Received: 28 July 2019 | Revised: 17 September 2019 | Accepted: 19 September 2019

DOI: 10.1002/arch.21628

RESEARCH ARTICLE

Archives of INSECT BIOCHEMISTRY AND PHYSIOLOGY WILEY

Genome-wide identification and analysis of genes encoding cuticular proteins in the endoparasitoid wasp *Pteromalus puparum* (Hymenoptera: Pteromalidae)

Jiale Wang[1] | Hongxia Jin[1] | Lei Yang[1,4] | Xinhai Ye[1] | Shan Xiao[1] | Qisheng Song[2] | David Stanley[3] | Gongyin Ye[1] | Qi Fang[1]

[1]State Key Laboratory of Rice Biology & Ministry of Agriculture and Rural Affairs Key Laboratory of Molecular Biology of Crop Pathogens and Insects, Institute of Insect Sciences, Zhejiang University, Hangzhou, China

[2]Division of Plant Sciences, University of Missouri, Columbia, Missouri

[3]USDA Agricultural Research Service, Biological Control of Insects Research Laboratory, Columbia, Missouri

[4]State Key Laboratory Breeding Base for Zhejiang Sustainable Pest and Disease Control, Zhejiang Academy of Agricultural Sciences, Hangzhou, China

Correspondence
Qi Fang, Institute of Insect Sciences, Zhejiang University, 310058 Hangzhou, China.
Email: fangqi@zju.edu.cn

Funding information
Natural Science Foundation of Zhejiang Province, Grant/Award Number: LY18C140001; Program for Chinese Outstanding Talents in Agricultural Scientific Research of Ministry of Agriculture and Rural Affairs of the People's Republic of China; National Key R&D Program of China, Grant/Award Number: 2017YFD0200400; Program for Chinese Innovation Team in Key Areas of Science and Technology of Ministry of Science and Technology of the People's Republic of China, Grant/Award Number: 2016RA4008; National Natural Science Foundation of China, Grant/Award Numbers: 31620103915, 31830074

Abstract

The multifunctional insect cuticle serves as the exoskeleton, determines body shape, restricts water loss, provides attachment sites for muscles and internal organs and is a formidable barrier to invaders. It is morphologically divided into three layers, including envelope, epicuticle, and procuticle and is composed of chitin and cuticular proteins (CPs). Annotation of CPs and their cognate genes may help understand the structure and functions of insect cuticles. In this paper, we interrogated the genome of *Pteromalus puparum*, an endoparasitoid wasp that parasitizes *Pieris rapae* and *Papilio xuthus* pupae, and identified 82 genes encoding CPs belonging to six CP families, including 62 in the CPR family, 8 in CPAP3, 5 in CPF/CPFL, 2 low complexity proteins, 2 in TWDL, and 3 in Apidermin. We used six RNA-seq libraries to determine CP gene expression profiles through development and compared the cuticle hydrophobicity between the *P. puparum* and the ectoparasitoid *Nasonia vitripennis* based on GRAVY values of CPR sequences. In the *Nasonia-Pteromalus* comparison, we found

in both *N. vitripennis* and *P. puparum*, the peak of their CPR hydrophobicity displayed at their pupal stage, whereas their adult stage showed the lowest level. Except at the adult stage, the CPR hydrophobicity in *N. vitripennis* is always higher than *P. puparum*. Finally, we identified three novel Apidermin genes, a family found solely in Hymenoptera and revealed a new sequence feature of this family. This new information contributes to a broader understanding of insect CPs generally.

KEYWORDS
apidermin, cuticular proteins, expression profile, hydrophobicity, *Pteromalus puparum*

1 | INTRODUCTION

Insect cuticle is a kind of complex extracellular biological composite, secreted by a single layer of epidermal cells. It makes up the exoskeleton, which serves as a support structure, attachment sites for muscles and ligaments, determines the shape, protects the internal organs and is a physical barrier to protect against desiccation and prevent from invaders, such as parasites and pathogens. Morphological studies show, insect cuticle consists of three distinct layers, including the waterproofing envelope (outermost), the protein-rich epicuticle and the chitin-rich procuticle (innermost; Locke, 2001; Moussian, 2010). It is mainly composed of chitin and cuticular proteins (CPs), which form a stabilized cuticular structure that maintains the elasticity and other physical properties of insect surfaces (Vincent & Wegst, 2004). According to the assumption that few differences in chitin chain length and acetylation among various cuticle types, CPs exert considerable influence on the structure and properties of insect cuticle (Andersen, Hojrup, & Roepstorff, 1995). In the silkworm, *Bombyx mori*, the number of genes encoding CPs exceeds 1% of the protein-coding genes (Futahashi et al., 2008), from which we infer analysis of CPs and their cognate genes can help understand cuticle structure and physiological functions.

The number of known CP sequences has increased by about 90% (from <150 to >1,400) since 2005, now assorted into 13 families (Willis, 2010). Among all CP gene families, six are found in Hymenoptera. As the most common family of cuticular proteins, the CPR family features a nearly 63-amino acid-long, conserved sequence, named the Rebers and Riddiford Consensus (R&R; Rebers & Riddiford, 1988). It is sub-classified into three forms (RR-1, RR-2, and RR-3; Andersen, 1998, 2000). RR-1 proteins have been isolated from flexible cuticles and are present in endocuticle. RR-2 proteins are associated with hard cuticles and exocuticle (Andersen, 2000). Compared with the RR-1 consensus region, the RR-2 counterpart is much more conserved in length and sequence. An RR-3 protein features a 75-amino acid region with sequence similarities and a variant of the R&R consensus sequence in the C-terminal end, but no exact definition has been established (Ioannidou, Theodoropoulou, Papandreou, Willis, & Hamodrakas, 2014). As is mentioned, few related proteins were assigned from RR-3 subfamily, and its definition is still unclear in many other species (Ioannidou et al., 2014). Therefore, like other reported CPR annotation manuscripts, the results related to the RR-3 subfamily of *P. puparum* will not be present and further discussed in this paper. The families with one and three of Chitin-binding domain type 2 (ChtBD2) are defined as Cuticular Proteins Analogous to Peritrophins (CPAP)1 and CPAP3, respectively (Jasrapuria et al., 2010). The CPF family corresponds to a conserved motif of about 42 amino acids and appears similar in the amino acids near their carboxyl-termini, while the other CPs, lacking this defining consensus in the C-terminal region, are named CPF-like

(CPFL; Togawa, Augustine Dunn, Emmons, & Willis, 2007). Members of cuticular protein of low complexity with proline-rich (CPLCP), a family widespread throughout insects and frequently occurs in two mosquitos genera *Aedes* and *Culex* in particular (Cornman & Willis, 2009), feature a high density of amino acids PV and PY pairs. The *Tweedle* family (TWDL) had been identified in *Drosophila melanogaster* and named after the tubby phenotype of one of its mutants. With four conserved regions defined, TWDL proteins are widely found in insects and contribute to the maintaining insect body shape during development. Apidermin (APD), presumably restricted to Hymenoptera, is recorded in *Apis mellifera* and *Nasonia vitripennis* and recognized by chromosomal linkage. Some motifs that do not symbolize families were found in CPs, including more than three copies of the most common short motif A-A-P-(A/V/L; Andersen et al., 1995; Willis, 2010), three repeats of an 18-amino acid motif found in a *Bombyx mori* protein (Nakato, Izumi, & Tomino, 1992), short motifs G-Y-R and Y-L-P recorded in a number of *Drosophila* species in the past decade (Cornman, 2009).

Parasitoid wasps, the most abundant and diverse species in Hymenoptera, usually lay eggs and develop larvae in (endoparasitoids) or on (ectoparasitoids) the bodies of their arthropod hosts. Cuticles from endoparasitoids directly interact with their hosts and may act in evading host immunity, diffusing oxygen and absorbing nutrition. *Pteromalus puparum* is an effective biocontrol agent that lays eggs in lepidopterans, particularly *Pieris rapae* and *Papilio xuthus*. In this paper, we report on genes encoding the *P. puparum* CPs.

2 | MATERIALS AND METHODS

2.1 | Annotation of CP genes

The total size of the assembled genome of *P. puparum* is 338.1 Mb with contig and scaffold N50 sizes of 38.7 Kb and 1.16 Mb, respectively. A *tBLASTn* analysis was performed using 12 recognized CP motifs and 3 CP hymenopteran sequences (Amel*CPR24*, NP_001167616.1, NP_00116858.1) as probes (File S1). With the R&R Consensus of chitin_bind_4 (PF00379/ IPR000618), the CPR family was identified using HMMER v3.1b2 (http://www.hmmer.org/) based on Pfam31.0 (Eddy, 1995; Finn et al., 2016). We used the website CuticleDB (http://bioinformatics.biol.uoa.gr/cuticleDB) to determine whether each CPR family member is RR-1 or RR-2 with default E-value cutoff $5.0e^{-07}$ (Karouzou et al., 2007). CPF (including CPFL), CPAP, and TWDL families were identified using a 44-amino acid consensus sequence (PF11018/ IPR022727), ChtBD2 domains (PF01607/IPR002557) and a TWDL motif (PF03103/ IPR004145), respectively (Willis, Papandreou, Iconomidou, & Hamodrakas, 2012). We took the remaining putative CP genes from *tBLASTn* results with highly conserved regions as queries to determine whether they were members of other CP families. All genes identified on sequence similarity were checked manually with online *BLASTp*.

2.2 | Phylogenetic analysis

All *A. mellifera* CPR sequences were downloaded from CuticleDB and combined with the *P. puparum* sequences for analysis. The sequences were aligned using Muscle v3.8.31 and adjusted by trimAl v1.2 (Capella-Gutierrez, Silla-Martinez, & Gabaldon, 2009; Edgar, 2004). The phylogenetic trees were constructed by the Maximum likelihood (ML) method and by RAxML with 1000 bootstrap replications (Stamatakis, 2014) respectively. The best model of amino acid replacement was selected by ProtTest v3.4 with a lowest Bayesian Information Criterion (BIC) score (Abascal, Zardoya, & Posada, 2005; Darriba, Taboada, Doallo, & Posada, 2011). We used iTOL v3 for coloring the dendrograms (Letunic & Bork, 2016).

2.3 | Expressions profile analysis

Six RNA-seq libraries were prepared from selected *P. puparum* developmental stages and sequenced by Beijing Genomics Institute (BGI, Shenzhen, China). The transcriptomic raw data was assembled using Trinity v2013-02-16

and the expression level calculated by Cufflinks (Trapnell et al., 2012). In RNA-Seq, the relative expression of transcript is proportional to the number of cDNA fragments that originate from it. FPKM representing as average Fragments Per Kilobase of exon per Million fragments mapped was used to reflect gene relative expression level of the transcript. Hierarchical clustering of CP expression profiles was performed using a heat-map tool on the platform BMKCloud (www.biocloud.net) and all FPKM values showing on heatmaps were normalized by rows.

2.4 | Real-time quantitative polymerase chain reaction and qPCR

P. puparum samples from four developmental stages (embryos, larvae, pupae, and adults) were collected and the pupae and adults were further divided into female and male individually. Individuals of these six groups were washed and pooled into a centrifuge tube with TRIzol reagent (Invitrogen), respectively, to extract the total RNA. Reverse transcription was performed by using PrimeScript™ One-Step RT-PCR Kit (Takara, Japan) and Real-time quantitative PCR was performed with 25 μl reaction mixture by using SYBR Green Supermix Kits (Takara) on BIO-RAD CFX96™ Real-Time System. Following the manufacturer's instructions, the process was programmed for 95°C for 30 s, 40 cycles with 95°C for 5 s to denature DNA, and 60°C for 34 s to anneal. The relative developmental-specific expression levels of the indicated genes were conducted according to the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen, 2001), taking a stably expressed reference gene 18 s rRNA as an internal control. The primers designed on website Primer 3 are listed in Table S1.

2.5 | Calculation of the grand average of hydrophobicity (GRAVY)

Using the online program Protein GRAVY, which is developed on the method of Kyte & Doolittle (www.bioinformatics.org/sms2/protein_gravy.html), the GRAVY value for each identified CPR protein was calculated as the sum of hydropathy values of all amino acids divided by the length of the deduced sequence (Kyte & Doolittle, 1982). For most proteins, GRAVY values are in the range from −2 to +2. The positively rated protein is hydrophobical, whereas the negatively rated protein is hydrophilical.

3 | RESULTS AND DISCUSSION

3.1 | Identification and classification of CP genes in *P. puparum*

We identified 82 genes encoding CPs in *P. puparum* belonging to six groups: CPR, CPAP3, CPF/CPFL, CPLCP, TWDL, and APD (Table S2). Some of the sequences are similar and may be isoforms of the same gene. They may be tandemly arranged on the same chromosome and form a phylogenetic cluster. For example, PPU02386, PPU02387, and PPU02389 are located on scaffold_165 and are 89–90% identical at the amino acid level to a predicted *N. vitripennis* larval cuticle protein (XP_001602894.2). PPU16075 and PPU16077 share identity with an *N. vitripennis* endocuticle structural glycoprotein ABD-5 (XP_001602743.1). PPU10777, PPU10778, and PPU10779 aligned with a predicted *N. vitripennis* pupal CP C1B-like (XP_001603496.1).

We found that 62 genes belong to the CPR family, the most abundant and widespread CP family that features a R&R Consensus. Some sequences with a R&R Consensus but calculated E-values higher than $5.0e^{-07}$ for RR-1 and RR-2 were named RR-unclassified (RR-UC). Thus, 24 RR-1, 32 RR-2, and 6 RR-UC were initially determined in *P. puparum*. Most of the RR-1 and RR-2 proteins were clearly separated, and several RR-UC proteins fell in the RR-1 and RR-2 regions, which may be reassigned into RR-1 and RR-2 (Figure 1). We revised the classification by embedding PPU03747 and PPU05069 in the RR-2 group and removing PPU07126 from RR-2 group to RR-UC. A cuticular protein from *A. mellifera* (Amel) named AmelCPR26 was assigned a member of CPR-RR2 in the CuticleDB database, but in the Hidden Markov Modeling its E-value cutoff ($8.4e^{-07}$) is too high to support assignment. Amel*CPR26* and PPU07079 occur in a closely related clade, and we classified PPU07079 as a member of the RR-UC.

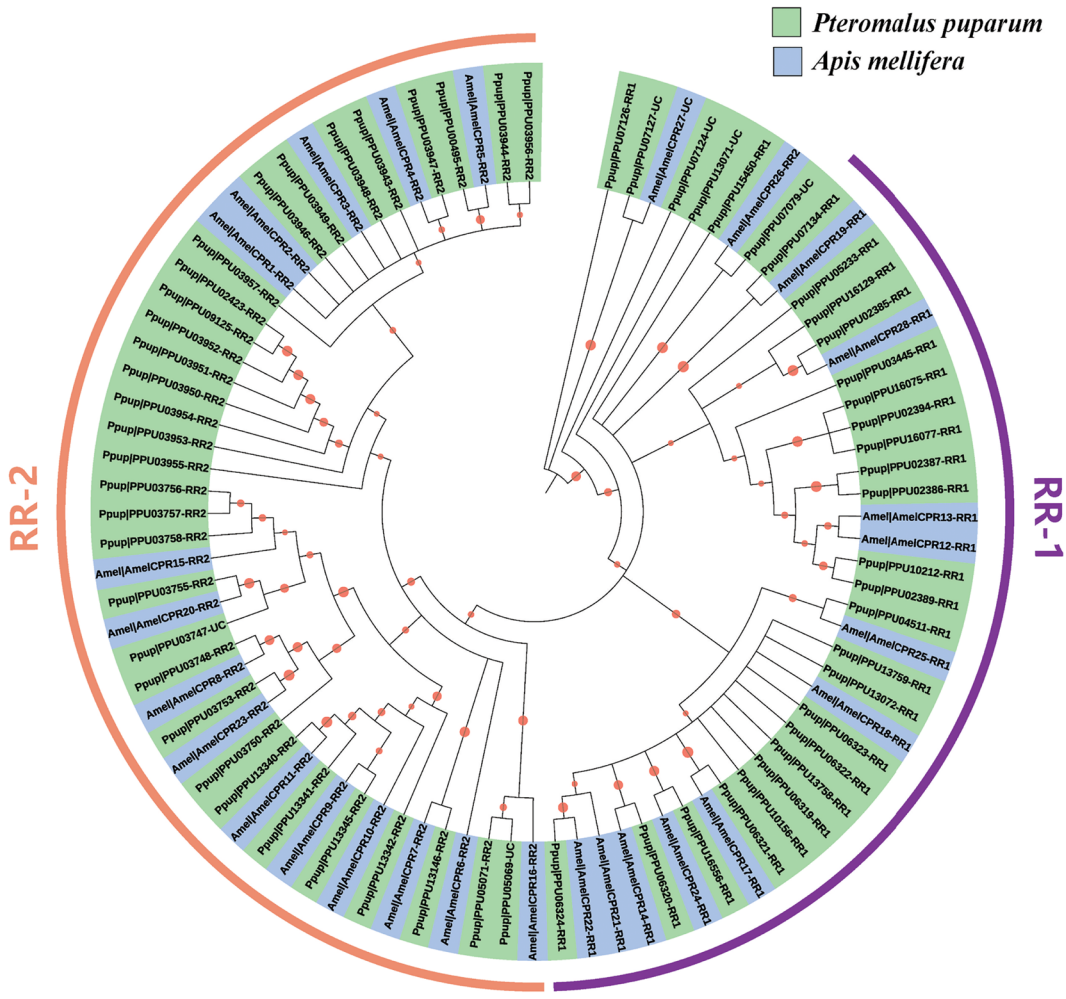

FIGURE 1 Phylogenetic analysis of CPR proteins among *Apis mellifera* (Amel) and *Pteromalus puparum* (Ppup). All sequences were adjusted by Trimal v1.2 and the branches with a bootstrap value below 20% (of 1,000 replicates) were deleted. Three predicted types of CPRs including RR-1, RR-2, and RR-UC (unclassified) were indicated

As the second-largest group of CPs, CPAPs were classified according to the number of peritrophin A-type chitin-binding domains (ChtBD2) they contain (Jasrapuria et al., 2010). All the sequences containing a ChtBD2 were taken from the *P. puparum* genome and submitted to another CP family prediction tool CutProtFam-Pred (http://aias.biol.uoa.gr/CutProtFam-Pred/home.php; Ioannidou et al., 2014), searching against the CPAP1 and CPAP3 families separately. We obtained the default scores of 50 for CPAP1s and 77.5 for CPAP3s. This confirmed our classification. Eight sequences belong to CPAP3, which were first recognized in *D. melanogaster* as *gasp* (gene analogous to small peritrophins) and *obstructor* (Barry, Triplett, & Christensen, 1999; Behr & Hoch, 2005), but no CPAP1 was found. Six of the putative CPAP3s were tandemly arranged on scaffold_21 and another two were located on scaffold_56 and _0 separately. Although the latter gene, PPU03978, was 98% identical at the amino acid level to an uncharacterized protein from *N. vitripennis* (XP_001605920.3), it had three ChtBD2 domains and was 91% identical at the amino acid level to a CP analogous to peritrophins 3-C precursor from *A. mellifera* (NP_001165860.1).

All the remaining *P. puparum* CPs had orthologs in *N. vitripennis*, including five CPF/CPFL genes, two CPLCPs, two TWDLs and three APDs. The total number of CPs varies across insect species, ranging from 50 in *A. mellifera* to

TABLE 1 Comparison between the numbers of CPs in different CP families manually annotated in insect genomes

	CPR	CPAP1	CPAP3	CPF + CPFL	CPLCP	TWDL	ADP	CPLCA	CPLCG	CPLCW	CPCFC	CPG	Other	Total
Pteromalus puparum	62	0	8	5	2	2	3	0	0	0	0			82
Apis mellifera	32	0	5	3	2	2	6	0	0	0	0			50
Nasonia vitripennis	62	0	6	4	3	2	3	0	0	0	0			80
Bombyx mori	155	0	1	1 + 4	7	4	0	2	0	0	1	23	33	226
Dendrolimus. punctatus	147	9	13	1 + 2	7	4	0	2	0	0	1	9	22	214
Manduca sexta	205	15	10	1 + 6		4					1			235
Spodoptera litura	193	13	9	1 + 7		5		4			1	28	26	279
Drosophila melanogaster	101	2	6	3	5	27	0	11	3	0	1			159
Anopheles gambiae	156	0	7	4 + 7	28	12	0	3	27	9	1		10	253
Nilaparvate lugens	96	17	8	3	6	3						4		137
Tribolium castaneum	101	10	7	5 + 3	4	3	0	0	2	0	2			129

Note: (Cornman & Willis, 2009; Dittmer et al., 2015; Ioannidou et al., 2014; Vannini, Bowen, Reed, & Willis, 2015; Liu et al., 2019; Werren et al., 2010; Pan et al., 2018).

279 in *Spodoptera litura* (Table 1). Numbers of CP genes increase with the number of CPR genes in any given species. The numbers of CPR genes in *P. puparum*, *A. mellifera*, and *N. vitripennis* are relatively small, possibly due to their protected lifestyle within hives or hosts during immature stages (Willis, 2010). In contrast, the CPR genes in lepidopteran species like *B. mori* (155; Futahashi et al., 2008; Liang, Zhang, Xiang, & He, 2010), *Dendrolimus punctatus* (147; Yang et al., 2017), *Manduca sexta* (205; Dittmer et al., 2015) and *S. litura* (193; Liu et al., 2019) are significantly enriched. These genes may contribute to success in their open and relatively dangerous habitats.

3.2 | Expression analysis of putative CP genes

We developed six RNA-seq libraries, including embryos, larvae (1-day), female pupae, male pupae, female adults and male adults. The average fragments per kilobase of exon per million fragments mapped (FPKM) of CPs expressed across all libraries was 95.90, with a minimum of 0 and a maximum of 9,281.93 (Table S3).

The CP genes with the highest and the second highest FPKM (PPU13758 with 9,281.93 and PPU13759 with 8,320.38) were from the larval library, belonging to CPR-RR1 family. Three other CPs with FPKM values over 1,000 in the larvae library were all in the CPR-RR1 family: PPU05233 (7,368.04), PPU10156 (5,337.44), and PPU02386 (3,490.91). One CP gene, PPU10499, in the male and female pupal libraries had similar FPKM values, 2,634.78 for females and 2,395.68 for males. We identified four other CPs in pupae with FPKM over a hundred (PPU10493, PPU10495, PPU03956, PPU05069). The expressions of most CPs in female and male adults were fairly consistent, with a few exceptions.

A variety of CPRs act in different developmental stages. All CPRs clustered into four groups, Clusters A to D represent upregulated genes in three developmental stages (Figure S1A). Andersen (2000) concluded that RR-1 proteins were from the soft, flexible cuticle whereas RR-2 proteins were prevalent in hard, rigid cuticles. We used the ratio of RR-1/RR-2 to reflect the cuticle features in across developmental stages, the higher the softer and vice versa. The ratio of pupal RR-1/RR-2 = 4 in Cluster A. In Cluster B adult RR-1/RR-2 = 0.67 and in C, larval RR-1/RR-2 = 0.91, suggesting that the *P. puparum* larval and pupal cuticles are more likely to be harder compared to adults. However, some exceptions have been found against Andersen hypothesis in *An. gambiae* (Vannini & Willis, 2017). It indicates that the localization of CPR groups depends more on the properties of individual proteins, rather than the physical properties of the cuticle (hard, soft). The physical properties may not be the sole factor.

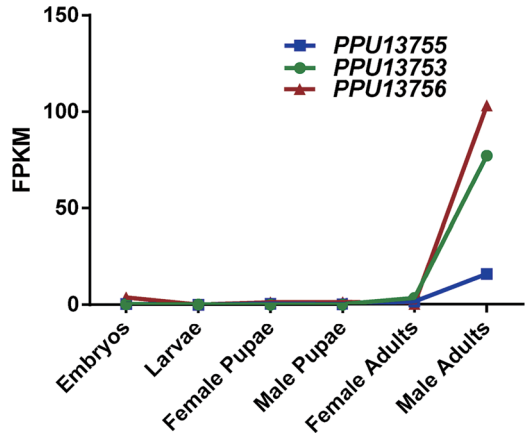

FIGURE 2 Transcriptional level of three APD genes in six RNA-seq libraries. Reflected by FPKM, all Apidermin genes in *Pteromalus puparum* were found only highly expressed in male adults. APD, apidermin; FPKM, Fragments Per Kilobase of exon per Million

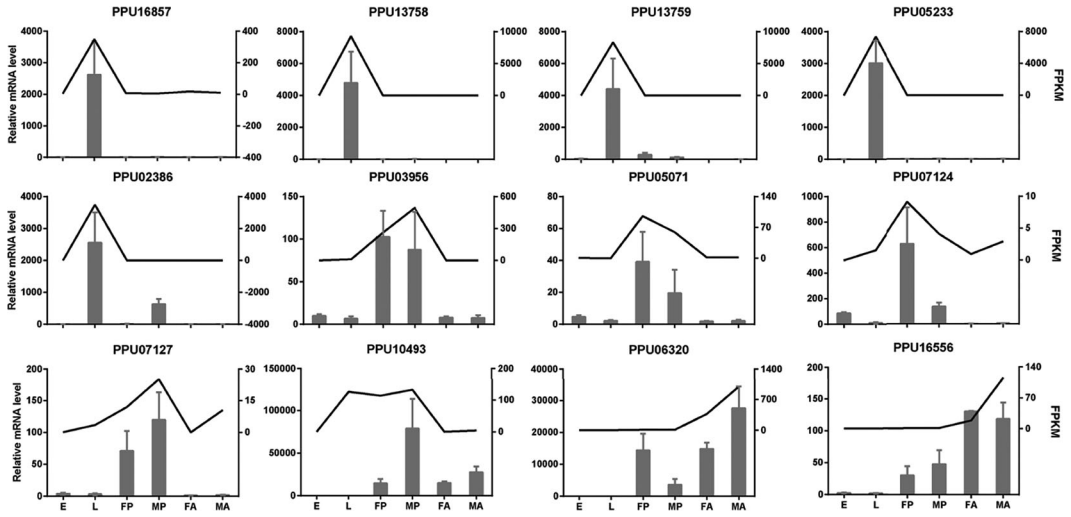

FIGURE 3 The relative developmental-specific expression levels of *Pteromalus puparum* CP genes. Total RNA was extracted from *P. puparum* embryos (E), larvae (L), female pupae (FP), male pupae (MP), female adults (FA), and male adult (MA). The bars are showing the relative mRNA levels plotting on the left Y axis while the polylines are showing the FPKM values plotting on the right Y axis. The standard errors of the mean from three biological replicates were represented by the error bars. FPKM, Fragments Per Kilobase of exon per Million

In *Tribolium castaneum*, most CPAP genes were expressed throughout the life stages (Jasrapuria, Specht, Kramer, Beeman, & Muthukrishnan, 2012), however, the *P. puparum* CPAP3s are particularly expressed in larvae and pupae (Figure S1B).

Three CP genes from the APD family were highly expressed in male adults (Figure 2). We speculated that APDs in *P. puparum* act in the maturing of adult cuticle as APDs were expressed in drone heads of *A. mellifera* (Sun et al., 2012; Sun, Huang, Zheng, & You, 2013). Most APDs from *A. mellifera* (Amel), such as Amel*apd-3-like*, Amel*apd-like*, and Amel*apd-2*, are highly expressed in male adults, as just seen in *P. puparum*. However, the function of those APDs remains unclear. To verify the expression patterns of CP genes, 12 CP genes highly expressed in different developmental stages were chosen for qPCR analysis. The results were basically consistent with the transcriptome data (Figure 3).

3.3 | Comparing amino acid components and hydrophobicity of *P. puparum* and *N. vitripennis* CPR proteins

Insect CPRs typically feature low levels of cysteine and methionine (Andersen et al., 1995; Willis, Iconomidou, Smith, & Hamodrakas, 2005), possibly due to frequent appearances of the AAPA/V repeats in CPs (Andersen et al., 1995). In CPRs of *P. puparum*, alanine, glycine, proline, and serine are the main amino acid residues. RR-1 contained significantly higher proportions of asparagine, glutamine, and glutamic acid, and lower proportions of histidine compared to RR-2 (Figure S2). Among *Cimex lectularius*, *Anopheles gambiae*, *D. punctatus*, and *P. puparum* (Cornman et al., 2008; Koganemaru, Miller, & Adelman, 2013; Yang et al., 2017), all RR-2s had higher proportions of histidine than RR-1s, which may act in sclerotization of hard cuticles through cross-linking with chitin and polyphenols (Andersen, 2010; Iconomidou, Willis, & Hamodrakas, 2005).

We compared the hydrophobicity of CPs between *P. puparum* and *N. vitripennis*, using grand average of hydropathicity (GRAVY), a modified hydrophobicity algorithm based on kd-Hydrophobicity according to Kyte and Doolittle (1982; www.bioinformatics.org/sms2/protein_gravy.html). Combining the expression profiles through the developmental stages, we selected the CPR proteins with the FPKM value above 1 from larval, pupal and adult

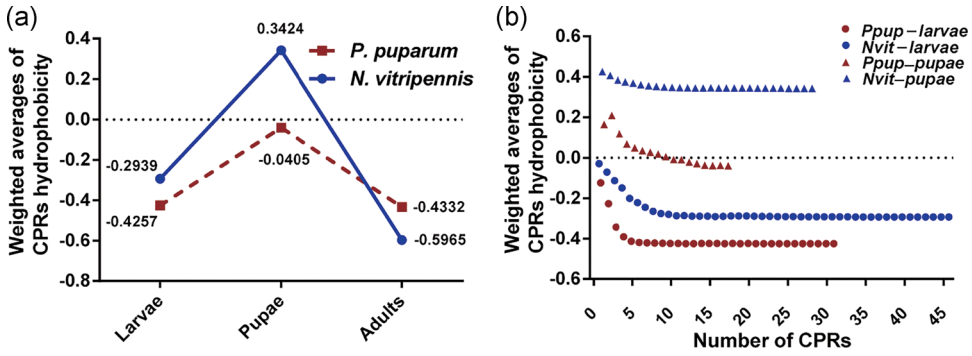

FIGURE 4 The weighted average of all CPRs whose FPKM values were above one in over developmental stages. (a) Developmental cuticle hydrophobicity comparison between an endoparasitoid *Pteromalus puparum* and an ectoparasitoid *Nasonia vitripennis*, indicating the relationship between parasitizing environments and hydrophobicity. (b) The changing rate of calculated cuticle hydrophobicity varies significantly when the number of CPRs considered increase by a positive order of FPKM, revealing the number of main hydrophobicity-affecting CPRs. FPKM, Fragments Per Kilobase of exon per Million

RNA-seq libraries separately, and sorted them in ascending order by the female expression levels if data from both genders were available. We found 31 CPRs expressed in larvae, 17 in pupae, and 10 in adults from *P. puparum* and 46 CPRs in larvae, 28 in pupae and 12 in adults from *N. vitripennis*. We acquired the cuticle hydrophobicity value of each species through development by calculating the weighted average of its CPRs hydrophobicity with FPKM as the weight (Figure 4a; Table S4). The weighted CPR hydrophobicity averages of *P. puparum* (−0.4257 in larvae, −0.0405 in pupae) were lower compared to *N. vitripennis* (−0.2939 in larvae, 0.3424 in pupae; the higher number indicates more hydrophobicity) possibly due to a more humid environment inside the host. The weighted average of CPR hydrophobicity values is higher at pupal stage than at larval stage in both ecto- and endoparasitoid species. As parasitoids leave their hosts after eclosion, the weighted averages of CPR hydrophobicity values of both adults were low, −0.4332 in *P. puparum* and −0.5965 in *N. vitripennis*. It likely implies that the cuticle in pupae is more hydrophobic than in larvae and adults. However, the factors influencing the hydrophobicity of insect cuticle are complex. All recognized cuticular layers and their numerous types of chemical compounds are probably involved in. Moreover, the hydrophobicity is related to the degree of sclerotization. The hardened incest cuticle tends to have more hydrophobic proteins (Hillerton & Vincent, 1983). Hence, the cuticle hydrophobicity at each developmental stage of the parasitoid wasps needs to be further experimentally determined. We also calculated the weighted means of CPR hydrophobicity values of both species through development (Figure 4b). In larvae, there are more hydrophobicity-affecting CPRs in *P. puparum* than *N. vitripennis*, which is reversed in pupae.

3.4 | Common characteristics of APD family CPs among four hymenopterans

Their names indicate the APD genes were first identified in *A. mellifera* epidermal cells. The APD family has 10 proteins known solely in Hymenoptera, including APD-1 to APD-3, APD-like, APD-2-like, and APD-3-like from *A. mellifera* (Kucharski, Maleszka, & Maleszka, 2007; Sun et al., 2013; Sun et al., 2012), APD-1 to APD-3 from *N. vitripennis* (Willis, 2010) and APD-2 from *Apis cerena cerena* (Sun et al., 2012). RT-PCR results indicate that the expression levels of APP-1 to APD-3 genes in *A. mellifera* were upregulated when its adult cuticle was being deposited (Miranda Soares et al., 2013). We downloaded the sequences from GenBank (http://www.ncbi.nlm.nih.gov). Based on putative orthologs to the *N. vitripennis* genes, we identified three *P. puparum* APDs by *tBLASTn*, named APD-1 to APD-3 (PPU13755, PPU13753, and PPU13756; Table S5).

We used 13 hydrophobic APD protein sequences, with a high percentage of alanine content (14.5–33.6%), to construct a phylogenetic tree (Figure 5a; Kumar, Stecher, & Tamura, 2016). The APD family proteins were assorted

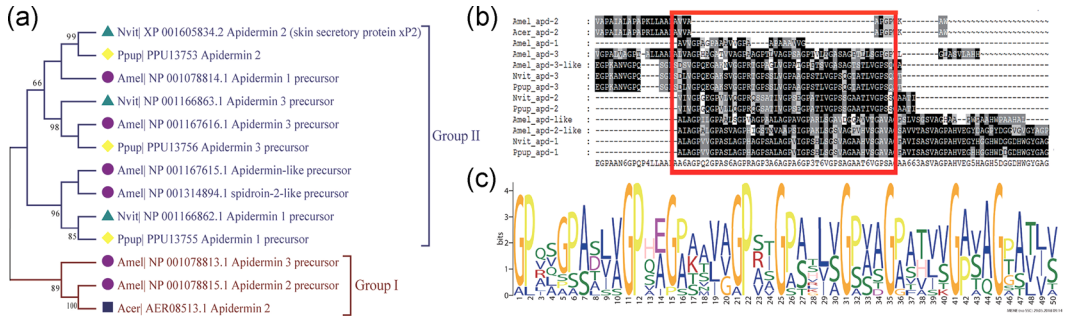

FIGURE 5 Features of three putative Apderimin genes in *Pteromalus puparum* and ten recognized Apidermin genes from other three species in Hymenoptera. (a) Phylogenetic tree. The tree was built in MEGA7 by the Maximum likelihood (ML) method with a JTT model (1,000 bootstrap replications) and the bootstrap values below 50% have been eliminated, revealing all Apidermin genes could be divided into Group I and Group II based on the length of the amino acid sequence. (B) Alignment of 13 amino acid sequences from Apidermin gene family. The highly conserved region in ~50 aa length was displayed with a box. (c) Sequence logo for Apidermin gene family. Thirteen sequences from four species of Hymenoptera were used. Note ten repeats of the highly conserved GP combination

into two groups. Three shorter APDs from Apidae (7.7–10.7 kDa) fell into group I and the other, relatively longer APDs (10.9–39.5 kDa), into group II including all APDs from Pteromalidae which were larger (21.6–39.5 kDa).

Due to the limited amount of recognized APD protein sequences, no defined motif or domain of this family has previously been revealed. Our amino acid alignments show a conserved region of about 50 amino acids, except for two APDs shorter than 80 amino acids from Apidae (NP_001078815.1 and AER08513.1; Figure 5b). We found ten Gly–Pro combinations within the 50 amino acid stretch, which may be a novel motif (Figure 5c). We propose that the Gly–Pro combinations may be an additional feature for identifying the APDs.

4 | CONCLUSION

Cuticular proteins (CPs) are essential elements of insect biology. They have evolved features that have many forms and functions that underlie the vast adaptive radiation of insects generally and, in the context of this special issue, parasitoids in particular. According to the position where they lay eggs and develop larvae, parasitoid wasps are classified into either endoparasitoids or ectoparasitoids. Compared to ectoparasitoids, endoparasitoids consume tissues inside of their arthropod hosts and evolved completely different niches, supported by the evolution of substantially differing CPs. Cuticles from endoparasitoids directly interact with their hosts and may have a different role from ectoparasitoids in evading host immunity, diffusing oxygen and absorbing nutrition. Therefore, understanding CPs is necessary to appreciate the many types of parasitoid-host interactions, each of which represents separate co-evolutionary host–parasitoid systems.

Here, the differences in cuticle protein compositions and hydrophobicity were compared using an endoparasitoid *P. puparum* and a model ectoparasitoid *N. vitripennis*, belonging to the same family Pteromalidae. Genome-wide analysis revealed 82 putative CPs in *P. puparum*. The number is similar in *N. vitripennis* (80 putative CPs; Werren et al., 2010), staying at a low level of total amounts of CPs compared with many other insect species. It may be related to the parasitic lifestyle. Especially for endoparasitoid larvae, a strongly defensive exoskeleton may not be required, as the cuticle of their host probably provide with a protective layer against infections from the environment. The GRAVY results of the CPR proteins reveal the cuticle hydrophobicity on the protein level. CPRs in endoparasitoids *P. puparum* are relatively hydrophilic whereas those in ectoparasitoids *N. vitripennis* are relatively hydrophobic. It implies that the lower cuticle hydrophobicity (on protein levels) in endoparasitoids may lead to

permeability to absorb nutrition from host hemolymph where endoparasitoids submerge during the early developmental stages.

Among parasitoid species, many have evolved specialized parasitic relationships with eggs, larvae, pupae, and adults. All CPs investigated in this paper are from pupal parasitoids *P. puparum* and *N. vitripennis*, but we predict that future studies will find CPs have evolved in separate ways in adaptations to each host life stage. The evolution and biological significance of CPs helps understand, in broad terms, why the parasitoid radiation led to their extraordinarily large numbers of species. This may be related to the many kinds of host species. Although insects are fairly similar with respect to their internal body compartments, such as hemolymph, organ and tissue qualities, they differ in subtle ways across and within orders. These differences create subtly differing niches available for new species. We speculate changes in CPs would be part of evolving adaptations to new species. The CP database is growing and we look forward to the coming new insights into this database will yield.

ACKNOWLEDGMENTS

The study is supported by Major International (Regional) Joint Research Project of National Natural Science Foundation of China (31620103915), Key Program of National Natural Science Foundation of China (31830074), National Key R&D Program of China (2017YFD0200400), the Program for Chinese Innovation Team in Key Areas of Science and Technology of Ministry of Science and Technology of the People's Republic of China (2016RA4008), Program for Chinese Outstanding Talents in Agricultural Scientific Research of Ministry of Agriculture and Rural Affairs of the People's Republic of China, and Natural Science Foundation of Zhejiang Province of China (LY18C140001). The authors also thanks professor John H. Werren from University of Rochester for his contributions to this manuscript. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. All programs and services of the U.S. Department of Agriculture are offered on a nondiscriminatory basis without regard to race, color, national origin, religion, sex, age, marital status, or handicap.

ORCID

Gongyin Ye http://orcid.org/0000-0002-5983-3061
Qi Fang http://orcid.org/0000-0003-4150-3408

REFERENCES

Abascal, F., Zardoya, R., & Posada, D. (2005). ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics*, *21*, 2104–2105.

Andersen, S. O. (1998). Amino acid sequence studies on endocuticular proteins from the desert locust, *Schistocerca gregaria*. *Insect Biochemistry and Molecular Biology*, *28*, 421–434.

Andersen, S. O. (2000). Studies on proteins in post-ecdysial nymphal cuticle of locust, *Locusta migratoria*, and cockroach, *Blaberus craniifer*. *Insect Biochemistry and Molecular Biology*, *30*, 569–577.

Andersen, S. O. (2010). Insect cuticular sclerotization: A review. *Insect Biochemistry and Molecular Biology*, *40*, 166–178.

Andersen, S. O., Hojrup, P., & Roepstorff, P. (1995). Insect cuticular proteins. *Insect Biochemistry and Molecular Biology*, *25*, 153–176.

Barry, M. K., Triplett, A. A., & Christensen, A. C. (1999). A peritrophin-like protein expressed in the embryonic tracheae of *Drosophila melanogaster*. *Insect Biochemistry and Molecular Biology*, *29*, 319–327.

Behr, M., & Hoch, M. (2005). Identification of the novel evolutionary conserved obstructor multigene family in invertebrates. *FEBS Letters*, *579*, 6827–6833.

Capella-Gutierrez, S., Silla-Martinez, J. M., & Gabaldon, T. (2009). TrimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*, 1972–1973.

Cornman, R. S. (2009). Molecular evolution of *Drosophila* cuticular protein genes. *PLoS One*, *4*. e8345.

Cornman, R. S., Togawa, T., Dunn, W. A., He, N., Emmons, A. C., & Willis, J. H. (2008). Annotation and analysis of a large cuticular protein family with the R&R Consensus in *Anopheles gambiae*. *BMC Genomics*, *9*, 22.

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

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics*, *27*, 1164–1165.

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 Biochemistry and Molecular Biology*, *62*, 100–113.

Eddy, S. R. (1995). Multiple alignment using hidden Markov models. *Proceedings. International Conference on Intelligent Systems for Molecular*, *3*, 114-120.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*, D279–D285.

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 Biochemistry and Molecular Biology*, *38*, 1138–1146.

Hillerton, J. E., & Vincent, J. F. V. (1983). Consideration of the importance of hydrophobic interactions in stabilizing insect cuticle. *International Journal of Biological Macromolecules*, *5*, 163–166.

Iconomidou, V. A., Willis, J. H., & Hamodrakas, S. J. (2005). Unique features of the structural model of 'hard' cuticle proteins: Implications for chitin–protein interactions and cross-linking in cuticle. *Insect Biochemistry and Molecular Biology*, *35*, 553–560.

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 Biochemistry and Molecular Biology*, *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 Biochemistry and Molecular Biology*, *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.

Karouzou, M. V., Spyropoulos, Y., Iconomidou, V. A., Cornman, R. S., Hamodrakas, S. J., & Willis, J. H. (2007). Drosophila cuticular proteins with the R&R consensus: Annotation and classification with a new tool for discriminating RR-1 and RR-2 sequences. *Insect Biochemistry and Molecular Biology*, *37*, 754–760.

Koganemaru, R., Miller, D. M., & Adelman, Z. N. (2013). Robust cuticular penetration resistance in the common bed bug (*Cimex lectularius* L.) correlates with increased steady-state transcript levels of CPR-type cuticle protein genes. *Pesticide Biochemistry and Physiology*, *106*, 190–197.

Kucharski, R., Maleszka, J., & Maleszka, R. (2007). Novel cuticular proteins revealed by the honey bee genome. *Insect Biochemistry and Molecular Biology*, *37*, 128–134.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*, 1870–1874.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*, 105–132.

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, *44*, W242–W245.

Liang, J., Zhang, L., Xiang, Z., & He, N. (2010). Expression profile of cuticular genes of silkworm, *Bombyx mori*. *BMC Genomics*, *11*, 173.

Liu, J., Li, S., Li, W., Peng, L., Chen, Z., Xiao, Y., … 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.

Locke, M. (2001). The Wigglesworth lecture: Insects for studying fundamental problems in biology. *Journal of Insect Physiology*, *47*, 495–507.

Miranda Soares, M. P., Barchuk, A. R., Quirino Simoes, A. C., Cristino, A. D. S., de Paula Freitas, F. C., Canhos, L. L., & Gentile Bitondi, M. M. (2013). Genes involved in thoracic exoskeleton formation during the pupal-to-adult molt in a social insect model, *Apis mellifera*. *BMC Genomics*, 14.

Moussian, B. (2010). Recent advances in understanding mechanisms of insect cuticle differentiation. *Insect Biochemistry and Molecular Biology*, *40*, 363–375.

Nakato, H., Izumi, S., & Tomino, S. (1992). Structure and expression of gene coding for a pupal cuticle protein of *Bombyx mori*. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, *1132*, 161–167.

Pan, P., Ye, Y., Lou, Y., Lu, J., Cheng, C., Shen, Y. … Zhang, C. (2018). A comprehensive omics analysis and functional survey of cuticular proteins in the brown planthopper. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 5175–5180.

Rebers, J. E., & Riddiford, L. M. (1988). Structure and expression of a manduca-sexta larval cuticle gene homologous to *Drosophila* cuticle genes. *Journal of Molecular Biology*, *203*, 411–423.

Shi, Y. Y., Sun, L. X., Huang, Z. Y., Wu, X. B., Zhu, Y. Q., Zheng, H. J., & Zeng, Z. J. (2013). A SNP based high-density linkage map of *Apis cerana* reveals a high recombination rate similar to *Apis mellifera*. *PLoS One*, *8*, 76459.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313.

Sun, L. X., Huang, Z. Y., Zheng, H. J., Ge, Q. X., Gong, L. P., & Chen, H. Y. (2012). Characterization of three new members of the Apidermin (APD) gene family from honeybees and sequence analysis of the insect APD family. *Acta Entomologica Sinica*, *1*, 12–23.

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 Biochemistry and Molecular Biology*, *37*, 675–688.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*, 562–578.

Vannini, L., Bowen, J. H., Reed, T. W., & Willis, J. H. (2015). The CPCFC cuticular protein family: Anatomical and cuticular locations in *Anopheles gambiae* and distribution throughout Pancrustacea. *Insect Biochemistry and Molecular Biology*, *65*, 57–67.

Vannini, L., & Willis, J. H. (2017). Localization of RR-1 and RR-2 cuticular proteins within the cuticle of *Anopheles gambiae*. *Arthropod structure & development*, *46*, 13–29.

Vincent, J. F., & Wegst, U. G. (2004). Design and mechanical properties of insect cuticle. *Arthropod Structure & Development*, *33*, 187–199.

Werren, J. H., Richards, S., Desjardins, C. A., Niehuis, O., Gadau, J., Colbourne, J. K., … Gibbs, R. A. (2010). Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science*, *327*, 343–348.

Willis, J. H. (2010). Structural cuticular proteins from arthropods: Annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochemistry and Molecular Biology*, *40*, 189–204.

Willis, J. H., Iconomidou, V. A., Smith, R. F., & Hamodrakas, S. J. (2005). Cuticular proteins. In L. I. Gilbert (Ed.), *Comprehensive Molecular Insect Science* (pp. 79–109). Chapel Hill, NC: Elsevier.

Willis, J. H., Papandreou, N. C., Iconomidou, V. A., & Hamodrakas, S. J. (2012). Cuticular proteins. In L. I. Gilbert (Ed.), *Insect Molecular Biology and Biochemistry* (pp. 134–166). Chapel Hill, NC: Academic Press.

Yang, C. H., Yang, P. C., Zhang, S. F., Shi, Z. Y., Kang, L., & Zhang, A. B. (2017). Identification, expression pattern, and feature analysis of cuticular protein genes in the pine moth *Dendrolimus punctatus* (Lepidoptera: Lasiocampidae). *Insect Biochemistry and Molecular Biology*, *83*, 94–106.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Wang J, Jin H, Yang L, et al. Genome-wide identification and analysis of genes encoding cuticular proteins in the endoparasitoid wasp *Pteromalus puparum* (Hymenoptera: Pteromalidae). *Arch Insect Biochem Physiol*. 2019;e21628. https://doi.org/10.1002/arch.21628