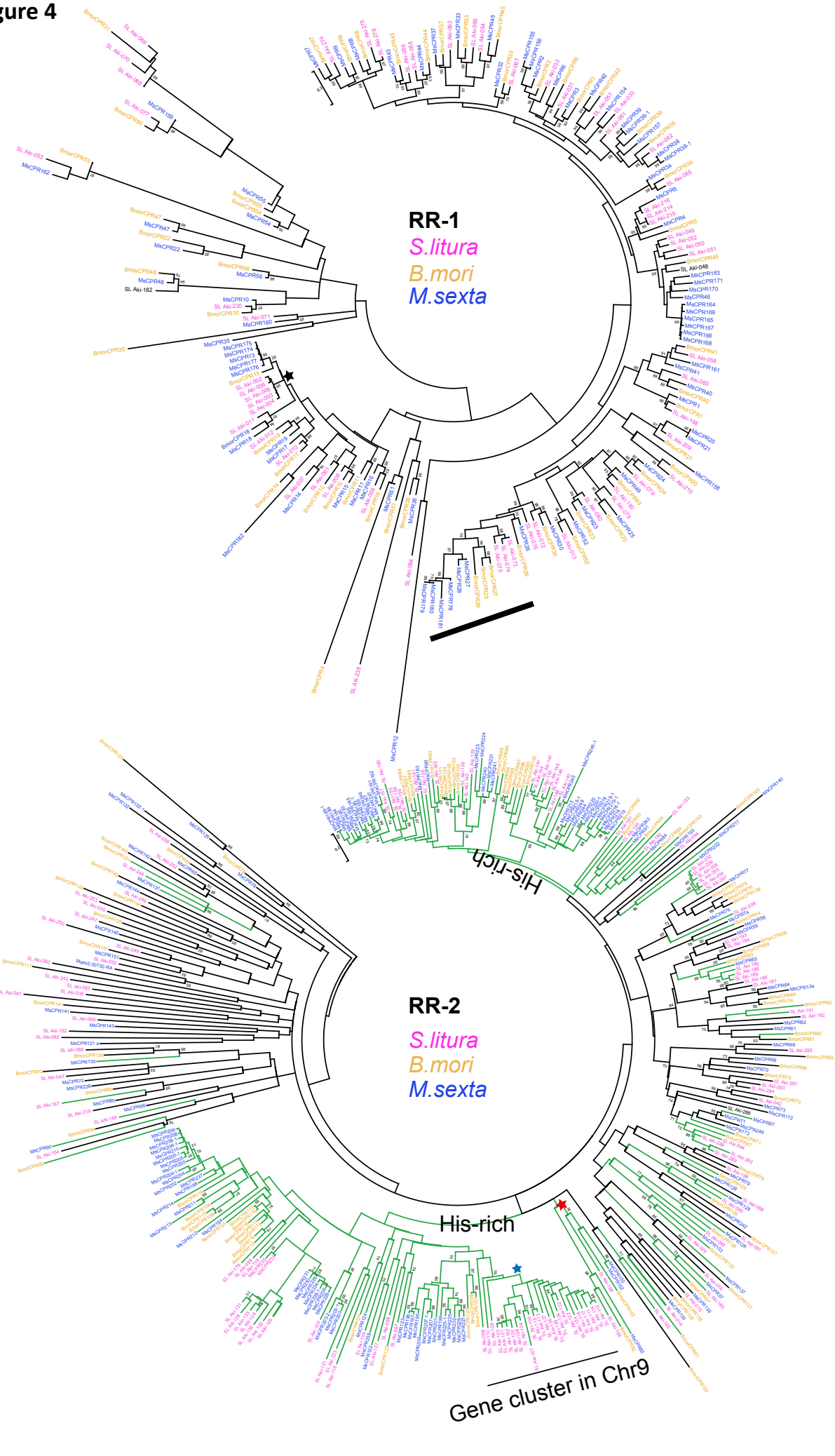

RR-1
S.litura
B.mori
M.sexta
His-rich
RR-2
S.litura
B.mori
M.sexta
His-rich
Gene cluster in Chr9

Figure 5

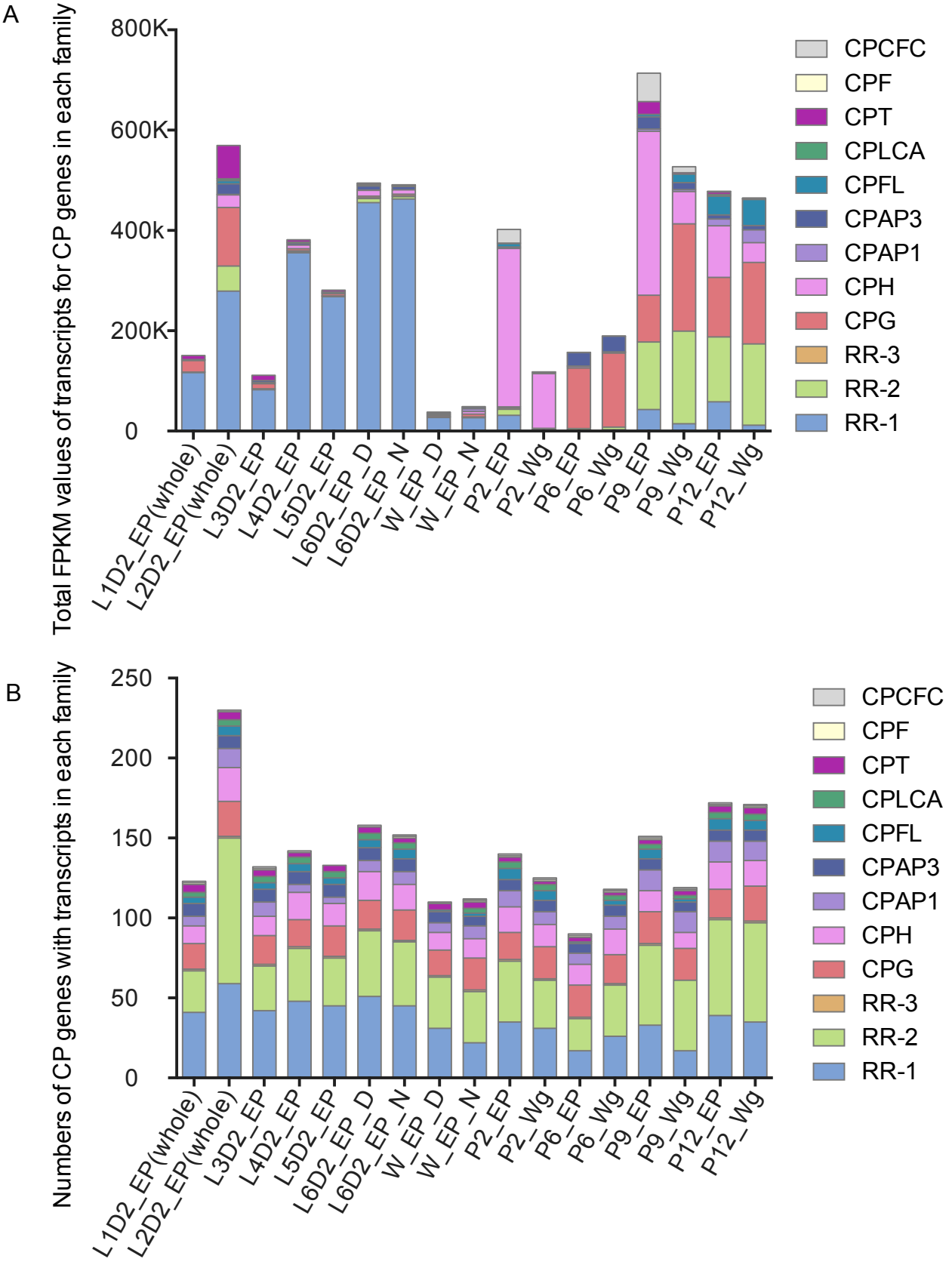

A
Total FPKM values of transcripts for CP genes in each family
800K
600K
400k
200K
0
L1D2_EP(whole)
L2D2_EP(whole)
L3D2_EP
L4D2_EP
L5D2_EP
L6D2_EP_D
L6D2_EP_N
W_EP_D
W_EP_N
P2_EP
P2_Wg
P6_EP
P6_Wg
P9_EP
P9_Wg
P12_EP
P12_Wg
CPCFC
CPF
CPT
CPLCA
CPFL
CPAP3
CPAP1
CPH
CPG
RR-3
RR-2
RR-1
B
Numbers of CP genes with transcripts in each family
250
200
150
100
50
0
L1D2_EP(whole)
L2D2_EP(whole)
L3D2_EP
L4D2_EP
L5D2_EP
L6D2_EP_D
L6D2_EP_N
W_EP_D
W_EP_N
P2_EP
P2_Wg
P6_EP
P6_Wg
P9_EP
P9_Wg
P12_EP
P12_Wg
CPCFC
CPF
CPT
CPLCA
CPFL
CPAP3
CPAP1
CPH
CPG
RR-3
RR-2
RR-1

Figure 6

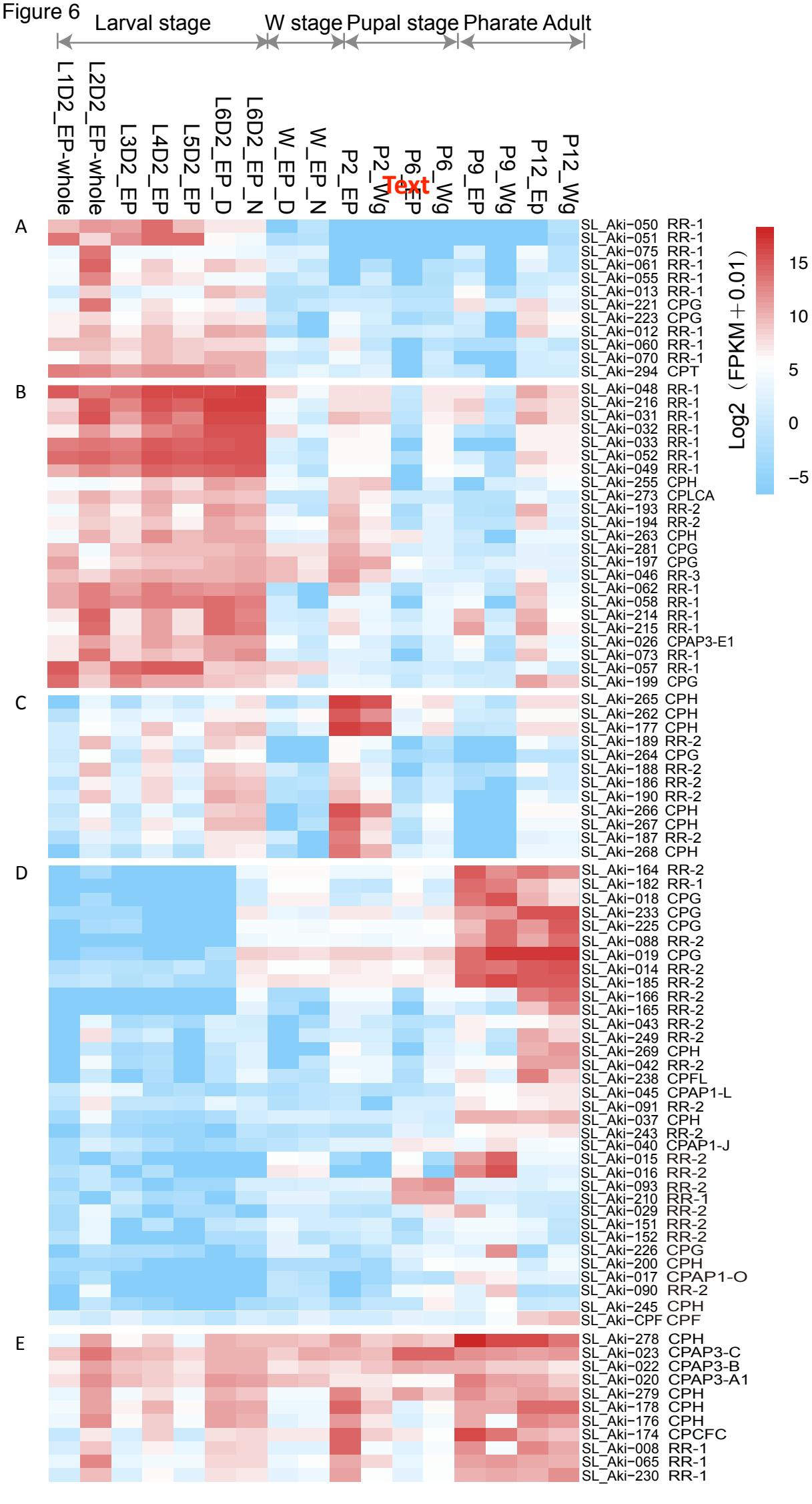

Larval stage
W stage
Pupal stage
Pharate Adult
L1D2_EP-whole
L2D2_EP-whole
L3D2_EP
L4D2_EP
L5D2_EP
L6D2_EP_D
L6D2_EP_N
W_EP_D
W_EP_N
P2_EP
P2_Wg
P6_EP
P6_Wg
P9_EP
P9_Wg
P12_Ep
P12_Wg
Text
A
B
C
D
E
SL_Aki-050 RR-1
SL_Aki-051 RR-1
SL_Aki-075 RR-1
SL_Aki-061 RR-1
SL_Aki-055 RR-1
SL_Aki-013 RR-1
SL_Aki-221 CPG
SL_Aki-223 CPG
SL_Aki-012 RR-1
SL_Aki-060 RR-1
SL_Aki-070 RR-1
SL_Aki-294 CPT
SL_Aki-048 RR-1
SL_Aki-216 RR-1
SL_Aki-031 RR-1
SL_Aki-032 RR-1
SL_Aki-033 RR-1
SL_Aki-052 RR-1
SL_Aki-049 RR-1
SL_Aki-255 CPH
SL_Aki-273 CPLCA
SL_Aki-193 RR-2
SL_Aki-194 RR-2
SL_Aki-263 CPH
SL_Aki-281 CPG
SL_Aki-197 CPG
SL_Aki-046 RR-3
SL_Aki-062 RR-1
SL_Aki-058 RR-1
SL_Aki-214 RR-1
SL_Aki-215 RR-1
SL_Aki-026 CPAP3-E1
SL_Aki-073 RR-1
SL_Aki-057 RR-1
SL_Aki-199 CPG
SL_Aki-265 CPH
SL_Aki-262 CPH
SL_Aki-177 CPH
SL_Aki-189 RR-2
SL_Aki-264 CPG
SL_Aki-188 RR-2
SL_Aki-186 RR-2
SL_Aki-190 RR-2
SL_Aki-266 CPH
SL_Aki-267 CPH
SL_Aki-187 RR-2
SL_Aki-268 CPH
SL_Aki-164 RR-2
SL_Aki-182 RR-1
SL_Aki-018 CPG
SL_Aki-233 CPG
SL_Aki-225 CPG
SL_Aki-088 RR-2
SL_Aki-019 CPG
SL_Aki-014 RR-2
SL_Aki-185 RR-2
SL_Aki-166 RR-2
SL_Aki-165 RR-2
SL_Aki-043 RR-2
SL_Aki-249 RR-2
SL_Aki-269 CPH
SL_Aki-042 RR-2
SL_Aki-238 CPFL
SL_Aki-045 CPAP1-L
SL_Aki-091 RR-2
SL_Aki-037 CPH
SL_Aki-243 RR-2
SL_Aki-040 CPAP1-J
SL_Aki-015 RR-2
SL_Aki-016 RR-2
SL_Aki-093 RR-2
SL_Aki-210 RR-1
SL_Aki-029 RR-2
SL_Aki-151 RR-2
SL_Aki-152 RR-2
SL_Aki-226 CPG
SL_Aki-200 CPH
SL_Aki-017 CPAP1-O
SL_Aki-090 RR-2
SL_Aki-245 CPH
SL_Aki-CPF CPF
SL_Aki-278 CPH
SL_Aki-023 CPAP3-C
SL_Aki-022 CPAP3-B
SL_Aki-020 CPAP3-A1
SL_Aki-279 CPH
SL_Aki-178 CPH
SL_Aki-176 CPH
SL_Aki-174 CPCFC
SL_Aki-008 RR-1
SL_Aki-065 RR-1
SL_Aki-230 RR-1
Log2（FPKM+0.01）
15
10
5
0
−5

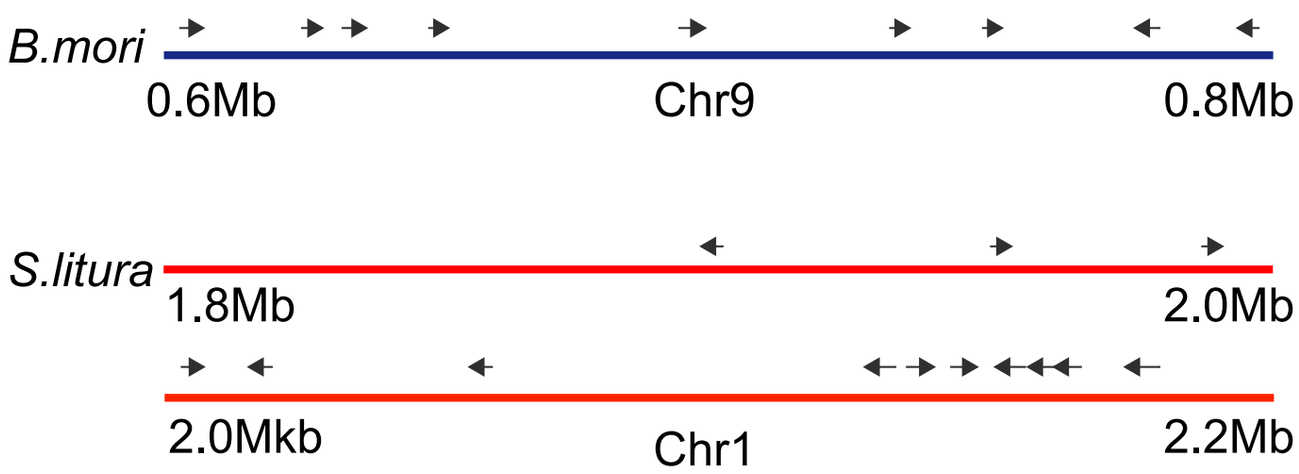
Table1. Size of each cuticular protein family in Spodoptera litura and Bombyx mori		
Motif	***S.litura***	***B.mori***
RR-1	63	56
RR-2	129	93(4)[a]
RR-3	1	3
Tweedle	5	4
CPF	1	1
CPFL	7	4
CPLCA	4	2(BmorCPH6,7)[b]
CPCFC	1	1(BmorCPH1)[c]
CPAP1	13	14
CPAP3	9	9
Glycine-Rich	28	29
CPH	26	31
Total	287	247
[a, b, c] items in parentheses refer to CPs reported in Futahashi et al. (2008)		
[a] 4 more cuticular RR-2 genes were identified in S. litura than B. mori		

Supplementary Table 1. Classification of CP coding genes.

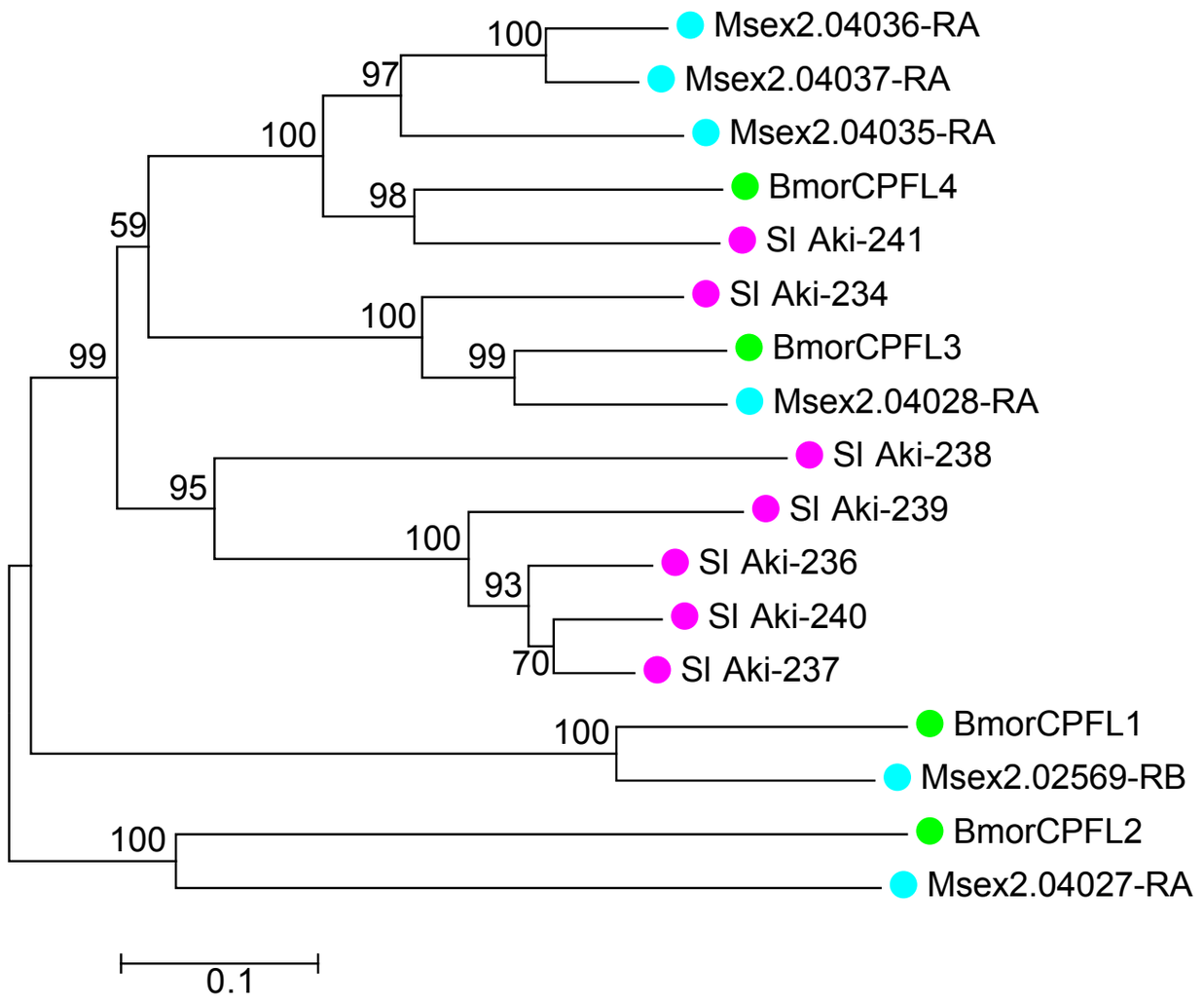
Supplementary Table 2. Numbers of genes from each CP group expressed per library																	
	L1D2_EP(whole)	L2D2_EP(whole)	L3D2_EP	L4D2_EP	L5D2_EP	L6D2_EP_D	L6D2_EP_N	W_EP_D	W_EP_N	P2_EP	P2_Wg	P6_EP	P6_Wg	P9_EP	P9_Wg	P12_EP	P12_Wg
RR-1(63)	41	59	42	48	45	51	45	31	22	35	31	17	26	33	17	39	35
RR-2(129)	26	91	28	33	30	41	40	32	32	38	30	20	32	50	44	60	62
RR-3(1)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
CPG(28)	16	22	18	17	19	18	19	16	20	17	20	20	18	20	20	18	22
CPH(26)	11	21	12	17	14	18	16	11	12	16	14	13	16	13	10	17	16
CPAP1(13)	6	12	9	5	4	7	8	6	8	10	8	7	8	13	13	13	12
CPAP3(9)	8	8	8	8	8	8	8	7	6	7	7	6	7	7	6	7	7
CPFL(7)	4	6	4	5	4	5	6	1	2	7	6	1	3	6	2	7	6
CPLCA(4)	3	4	4	4	4	4	4	0	3	4	4	0	3	3	2	4	4
CPT(5)	5	5	4	3	4	4	3	4	4	3	2	3	2	3	3	4	4
CPF(1)	1	0	1	0	0	0	1	0	1	1	1	1	1	1	1	1	1
CPCFC(1)	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
Total(287)	123	230	132	142	133	158	152	110	112	140	125	90	118	151	119	172	171

Supplementary Table 3. FPKM values of CP groups expressed per library

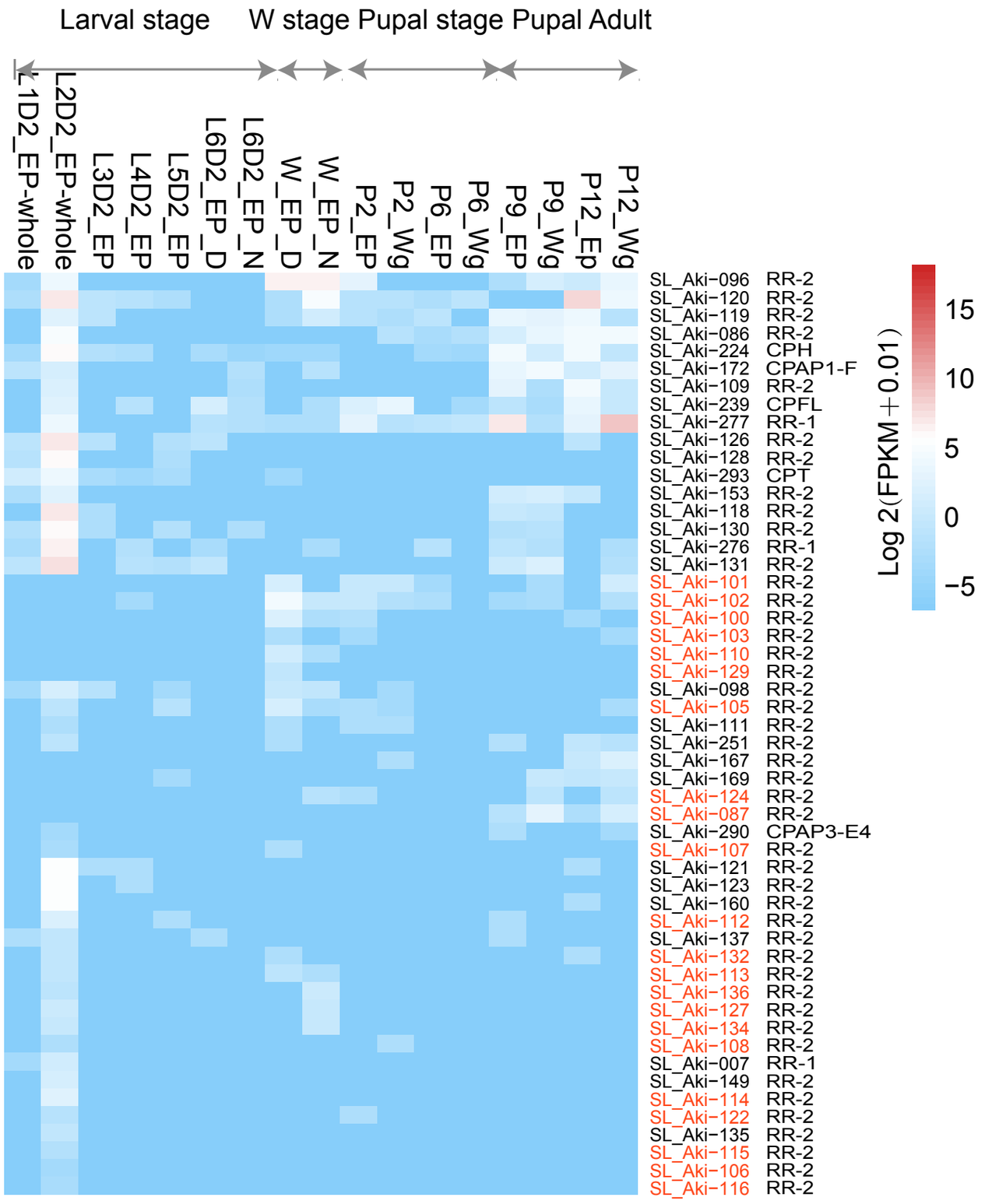
	L1D2_EP(whole)	L2D2_EP(whole)	L3D2_EP	L4D2_EP	L5D2_EP	L6D2_EP_D	L6D2_EP_N	W_EP_D	W_EP_N	P2_EP	P2_Wg	P6_EP	P6_Wg	P9_EP	P9_Wg	P12_EP	P12_Wg
RR-1	116705	278947	82667	355390	268002	455408	462561	26954	26775	31679	1660	1890	2669	42854	14808	58283	11567
RR-2	348	49969	569	2478	377	8517	5084	992	999	11732	1051	3025	5340	134920	184337	129479	162394
RR-3	642	303	1087	1332	949	1245	1443	635	207	2541	293	10	4	4	0	85	7
CPG	22935	116540	10206	4550	4322	3499	3163	3068	5817	2265	2335	120649	147493	93050	214203	118529	162159
CPH	149	25029	647	7192	1269	11518	8751	799	5365	315771	109135	3238	1526	327068	64099	102801	39777
CPAP1	203	882	1494	192	135	221	226	2348	5271	339	147	646	746	4272	4208	13776	25131
CPAP3	1271	21586	3937	3601	1572	8445	6543	2515	2740	2654	2432	27230	31503	24499	13611	81304	8913
CPFL	17	6231	29	181	18	1072	793	39	47	6635	653	35	143	4787	16926	37249	51676
CPLCA	129	3625	325	1900	713	2035	1356	0	18	1334	161	0	12	42	15	3628	233
CPT	8404	66386	10511	4451	3861	2222	934	65	1615	16	16	82	21	25446	2762	4573	1563
CPF	4	0	10	0	0	0	3	0	8	2	2	8	5	4	29	421	856
CPCFC	0	145	3	382	0	177	226	136	117	27017	327	144	40	56928	11963	1009	456
Total	150807	569643	111485	381649	281218	494359	491083	37551	48979	401985	118212	156957	189502	713874	526961	551137	464732



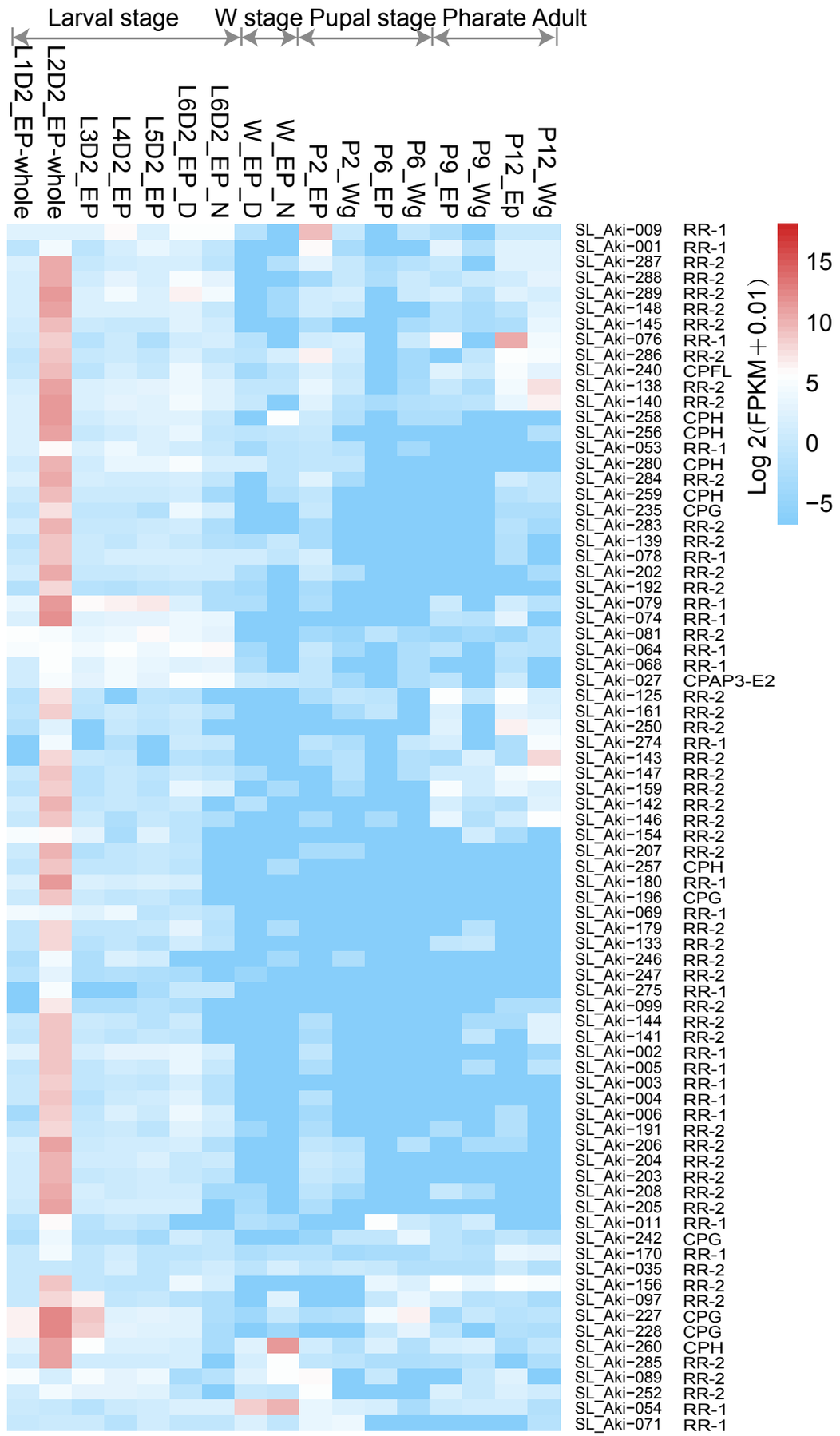
Supplementary Figure1. Expansion of *RR-1 CP* gene cluster on Chr1 in *S. litura*.The orthologous RR-1CPs are clustered in Chr9 in *B. mori.*



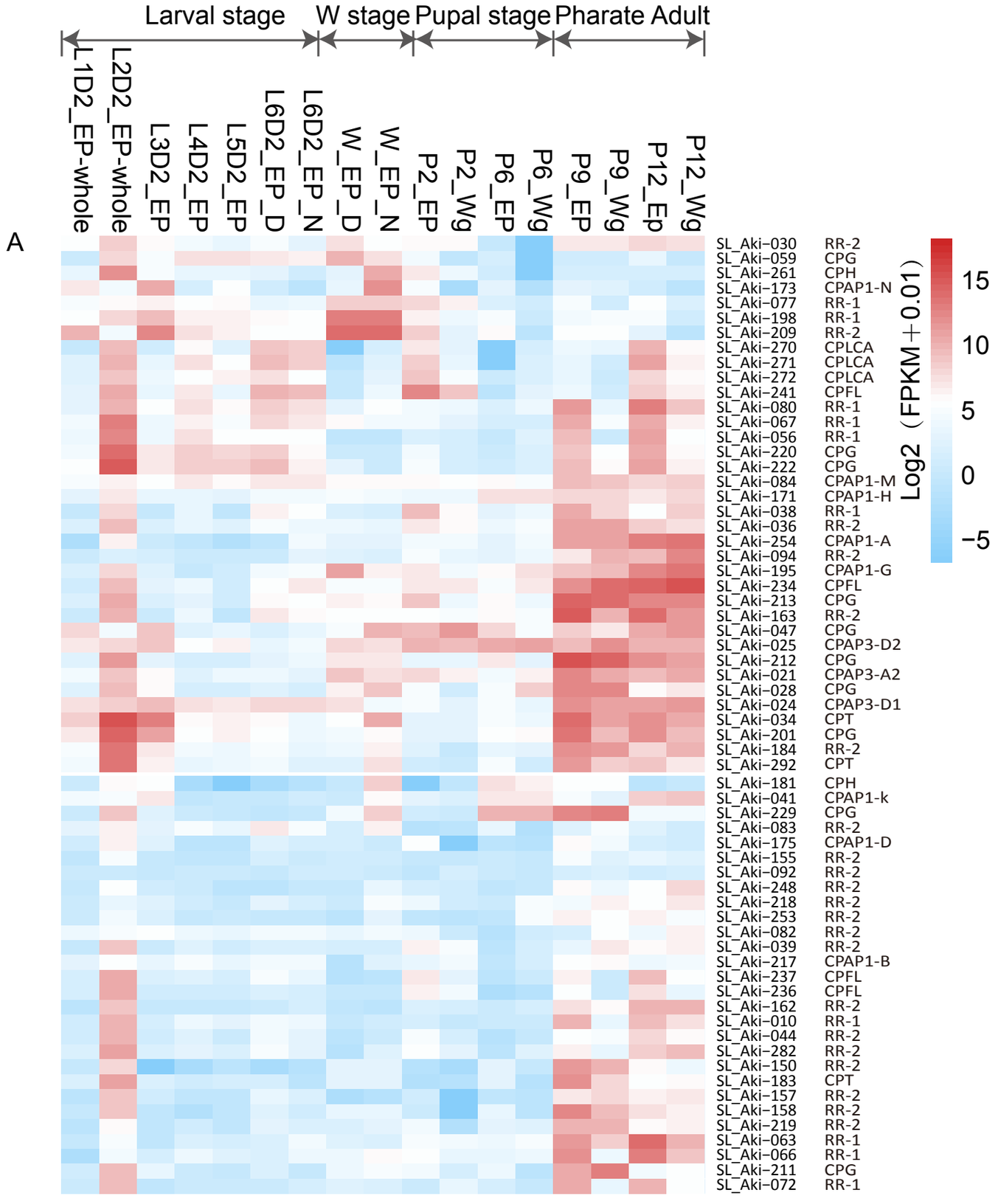
Supplementary Figure 2. The phylogenetic tree of CPFL CPs of *S. litura*, *M. sexta* and *B. mori*. Five CPFL CPs formed a species specific clade in *S. litura*.



Supplementary Figure 3. *CPs* showing low expresson in most RNA-seq libraries. Gene names colored red belong to the largest species-specific clade in the phylogenetic tree.



Supplementary Figure 4. *S. litura* CP transcripts heatmap. *S. litura CPs* were highly abundant in the 2nd instar larval epidermis exclusively.



Supplementary Figure 5. Heatmap of transcripts of all the other *CP* genes of *S. litura* not shown in Fig. 6, Supplementary Fig. S3 and S4.