Insect Biochemistry and Molecular Biology 110 (2019) 90–97

Contents lists available at ScienceDirect

Insect Biochemistry and Molecular Biology

journal homepage: www.elsevier.com/locate/ibmb

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

Jianqiu Liu[a,b], Shenglong Li[a,b], Wanshun Li[a,b], Li Peng[a], Zhiwei Chen[a,b], Yingdan Xiao[a,b], Huizhen Guo[a,b], Jiwei Zhang[a,b], Tingcai Cheng[a,b], Marian R. Goldsmith[a,b,c], Kallare P. Arunkumar[a,b,d], Qingyou Xia[a,b], Kazuei Mita[a,b,*]

[a] State Key Laboratory of Silkworm Genome Biology, Biological Science Research Center, Southwest University, Chongqing, 400716, China
[b] Chongqing Key Laboratory of Sericultural Science, Chongqing Engineering and Technology Research Center for Novel Silk Materials, Southwest University, Chongqing, 400716, China
[c] University of Rhode Island, Kingston, 02881, USA
[d] Central Muga Eri Research and Training Institute, (CMER&TI), Central Silk Board, Lahdoigarh, Jorhat, 785700, India

ARTICLE INFO

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

* Corresponding author. State Key Laboratory of Silkworm Genome Biology, Biological Science Research Center, Southwest University, Chongqing, 400715, China.
E-mail address: mitakazuei@gmail.com (K. Mita).

https://doi.org/10.1016/j.ibmb.2019.04.012
Received 27 December 2018; Received in revised form 11 April 2019; Accepted 13 April 2019
Available online 19 April 2019

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., 2004, 2007). Three distinct forms of this consensus have been classified as RR-1, RR-2 and RR-3 (Andersen, 1998, 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.

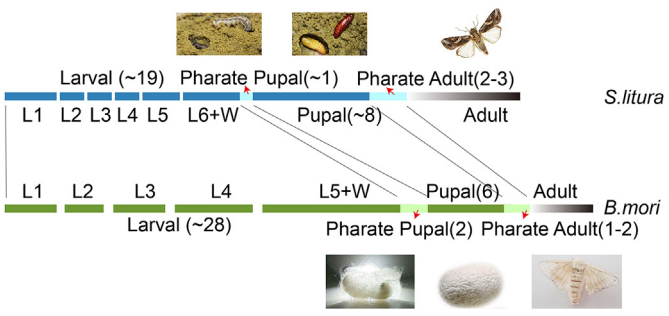

Fig. 1. Differences in behavior and life cycle between *S. litura* and *B. mori.* L1-L6 refer to larval instars 1–6; W is the wandering stage; numbers in parentheses represent days at a particular stage.

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 12 h light/12 h 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

Table 1
Size of each cuticular protein family in S.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 gene were identified in S. litura than *B. mori*.

annotated in the *S. litura* genome (Table 1) (Tetreau et al., 2015).

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

CPF has a conserved region with 44 amino acids (Togawa et al., 2007). One putative *CP* gene belonging to the CPF family in the *S. litura* genome (SWUSl0111680) was identified which we named *Sl_Aki-CPF*. Seven *CPFL* genes, which have a conserved C-terminal region similar to CPF (Togawa et al., 2007), were found. All of the *CPFL* genes together with two other *CPG* genes (*Sl_Aki-235* and *242*) formed a cluster on Chr25. This genomic structure was similar to *B. mori*, in which three CPFL genes (*BmorCPFL2*, *BmorCPFL3* and *BmorCPFL4*) and two CPG genes (*BmorCPG23* and *BmorCPG42*) form a gene cluster on Chr23. In addition, *Sl_Aki-235* and *Sl_Aki-242* were orthologous to *BmorCPG23* and *BmorCPG42,* respectively.

We identified five genes with a Tweedle motif, among which *Sl_Aki-292, 293* and *294* formed a cluster on ChrZ; the other two *CPT* genes were not linked to other *CPs*.

The CPG family contains GG repeats but does not have any definite motif described in other CP families (Futahashi et al., 2008; Willis et al., 2012). However, although some glycine-rich cuticular proteins also had an R&R Consensus or Tweedle motif (i.e., *Sl_Aki-008* (RR-1), *Sl_Aki-185* (RR-2), and *Sl_Aki-34* (CPT)), we left them in their well defined families. In *S. litura*, 28 putative *CP* genes were classified as CPGs.

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

3.2. Characterization of histidine–rich CPs

The amino acid composition of the 287 CPs revealed 112 His-rich CPs in which histidine residues ranged from 6.00 to 30.14% (Fig. 2). Interestingly, most His-rich CPs belonged to the RR-2 family, which is

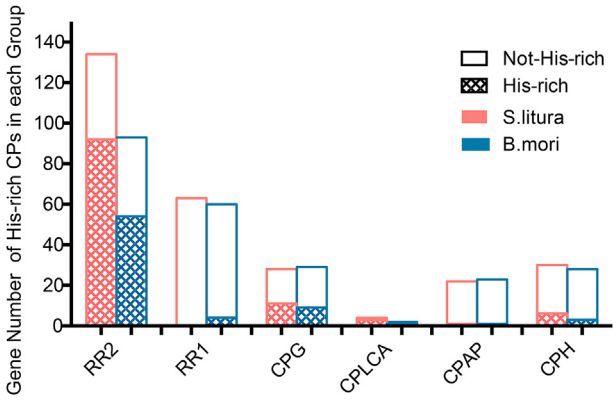

Fig. 2. The number of His-rich CPs in each CP group. His-rich RR-2 CPs in *S. litura*, 91; *B. mori*, 54; His-rich RR-1 CPs in *S. litura*, 0; *B. mori*: 2; His-rich CPGs in *S. litura*, 11; *B. mori* 9; His-rich CPLCA in *S. litura*, 3; *B. mori*, 1; His-rich CPAP in *S. litura*, 3; *B. mori*, 1; His-rich CPH in *S. litura*, 4; *B. mori*, 2. Blue, *B. mori* CPs; red, *S. litura* CPs. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

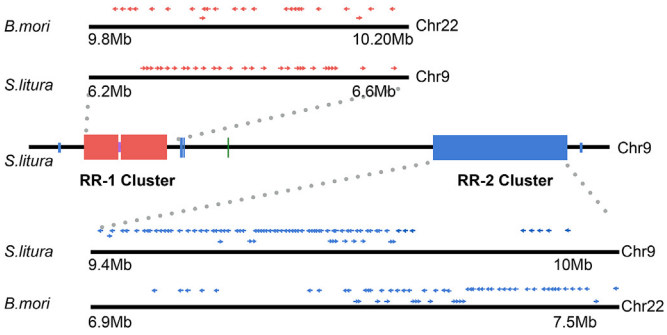

Fig. 3. Chromosomal location of the largest cluster of CPR genes in *S. litura* compared with *B. mori*. RR-1 genes (red) and RR-2 genes (blue) are located primarily in separate clusters on chromosome 9. Arrows indicate direction of transcription. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

common to both *B. mori* and *S. litura*. *B. mori* has one extremely high His-rich CP, BmorCPR152, of 45%, while the orthologous CP with the highest His content in *S. litura* is Sl_Aki-166 with 30.14%. Fig. 3 shows a large expansion of His-rich *CPs* of the RR-2 family in the *S. litura* genome. As Andersen and Roepstorff (2007) stated, histidine and several other amino acid residues in CPs could be involved in cuticular adduct formation based on biochemical analysis of the cuticular hydrolysates from different cuticular types of *M. sexta*, desert locust *Schistocerca gregaria,* and yellow mealworm, *Tenebrio molitor* (Andersen, 2007; Andersen and Roepstorff, 2007; James L.Kerwin et al., 1999Andersen et al., 1997). EM immunolocalization utilized by Vannini and Willis (2017) support the hypothesis about the deployment of RR-1 and RR-2 localization first proposed by Andersen (1998). Namely, the RR-1 CPs are mostly found in soft cuticle like inter-segmental membranes, whereas RR-2s are restricted to hard cuticles in *A. gambiae* (Vannini and Willis, 2017). Histidino-β-dopamine is the dominating adduct in hard cuticles like those of adult beetle cuticle, lepidopteran pupae and dipteran puparia (Andersen, 2008). Consistent with these reports is that His-rich CPs were mostly found in RR-2, but not RR-1 CPs of *S. litura*.

3.3. Major clusters of CPR genes in S. litura

Compared to *B. mori* which has a cluster of CPR genes on chromosome 22, major expansions of *S. litura CPR* genes derived from this family were on *S. litura* Chr9. Fig. 3 shows that 34 RR-1 (*Sl-Aki-48 - 80*)

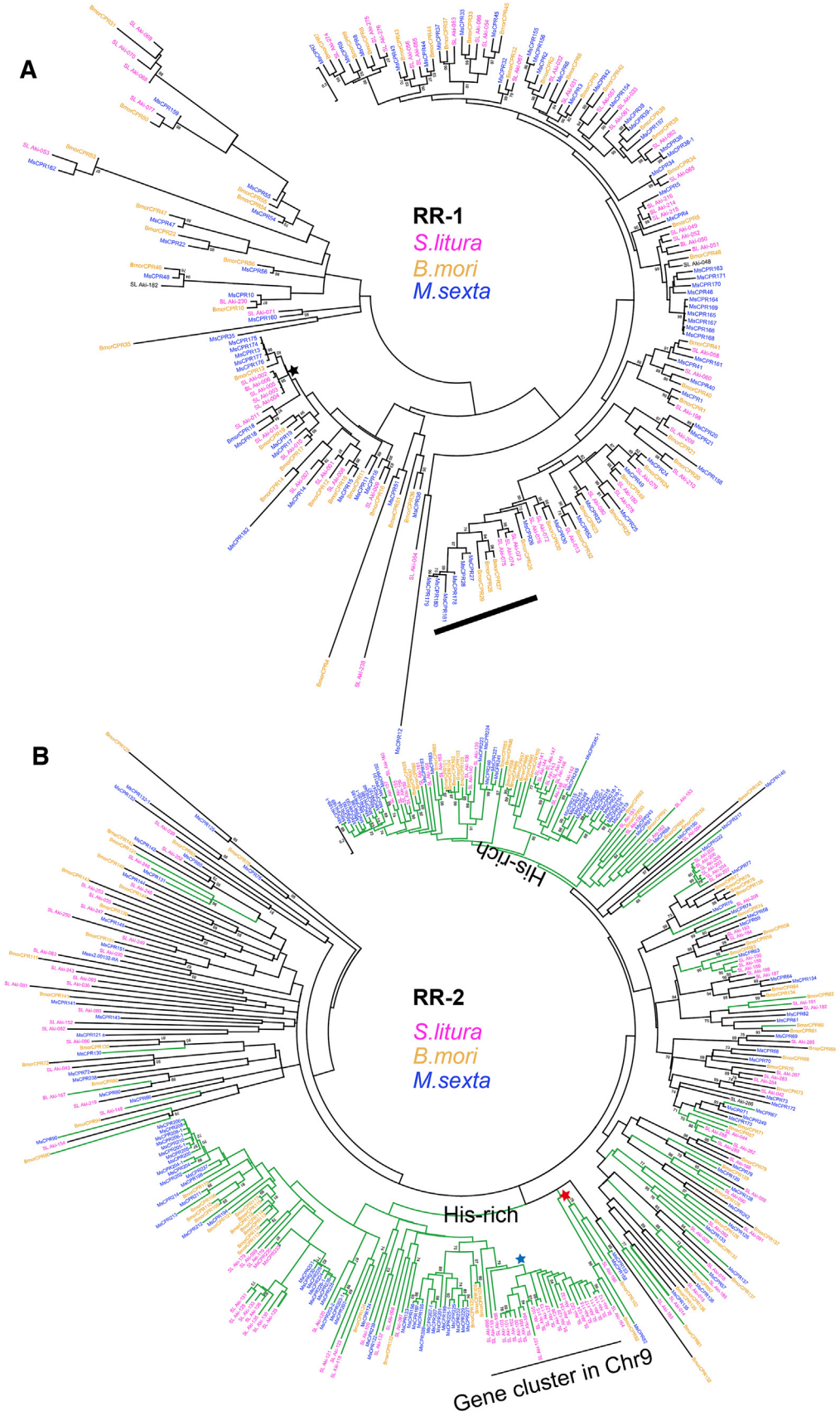

Fig. 4. Phylogenetic trees of annotated RR-1 and RR-2 proteins among *S. litura, B. mori* and *M. sexta*. (A) RR-1 protein genes. Black star, BmorCPR13; thick bar, species-specific clade. (B) RR-2 protein genes. blue, *M. sexta* CPs; yellow, *B. mori* CPs; magenta, *S. litura* CPs; blue star, 25 RR-2 CPs forming the biggest *S. litura* specific clade; red star, highest His-content RR-2 CP among the three species, forming a clade. Green branch, His-rich CP genes. Bootstrap values of 70 or higher are shown in the branches. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

genes and 82 RR-2 (*Sl-Aki-86, 88–94, 96–169*) genes were present as two large clusters located on different regions of Chr9. Intriguingly, all of the RR-2 *CP* members belonged to the cluster encoding His-rich CPs, whereas none of the 34 RR-1 *CPs* in the large cluster on Chr9 were His-rich. Chr22 in *B. mori* also has the orthologs of these RR-1 genes, but none of them were found to be His-rich (Fig. 3). Although much smaller than *S. litura*, there was a separate His-rich RR-2 *CP* cluster (*BmorCPR79-129*) on Chr22 in *B. mori.* Thirteen RR-1 (*Sl-Aki-001 - 013*) genes were also localized on *S. litura* Chr1 as a cluster (Supplementary Fig. 1). Their orthologs in *B. mori* also formed a cluster on Chr9.

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

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

In sharp contrast to the situation with RR-1s, more than half of the RR-2 CPs formed species-specific clades (Fig. 4B), indicating species-specific expansions by gene duplication events. Intriguingly, all CP members of *S. litura*, *M. sexta* and *B. mori* belonging to large species specific clades (green branch in Fig. 4B) were His-rich. Twenty-five RR-2 CPs (Sl_Aki-100-108, Sl_Aki-110-117, Sl_Aki-122, 124, 127, 129, 132, 134 136) formed the biggest *S. litura* specific clade (Blue star in Fig. 4B), all of which belonged to the *RR-2 CP* gene cluster in chr9. BmorCPR152, MsCPR152, 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 adults (Dittmer et al., 2015; Futahashi et al., 2008; Vannini and Willis, 2017).

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

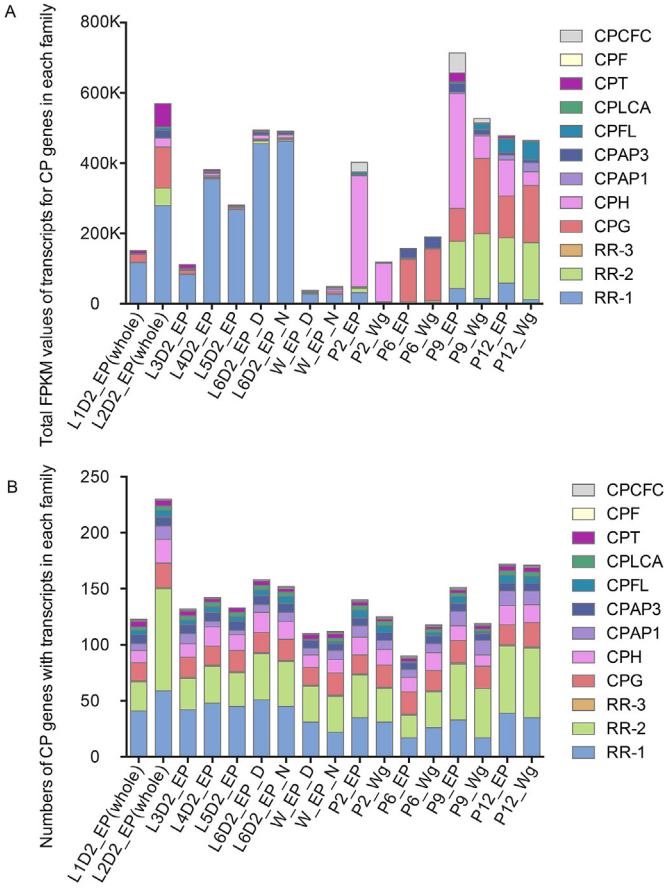

Fig. 5. Transcript distribution of *CP* genes in RNA-Seq libraries. (A) Total FPKM values of transcripts for CP genes in each family. (B) Numbers of CP genes with transcripts in each family. EP, Epidermis; D, daytime; N, nighttime; Wg, wing.

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

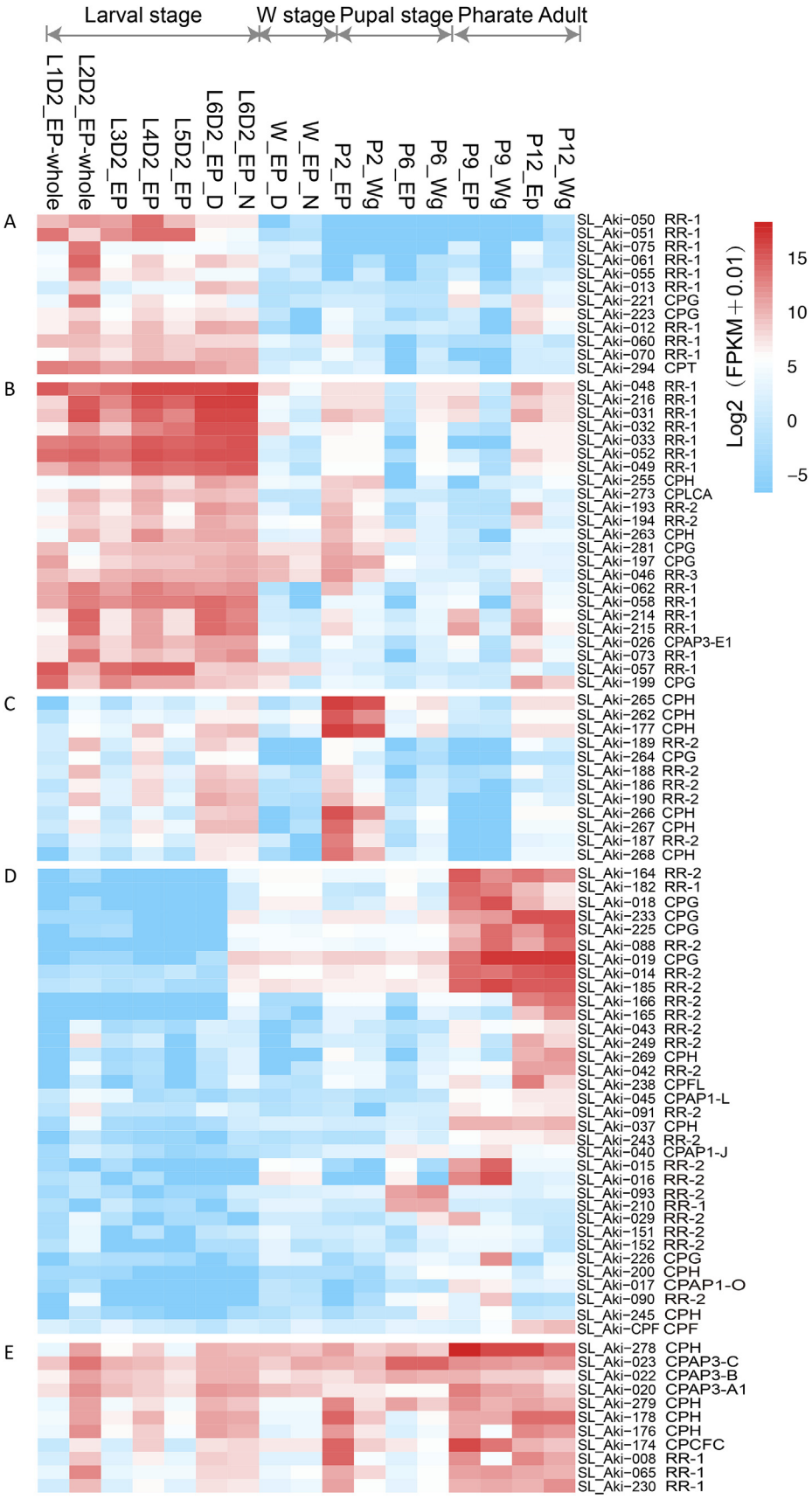

Fig. 6. Heatmap of transcripts of *S. litura CP* genes grouped into five distinct patterns. (A) *CP* transcripts mainly found in the larval epidermis. (B) *CP* transcripts expressed highly in larval and pharate adult stages. (C) *CP* transcripts expressed highly in epidermis of late larval to early pupal stages. (D) *CP* transcripts found mainly in the pharate adult stage. (E) Transcripts of *CP* genes highly abundant through the larval to pupal stage.

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

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

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ibmb.2019.04.012.

References

Andersen, S.O., 1998. Amino acid sequence studies on endocuticular proteins form the desert locust *Schistocerca gregaria*. Insect Biochem. Mol. Biol. 28, 421.
Andersen, S.O., 2000. Studies on proteins in post-ecdysial nymphal cuticle of locust, *Locusta migratoria*, and cockroach, *Blaberus craniifer*. Insect Biochem. Mol. Biol. 30, 569–577.
Andersen, S.O., 2007. Involvement of tyrosine residues, N-terminal amino acids, and beta-alanine in insect cuticular sclerotization. Insect Biochem. Mol. Biol. 37, 969–974.
Andersen, S.O., 2008. Quantitative determination of catecholic degradation products from insect sclerotized cuticles. Insect Biochem. Mol. Biol. 38, 877–882.
Andersen, S.O., 2010. Insect cuticular sclerotization: a review. Insect Biochem. Mol. Biol. 40, 166–178.
Andersen, S.O., 2012. Cuticular sclerotization and tanning. In: Gilbert, L.I., Iatrou, K., Gill, S. (Eds.), Comprehensive Molecular Insect Science. vol. 4. Elsevier Press, Oxford, UK, pp. 145–170.
Andersen, S.O., Roepstorff, P., 2007. Aspects of cuticular sclerotization in the locust, *Scistocerca gregaria*, and the beetle, *Tenebrio molitor*. Insect Biochem. Mol. Biol. 37, 223–234.
Andersen, S.O., Hojrup, P., Roepstorff, P., 1995. Insect cuticular proteins. Insect Biochem. Mol. Biol. 25, 153–176.
Andersen, S.O., Rafn, K., 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.
Baton, L.A., Robertson, A., Warr, E., Strand, M.R., Dimopoulos, G., 2009. Genome-wide transcriptomic profiling of *Anopheles gambiae* hemocytes reveals pathogen-specific signatures upon bacterial challenge and Plasmodium berghei infection. BMC Genomics 10, 257.

Cheng, T., Wu, J., Wu, Y., Chilukuri, R.V., Huang, L., Yamamoto, K., Feng, L., Li, W., Chen, Z., Guo, H., Liu, J., Li, S., Wang, X., Peng, L., Liu, D., Guo, Y., Fu, B., Li, Z., Liu, C., Chen, Y., Tomar, A., Hilliou, F., Montagne, N., Jacquin-Joly, E., d'Alencon, E., Seth, R.K., Bhatnagar, R.K., Jouraku, A., Shiotsuki, T., Kadono-Okuda, K., Promboon, A., Smagghe, G., Arunkumar, K.P., Kishino, H., Goldsmith, M.R., Feng, Q., Xia, Q., Mita, K., 2017. Genomic adaptation to polyphagy and insecticides in a major East Asian noctuid pest. Nat Eco Evol 1, 1747–1756.

Cornman, R.S., Willis, J.H., 2009. Annotation and analysis of low-complexity protein families of *Anopheles gambiae* that are associated with cuticle. Insect Mol. Biol. 18, 607–622.

Dittmer, N.T., Tetreau, G., Cao, X., Jiang, H., Wang, P., Kanost, M.R., 2015. Annotation and expression analysis of cuticular proteins from the tobacco hornworm, *Manduca sexta*. Insect Biochem. Mol. Biol. 62, 100–113.

Dong, Z., Zhang, W., Zhang, Y., Zhang, X., Zhao, P., Xia, Q., 2016. Identification and characterization of novel chitin-binding proteins from the larval cuticle of silkworm, *Bombyx mori*. J. Proteome Res. 15, 1435–1445.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Futahashi, R., Okamoto, S., Kawasaki, H., Zhong, Y.-S., Iwanaga, M., Mita, K., Fujiwara, H., 2008. Genome-wide identification of cuticular protein genes in the silkworm, *Bombyx mori*. Insect Biochem. Mol. Biol. 38, 1138–1146.

Gu, S., Willis, J.H., 2003. Distribution of cuticular protein mRNAs in silk moth integument and imaginal discs. Insect Biochem. Mol. Biol. 33, 1177–1188.

He, N., Botelho, J.M., McNall, R.J., Belozerov, V., Dunn, W.A., Mize, T., Orlando, R., Willis, J.H., 2007. Proteomic analysis of cast cuticles from *Anopheles gambiae* by tandem mass spectrometry. Insect Biochem. Mol. Biol. 37, 135–146.

Ioannidou, Z.S., Theodoropoulou, M.C., Papandreou, N.C., Willis, J.H., Hamodrakas, S.J., 2014. CutProtFam-Pred: detection and classification of putative structural cuticular proteins from sequence alone, based on profile hidden Markov models. Insect Biochem. Mol. Biol. 52, 51–59.

Jasrapuria, S., Arakane, Y., Osman, G., Kramer, K.J., Beeman, R.W., Muthukrishnan, S., 2010. Genes encoding proteins with peritrophin A-type chitin-binding domains in *Tribolium castaneum* are grouped into three distinct families based on phylogeny, 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.

Kelly, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858.

Kerwin, J.L., Turecek, F., Xu, R., Kramer, K.J., Hopkins, T.L., Gatlin, C.L., John, R., Yates III, 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 Bioinf. 5, 138.

Nei, N.S.M., 1987. The neighbor-joining method- a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

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.

Petkau, G., Wingen, C., Jussen, L.C.A., Radtke, T., Behr, M., 2012. Obstructor-a is required for epithelial extracellular matrix dynamics, exoskeleton function, and tubulogenesis. J. Biol. Chem. 287, 21396–21405.

Rebers, J.E., Willis, J.H., 2001. A conserved domain in arthropod cuticular proteins binds chitin. Insect Biochem. Mol. Biol. 31, 1083–1093.

Suetsugu, Y., Futahashi, R., Kanamori, H., Kadono-Okuda, K., Sasanuma, S., Narukawa, J., Ajimura, M., Jouraku, A., Namiki, N., Shimomura, M., Sezutsu, H., Osanai-Futahashi, M., Suzuki, M.G., Daimon, T., Shinoda, T., Taniai, K., Asaoka, K., Niwa, R., Kawaoka, S., Katsuma, S., Tamura, T., Noda, H., Kasahara, M., Sugano, S., Suzuki, Y., Fujiwara, H., Kataoka, H., Arunkumar, K.P., Tomar, A., Nagaraju, J., Goldsmith, M.R., Feng, Q., Xia, Q., Yamamoto, K., Shimada, T., Mita, K., 2013. Large scale full-length cDNA sequencing reveals a unique genomic landscape in a lepidopteran model insect. *Bombyx mori*. G3. 3, 1481–1492.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: 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., 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.

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.

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

Willis, J.H., 2010. Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genome era. Insect Biochem. Mol. Biol. 40, 189–204.

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