Sociobiology 66(2): 348-357 (June, 2019) DOI: 10.13102/sociobiology.v66i2.4239

Sociobiology
An international journal on social insects

RESEARCH ARTICLE - BEES

# Preliminary analysis of PGRP-LC gene and structure characteristics in bumblebees

M Chen[1], N Ye[1], Y Liu[2], J An[2]

1 - Institute of Pharmaceutical Biotechnology and Engineering, College of Biological Science and Technology, Fuzhou University, Fuzhou, Fujian, China

2 - Key Laboratory of Insect-Pollinator Biology of the Ministry of Agriculture, Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, China

**Article History**

**Edited by**
Marco Costa, UESC, Brazil
Received 19 November 2018
Initial acceptance 02 April 2019
Final acceptance 29 May 2019
Publication date 20 August 2019

**Keywords**
Bumblebees, PGRP-LC, phylogenetic tree, purification.

**Corresponding author**
Yanjie Liu
Institute of Apicultural Research
Chinese Academy of Agricultural Sciences, Beijing, China.
E-Mail: liuyanjie@caas.cn

**Abstract**

Peptidoglycan recognize protein LC (PGRP-LC) is a significant pattern recognition receptor of the insect innate immune system that can recognize peptidoglycans and activate immune signaling pathways regulating the expression and release of antimicrobial peptides against infection. We for the first time analyzed the phylogenetic tree, purification and structure of bumblebee PGRP-LC. The results showed high conservation of bumblebee PGRP-LC among the 16 bumblebee species, and further phylogenetic analysis showed that the PGRP-LC phylogeny of different subgenera (*Subterraneobombus*, *Megabombus*, *Melanobombus*, *Bombus*) is consistent with that of the COI gene. Additionally, the phylogeny of PGRP-LCs among *Bombus*, *Apis* and the solitary bee *Megachile rotundata* coincides with the sociality evolution of bees. Moreover, bumblebee PGRP-LC (Bl-PGRP-LC) shares the *Drosophila* PGRP-LCx and PGRP-LCa topology, retaining conserved disulfide bonds and 80% binding residues involved in the interaction between tracheal cytotoxin (TCT) and PGRP-LCx. Therefore, Bl-PGRP-LC might share some similar binding characteristics with *Drosophila* PGRP-LCx. In addition, Bl-PGRP-LC has shorter β5 and β1 sheets, longer β2, β3, and β4 sheets and a shallow binding groove. To determine the characteristics of Bl-PGRP-LC, high-purity PGRP-LC inclusion bodies, soluble GST-tag Bl-PGRP-LC fusion protein and soluble pure Bl-PGRP-LC were obtained in vitro. The results will be helpful for further study of the function and structure of Bl-PGRP-LC.

## Introduction

Bumblebees are important pollinators and play a key role in agricultural and natural ecosystems (An et al., 2014; P Williams, Colla, & Xie, 2009). However, bumblebees are facing various threats, such as climate change, pesticides and pathogens, which result in the decline of bumblebee populations in many areas of the world (Goulson, Lye, & Darvill, 2008; Meeus, Brown, De Graaf, & Smagghe, 2011). Moreover, the development of commercial bumblebees is enhancing the spread of pathogens (Meeus et al., 2011). Thus, research on closely related immune responses to pathogens is necessary. For insects, resistance to pathogens occurs through innate immunity, and responses to pathogens are mainly triggered by insect pattern recognition receptors (PRRs) recognition of pathogen-associated molecular pattern (PAMPs), such as lipopolysaccharide, lipteichoic acid, peptidoglycans (PGNs), and β-1,3- dextran, which are produced by most pathogenic microorganisms (Yu, Zhu, Ma, Fabrick, & Kanost, 2002). PGNs are among the most important constituents of gram-negative and -positive bacterial cell walls and can be divided into Dap-type PGN and Lys-type PGN according to differences in the third amino acid of the short peptide (Schleifer & Kandler, 1972). Dap-type PGN is largely found in all gram-negative bacteria and some gram-positive bacteria, e.g., *Bacillus* and *Clostridium* spp., whereas Lys-type PGN is mainly found in gram-positive bacteria and is absent in gram-negative bacteria (Kurata, 2014).

PGNs are mainly recognized by peptidoglycan recognition proteins (PGRPs), which activate the expression and release of antimicrobial peptides through two main immune pathways: the Toll pathway and the Imd pathway (Steiner, 2004).

As important PRRs, PGRPs were first isolated from *Bombyx mori* (Yoshida, Kinoshita & Ashida, 1996), and PGRP genes were later found to be widely distributed among various animals, except for nematodes, but are absent in plants (Kang, Liu, Lundstrom, Gelius & Steiner, 1998; Royet, Gupta & Dziarski, 2011). Four PGRPs, PGLYRP1, PGLYRP2, PGLYRP3 and PGLYRP4, exist in mammals, play a role in bactericidal activity against both gram-positive and gram-negative bacteria, and share a functional domain with N-acetylmuramoyl-l-alanine amidase (Kang et al., 1998; Liu, Xu, Gupta & Dziarski, 2001; Lu et al., 2006; Wang et al., 2007). In contrast to mammals, insects have more PGRP genes, which can be divided into long (PGRP-L) and short (PGRP-S) types based on the size of the transcript (Werner et al., 2000). For example, the *Drosophila* genome contains 13 PGRP genes, including 7 PGRP-Ss (PGRP-SA, PGRP-SB, PGRP-SC (A/b), PGRP-SC2, PGRP-SD) and 6 PGRP-Ls (PGRP-LC, PGRP-LF, PGRP-LE, PGRP-LD, PGRP-LA, PGRP-LB) (Kurata, 2014; Werner et al., 2000). PGRP-SA and PGRP-SD bind to Lys-type PGN and Dap-type PGN, respectively, and mediate activation of the Toll pathway (Bischoff et al., 2004; Michel, Reichhart, Hoffmann & Royet, 2001). Conversely, PGRP-LB, PGRP-LA, PGRP-LE and PGRP-LC activate and regulate the Imd pathway (Gendrin et al., 2013; Goto et al., 2010; Zaidman-Remy et al., 2006). *Drosophila* PGRP-LC is expressed as three types of splicing proteins (PGRP-LC a/x/y) with the same intracellular domain but with different extracellular domains (Choe, Werner, Stoven, Hultmark & Anderson, 2002; Royet & Dziarski, 2007; Werner, Borge-Renberg, Mellroth, Steiner & Hultmark, 2003). PGRP-LCx mainly recognizes Dap-PGN and activates the Imd pathway with the assistance of PGRP-LCa, and the PGRP-LCx-LCa-TCT (tracheal cytotoxic) complex identifies assisted residues and contact residues (Chang, Chelliah, Borek, Mengin-Lecreulx & Deisenhofer, 2006; Chang et al., 2005). However, there is no research to date on PGRP-LC from other insects. In this study, the characteristics of the mature *Bombus* PGRP-LC protein were determined, and phylogenetic analysis was performed. In addition, pure *B. lantschouensis* PGRP-LC (Bl-PGRP-LC) inclusion bodies, soluble GST-Bl-PGRP-LC fusion protein and soluble Bl-PGRP-LC protein were obtained in vitro. The results provide an important basis for further research on the function and structure of bumblebee PGRP-LC.

## Materials and Methods

### *Bumblebee PGRP-LC cloning and genome schematic diagram*

Samples of worker bees of 16 bumblebee species (including *Bombus impetuosus*, *B. melanurus*, *B. difficillimus, B. lemniscatus*, *B. ladakhensis*, *B. friseanus*, *B. keriensis*, *B. pyrosoma*, *B. koreanus*, *B. ignitus*, *B. lantschouensis*, *B. longipennis*, *B. consobrinu*, *B. picipes, B. patagiatus* and *B. kashmirensis*) were collected in northern China in summer between 2016 and 2018. All samples were collected and stored at -80 °C for RNA extraction. Total RNA was isolated from the 16 species workers using TRIzol (Invitrogen, USA) according to the manufacturer's instructions. cDNA was synthesized according to the High-capacity cDNA Reverse Transcription kit (Thermo Fisher Scientific Inc., MA, USA). *PGRP-LC* genes were amplified using 1 μL cDNA, 1 μL primers (cloning primers and expression primers), 25 μL PCR enzyme and 23 μL nuclease-free water under the following conditions: step 1, 94 °C for 3 min; step 2, 94 °C for 30 sec, 55 °C for 30 sec, and 72 °C for 1 min (32 cycle); step 3, 72 °C for 7 min. *PGRP-LC* genes were aligned to the genome of the European bumblebee *Bombus terrestris* to identify genome structure. In addition, 3'RACE was performed to detected the spliceosome of *PGRP-LC* using the 3'-Full RACE Core Set with PrimeScript™ RTase kit (Takara Bio Inc.). The forward primer (RACE-F) was designed according to the *PGRP-LC* genes of the America bumblebee *B. impatiens* and the European bumblebee *B. terrestris* (GenBank NO. XP_012238222.1 and XP_020719583.1) (Table 1).

Cloning primers were designed according to the sequence of PGRP-LC from the 3'RACE results (Table 1). Primers for analyzing expression were designed by adding restriction enzymes for different expression vectors (*EcoR* I, *Xho* I for pET21a and *BamH* I, *Xho* I for pGEX-6p-1) (Table 1). PCR products were digested and ligated into the cloning vector pMD18-T and expression vectors pET21a and pGEX-6P-1 using T4 DNA ligase at 16 °C. The ligated constructs were transformed into *Escherichia coli* strain DH5α and cultured overnight at 37 °C with LB medium. These plasmids were extracted using Axygen Plasmid Kit (Axygen Scientific Inc., CA, USA).

### *Sequence alignment and phylogenetic analysis*

Alignment of bumblebee PGRP-LC amino acid (AA) sequences was performed with the Clustal Omega server (https://www.ebi.ac.uk/Tools/msa/clustalo/) and Jalview Desktop (http://www.jalview.org/). Phylogenetic analysis of PGRP-LC AA sequences from *Megachile rotundata*, *Eufriesea mexiacana*, *Dufourea novaeangliae*, *Apis mellifera*, *Apis cerana*, *Apis florea*, *Apis dorsata* and the 16 bumblebee species from seven subgenera was performed using the maximum likelihood method based on the Whelan and Goldman model (Whelan & Goldman, 2001) and MEGA version 7.0.26 (Kumar, Stecher, & Tamura, 2016) with a JTT+G model.

### *Preparation of PGRP-LC protein, analysis of hydrophobic properties and structural modeling*

In this study, the *B. lantschouensis* PGRP-LC (Bl-PGRP-LC) gene was used to construct expression plasmids (pET21a/PGRP-LC and pEGX-6p-1/PGRP-LC). These expression plasmids were transformed into *E. coli* strain Transseta (DE3), and expression in ampicillin LB medium at 37 °C and 16 °C

**Table 1.** Primers used in this study.

| Primer name | Primer sequences | Purpose |
|---|---|---|
| RACE-F | ATGGTGACATTGGTGCAACAGC | 3'RACE |
| 3'RACE Outer Primer | TACCGTCGTTCCACTAGTGATTT | |
| PGRP-LC-F | ATGGTGACATTGGTGCAACAGC | Cloning primer |
| PGRP-LC-R | TTATGGCGATGGTGCCAATGGG | |
| LC-EF1 | CGGAATTCACTACCACGACCTCTGTGGTTTTTC | Expression primers |
| LC-ER1 | CCGCTCGAGTTATGGTGACCAATGGG | |
| LC-EF2 | CGGGATCCACTACCACGACCTCTGTGGTTTTTC | |

(for pEGX-6p-1/PGRP-LC) was induced by 0.1, 0.2, 0.4, 0.6, 0.8, or 1 mM IPTG at OD = 0.4, 0.6, 0.8, or 1.0. The bacteria were harvested by centrifugation at 8000×g for 3 min at 4 °C and then resuspended in cold phosphate-buffered saline (PBS). After sonication, the samples were centrifuged at 11,000 ×g, and the supernatants and pellets were collected for SDS-PAGE analysis. Inclusion bodies were washed three times with a solution consisting of 20 mM Tris (pH 8.0), 100 mM NaCl, 1 mM EDTA, 1 mM DTT and 0.5% Triton X-100. The supernatants were filtered through a 0.22-μm filter and then purified on a GST column using binding buffer (140 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$, pH 7.3) and elution buffer (50 mM Tris (pH 8.0), 10 mM reduce glutathione) and a Superdex 200 10/300 increase column with 50 mM Tris (pH 8.0)/150 mM NaCl. Purified proteins were collected and digested with the PSP enzyme and analyzed by SDS-PAGE. Protein functional domains and hydrophobic properties were predicted by SMART (http://smart.embl-heidelberg.de/) and HCA (http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=HCA#forms::HCA). The structure of Bl-PGRP-LC was modeled by SWISS-MODEL (https://swissmodel.expasy.org/) and evaluated by the GMEQ and QMEAN values (Biasini et al., 2014; Bienert et al., 2017; Bordoli et al., 2009; Guex, Peitsch, & Schwede, 2009; Waterhouse et al., 2018) with template (PDB ID: 2f2l). Structural analysis was performed using PyMOL (DeLano Scientific LLC).

## Results

### *Bumblebee PGRP-LC sequence alignment and phylogenetic analysis*

*PGPR-LC* genes encoding the complete mature peptide for the 16 bumblebee species were cloned from cDNA, and an expected length of 1254 bp was observed by agarose gel electrophoresis (Fig 1A). Based on alignment to the *B. terrestris* and *B. impatients* genomes, *PGRP-LC* consists of four exons (Exon 1: 6645776 nt-6646784 nt, Exon 2: 6647933 nt-6648518 nt, Exon 3: 6649718 nt-6649876 nt, Exon 4: 6650035 nt-6650667 nt), with a length of 2742 bp (Fig 1B). The 3'RACE of *B. lanschouensis PGRP-LC* showed only one pure band based on the template RNA (Fig 1C). This band was extracted and cloned into T vectors followed by sequencing, and the results showed that the sequences were same between different clones (Fig S1; see supplemmentary material - doi: 10.13102/sociobiology.v66i2.4239.s2269). AA alignment of the bumblebee PGRP-LCs revealed high conservation, with identity above 90% (Fig 2). However, diverse non-conserved sites distributed on different domains, including the intracellular domain, transmembrane domain and PGRP domain, consisting of α helix, β sheets and loops, were found among the different species (Fig 2 and S2A). Additionally, bumblebees from the same subgenus share fewer diverse sites compared with those from different subgenera. Phylogenetic tree analysis showed that bumblebees from the same subgenus cluster into the same clade and that different subgenera group into different clades, with the following evolutionary order of different bumblebee species PGRP-LCs: *Subterraneobombus*, *Megabombus*, *Melanobombus*, *Bombus* (Fig 3). Bumblebees and honeybees were distributed in the two parallel clades, and compared to honeybees and bumblebees, *Megachile rotundata* and *Dufourea novaeangliae* were located on a more distant clade (Fig 3).

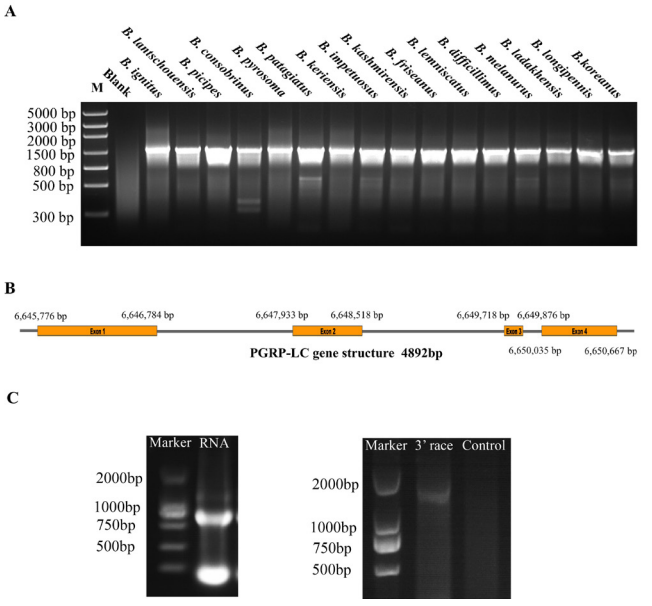


**Fig 1**. Cloning and genome schematic diagram of *PGRP-LC* gene in bumblebees. A. Cloning of the *PGRP-LC* gene from 16 bumblebee species with template cDNA. B. Schematic diagram of bumblebee *PGRP-LC* genes based on alignments to the *B. terriestris* and *B. impatients* genomes; exons are marked by orange bars, introns are marked by a black line, and the number represents the position in the genome. C. *B. lantschouensis* RNA extraction and 3'RACE results of the *PGRP-LC* gene with template *B. lantschouensis* RNA.

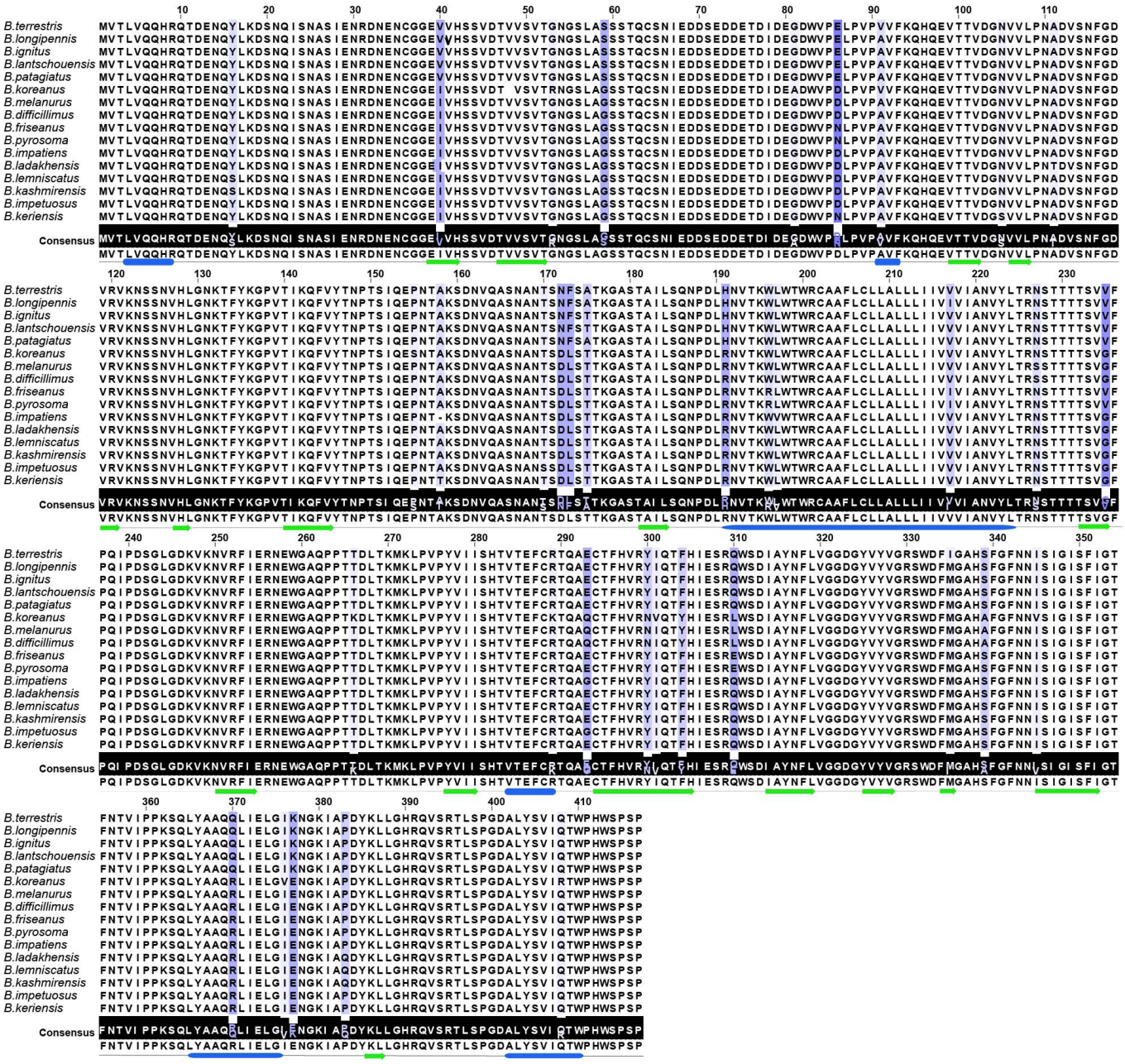


**Fig 2**. Amino acid alignment of the PGRP-LC protein from 16 bumblebee species. GenBank accession numbers are as follows: *B. terrestris,* NO.XP_020719583.1; *B. impatiens*, NO.XP_012238222.1; *B. impetuosus,* NO.MK061352; *B. melanurus*, NO.MK061347; *B. difficillimus,* NO.MK061348; *B. lemniscatus,* NO.MK061349; *B. ladakhensis,* NO.MK061346; *B. friseanus,* NO.MK061350; *B. keriensis,* NO.MK061353; *B. pyrosoma,* NO.MK061355; *B. koreanus,* NO.MK061343; *B. ignitus,* NO.MK061342; *B. lantschouensis*, NO.MK061344; *B. longipennis,* NO.MK061345; *B. patagiatus,* NO.MK061354; *B. kashmirensis* NO.MK061351. α helices, β sheets and continuous loops are marked by blue bars, green arrows and gray lines respectively.

*Expression and purification of the PGRP-LC protein of Bombus lantschouensis*

The PGRP domain of the mature PGRP-LC protein of *B. lantschouensis* (Bl-PGRP-LC) was expressed in *E. coli*. After induction, pET21a/Bl-PGRP-LC was found in inclusion bodies at OD 0.4, 0.6, 0.8, and 1.0 and 0.4 mM, 0.6 mM, 0.8 mM, and 1.0 mM IPTG. A clearly defined band appeared at a predicted mass weight of approximately 21 kDa after expression at 37 °C (Fig 4A). The yield of protein purified from inclusion bodies was 50% (Fig 4B). Property analysis showed that Bl-PGRP-LC contains several hydrophobic clusters, which might prevent the correct refolding of Bl-PGRP-LC and expression of a soluble protein (Fig S2B). pGEX-6p-1/Bl-PGRP-LC was also induced using various conditions and found in inclusion bodies, except for OD 0.8 and 0.6 mM IPTG at 16 °C (Fig 5A). The mass weight was identical to the predicted 48 kDa. The lysate supernatant was purified using

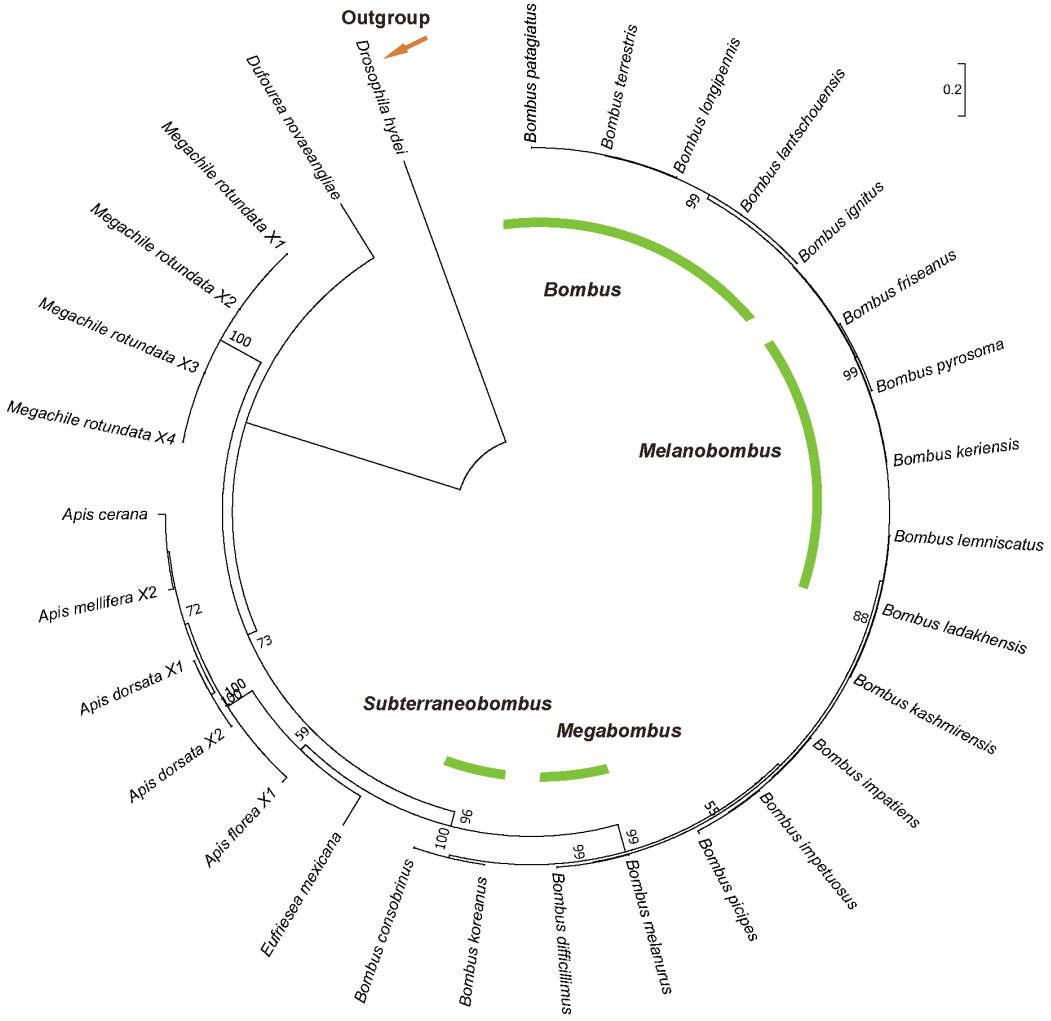


**Fig 3**. Phylogenetic analysis of PGRP-LC from *Bombus* and other species, including *Apis, Megachile* and *Drosophila* (outgroup), reconstructed based on amino acids using the neighbor-joining method in MEGA7. The phylogeny assessment was carried out using a bootstrap analysis of 1000 replications, and the bootstrap values are shown for each node of the tree. GenBank accession numbers are as follows: *Eufresea Mexicana*, XP_017759657.1; *Apis florea* X1, XP_003693124.1; *Apis florea* X2, XP_012341056.1; *Apis dorsata* X1, XP_006618726.1; *Apis dorsata* X2, XP_006618727.1; *Apis cerana*, XP_016914774.1; *Apis mellifera* X1, XP_392452.2; *Apis mellifera* X2, XP_026297747.1; *Megachile rotundata* X1, XP_012138565.1; *Megachile rotundata* X2, XP_012138566.1; *Megachile rotundata* X3, XP_012138567.1; *Megachile rotundata* X4, XP_003702406.2; *Dufourea novaeangliae*, XP_015429839.1; *Drosophila hydei*, XP_023178744.1.

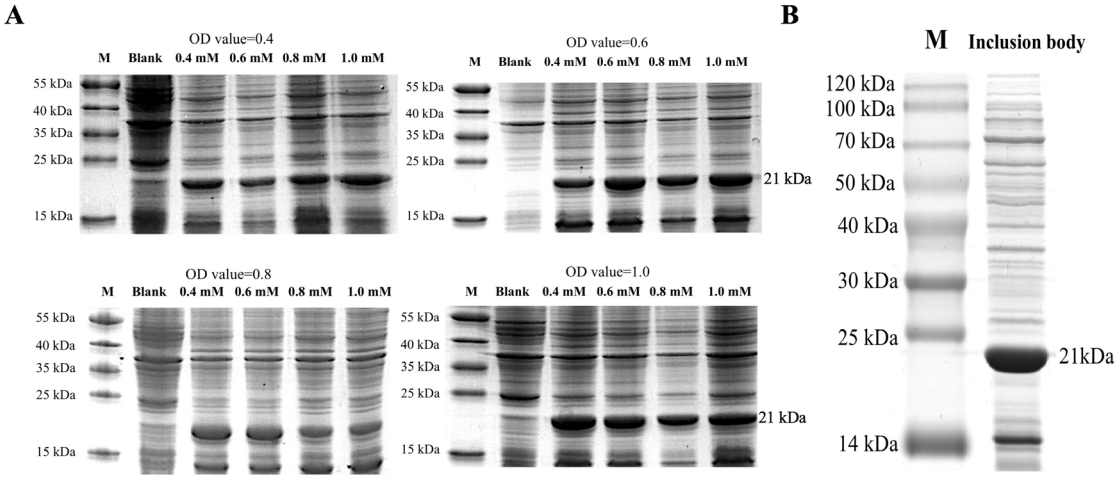


**Fig 4.** Optimized expression and purification of pET21a/Bl-PGRP-LC. A. Optimization of the expression conditions for pET21a/Bl-PGRP-LC, including OD values and IPTG concentrations at 37 °C for 5 h. M: Protein molecular weight marker; OD values: 0.4, 0.6, 0.8, 1.0; IPTG concentrations: 0.4 mM, 0.6 mM, 0.8 mM, 1.0 mM. B. Purification of Bl-PGRP-LC inclusion bodies; M: protein molecular weight marker.

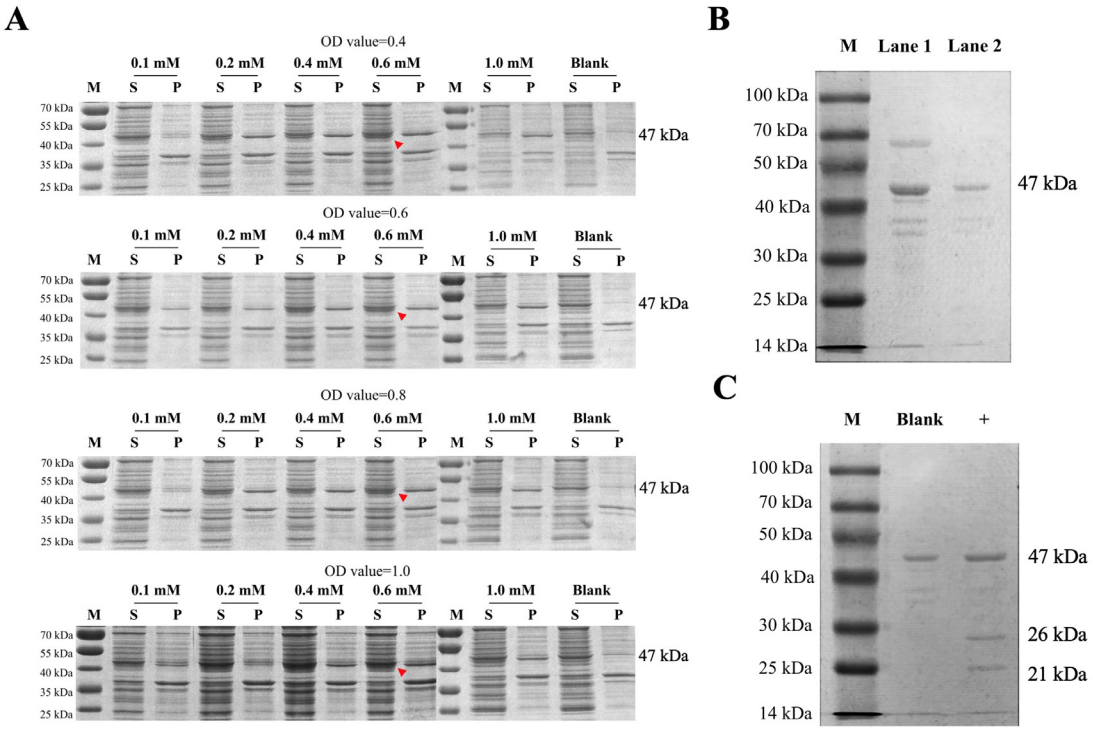


**Fig 5**. Sequence alignment and structural overlap among Bl-PGRP-LC, Dm-PGRP-LCx and Dm-PGRP-LCa. A. Optimization of the expression conditions for pGEX-6p-1/Bl-PGRP-LC, including OD values and IPTG concentrations at 16 °C for 4 h. M: Protein molecular weight marker; S: supernatant; P: pellet; OD values: 0.4, 0.6, 0.8, 1.0; IPTG concentrations: 0, 0.1, 0.2, 0.4, 0.6, 1.0 mM; the target proteins are marked by a red triangle. B. Purification of the GST-Bl-PGRP-LC fusion protein and Bl-PGRP-LC protein. M: Protein molecular weight marker. Lane 1: Collected supernatants were purified using a GST column; Lane 2: GST-Bl-PGRP-LC purified using a Superdex200 10/300 increase column after the GST column. C. Cleavage of the GST-tag from GST-Bl-PGRP-LC using the PSP enzyme (kept in our lab), +: add PSP enzyme, Blank: without PSP enzyme.

a GST column followed by the Superdex200 10/300 increase column, and a pure band was observed by SDS-PAGE (Fig 5B). Furthermore, the GST-tag was successfully cleaved by the PSP enzyme (purified by our lab), and pure GST-tag and Bl-PGRP-LC bands were visualized by SDS-PAGE (Fig 5C).

*Structural characteristics of Bl-PGRP-LC*

The structure of Bl-PGRP-LC was modeled with a reliable value (GMEQ=0.32 and QMEAN=-0.42). Bl-PGRP-LC exhibits a conserved overall structure with that of *Drosophila* PGRP-LCx and PGRP-LCa and consists of four α helices, α1, α2, α3, and α4, five sheets and loops, β1, β2, β3, β4, and β5 (Fig 6A). Bl-PGRP-LC retains conserved cysteines and shares 46.78% and 43.27% identity with *Drosophila* PGRP-LCx and PGRP-LCa, respectively (Fig 6A and B). Moreover, nine of the eleven contact residues of *Drosophila* PGRP-LCx involved in the binding of TCT are conserved in Bl-PGRP-LC (Fig 6B and C).

## Discussion

PGRPs are the key pattern recognition receptors in the insect innate immune system and are divided into two types: PGRP-S and PGRP-L (Dziarski & Gupta, 2005, 2006). As a PGRP-L protein, PGRP-LC of *Drosophila* has been shown to bind Dap-type PGN and trigger the Imd immune pathway (Hoffmann, 2003; Kaneko et al., 2006; Takehana et al., 2004). Structural analysis of the *Drosophila* TCT/PGRP-LCx-LCa complex has revealed the AA of PGRP-LCx contributing to the binding of TCT as well as the assist binding residues of PGRP-LCa (Chang et al., 2006). Interestingly, bees, including bumblebees and honeybees, share limited immune genes, including PGRPs, compared to *Drosophila* and mosquitos (Barribeau et al., 2015; Evans et al., 2006; Honeybee Genome Sequencing, 2006; Sadd et al., 2015). The 3'RACE results of *B. lantschouensis* PGRP-LC proved that only one type of PGRP-LC existed in bumblebees, which differed from the three splicing proteins (PGRP-LCx, a, y) of *Drosophila*, consisting of four exons (Fig 1C and S1). The length of the gene is 2742 bp and produces a transcript of 1254 bp (Fig 1B) (Sadd et al., 2015). In the current study, *PGRP-LC* genes from 16 bumblebee species of seven subgenera were examined for phylogenetic relationships (Fig 1A). The alignment showed that PGRP-LCs share high identity among bumblebees and that species belonging to the same subgenus exhibit higher identity, with several non-conserved sites mainly located on the intracellular domain and PGRP domain in different species (Fig 2). Further phylogenetic analysis showed that the evolutionary relationship of PGRP-LCs from different subgenera (*Subterraneobombus*, *Megabombus*,

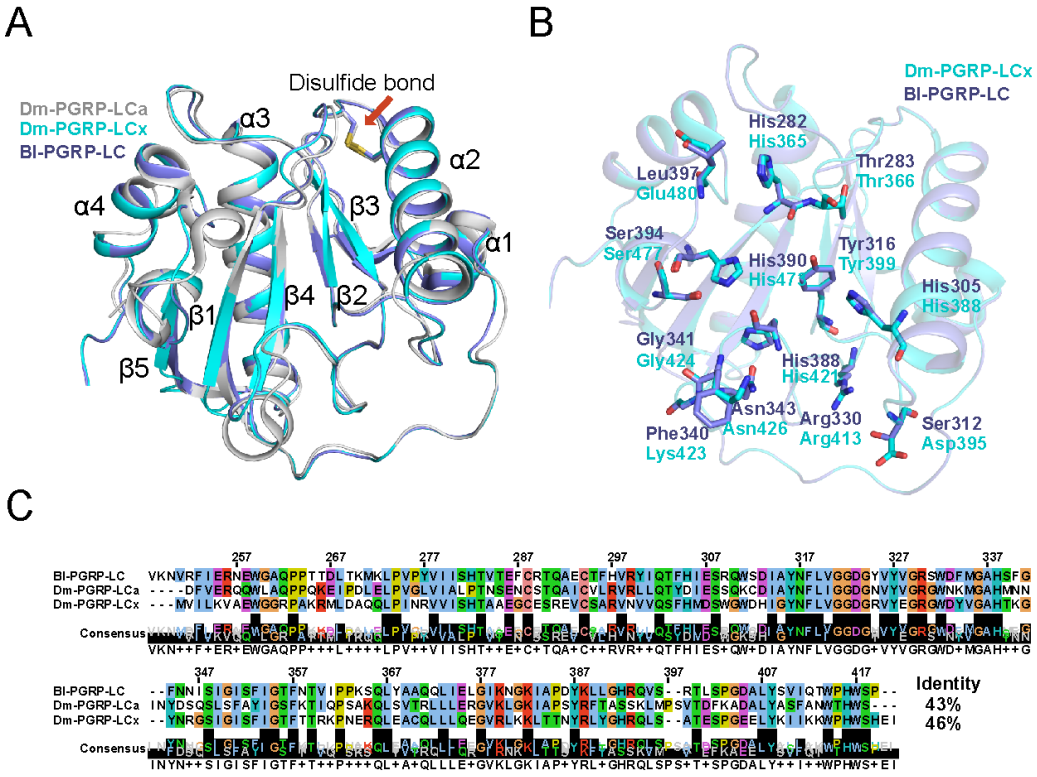


**Fig 6**. Sequence alignment and structural overlap among Bl-PGRP-LC, Dm-PGRP-LCx and Dm-PGRP-LCa. A. Superimposition of the modeled structure of Bl-PGRP-LC using the Swiss-Model sever (https://www.swissmodel.expasy.org/) and Dm-PGRP-LCx and Dm-PGRP-LCa; a disulfide bond is marked by sticks. B. Comparison of the contact residues involved in the binding of TCT by Dm-PGRP-LCx and Bl-PGRP-LC; residues are marked by sticks and colored by cyan and blue for *Drosophila* and *Bombus*, respectively. C. Sequence alignment of PGRP domains between Bl-PGRP-LC and Dm-PGRP-LC(x/a).

*Melanobombus*, *Bombus*) was consistent with the phylogeny obtained for bumblebee subgenera using the COI gene (Fig 3) (Paulh. Williams et al., 2012). Moreover, *Bombus* and *Apis* were found in parallel clades, distant from the solitary bee (*Megachile rotundata*) clade, coinciding with the social evolution of bees (Fig 3). Hence, we hypothesize that the immune gene *PGRP-LC* possesses a similar evolutionary tendency as the evolution of sociality among bee species.

Bl-PGRP-LC shares greater than 45% identity with *Drosophila* PGRP-LC and higher identity with LCx than LCa (Fig 6B). Structural analysis showed that Bl-PGRP-LC retains the topology of *Drosophila* PGRP-LCx and PGRP-LCa, with conserved disulfide bonds, and that the binding residues contributing to interaction in the TCT/PGRP-LCx-LCa complex are highly conserved, except for Leu$^{397}$ and Ser$^{312}$ (Fig 6A and C). As the 16 bumblebee species share these conserved residues (Fig 2), Bl-PGRP-LC might possess a binding function similar to that of *Drosophila* PGRP-LCx with Dap-type PGNs. However, Bl-PGRP-LC has shorter β5 and β1 sheets and longer β2, β3, and β4 sheets; additionally, the β2 and β3 sheets point toward the center of the PGN binding groove (Fig 6A). Based on this, Bl-PGRP-LC also might have a narrow binding groove, which might impede binding of certain PGNs. Therefore, expression and purification of Bl-PGRP-LC was performed in vitro. Bl-PGRP-LC was expressed as inclusion bodies under conditions of diverse OD values and IPTG concentrations using the pET21a vector, with a predicted mass weight of 21 kDa (Fig 4). However, using the pGEX-6p-1 vector, Bl-PGRP-LC was expressed as a mixture of inclusion bodies and soluble protein, and the optimal expression conditions were OD=0.8, 0.6 mM IPTG, 16 °C (Fig 5A). After further purification using a GST column and a Superdex 200 10/300 increase column, we obtained a pure GST-Bl-PGRP-LC fusion protein (Fig 5B). This fusion protein will be used for pull-down of upstream and downstream proteins interacting with Bl-PGRP-LC, which may be involved in the immune response. In addition, pure Bl-PGRP-LC without tags was obtained via PSP enzyme digestion on the GST column and further Sephadex chromatography (Fig 5C). Pure Bl-PGRP-LC may be used for preparation of a monoclonal antibody and crystallization.

In conclusion, this study highlights the evolutionary relationship of PGRP-LC among *B. impetuosus*, *B. melanurus*, *B. difficillimus, B. lemniscatus*, *B. ladakhensis*, *B. friseanus*, *B. keriensis*, *B. pyrosoma*, *B. koreanus*, *B. ignitus*, *B. lantschouensis*, *B. longipennis*, *B. consobrinus*, *B. picipes, B. patagiatus*, *B. kashmirensis, B. terrestris, B. impatiens, E. mexicana, A. florea, A. dorsata, A. cerana, A. mellifera, M. rotundata* and *D. novaeangliae* in Apidae and their interactions with Dap-type PGNs. In addition, soluble GST-tag Bl-PGRP-LC and Bl-PGRP-LC were successfully expressed and purified.

This material will be very important for further study on the function and structure of bumblebee PGRP-LC.

**Acknowledgments**

This work was supported by the Agricultural Science and Technology Innovation Programme (CAASASTIP-2015-IAR). The funders had no role in study design, data collection, and analysis, the decision to publish, or the preparation of the manuscript.

**Authors contribution**

The experiments were designed by Yanjie Liu and were finished by Minming Chen. The manuscript was drafted by Yanjie Liu and revised by Nanhui Ye and Jiandong An.

**Reference**

An, J., Huang, J., Shao, Y., Zhang, S., Wang, B., Liu, X., Wu, J. & Williams, P.H. (2014). The bumblebees of North China (Apidae, *Bombus* Latreille). Zootaxa, 3830: 1-89. doi: 10.11646/zootaxa.3830.1.1.

Barribeau, S.M., Sadd, B.M., Du Plessis, L., Brown, M.J.F., Buechel, S.D., Cappelle, K., Carolan, J.C., Christiaens, O., Colgan, T,J., Erler, S., Evans, J., Helbing, S., Karaus, E., Lattorff, H.M., Marxer, M., Meeus, I., Näpflin, K., Niu, J., Schmid-Hempel, R., Smagghe, G., Waterhouse, R.M., Yu, N., Zdobnov, E.M. & Schmid-Hempel, P. (2015). A depauperate immune repertoire precedes evolution of sociality in bees. Genome Biology, 16: 83-83. doi: 10.1186/s13059-015-0628-y.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L. & Schwede, T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research, 42: W252-258. doi: 10.1093/nar/gku340. Epub 2014 Apr 29.

Bienert, S., Waterhouse, A., de Beer, T.A., Tauriello, G., Studer, G., Bordoli, L. & Schwede T. (2017). The SWISS-MODEL Repository-new features and functionality. Nucleic Acids Research, 45: D313-D319. doi: 10.1093/nar/gkw1132. Epub 2016 Nov 29.

Bischoff, V., Vignal, C., Boneca, I.G., Michel, T., Hoffmann, J.A. & Royet, J. (2004). Function of the *Drosophila* pattern-recognition receptor PGRP-SD in the detection of Gram-positive bacteria. Nature Immunology, 5: 1175-1180. doi: 10.1038/ni1123.

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J. & Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. Nature Protocol, 4: 1-13. doi: 10.1038/nprot.2008.197.

Chang, C.I., Chelliah, Y., Borek, D., Mengin-Lecreulx, D. & Deisenhofer, J. (2006). Structure of tracheal cytotoxin in complex with a heterodimeric pattern-recognition receptor. Science, 311: 1761-1764. doi: 10.1126/science.1123056.

Chang, C.I., Ihara, K., Chelliah, Y., Mengin-Lecreulx, D., Wakatsuki, S. & Deisenhofer, J. (2005). Structure of the ectodomain of *Drosophila* peptidoglycan-recognition protein LCa suggests a molecular mechanism for pattern recognition. Proceedings of the National Academy of Sciences of the United States of America, 102: 10279-10284. doi: 10.1073/pnas.0504547102.

Choe, K.M., Werner, T., Stoven, S., Hultmark, D. & Anderson, K.V. (2002). Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in *Drosophila*. Science, 296: 359-362. doi: 10.1126/science.1070216.

Dziarski, R. & Gupta, D. (2005). Peptidoglycan recognition in innate immunity. Journal of Endotoxin Research, 11: 304-310. doi: 10.1179/096805105X67256.

Dziarski, R. & Gupta, D. (2006). The peptidoglycan recognition proteins (PGRPs). Genome Biology, 7: 232. doi: 10.1186/gb-2006-7-8-232.

Evans, J.D., Aronstein, K., Chen, Y.P., Hetru, C., Imler, J.L., Jiang, H., Kanost, M., Thompson, G.J., Zou, Z. & Hultmark, D. (2006). Immune pathways and defence mechanisms in honey bees *Apis mellifera*. Insect Molecular Biology, 15: 645-656. doi: 10.1111/j.1365-2583.2006.00682.x.

Gendrin, M., Zaidman-Remy, A., Broderick, N.A., Paredes, J., Poidevin, M., Roussel, A. & Lemaitre, B. et al. (2013). Functional analysis of PGRP-LA in *Drosophila* immunity. PLoS One, 8: e69742. doi: 10.1371/journal.pone.0069742.

Goto, A., Yano, T., Terashima, J., Iwashita, S., Oshima, Y. & Kurata, S. (2010). Cooperative regulation of the induction of the novel antibacterial Listericin by peptidoglycan recognition protein LE and the JAK-STAT pathway. Journal of Biological Chemistry, 285: 15731-15738. doi: 10.1074/jbc.M109.082115.

Goulson, D., Lye, G.C. & Darvill, B. (2008). Decline and conservation of bumble bees. Annual Review of Entomology, 53: 191-208. doi: 10.1146/annurev.ento.53.103106.093454.

Guex, N., Peitsch, M.C. & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. Electrophoresis, 30 Suppl 1: S162-173. doi: 10.1002/elps.200900140.

Hoffmann, J.A. (2003). The immune response of *Drosophila.* Nature, 426: 33-38. doi: 10.1038/nature02021.

Honeybee Genome Sequencing, C. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera.* Nature, 443: 931-949. doi: 10.1038/nature05260.

Kaneko, T., Yano, T., Aggarwal, K., Lim, J.H., Ueda, K., Oshima, Y., Peach, C., Erturk-Hasdemir, D., Goldman, W.

E., Oh, B.H., Kurata, S. & Silverman, N. (2006). PGRP-LC and PGRP-LE have essential yet distinct functions in the *Drosophila* immune response to monomeric DAP-type peptidoglycan. Nature Immunology, 7: 715-723. doi: 10.1038/ni1356.

Kang, D., Liu, G., Lundstrom, A., Gelius, E. & Steiner, H. (1998). A peptidoglycan recognition protein in innate immunity conserved from insects to humans. Proceedings of the National Academy of Sciences of the United States of America, 95: 10078-10082. doi: 10.1073/pnas.95.17.10078.

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. doi: 10.1093/molbev/msw054.

Kurata, S. (2014). Peptidoglycan recognition proteins in *Drosophila* immunity. Developmental and Comparative Immunology, 42: 36-41. doi: 10.1016/j.dci.2013.06.006.

Liu, C., Xu, Z., Gupta, D. & Dziarski, R. (2001). Peptidoglycan recognition proteins: a novel family of four human innate immunity pattern recognition molecules. Journal of Biological Chemistry, 276: 34686-34694. doi: 10.1016/j.dci.2013.06.006.

Lu, X., Wang, M., Qi, J., Wang, H., Li, X., Gupta, D. & Dziarski, R. (2006). Peptidoglycan recognition proteins are a new class of human bactericidal proteins. Journal of Biological Chemistry, 281: 5895-5907. doi: 10.1074/jbc.M511631200.

Meeus, I., Brown, M.J., De Graaf, D.C. & Smagghe, G. (2011). Effects of invasive parasites on bumble bee declines. Conservation Biology, 25: 662-671. doi: 10.1111/j.1523-1739.2011.01707.x.

Michel, T., Reichhart, J.M., Hoffmann, J.A. & Royet, J. (2001). *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. Nature, 414: 756-759. doi: 10.1038/414756a.

Royet, J. & Dziarski, R. (2007). Peptidoglycan recognition proteins: pleiotropic sensors and effectors of antimicrobial defences. Nature Reviews Microbiology, 5: 264-277. doi: 10.1038/nrmicro1620.

Royet, J., Gupta, D. & Dziarski, R. (2011). Peptidoglycan recognition proteins: modulators of the microbiome and inflammation. Nature Reviews Immunology, 11: 837-851. doi: 10.1038/nri3089.

Sadd, B.M., Barribeau, S.M., Bloch, G., De Graaf, D.C., Dearden, P., Elsik, C.G., Gadau, J., Grimmelikhuijzen, C.J., Hasselmann, M., Lozier, J.D., Robertson, H.M., Smagghe, G., Stolle, E., Van Vaerenbergh, M., Waterhouse, R.M., Bornberg-Bauer, E., Klasberg, S., Bennett, A.K., Câmara, F., Guigó, R., Hoff, K., Mariotti, M., Munoz-Torres, M., Murphy, T., Santesmasses, D., Amdam, G.V., Beckers, M., Beye, M., Biewer, M., Bitondi, M.M., Blaxter, M. L., Bourke, A. F., Brown, M.J., Buechel, S.D., Cameron, R., Cappelle, K., Carolan, J.C., Christiaens, O., Ciborowski, K.L., Clarke, D. F., Colgan, T.J., Collins, D.H., Cridge, A.G., Dalmay, T., Dreier, S., du Plessis, L., Duncan, E., Erler, S., Evans, J., Falcon, T., Flores, K., Freitas, F.C., Fuchikawa, T., Gempe, T., Hartfelder, K., Hauser, F., Helbing, S., Humann, F.C., Irvine, F., Jermiin, L.S., Johnson, C.E., Johnson, R.M., Jones, A.K., Kadowaki, T., Kidner, J.H., Koch, V., Köhler, A., Kraus, F.B., Lattorff, H.M., Leask, M., Lockett, G.A., Mallon, E.B., Antonio, D.S., Marxer, M., Meeus, I., Moritz, R.F., Nair, A., Näpflin, K., Nissen, I., Niu, J., Nunes, F.M., Oakeshott, J.G., Osborne, A., Otte, M., Pinheiro, D.G., Rossié, N., Rueppell, O., Santos, C.G., Schmid-Hempel, R., Schmitt, B. D., Schulte, C., Simões, Z.L., Soares, M.P., Swevers, L., Winnebeck, E.C., Wolschin, F., Yu, N., Zdobnov, E.M., Aqrawi, P. K., Blankenburg, K. P., Coyle, M., Francisco, L., Hernandez, A.G., Holder, M., Hudson, M.E., Jackson, L., Jayaseelan, J., Joshi, V., Kovar, C., Lee, S.L., Mata, R., Mathew, T., Newsham, I.F., Ngo, R., Okwuonu, G., Pham, C., Pu, L.L., Saada, N., Santibanez, J., Simmons, D., Thornton, R., Venkat, A., Walden, K.K., Wu, Y.Q., Debyser, G., Devreese, B., Asher, C., Blommaert, J., Chipman, A.D., Chittka, L., Fouks, B., Liu, J., O'Neill, M. P., Sumner, S., Puiu, D., Qu, J., Salzberg, S.L., Scherer, S.E., Muzny, D.M., Richards, S., Robinson, G.E., Gibbs, R.A., Schmid-Hempel, P. & Worley, K.C. (2015). The genomes of two key bumblebee species with primitive eusocial organization. Genome Biology, 16: 76. doi: 10.1186/s13059-015-0623-3.

Schleifer, K.H. & Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriological Reviews, 36: 407-477.

Steiner, H. (2004). Peptidoglycan recognition proteins: on and off switches for innate immunity. Immunological Reviews, 198: 83-96. doi: 10.1111/j.0105-2896.2004.0120.x.

Takehana, A., Yano, T., Mita, S., Kotani, A., Oshima, Y. & Kurata, S. (2004). Peptidoglycan recognition protein (PGRP)-LE and PGRP-LC act synergistically in *Drosophila* immunity. EMBO Journal, 23: 4690-4700. doi: 10.1038/sj.emboj.7600466.

Wang, M., Liu, L.H., Wang, S., Li, X., Lu, X., Gupta, D. & Dziarski, R. (2007). Human peptidoglycan recognition proteins require zinc to kill both gram-positive and gram-negative bacteria and are synergistic with antibacterial peptides. Journal of Immunology, 178: 3116-3125. doi: 10.4049/jimmunol.178.5.3116.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R. & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research, 46: W296-W303. doi: 10.1093/nar/gky427.

Werner, T., Borge-Renberg, K., Mellroth, P., Steiner, H. & Hultmark, D. (2003). Functional diversity of the *Drosophila* PGRP-LC gene cluster in the response to lipopolysaccharide and peptidoglycan. Journal of Biological Chemistry, 278:

26319-26322. doi: 10.1074/jbc.C300184200.

Werner, T., Liu, G., Kang, D., Ekengren, S., Steiner, H. & Hultmark, D. (2000). A family of peptidoglycan recognition proteins in the fruit fly *Drosophila* melanogaster. Proceedings of the National Academy of Sciences of the United States of America, 97: 13772-13777. doi: 10.1073/pnas.97.25.13772

Whelan, S. & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Molecular Biology and Evolution, 18: 691-699. doi: 10.1093/oxfordjournals.molbev.a003851.

Williams, P.H., Brown, M.F., Carolan, J., Jian Dong, A., Goulson, D., Murat Aytekin, A., Lincoln, R., Best Alexandr, M., Björn Cederberg, B., Dawson, R., Jia Xing, Huang., Ito, M., Monfared, A., H. Raina, R., Schmid-Hempel, P., S. Sheffield, C., Šima, P. & Zenghua, X. (2012). Unveiling cryptic species of the bumblebee subgenus *Bombus* s. str. worldwide with COI barcodes (Hymenoptera: Apidae). Systematics and Biodiversity, 10: 21-56. doi: 10.1080/14772000.2012.664574.

Williams, P., Colla, S. & Zheng Hua, X. (2009). Bumblebee vulnerability: common correlates of winners and losers across three continents. Conservation Biology, 23: 931-940. doi: 10.1111/j.1523-1739.2009.01176.x.

Yoshida, H., Kinoshita, K. & Ashida, M. (1996). Purification of a peptidoglycan recognition protein from hemolymph of the silkworm, *Bombyx mori*. Journal of Biological Chemistry, 271: 13854-13860. doi: 10.1074/jbc.271.23.13854

Yu, X.Q., Zhu, Y.F., Ma, C., Fabrick, J.A. & Kanost, M.R. (2002). Pattern recognition proteins in *Manduca sexta* plasma. Insect Biochemistry and Molecular Biology, 32: 1287-1293. doi: 10.1016/S0965-1748(02)00091-7.

Zaidman-Remy, A., Herve, M., Poidevin, M., Pili-Floury, S., Kim, M. S., Blanot, D., Oh, B. H., Ueda, R., Mengin-Lecreulx, D. & Lemaitre, B. (2006). The *Drosophila* amidase PGRP-LB modulates the immune response to bacterial infection. Immunity, 24: 463-473. doi: 10.1016/j.immuni.2006.02.012.