


# SCIENTIFIC REPORTS



OPEN

## Identification and characterization of serine protease inhibitors in a parasitic wasp, *Pteromalus puparum*

Lei Yang<sup>1</sup>, Yaotian Mei<sup>1</sup>, Qi Fang<sup>1</sup>, Jiale Wang<sup>1</sup>, Zhichao Yan<sup>1</sup> , Qisheng Song<sup>2</sup>, Zhe Lin<sup>3</sup> & Gongyin Ye<sup>1</sup>

Serine protease inhibitors (SPIs) regulate protease-mediated activities by inactivating their cognate proteinases, and are involved in multiple physiological processes. SPIs have been extensively studied in vertebrates and invertebrates; however, little SPI information is available in parasitoids. Herein, we identified 57 SPI genes in total through the genome of a parasitoid wasp, *Pteromalus puparum*. Gene structure analyses revealed that these SPIs contain 7 SPI domains. Depending on their mode of action, these SPIs can be categorized into serpins, canonical inhibitors and alpha-2-macroglobulins (A2Ms). For serpins and canonical inhibitors, we predicted their putative inhibitory activities to trypsin/chymotrypsin/elastase-like enzymes based on the amino acids in cleaved reactive sites. Sequence alignment and phylogenetic tree indicated that some serpins similar to known functional inhibitory serpins may participate in immune responses. Transcriptome analysis also showed some canonical SPI genes displayed distinct expression patterns in the venom gland and this was confirmed by quantitative real-time PCR (qPCR) analysis, suggesting their specific physiological functions as venom proteins in suppressing host immune responses. The study provides valuable information to clarify the functions of SPIs in digestion, development, reproduction and innate immunity.

Serine proteases (SPs), account for almost one-third of all proteases, and serve as inevitable components in catalyzing hydrolytic reactions both intra- and extracellularly. SPs participate in various physiological processes such as food digestion, embryo development and immune defense<sup>1</sup>. Although SPs are of great importance, sometimes they may be potentially hazardous to their enzymatic environments when not being properly controlled<sup>2</sup>. The hazard of excessive peptidase activities includes inappropriate coagulation, melanization, tissue damage, etc.<sup>3,4</sup>. In fact, the action of SPs is tightly regulated by their inhibitors, including serine protease inhibitors (SPIs)<sup>5-7</sup>. When SPs finish their work, they are inactivated by SPIs and then move out of the circulation, to keep the homeostasis of living body. The roles of SPIs include blood coagulation, reproduction, complement system and innate immune response<sup>8,9</sup>.

Based on the mechanisms of action, SPIs can be classified into serpins,  $\alpha$ -macroglobulins (A2Ms), canonical inhibitors and non-canonical inhibitors<sup>5,10</sup>. Most of the serpins share identical structures with seven or nine  $\alpha$ -helices and three  $\beta$ -sheets. Serpin features the property of a reactive center loop (RCL) that binds to objective protein near the C terminus of the sequence<sup>11</sup>. Serpin RCL possesses a scissile bond between residues P1 (N-terminus of cleavage event) and P1' (C-terminus of cleavage event). Once scissile bond is cleaved, a dramatic conformational change is triggered. Serpin together with its peptide ultimately forms a permanent covalent complex. In comparison to serpins, the canonical protein inhibitors usually occur as single, small proteins and are composed of several disulfide bonds. Through an exposed RCL, serpin can interact with the protease reactive center region and defunctionalize the activity of the proteases. The mechanism of canonical protein inhibitors is

<sup>1</sup>State Key Laboratory of Rice Biology & Ministry of Agriculture Key Lab of Molecular Biology of Crop Pathogens and Insects, Institute of Insect Sciences, Zhejiang University, Hangzhou, 310058, China. <sup>2</sup>Division of Plant Sciences, College of Agriculture, Food and Natural Resources, University of Missouri, Columbia, Missouri, USA. <sup>3</sup>State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing, 100101, China. Correspondence and requests for materials should be addressed to G.Y. (email: [chu@zju.edu.cn](mailto:chu@zju.edu.cn))

similar to that in serpins, since both of them can bind to the objective proteins in protease-substrate interaction mode. The difference between them is that canonical protein inhibitors are kept in reversible tight-binding interactions. Canonical inhibitors can be divided into about 20 protein families such as Kazal, trypsin inhibitor-like (TIL), Pacifastin and Kunitz serine protease inhibitors<sup>5</sup>. The A2Ms are much larger proteins discovered in living organisms and vary widely from invertebrates to vertebrates. The A2Ms are members of thioester-containing proteins (TEPs)<sup>12</sup>. There are six members in TEPs family, including C3/C4/C5, A2M, pregnancy zone protein (PZP), CD109, insect TEP (iTEP), and the complement 3 and PZP-like A2M domain-containing 8 (CPAMD8)<sup>13,14</sup>. Other than the inhibitors mentioned above, A2M could trap the target proteases by a  $\alpha_2$ -macroglobulin cage to form a reversible complex. As a result, proteases from different classes such as thiol-, carboxyl-, serine- and metalloprotease can be enfolded by A2Ms<sup>15</sup>.

Genetic analyses have provided fundamental understanding of functions of these SPIs in immunity, embryo development and reproduction. It is widely studied that serpins take effect in inhibiting Spätzle-processing enzyme in Toll signaling pathway and PPO activation cascade<sup>16</sup>. In *Manduca sexta*, serpin-4 (MsSRPN4) and serpin-5 (MsSRPN5) are inhibitors of hemolymph proteinase-1 (MsHP-1) and hemolymph proteinase-6 (MsHP-6), and function in protease cascades of PPO pathway. Thus, MsSRPN4 and MsSRPN5 may be involved in melanization processes<sup>17</sup>. *Manduca sexta* serpin-1J (MsSRPN1J) and serpin-6 (MsSRPN6), capable of forming covalent complexes with prophenoloxidase-activating proteinase-3 (MsPAP3) and hemolymph proteinase-8 (MsHP-8), may participate in the Toll pathway and PPO activating cascades<sup>18,19</sup>. Studies also showed that serpins mediate several anti-inflammatory effects in host-pathogen interactions<sup>20</sup>. In addition, some serpins even inhibit cysteine proteinase through an alternative mechanism<sup>21</sup>. Compared to serpins, there are fewer molecular studies on canonical SPIs. A Kazal-type serine protease inhibitor named oryctin inhibits  $\alpha$ -chymotrypsin, endopeptidase K, subtilisin and elastase, indicating it may be related to digestion<sup>22</sup>. Pacifastin-like peptides can induce significant growth retardation in a desert locust<sup>23</sup> and can defend against entomopathogenic fungi in the silkworm<sup>24</sup>. In some mosquitos and ticks, canonical SPIs secreted by salivary or venom gland possess anticoagulation activity, which is vital for these blood-feeding insects to successfully manipulate the physiology of hosts<sup>8,25</sup>.

Recently, genome-wide analyses have helped us to identify serpin genes from several insect species, including *Anopheles gambiae* and *Aedes aegypti*<sup>26</sup>, 12 *Drosophila* species<sup>27</sup>, *Apis mellifera*<sup>28,29</sup>, *Tribolium castaneum*<sup>30</sup>, *Bombyx mori*<sup>31</sup>, and *Plutella xylostella*<sup>32</sup>. Compared to serpins, the comprehensive analyses of all the SPI genes based on genome databases have only been illustrated in two lepidopteran species (*B. mori* and *P. xylostella*) so far<sup>26,33</sup>. Since all of the SPIs play significant roles in several biological processes, identification of these regulator molecules is necessary for better understanding of the biochemical and molecular functions of SPIs.

In this study, *Pteromalus puparum* SPI genes were identified and characterized based on newly sequenced *P. puparum* genome. *P. puparum* is a pupal endoparasitoid wasp and plays a critical role in biological control of certain pierid species, especially the small white butterfly, *Pieris rapae*. Since *P. rapae* is an important pest of the crucifer and caper families, studying on this parasitic wasp is of great significance. Besides, we have finished the whole *P. puparum* genome and several transcriptomes (unpublished data). Serine proteases and their homologs, serving as the putative targets of SPIs, have already been characterized in *P. puparum* genome<sup>34</sup>. Thus, this information provided us a solid foundation to explore SPI genes of *P. puparum*. In the present study, *P. puparum* SPIs and their putative functions were identified and predicted. Meanwhile, the temporospatial expression analyses of SPI genes were also performed. These findings give us a detailed overview of the *P. puparum* SPIs for further clarifying their functions in digestion, innate immunity, development and reproduction.

## Results

**Identification of SPI genes in *P. puparum*.** We obtained 57 SPI genes in total through BLAST searches in *P. puparum* genome database (Table 1 & Supplementary Table S1). All of the SPI amino sequences were summarized in Supplementary Table S2. The SPI genes of an ectoparasitoid wasp named *Nasonia vitripennis* were also listed for comparative purpose (Supplementary Table S3). These 57 *P. puparum* SPIs (PpSPIs) contain 7 inhibitor domains and can be divided into three types: serpin, A2M and canonical SPIs (Table 1).

**Serpins.** A total of 10 serpin genes of *P. puparum* were characterized in this study. BLAST searches for serpin genes were also conducted in *N. vitripennis* genome, which revealed a similar number of serpins (9) (Supplementary Table S3). The length of *P. puparum* serpins varies from 150 to 830 amino residues, with a typical mature serpin about 300–400 amino acids. PpSPI10 is incomplete in part of the amino acid sequence whereas PpSPI3 contains two intact serpin domains, resulting in abnormal sequence length (Supplementary Table S1). The structures of serpins determine their mechanisms in protease inhibition. The most important regions involved in the conformational changes of serpin in proteinase inhibition are beta-sheet A and RCL. We aligned *P. puparum* serpins with structurally well-defined inhibitory serpins (*Manduca sexta* serpin 1 K and human alpha-1 antitrypsin)<sup>35,36</sup>, and marked the structures, beta-sheets, and the hinge, breach, shutter and gate regions accordingly (Supplementary Figure S1). Analysis of the serpin molecular structure showed that three alpha-helices and nine beta-sheets are conserved in most of *P. puparum* serpins. The sequence of PpSPI6 is distinguished from other *P. puparum* serpins by missing the definitive amino acids in hinge, breach, shutter and gate regions. PpSPI4 contains nearly 80 residues between helix I and strand 5 A while PpSPI10 loses the structure from helix C to strand 3 C.

The hinge region, which stabilizes the metastable native serpin conformation, is located at N-terminal portion of the RCL region. The consensus patterns of specific amino acids in the hinge region were observed in most of the *P. puparum* serpins except PpSPI6. The amino acid at P1 position determines the enzyme inhibition specificity of the serpin. Serpin with Arg or Lys at the P1 position may prohibit trypsin-like SPIs, may have the inhibitory activity to chymotrypsin-like SPIs with Phe, Tyr, Leu or Ile at P1 position, and may participate in inhibiting elastase-like enzyme with P1 position of Ala or Val<sup>26,33</sup>. We predicted the position of the scissile bond of the P.

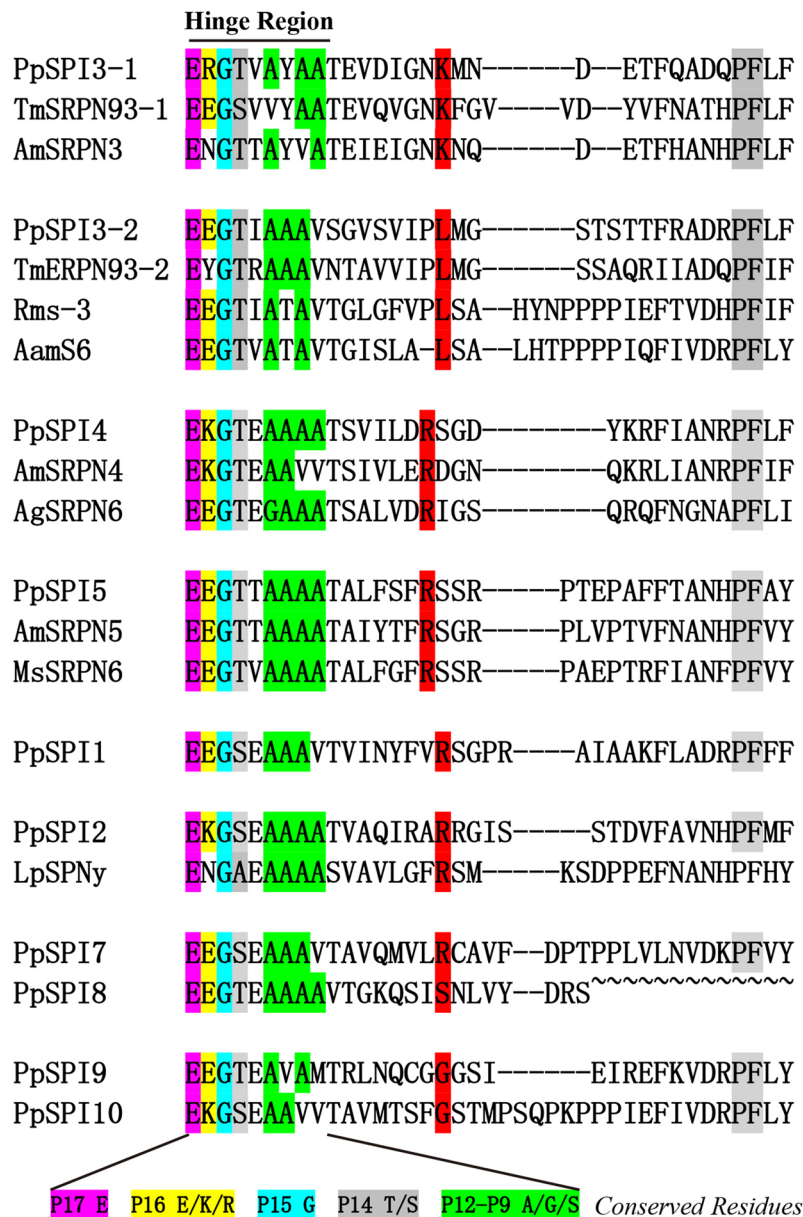
Gene Name	SPI domain	Number of Domain	Category
PpSPI1~PpSPI2	serpin	1	serpin
PpSPI3		2	
PpSPI4~PpSPI10		1	
PpSPI11~PpSPI12	Kazal	1	Canonical SPIs
PpSPI13		3	
PpSPI14~PpSPI26		1	
PpSPI27		11	
PpSPI28		3	
PpSPI29~PpSPI30		1	
PpSPI31		8	
PpSPI32		1	
PpSPI33~PpSPI34	Pacifastin	1	
PpSPI35		7	
PpSPI36		8	
PpSPI37		5	
PpSPI38		9	
PpSPI39		4	
PpSPI40		2	
PpSPI41~PpSPI42		1	
PpSPI43		2	
PpSPI44		1	
PpSPI45		2	
PpSPI46~PpSPI50	TIL	5	
PpSPI51~PpSPI53	Kunitz_BPTI	3	
PpSPI54	Kunitz_BPTI/WAP	11/1	
PpSPI55~PpSPI57	A2M	1	A2M

**Table 1.** Serine protease inhibitor (SPI) domains in *Pteromalus puparum*.

*puparum* serpins (Fig. 1). PpSPI1, 2, 3-1, 4, 5 and 7 are anticipated to inhibit trypsin-like SPs with P1 position containing Arg or Lys. PpSPI3-2 has a Leu residue at P1 position, and may serve as a chymotrypsin inhibitor.

Phylogenetic tree was constructed using *P. puparum* serpins along with 5 *A. mellifera* serpins and 24 known function serpins from 10 insect species (Fig. 2). The serpins from *P. puparum* form 7 clades. PpSPI3 is a twin-domain serpin, with the first domain PpSPI3-1 clusters with AmSRPN3 and TmSPN93-1, the second PpSPI3-2 forms clades with AamS6, Rms-3 and TmSPN93-2. PpSPI3-1 and PpSPI3-2 shared the same amino acids at P1 position (Lys) with the corresponding genes, respectively. PpSPI4 shares a high similarity with AmSRPN4 and AgSRPN6. Sequences alignment showed that the hinge region and the residue at P1 position (Arg) of AgSRPN6 are identical to that of PpSPI4 (Fig. 1). We examined the microbe-induced expressions of *PpSPI4* using Gram-negative bacterium *Escherichia coli*, Gram-positive bacterium *Micrococcus luteus* and entomopathogenic fungus *Beauveria bassiana* (Supplementary Figure S2). The expression of *PpSPI4* could be induced by the treatment of *B. bassiana* 24 h post infection (p. i.). PpSPI5, AmSRPN5 and MsSRPN6 form a clade with supported bootstrap values no less than 95%. PpSPI5 and MsSRPN6 share the same amino acids at P1 position (Arg) and hinge position (Fig. 1). Quantitative real-time PCR (qPCR) analyses indicated a dramatically increased expression of *PpSPI5* in all of three microbes-infected cases. The induction was stable after *E. coli* infection whereas the expression of *PpSPI5* was highly induced in the *M. luteus*-infected and *B. bassiana*-infected samples 6 h and 24 h p. i., respectively (Supplementary Figure S2).

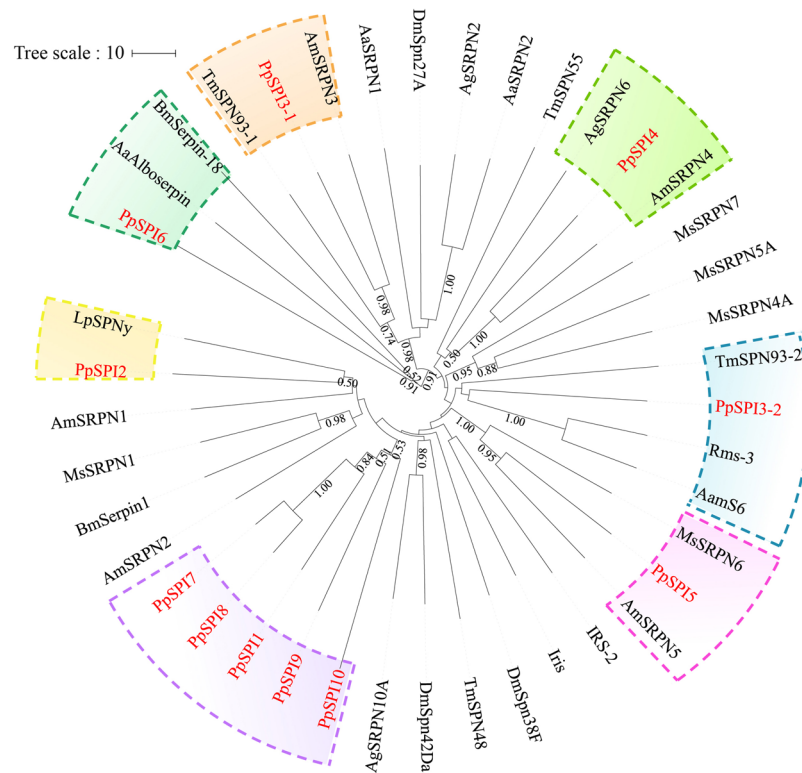
**Canonical serine protease inhibitors.** There were 44 canonical SPI genes identified in *P. puparum*. The canonical SPIs in *P. puparum* can be divided into 5 different families: Kazal, TIL, Pacifastin, Kunitz and whey acidic protein (WAP) (Table 1). Most of these canonical SPIs only contain one kind of SPI domains except PpSPI54. Canonical SPIs are cysteine-rich peptides and common with multiple disulfide bonds buried inside the molecule to offer structure stability<sup>5</sup>. Multiple sequence alignments in the same SPI family showed conservation in the number and arrangement of the cysteine residues (Supplementary Figure S3). Kazal, Pacifastin and Kunitz domain contain 3 pairs of disulfide bonds (with the combination form of 1-5, 2-4 and 3-6 in Kazal, 1-6, 2-4 and 3-5 in Kunitz, 1-4, 2-6 and 3-5 in Pacifastin), 4 pairs in WAP (linking 1-6, 2-7, 3-5, 4-8) and 5 pairs in TIL (ranging from 1-7, 2-6, 3-5, 4-10 and 8-9). SPIs from the same family show gene cluster phenomenon in scaffolds. There are 4, 5 and 5 Kazal genes clustered in scaffold\_111, \_17 and \_19, respectively. The same things also happen in Pacifastins family. Three members in Pacifastin family are located in scaffold\_6 and another 4 members located in scaffold\_7. Studies also showed all of the 5 TIL genes are clustered in scaffold\_119, sharing similar molecular sizes and intron-exon structures (Fig. 3).



**Figure 1.** Multiple sequence alignment of the hinge and reactive center loop (RCL) regions of *Pteromalus puparum* serpins. The hinge and RCL regions of *P. puparum* serpins were aligned with those serpins with known functions. Predicted P1 residues are highlighted in red. *P. puparum* serpins are presented in numerical order and grouped with homologous serpins from other insects determined by phylogenetic analysis. The GenBank accession number for sequences used in phylogenetic analysis in Fig. 1: AamS6 (ABS87358.1, *Amblyomma americanum*); AmSRPN3, AmSRPN4, AmSRPN5 (XP\_001122067.2, XP\_003249882.1, XP\_006562425.1, *Apis mellifera*); AgSRPN6 (ABJ52806.1, *Anopheles gambiae*); LpSPNy (ACQ83466.1, *Leptopilina bouvardi*); MsSRPN6 (AAV91026.1, *Manduca sexta*); Rms-3 (AHC98654.1, *Rhipicephalus microplus*) and TmSPN93 (BAL03254.1, *Tenebrio molitor*).

Canonical SPI interacts with its substrate via a convex, exposed binding loop. The loop has an active P1-P1' scissile bond in the cleavage site of the SPI. P1 residue determines the specific inhibitory activities of the SPIs, which are similar to the roles of P1 in serpins<sup>6</sup>. We listed the putative P1 residue through the alignments of SPI domains (Supplementary Figure S3). Of 43 canonical SPIs, 28 are single domain proteins and their inhibition activities depend on the characteristics of single SPI domain. According to sequence alignments, PpSPI35, PpSPI46 and PpSPI57 may inhibit trypsin-like SPs (with Arg located at P1 position); PpSPI31 is anticipated to inhibit elastase-like SPs (with Ala located at P1 position); PpSPI28 and PpSPI48 may participate in inhibiting chymotrypsin enzymes (with Phe located at P1 position).

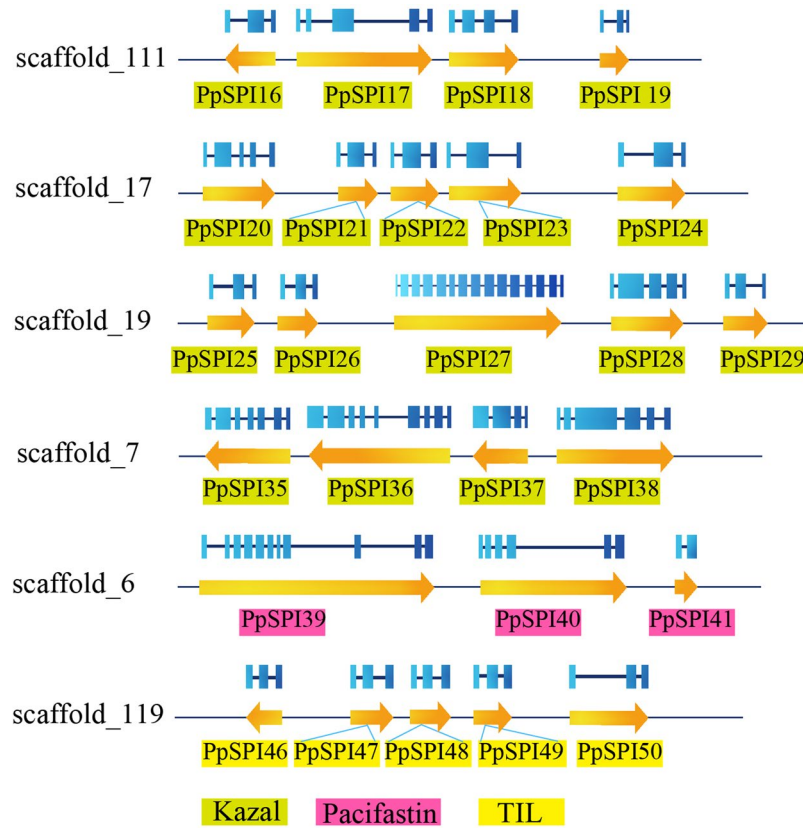
Fifteen canonical SPIs encompass multi domains. SPIs in Kazal, Kunitz and Pacifastin families contain several repeats of homogeneous SPI domains. SPIs with multiple SPI domains come along with varieties of P1 positions, so these SPIs exhibit inhibitory activities. PpSPI29 may inhibit all of the trypsin, chymotrypsin and elastase



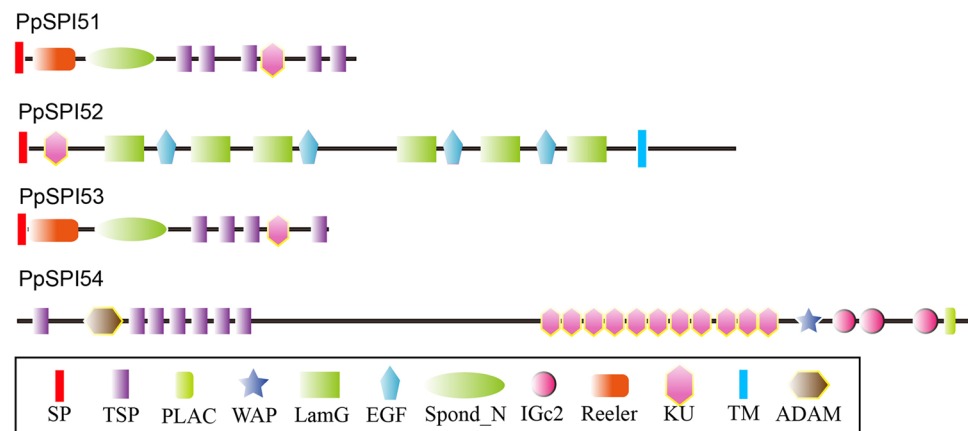
**Figure 2.** The phylogenetic tree includes 11 *Pteromalus puparum* serpins, 5 *Apis mellifera* serpins and 24 known function serpins from 10 other insect species. All of the serpins were named according to relevant literatures. Phylogenetic tree was constructed by Neighbor-joining method, using the program Mega 5.10. The bootstrap values greater than 0.5 are dotted on the nodes. The GenBank accession numbers for sequences used in phylogenetic analysis in Fig. 2: AaSRPN1, AaSRPN2, AaAlboserpin (XP\_001648011.1, XP\_001651231.2, AAC31158.1, *Aedes aegypti*); AamS6 (ABS87358.1, *Amblyomma americanum*); AgSRPN2, AgSRPN6, AgSRPN10A (ABJ52801.1, ABJ52806.1, XP\_314159.2, *Anopheles gambiae*); AmSRPN1, AmSRPN2, AmSRPN3, AmSRPN4, AmSRPN5 (GB17012<sup>33</sup>, XP\_016772722.1, XP\_001122067.2, XP\_003249882.1, XP\_006562425.1, *Apis mellifera*); BmSerp1, BmSerp1-18 (ACT36272.1, ACG61181.1, *Bombyx mori*); DmSpn27A, DmSpn38F, DmSpn42Da (NP\_652024.1, CAB63098.1, NP\_724512.1, *Drosophila melanogaster*); Iris, IRS-2 (CAB55818.2, ABI94056.2, *Ixodes ricinus*); LpSPNy (ACQ83466.1, *Leptopilina boulardi*); MsSRPN1, MsSRPN2, MsSRPN3, MsSRPN4, MsSRPN5, MsSRPN6, MsSRPN7 (AAC47341.1, AAS68503.1, AAS68507.1, AAV91026.1, ADM86478.1, *Manduca sexta*); Rms-3 (AHC98654.1, *Rhipicephalus microplus*); and TmSPN48, TmSPN55, TmSPN93 (BAI59108.1, BAI59107.1, BAL03254.1, *Tenebrio molitor*).

enzymes, due to Ile and Phe at the second and tenth P1 position, Ala and Val at the third and sixth P1 position, and Arg and Lys at the fourth and seventh P1 position. PpSPI37, 38, 39, 40, 41 could also inhibit both chymotrypsin and trypsin according to the corresponding amino acids at P1 positions (Supplementary Figure S3). We also analyzed the structure domains of *P. puparum* Pacifastins and predicted putative dibasic cleavage sites based on the liable fracture site (RR, RK or KK residues) (Supplementary Figure S4). Several smaller, single or multi domain inhibitors were obtained, and each can exert function alone. The SPIs in Kunitz and WAP family have the mixtures of SPI domains and non-inhibitor domains (Fig. 4). PpSPI51 and PpSPI53 show similarity in structural arrangement. PpSPI53 consists of 1 Reeler, 1 Spond\_N, 4 TSP and 1 Kunitz\_BPTI domains, whereas PpSPI51 has one added TSP domain in C-terminal amino sequences. PpSPI52 includes 1 Kunitz\_BPTI, 4 EGF and 6 Laminin G domains. PpSPI54 is composed of 7 TSP, 11 Kunitz\_BPTI, 3 immunoglobulin C-2, 1 ADAM-spacer1, 1 WAP and 1 PLAC domains.

**A2M.** The A2Ms belong to the TEP family. TEPs usually contain A2M-N, A2M-N2, A2M, A2M-COMP and A2M-RECEP domains. In the genome of *P. puparum*, we identified three SPIs containing typical featured domains of TEPs. Phylogenetic tree was constructed using three *P. puparum* SPIs and several TEP sequences from 22 different species by NJ method (Fig. 5). TEPs are separated into C3 and A2M subfamilies, both of which are supported by 96% bootstrap values. The subfamily of A2M is then categorized into four parts, A2M, CPAMD8, CD109 and iTEP. SPIs from *P. puparum* share high similarity with corresponding protein sequences from *N. vitripennis* and *A. mellifera* with 1:1:1 ratio of ortholog. In detail, PpSPI55 clusters with A2Ms whereas PpSPI57 clusters with insect TEPs. PpSPI56, together with corresponding SPIs in *N. vitripennis* and *A. mellifera*, forms a branch in A2M subfamily, distinct from other A2M subfamily genes in vertebrate and invertebrate. Based on the overall topology of the tree, PpSPI56 is most likely to be an iTEP. Therefore, we considered PpSPI55 as an A2M, PpSPI56 and PpSPI57 as iTEPs.



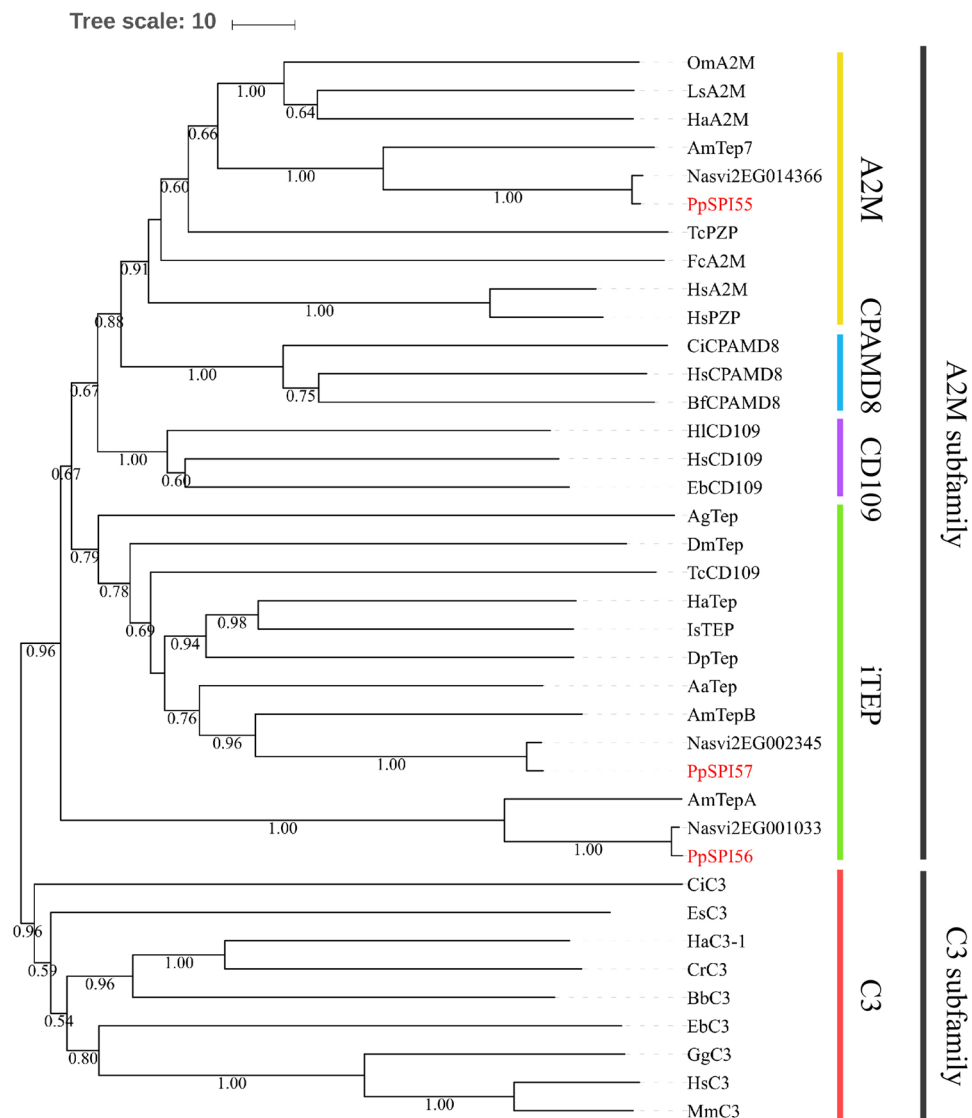
**Figure 3.** Structure and location of *Pteromalus puparum* canonical SPI genes on scaffolds. The arrows indicate the transcription orientations and gene sizes on scaffolds. The exons are shown with blue boxes.



SP, signal peptide; TSP, thrombospondin type 1 repeats; PLAC, protease and lacunin; LamG, Laminin G domain; EGF, EGF-like; Igc2, immunoglobulin C-2 Type; KU, BPTI/Kunitz; ADAM, ADAM-spacer1; TM, transmembrane domain.

**Figure 4.** Domain organizations of *Pteromalus puparum* SPIs with non-inhibitory units in the Kunitz family.

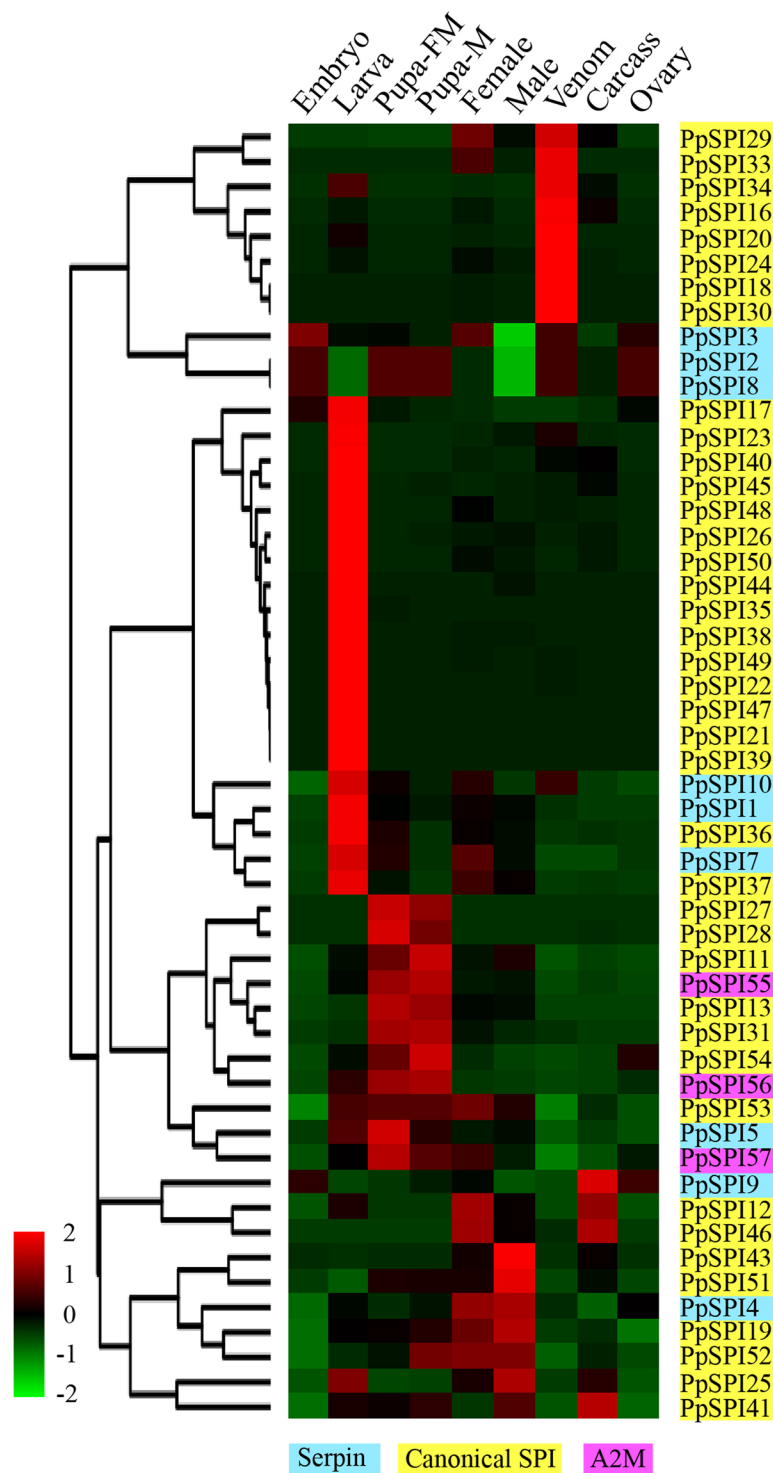
To classify the conserved and diverged residues of three *P. puparum* A2M and iTEP genes, multiple sequence alignments were performed. PpSPI55, PpSPI56 and PpSPI57 possess putative signal peptides in the N-terminus for secretion. In addition, four proteinase-binding A2M domains (A2M-N, A2M-N2, A2M and A2M-COMP) and one receptor binding A2M domain (A2M-RECEP) were characterized (Supplementary Figure S5). The sequences between *P. puparum* A2M and iTEP with corresponding *N. vitripennis* genes share significant similarity with the identities no less than 97%. The FPETW conserved in bait region of A2M family, also exists in the sequences of PpSPI55. The thioester motifs in PpSPI55 and PpSPI57 are GCGEQ, which are also identical to those of other TEPs. The corresponding position in the sequences of PpSPI56 is replaced by DCGEQ. His (H)



**Figure 5.** Phylogenetic analysis of thioester-containing proteins. Phylogenetic tree was constructed by Neighbor-joining method, using the program Mega 5.10. The bootstrap values greater than 0.51 are presented on the nodes. The first two letters in each of the serpins represent the acronym of scientific name for a given species. The GenBank accession number for sequences used in phylogenetic analysis in Fig. 5: AaTep (EAT39604.1, *Aedes aegypti*); AgTep (AAG00600.1, *Anopheles gambiae*); AmTepA, AmTepB, AmTep7 (XP\_397416.3, GB12605<sup>28</sup>, XP\_006565503.1, *Apis mellifera*); BbC3 (AB050668.1, *Branchiostoma belcheri*); BfCPAMD8 (XP\_002586872.1, *Branchiostoma floridae*), CiCPAMD8, CiC3 (XP\_009861615.2, NP\_001027684.1, *Ciona intestinalis*); CrC3 (AAQ08323.1, *Carcinoscorpius rotundicauda*); DmTep (NP\_523578.1, *Drosophila melanogaster*); DpTep (EFX86067.1, *Daphnia pulex*); EbCD109, EbC3 (BAD12264.1, P98094.1, *Eptatretus burgeri*); EsC3 (ACF04700.1, *Euprymna scolopes*); FcA2M (ABP97431.1, *Fenneropenaeus chinensis*); GgC3 (U16848.1, *Gallus gallus*); HaA2M, HaC3-1, HaTep (AB622470.1, AB622468.1, AB622471.1, *Hasarius adansoni*); HICD109 (AB481386.1, *Haliplanella lineata*); HsA2M, HsCD109, HsCPAMD8, HsC3, HsPZP (P01023.3, NP\_598000.2, NP\_056507.2, NP\_000055.2, CAA38255.1, *Homo sapiens*); IsTEP (XP\_002409560.1, *Ixodes scapularis*); LsA2M (BAA19844.1, *Limulus sp.*); MmC3 (NM\_009778.3, *Mus musculus*); Oma2M (AAN10129.1, *Ornithodoros moubata*) and TcPZP, TcCD109 (XP\_008195453.1, XP\_972838.2, *Tribolium castaneum*).

residue locates at approximately 100 amino acids downstream of the thioester motif, may influence the specific binding activity of the A2M<sup>27</sup>. However, Asp occupies this conserved site in PpSPI56. We also observed an excessive cysteine-rich C-terminal extension in PpSPI55 and corresponding gene in *N. vitripennis*.

**Developmental and sex-specific expression.** We analyzed the expression patterns of the *P. puparum* SPI genes in different developmental stages (embryo, larva, pupa and adult) and different tissues (ovary, venom gland and carcass without venom gland). The transcriptome data enable us to calculate FPKM values of the entire SPI genes (Supplementary Table S4). *PpSPI6*, *PpSPI14*, *PpSPI15*, *PpSPI32* and *PpSPI42* with very low FPKM values in

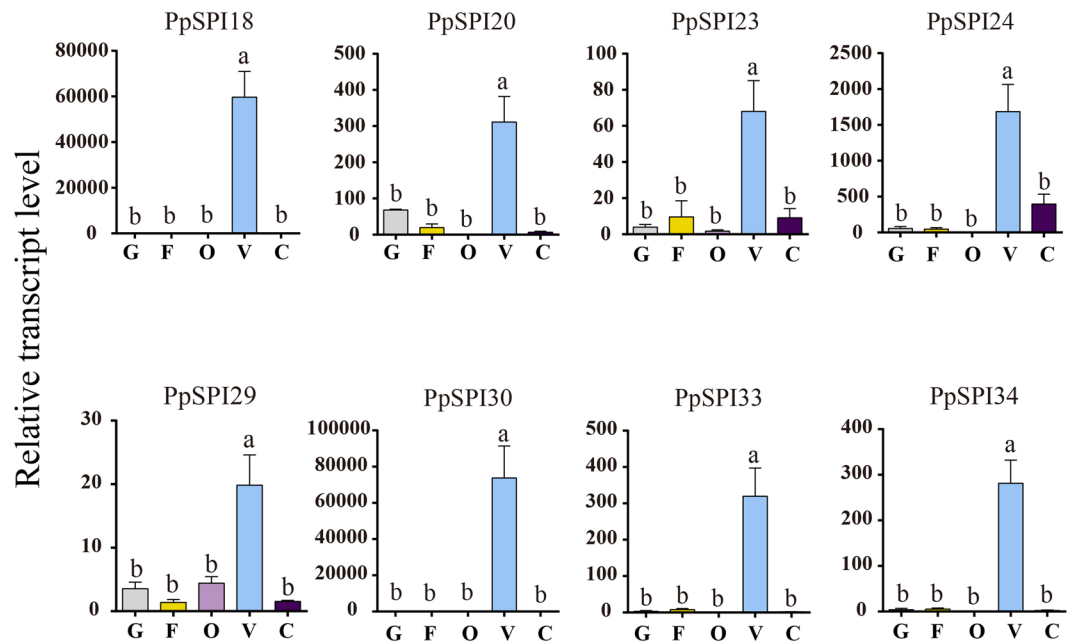


**Figure 6.** Expression profiles of *Pteromalus puparum* SPI genes across different developmental stages and tissues. Log<sub>2</sub> FPKM values for the SPIs are presented by bar colors where the darker red represent higher expression values, the darker green represent lower expression values.

all these existing databases may be pseudogenes. We retained these SPI genes in consideration of that some genes were only highly expressed in a specific tissue, but not or barely expressed in other tissues. The expression levels of 52 *PpSPI* genes (with FPKM values greater than 1) were profiled by all of nine transcriptomes (Fig. 6).

The results showed that *P. puparum* SPI genes exhibited stage-, sex- and tissue-specific expression patterns. The *PpSPI* gene expression patterns can be roughly divided into two groups. In the first group, 6 Kazals and 2 Pacifastins are most abundant in venom gland. The venom serves a decisive role in parasitism, and the expressions of SPI genes in venom gland are our major focus. We also used the same methods mentioned before<sup>38</sup> to





**Figure 7.** Tissue-specific expressions of *Pteromalus puparum* SPI genes. Total RNA was extracted from the gut (G), fat body (FB), ovary (O), venom (V) and carcass (C, the remaining body) of the adult female *P. puparum*, and used to analyze the expression patterns of these SPIs using qPCR. *P. puparum* 18 s rRNA was used as a housekeeping gene. Error bars represent the means  $\pm$  standard deviations from three biological replicates. A one-way ANOVA was used to determine the significant difference with different lowercase letter (a–c) ( $p < 0.05$ ).

define specifically expressed genes in the venom gland and the analysis yielded 8 venom-specific SPI genes, with *PpSPI16* replaced by *PpSPI23* (Supplementary Table S5). qPCR results of these venom gland specifically expressed genes are consistent with the transcriptome data (Fig. 7). Furthermore, four of eight SPIs also exist in the venom gland proteome and were considered as venom proteins (Supplementary Table S5). SPI genes in the second group form three clusters that were highly expressed in larval, pupae and adult stages, respectively. The SPI genes in the first cluster presented distinct expressions in larva stage, of which canonical SPI genes occupy 17 out of total 20 genes. Eleven SPI genes including all of three A2M-like genes showed relative high expressions in female and male pupae and belong to the second cluster. The last presented genes were mainly expressed in carcass, female adult and male adult. The expression profiles of SPI genes in different development stages were also verified by qPCR results, and the expression patterns were parallel to the transcriptome data (Supplementary Figure S6).

## Discussion

In the genome of *P. puparum*, total 57 SPI genes belonging to 7 families (serpin, Kazal, Pacifastin, TIL, Kunitz, WAP and A2M) were identified. The comprehensive analyses of SPIs have been conducted in two lepidopteran insects, with 80 SPI genes in *B. mori* and 61 in *P. xylostella*<sup>26,33</sup>. The number of *P. puparum* SPI genes is less than that in *B. mori* or *P. xylostella*. The differences in the numbers of canonical SPI families are also showed between *P. puparum* and lepidoptera insects. Four families of canonical SPI genes (WAP, WAP/Antistatin/Kunitz, amfpi and Bowman-Birk) identified in the lepidopteran insects are absent in *P. puparum* genome. SPI genes in *N. vitripennis* are precisely similar to *P. puparum* SPI genes in the quantity and the ratio (Supplementary Table S3). The analysis of scaffold location revealed that tandem repeats of genes with similar structure are observed in *P. puparum* TIL, Pacifastin and Kazal genes. This is also a common event identified in several species, not just restricted to SPI gene families<sup>39</sup>. The expansion of SPI sequences in one scaffold may be the result of gene duplication during quick evolution.

Besides the gene duplication, diversity of the SPIs can also be generated post transcriptionally or post translationally. For some insect serpins, alternate splicing of an exon that encoded the RCL regions gives rise to several serpins with different inhibitory activities. *MsSRPN1* is a good example for this mechanism, which contains 14 copies of its 9th exons for different carboxyl-terminal residues<sup>40,41</sup>. The serpin-1 from *B. mori* orthologous to *MsSRPN1*, contains four alternative splices of the exon 9<sup>27,42</sup>. Similar situation can also be identified in the hymenoptera wasps, *PpSPI1*<sup>43</sup> with 16 putative splicing isoforms sharing the first seven exons, but differing in the RCL encoded exon 8, resulting in diverse inhibitors. Previous studies have reported that posttranslational modification is common in Pacifastins family<sup>24</sup>. Members in Pacifastin families generally contain one signal peptide followed by several Pacifastin-like domains<sup>44</sup>. These Pacifastins could be divided into several smaller inhibitor peptides by putative dibasic cleavage sites. The cleavages of multi-domain Pacifastin inhibitors and the structural organizations are similar to the mechanisms in typical endocrine or neuronal peptide precursors, which could be broken into small peptides by endoproteases through pairs of basic amino acids<sup>45,46</sup>. Further investigation should be

used to detect the putative enzyme and the possible physiological role of the cleavage of these multiply-domain Pacifastins. Analyses of Pacifastin sequences in *P. puparum* showed that they can transport extracellularly and be divided into several smaller peptides by dibasic cleavage sites.

Serpins are essential components in immune response, with their role being regulatory control over Toll pathway<sup>18,47</sup> and PPO activation by inhibiting serine protease cascades<sup>16</sup>. Limited studies have also shown that serpins play a role in development and reproduction<sup>48,49</sup>. A total of 10 serpins were identified in the genome of *P. puparum*. Most of these serpins contain conserved beta-sheet and alpha-helix structures. PpSPI6 distinguished with other *P. puparum* serpins in sequences. The phylogenetic tree showed PpSPI6 forms a clade distant from other serpins. These results indicated PpSPI6 may not be a typical inhibitor. Genetic and biological studies could be used to test the biochemical roles of PpSPI6 in further research. The rest seven of nine *P. puparum* serpins may function as serine protease inhibitors, conformed to the sequence alignments and putative dibasic cleavage sites. PpSPI3 consists of two serpin domains, sharing similarity in hinge regions and cleaved sites with a twin-domain serpin TmSPN93. It was reported that TmSPN93 isolated from a large beetle *Tenebrio molitor*, also contains two complete, tandemly arrayed serpin domains. The first serpin domain of TmSPN93 could bind to Spätzle-processing enzyme to form a permanent protein complex whereas the second serpin domain forms a stable complex with Spätzle activating enzyme, both of which were involved in the Toll signaling cascade<sup>47</sup>. Although *P. puparum* and *T. molitor* have distant correlation between relatives, they display the same dual-domain organizations, and similar amino acids in hinge regions and P1 position (Fig. 1). Therefore, we inferred PpSPI3 may modulate relative serine proteinases in the Toll activation pathway as TmSPN93. PpSPI4 shares a high similarity with AmSRPN4 and AgSRPN6. Previous studies indicated that AgSRPN6 could inhibit bovine pancreatic trypsin *in vitro*<sup>50</sup>. Since AgSRPN6 and PpSPI4 share the same residue at P1 position (Arg), PpSPI4 may serve as a trypsin-inhibitor. The experiment also showed the expression profile of AgSRPN6 was significantly induced after infection. Knockdown of the expression level of AgSRPN6 by RNA interference delayed progression of parasite lysis as well as increases number of melanized parasites<sup>50,51</sup>. We examined the microbe-induced expressions of PpSPI4. qPCR showed that PpSPI4 can be highly induced *p. i.* PpSPI5 clusters with MsSRPN6. In *M. sexta*, MsSRPN6 can covalently link with MsPAP3 or MsHP-8. MALDI-TOF-MS showed that the cleavage site of MsSRPN6 is between Arg and Ser<sup>19</sup>. PpSPI5 and MsSRPN6 share the identical amino acids (Arg and Ser) at P1 position of the cleavage site and hinge region (Fig. 1), which further indicated the result that PpSPI5 may inactivate its cognate enzyme in the scissile bond between residues Arg and Ser. An induction pattern of MsSRPN6 was observed after bacterial or fungus infection. We also detected the expression profiles of PpSPI5. The expression levels of PpSPI5 increased strikingly after immune challenges. Besides, previous studies have demonstrated that AgSRPN6 and MsSRPN6 participate in regulating the PPO activation pathway<sup>19,51</sup>. Based on sequence alignment, phylogenetic tree and expression patterns, we inferred that PpSPI4 and PpSPI5 may also participate in the PPO activation pathway.

Apart from serpins, canonical SPIs with single or multiply domains were characterized. Among them, three SPIs are considered as mixed type inhibitors for containing non-inhibitor domains. PpSPI52 is a protein with unknown function. PpSPI54 is similar to papilin, an extracellular matrix glycoprotein involved in cell rearrangements of different matrices in embryo development and may modulate metalloproteases<sup>52</sup>. PpSPI51 and PpSPI53 are homologs to F-spondin. Previous studies revealed F-spondin serves as an extracellular matrix that attached molecule to promote neuritis commissural and inhibit the motor axons<sup>53</sup>.

There were three SPI genes containing the characteristic of A2M domains identified in *P. puparum* genome. We made sequences alignment and phylogenetic tree, which would be useful in better understanding of this family. The thioester domain GCGEQ and His residue 100 amino acids downstream of the thioester are conserved residues in most of the TEPs. However, DCGEQ and Asn replace these two conserved sites in PpSPI56. Besides, phylogenetic tree showed PpSPI56 was divided into a branch far away from other A2Ms. The differences in the potentially important sites and topology of the tree indicated the function of PpSPI56 divergent from other A2Ms<sup>54</sup>. PpSPI55 possesses both of the conserved motifs. An excessive cysteine-rich C-terminal extension was also observed in PpSPI55. The structure difference in C-terminal of the A2M sequences is also popular in ant, wasp and honey bee A2M. However, it is not observed in other insect A2Ms<sup>55</sup>. Based on the typical thioester motif and the catalytic His residue, PpSPI55 may show a characteristic biological function as a typical A2M, but the function of unusual C-terminal extensions is difficult to predict.

RNA-seq databases enable us to detect gene temporospatial expression patterns. Gene expression profiles showed that 22 SPI genes were highly expressed in the larval stage, occupying for nearly 40% of the whole SPI genes. The expression patterns were extremely parallel to 53 out of 133 serine protease genes previously found in *P. puparum*<sup>34</sup>. The wasps secreted massive proteinases in food digestion for its rapid growth. The sufficient SPIs are necessary in immobilizing excessive enzyme activities. Insights into the distribution of adult female tissues, 8 *P. puparum* SPI genes (6 Kazals and 2 Pacifastins) were differentially expressed in the venom gland. Among them, 2 Kazals and 2 Pacifastins were detected in both of the venom gland transcriptome and the venom proteome (Supplementary Table S5). The other four Kazals were significantly expressed in the venom gland as well. However, we cannot identify these proteins in the proteomic data. These Kazals are small peptides with the average of the molecular mass at 8.6 kDa (Supplementary Table S1). These small peptides may not be retained by SDS-PAGE used for proteomic identification, and therefore missed in the proteomic data. Molecular and biological researches such as western blot are necessary to validate whether or not these Kazals were secreted into the hemocoel of the host. It was reported that venom proteins were recruited from non-venom proteins through several ways, including gene duplication<sup>56</sup>, alternative splicing<sup>43</sup>, single copy gene<sup>57</sup> and lateral transferred gene<sup>58</sup>. Kazals and Pacifastins specifically expressed in venom gland may also be recruited from non-venom proteins through gene duplications and probably co-function in inhibiting host immune response.

All in all, the study provides us with detailed information of serine protease inhibitors in *P. puparum*, including gene structures, domain organizations and putative reactive sites. This information could be helpful in

discovering the potential functions of *P. puparum* SPIs, which are crucial for wasps to keep the homeostasis themselves related to reproduction, development and immune response, as well as to evade their host immune system.

## Materials and Methods

**Insect rearing.** *P. puparum* colony and its host *P. rapae* were reared under the cases of 14: 10h (light: dark) photoperiod at  $25 \pm 1$  °C<sup>31</sup>. Pupated *P. rapae* and mated female wasp of *P. puparum* were placed together in a plastic tube. Once oviposition was observed, the wasp was taken out of the tube. The parasitized pupae were then reared under the conditions mentioned above. Once the new wasp adults emerged from the parasitized *P. rapae* pupae, they were collected immediately, held in plastic finger-type containers and fed with 20% (v/v) honey solution.

**Identification of SP/SPH genes from *P. puparum* genome.** We downloaded SPI genes from the genome or NCBI GeneBank of *D. melanogaster*, *M. sexta*, *A. mellifera*, *A. gambian*, *T. castaneum* and *B. mori* (<https://www.ncbi.nlm.nih.gov/genome>). These genes were used as queries against *P. puparum* genome database using local BLAST program (E-value  $1e^{-5}$ ). Identified genes were validated manually using NCBI online BLASTP. The domain architectures of these SPI proteins were calculated by PROSITE (<http://prosite.expasy.org/>) and SMART (<http://smart.embl.de/>). Theoretical isoelectric point (pI) and molecular weight (Mw) of these SPI proteins were carried out by pI/Mw ([http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/)). We also annotated the SPI genes of *N. vitripennis* genome<sup>59</sup> in silico for comparative analysis.

**Sequence alignment and phylogenetic analysis.** *P. puparum* SPI sequences were aligned with SPIs with known function or structure from other species. ClustalX 2.0<sup>60</sup> was used in multiple sequence alignment. Phylogenetic tree was constructed using neighbor-joining method by MEGA5.10<sup>61</sup> with 1000 bootstrap replicates.

**Samples collection, cDNA synthesis and qPCR analysis.** All of the RNA samples used for cDNA synthesis were collected as previous described<sup>34</sup>. The different developmental stages (embryos, larvae, female and male pupae and adults), tissues (fat body, gut, ovary, venom gland and the remaining carcass of 2-day aged *P. puparum* female wasps) and immune challenged samples (PBS, Gram-positive bacterium *M. luteus*, Gram-negative bacterium *E. coli* and fungus *B. bassiana*) were obtained, washed, and then pooled into centrifuge tube with Trizol reagent (Invitrogen, USA), respectively. Any genomic DNA in these RNA samples was removed by adding RNase-free Dnase (Promega). The total RNA was obtained followed by the manufacturer's instructions. PrimeScript™ One Step RT-PCR Kit (Takara, Japan) was used in cDNAs synthesis. Primers for target and control genes (Supplementary Table S6) were designed on website Primer 3<sup>62</sup>. The qPCR was conducted in 25 µl reaction mixture followed the protocol of SYBR Green Supermix Kits (Takara, Japan) on BIO-RAD CFX96™ Real-Time System. The qPCR reactions were set up as follows: enzyme activation at 95 °C for 30 secs, followed by 40 cycles with denaturation at 95 °C for 5 sec, annealing at 60 °C for 34 sec. Dissociation curves were checked at the end of PCR reactions. The relative expression profiles of SPI genes in different immune challenge stages took *P. puparum* *actin 1* as an internal control. Developmental- and tissues-specific expressions were normalized to a reference gene (18s rRNA). We calculated the relative expression levels in accordance with the  $2^{-\Delta\Delta Ct}$  method<sup>63</sup>. Error bars represent the means  $\pm$  standard deviations from three biological replicates. The expression profiles of SPI genes in different development stages and tissues distributions were performed by one-way analysis of variance (ANOVA) and the expression profiles of SPI genes in immune responses were conducted by two-way ANOVA.

**Analyses of RNA-seq Data.** RNAseq libraries from 6 developmental stage samples and 3 tissue distribution samples were previously completed in our laboratory. Various life stages include newly laid *P. puparum* embryos, larvae, female and male wasps in pupal stage and adult stage. For tissue analyses ovary, venom gland and carcass without venom gland were prepared from female adults of the wasps. We calculated the FPKM values of *P. puparum* SPI genes (Supplementary Table S4). The expression profiling of SPI genes at different development stages were visualized using the R statistical program version 3.1.3. Differentially expressed analyses between venom gland and carcass were conducted by R package DEGSeq v1.2.2<sup>64</sup> (Supplementary Table S5). Benjamini & Hochberg method was used to adjust the *p*-values. Corrected *p*-value  $< 0.001$ ,  $\log_2$  (FPKM\_VG/FPKM\_Carcass)  $> 1$  and FPKM\_VG (Venom gland)  $> 10$  were set as the threshold.

## References

- Kanost, M. R. & Jiang, H. B. Clip-domain serine proteases as immune factors in insect hemolymph. *Curr. Opin. in Insect Sci.* **11**, 47–55 (2015).
- Neurath, H. proteolytic processing and physiological regulation. *Trends Biochem. Sci.* **14**, 268–271 (1989).
- Gubb, D., Sanz-Parra, A., Barcena, L., Troxler, L. & Fullaondo, A. Protease inhibitors and proteolytic signalling cascades in insects. *Biochimie* **92**, 1749–1759 (2010).
- Eleftherianos, I. & Revenis, C. Role and importance of phenoloxidase in insect hemostasis. *J. Innate Immun.* **3**, 28–33 (2011).
- Krowarsch, D., Cierpicki, T., Jelen, F. & Otlewski, J. Canonical protein inhibitors of serine proteases. *Cell. Mol. Life Sci.* **60**, 2427–2444 (2003).
- Otlewski, J., Krowarsch, D. & Apostoluk, W. Protein inhibitors of serine proteinases. *Acta Biochim. Pol.* **46**, 531–565 (1999).
- Rawlings, N. D., Tolle, D. P. & Barrett, A. J. Evolutionary families of peptidase inhibitors. *Biochem. J.* **378**, 705–716 (2004).
- Watanabe, R. M. O. *et al.* A novel trypsin Kazal-type inhibitor from *Aedes aegypti* with thrombin coagulant inhibitory activity. *Biochim.* **92**, 933–939 (2010).
- Li, Y., Qian, Y. Q., Ma, W. M. & Yang, W. J. Inhibition mechanism and the effects of structure on activity of male reproduction-related peptidase inhibitor Kazal-Type (MRPINK) of *Macrobrachium rosenbergii*. *Mar. Biotechnol.* **11**, 252–259 (2009).
- Kanost, M. R. Serine proteinase inhibitors in arthropod immunity. *Dev. Comp. Immunol.* **23**, 291–301 (1999).
- Gettins, P. G. W. Serpin structure, mechanism, and function. *Chem. Rev.* **102**, 4751–4803 (2002).
- Blandin, S. & Levashina, E. A. Thioester-containing proteins and insect immunity. *Mol. Immunol.* **40**, 903–908 (2004).
- Rehman, A. A., Ahsan, H. & Khan, F. H. Alpha-2-macroglobulin: A physiological guardian. *J. Cell. Physiol.* **228**, 1665–1675 (2013).

14. Blandin, S. *et al.* Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell* **116**, 661–670 (2004).
15. Husted, L. B. *et al.* Localization of carbohydrate attachment sites and disulfide bridges in *Limulus* alpha(2)-macroglobulin - Evidence for two forms differing primarily in their bait region sequences. *J. Biol. Chem.* **277**, 43698–43706 (2002).
16. Chu, Y. *et al.* *Ostrinia furnacalis* serpin-3 regulates melanization cascade by inhibiting a prophenoloxidase-activating protease. *Insect Biochem. Mol. Biol.* **61**, 53–61 (2015).
17. Tong, Y. R. & Kanost, M. R. *Manduca sexta* serpin-4 and serpin-5 inhibit the prophenol oxidase activation pathway. *J. Biol. Chem.* **280**, 14923–14931 (2005).
18. An, C. J., Ragan, E. J. & Kanost, M. R. Serpin-1 splicing isoform J inhibits the prSpätzle-activating proteinase HP8 to regulate expression of antimicrobial hemolymph proteins in *Manduca sexta*. *Dev. Comp. Immunol.* **5**, 135–141 (2011).
19. Zou, Z. & Jiang, H. B. *Manduca sexta* serpin-6 regulates immune serine proteinases PAP-3 and HP8 - cDNA cloning, protein expression, inhibition kinetics, and function elucidation. *J. Biol. Chem.* **280**, 14341–14348 (2005).
20. Ibelli, A. M. G. *et al.* A blood meal-induced *Ixodes scapularis* tick saliva serpin inhibits trypsin and thrombin, and interferes with platelet aggregation and blood clotting. *Int. J. Parasitol.* **44**, 369–379 (2014).
21. Guo, P. C. *et al.* Structural insights into the unique inhibitory mechanism of the silkworm protease inhibitor serpin18. *Sci. Rep.* **5**, 11863 (2015).
22. Horita, S. *et al.* Isolation, cDNA cloning, and structure-based functional characterization of Oryctin, a hemolymph protein from the coconut rhinoceros beetle, *Oryctes rhinoceros*, as a novel serine protease inhibitor. *J. Biol. Chem.* **285**, 30150–30158 (2010).
23. Spit, J. *et al.* Growth-inhibition effects of pacifastin-like peptides on a pest insect: The desert locust. *Schistocerca gregaria*. *Peptides* **34**, 251–257 (2012).
24. Breugelmanns, B., Simonet, G., van Hoef, V., van Soest, S. & Vanden Broeck, J. Identification, distribution and molecular evolution of the pacifastin gene family in Metazoa. *BMC Evolutionary Biology* **9**, 13pp-13pp. (2009).
25. Dai, S. X., Zhang, A. D. & Huang, J. F. Evolution, expansion and expression of the Kunitz/BPTI gene family associated with long-term blood feeding in *Ixodes Scapularis*. *BMC Evol. Biol.* **12**, 4 (2012).
26. Waterhouse, R. M. *et al.* Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* **316**, 1738–1743 (2007).
27. Garrett, M., Fullaondo, A., Troxler, L., Micklem, G. & Gubb, D. Identification and analysis of serpin-family genes by homology and synteny across the 12 sequenced *Drosophilid* genomes. *BMC Genomics* **10**, 489 (2009).
28. Evans, J. D. *et al.* Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect Mol. Biol.* **15**, 645–656 (2006).
29. Zou, Z., Lopez, D. L., Kanost, M. R., Evans, J. D. & Jiang, H. Comparative analysis of serine protease-related genes in the honey bee genome: possible involvement in embryonic development and innate immunity. *Insect Mol. Biol.* **15**, 603–614 (2006).
30. Zou, Z. *et al.* Comparative genomic analysis of the *Tribolium* immune system. *Genome Biol.* **8**, R177 (2007).
31. Zou, Z., Zhao, P. C., Hua, W., Mita, K. & Jiang, H. B. A comparative analysis of serpin genes in the silkworm genome. *Genomics* **93**, 367–375 (2009).
32. Lin, H. *et al.* Characterization and expression profiling of serine protease inhibitors in the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). *BMC Genomics* **18**, 162 (2017).
33. Zhao, P. *et al.* Genome-wide identification and immune response analysis of serine protease inhibitor genes in the silkworm, *Bombyx mori*. *PLoS One* **7**, e31168 (2012).
34. Yang, L. *et al.* The genomic and transcriptomic analyses of serine proteases and their homologs in an endoparasitoid. *Pteromalus puparum*. *Dev. Comp. Immunol.* **77**, 56–68 (2017).
35. Li, J. P. *et al.* The structure of active serpin 1K from *Manduca sexta*. *Struct. Fold. Des.* **7**, 103–109 (1999).
36. Elliott, P. R., Abrahams, J. P. & Lomas, D. A. Wild-type alpha(1)-antitrypsin is in the canonical inhibitory conformation. *J. Mol. Biol.* **275**, 419–425 (1998).
37. Wu, C., Noonin, C., Jiravanichpaisal, P., Soderhall, I. & Soderhall, K. An insect TEP in a crustacean is specific for cuticular tissues and involved in intestinal defense. *Insect Biochem. Mol. Biol.* **42**, 71–80 (2012).
38. Yan, Z. C. *et al.* Insights into the venom composition and evolution of an endoparasitoid wasp by combining proteomic and transcriptomic analyses. *Sci. Rep.* **6**, 19604 (2016).
39. Zou, Z., Zhao, P., Hua, W., Mita, K. & Jiang, H. A comparative analysis of serpin genes in the silkworm genome. *Genomics* **93**, 367–375 (2009).
40. Meekins, D. A., Kanost, M. R. & Michel, K. Serpins in arthropod biology. *Semin. Cell Dev. Biol.* **62**, 105–119 (2017).
41. Jiang, H. B. *et al.* Organization of serpin gene-1 from *Manduca sexta* - Evolution of a family of alternate exons encoding the reactive site loop. *J. Biol. Chem.* **271**, 28017–28023 (1996).
42. Liu, H. F. *et al.* Alternative splicing of the antitrypsin gene in the silkworm. *Bombyx mori*. *Mol. Biol. Rep.* **38**, 2793–2799 (2011).
43. Yan, Z. C. *et al.* A venom serpin splicing isoform of the endoparasitoid wasp *Pteromalus puparum* suppresses host Prophenoloxidase cascade by forming complexes with host hemolymph proteinases. *J. Biol. Chem.* **292**, 1038–1051 (2017).
44. Simonet, G. *et al.* Characterization of two novel pacifastin-like peptide precursor isoforms in the desert locust (*Schistocerca gregaria*): cDNA cloning, functional analysis and real-time RT-PCR gene expression studies. *Biochem. J.* **388**, 281–289 (2005).
45. Simonet, G., Claeys, I., Franssens, V., De Loof, A. & Vanden Broeck, J. Genomics, evolution and biological functions of the pacifastin peptide family: a conserved serine protease inhibitor family in arthropods. *Peptides* **24**, 1633–1644 (2003).
46. Xu, G. *et al.* Identification and expression profiles of neuropeptides and their G protein-coupled receptors in the rice stem borer *Chilo suppressalis*. *Sci. Rep.* **6**, 28976 (2016).
47. Jiang, R. *et al.* 93-kDa twin-domain serine protease inhibitor (Serpin) has a regulatory function on the beetle Toll proteolytic signaling cascade. *J. Biol. Chem.* **286**, 35087–35095 (2011).
48. Ligoxygakis, P., Roth, S. & Reichhart, J. M. A serpin regulates dorsal-ventral axis formation in the *Drosophila* embryo. *Curr. Biol.* **13**, 2097–2102 (2003).
49. Coleman, S., Drahn, B., Petersen, G., Stolorov, J. & Kraus, K. A *Drosophila* male accessory-gland protein that is a member of the serpin superfamily of proteinase-inhibitors is transferred to females during mating. *Insect Biochem. Mol. Biol.* **25**, 203–207 (1995).
50. An, C. *et al.* Biochemical Characterization of *Anopheles gambiae* SRPN6, a malaria parasite invasion marker in mosquitoes. *PLoS One* **7**, e48689 (2012).
51. Abraham, E. G. *et al.* An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. *Proc. Natl. Acad. Sci. USA* **102**, 16327–16332 (2005).
52. Kramerova, I. A., Kramerov, A. A. & Fessler, J. H. Alternative splicing of papilin and the diversity of *Drosophila* extracellular matrix during embryonic morphogenesis. *Dev. Dyn.* **226**, 634–642 (2003).
53. Tzarfaty-Majar, V., Burstyn-Cohen, T. & Klar, A. F-spondin is a contact-repellent molecule for embryonic motor neurons. *Proc. Natl. Acad. Sci. USA* **98**, 4722–4727 (2001).
54. Vanleuven, F. Human alpha-2-macroglobulin - structure and function. *Trends Biochem. Sci.* **7**, 185–187 (1982).
55. Sekiguchi, R. & Nonaka, M. Evolution of the thioester-containing protein genes in the arthropoda. *Immunobiology* **217**, 1168–1169 (2012).
56. Colinet, D. *et al.* Identification of the main venom protein components of *Aphidius ervi*, a parasitoid wasp of the aphid model *Acyrtosiphon pisum*. *BMC Genomics* **15**, 342 (2014).

57. Martinson, E. O., Mrinalini, Kelkar, Y. D., Chang, C. H. & Werren, J. H. The evolution of venom by co-option of single-copy genes. *Curr. Biol.* **27**, 647–649 (2017).
58. Martinson, E. O., Martinson, V. G., Edwards, R., Mrinalini & Werren, J. H. Laterally transferred gene recruited as a venom in parasitoid wasps. *Mol. Biol. Evol.* **33**, 1042–1052 (2016).
59. Werren, J. H. *et al.* Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**, 343–348 (2010).
60. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
61. Tamura, K. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
62. Untergasser, A. *et al.* Primer3-new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
63. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_T}$  method. *Methods* **25**, 402–408 (2001).
64. Wang, L. K., Feng, Z. X., Wang, X., Wang, X. W. & Zhang, X. G. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* **26**, 136–138 (2010).

## Acknowledgements

The study is supported by grants from Major International (Regional) Joint Research Project of National Natural Science Foundation (Grant no. 31620103915, <http://www.nsf.gov.cn/>), National Natural Science Foundation of China (Grant no. 31472038, 31401804, 31272098, <http://www.nsf.gov.cn/>), the Program for Chinese Outstanding Talents in Agricultural Scientific Research, the Program for Chinese Innovation Team in Key Areas of Science and Technology (2016RA4008), and the Fundamental Research Funds for the Central Universities (2017FZA6013).

## Author Contributions

L.Y. and G.Y. conceived the assay design; L.Y., Y.M. and J.W. enforced the gene expression work; L.Y., Y.M., Q.F., Z.Y. and Z.L. dealt with the data; and L.Y., Q.S., and G.Y. prepared the first draft of the article. All authors contributed to the final paper.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-16000-5>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017