Journal of Insect Science, (2021) 21(4): 12; 1–20
https://doi.org/10.1093/jisesa/ieab054
Research

ENTOMOLOGICAL SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

OXFORD

New Insect Host Defense Peptides (HDP) From Dung Beetle (Coleoptera: Scarabaeidae) Transcriptomes

Germán Alberto Téllez Ramirez,[1] Juan Felipe Osorio-Méndez, Diana Carolina Henao Arias, Lily Johanna Toro S., Juliana Franco Castrillón, Maribel Rojas-Montoya, and Jhon Carlos Castaño Osorio

Center of Biomedical Research, Group of Molecular Immunology, Universidad del Quindío, Carrera 15 and Calle 12 Norte, Armenia, Quindío, Colombia and [1]Corresponding author, e-mail: gatellez@uniquindio.edu.co

Subject Editor: Daniel Swale

Received 5 April 2021; Editorial decision 15 July 2021

Abstract

The Coleoptera Scarabaeidae family is one of the most diverse groups of insects on the planet, which live in complex microbiological environments. Their immune systems have evolved diverse families of Host Defense Peptides (HDP) with strong antimicrobial and immunomodulatory activities. However, there are several peptide sequences that await discovery in this group of organisms. This would pave the way to identify molecules with promising therapeutic potential. This work retrieved two sources of information: 1) De-novo transcriptomic data from two species of neotropical Scarabaeidae (*Dichotomius satanas* and *Ontophagus curvicornis*); 2) Sequence data deposited in available databases. A Blast-based search was conducted against the transcriptomes with a subset of sequences representative of the HDP. This work reports 155 novel HDP sequences identified in nine transcriptomes from seven species of Coleoptera: *D. satanas* (n = 76; 49.03%), *O. curvicornis* (n = 23; 14.83%), (Trypoxylus *dichotomus*) (n = 18; 11.61%), (*Onthophagus nigriventris*) (n = 10; 6.45%), (*Heterochelus* sp) (n = 6; 3.87%), (Oxysternon *conspicillatum*) (n = 18; 11.61%), and (Popillia *japonica*) (n = 4; 2.58%). These sequences were identified based on similarity to known HDP insect families. New members of defensins (n = 58; 37.42%), cecropins (n = 18; 11.61%), attancins (n = 41; 26.45%), and coleoptericins (n = 38; 24.52%) were described based on their physicochemical and structural characteristics, as well as their sequence relationship to other insect HDPs. Therefore, the Scarabaeidae family is a complex and rich group of insects with a great diversity of antimicrobial peptides with potential antimicrobial activity.

Key words: antimicrobial cationic peptide, insect protein, computational biology, cecropin, defensin

One of the main effectors of an insect's immune response is the production of host defense peptides (HDP) or antimicrobial peptides. Families of these peptides have been identified in all taxonomic groups, thus, representing an ancient and efficient defense mechanism against pathogens. In insects, most HDP are synthesized as precursors or pro-proteins in the fat body and hemocytes (Cociancich et al. 1994, Hoffmann 1995, Hoffmann et al. 1996). HDP are cationic, amphipathic polypeptides, produced in all known genera of living organisms, and represent an ancient innate defense mechanism (Faruck et al. 2016, Alencar-Silva et al. 2018, Saikia et al. 2019). Once activated by post-translational proteolysis (Boman 1995, Hoffmann 1995, Zhang and Gallo 2016, Gómez et al. 2017), they act as effector molecules against a broad spectrum of pathogens, including Gram-positive and Gram-negative bacteria, protozoa, fungi, and viruses. They also have a low propensity for developing resistance. This efficiency is thought to be one of the biological attributes that would explain the evolutionary success of insects (Imler and Bulet 2005, Pasupuleti et al. 2012). Therefore, in recent years, they have attracted attention in the development of new antimicrobials with clinical applications (Xiao et al. 2013, Meher et al. 2017).

So far, the identification of insect HDP has focused on Hemiptera, Hymenoptera, Lepidoptera, and Diptera orders (Mylonakis et al. 2016). In Coleoptera, the most diverse order of insects, only a few HDP have been reported (Cociancich et al. 1994, Bulet et al. 2004, Mylonakis et al. 2016, Toro Segovia et al. 2017). Insect HDP have been classified in different families in accordance with their sequence, physico-chemical and structural properties. Representative families are cecropin, defensin, attacin, and coleoptericin. Other families, such as moricin and gloverin, have been identified only in Lepidoptera (Cociancich et al. 1994, Bulet et al. 1999, Wiesner and Vilcinskas 2010). Cecropins are the most abundant family of linear α-helical HDP in insects (Brady et al. 2019). They have been identified in Hexapoda orders, like

Coleoptera, (Diptera), and (Lepidoptera) (Yi et al. 2014), and it is characterized by low molecular weights (3–4 kDa) and positive net charges. Their structure is amphipathic and alpha helical (Boman 1998, Hultmark et al. 2005). Mature active cecropins are generated after the removal of a classical secretory signal peptide. A long hydrophobic C-terminal and a strongly basic N-terminal domain is presumptively required for the biological activity, mostly against Gram-negative bacteria (DeLucca et al. 1997, Cavallarin et al. 1998, Marshall and Arenas 2003, Wang et al. 2007).

Defensin is a large and ubiquitous family, with members expressed in almost all forms of life (Volkoff et al. 2003). In insects, they have been found in Diptera, (Hymenoptera), Coleoptera, (Trichoptera), (Hemiptera), and (Odonata) orders (Hoffmann and Hetru 1992, Bulet et al. 1999). They are active mainly against Gram-positive bacteria (Bulet et al. 2004). Once synthesized, pre-pro-defensins are proteolytically processed by the removal of the signal peptide to pro-defensin; then, an additional cut of the propeptide by a furin-like enzyme in an R-X-[RK]-R site produces an active mature peptide (Bulet et al. 1999, Lowenberger et al. 1999). The majority of mature defensins are cationic peptides composed of 18–45 amino acids with 6–8 conserved cysteine residues that pair through disulfide bridges (Araújo et al. 2006, Zhu and Gao 2013, Wu et al. 2018). Insect defensins have the same cysteine pairing: Cys1–Cys4, Cys2–Cys5, and Cys3–Cys6. This structural topology is known as cysteine-stabilized αβ motif (CSαβ) and is common among defensin peptides across different organisms (Cornet et al. 1995, Hultmark et al. 2005, Hwang et al. 2009, Dias and Franco 2015, Tarr 2016, Shafee et al. 2017).

Attacins are larger peptides with molecular weights of 20–23 kDa. Acidic (pI ~ 5.7) and basic (pI ~ 8.3) isoforms have been described (Hultmark et al. 1983). They are synthesized as pre-pro-proteins containing a signal peptide, and a conserved furin-like cutting site. The mature peptide is composed of an N-terminal attacin domain, followed by a glycine-rich segment (Hultmark et al. 1983, Kockum et al. 1984, Gunne et al. 1990, Shin and Park 2019). Their secondary structure is composed of a hydrophobic alpha helix similar to glycine-rich peptides (Wang et al. 2008). They can inhibit growth of Gram-negative bacteria and synthesis of bacterial proteins, like OmpC, OmpF, OmpA, and LamB (Engström et al. 1984b, Carlsson et al. 1991).

Coleoptericin peptides are found exclusively in the Coleoptera order. The first coleoptericin was described in the (Tenebrionidae) beetle (*Zophobas atratus*) (Bulet et al. 1991). This family is characterized by a signal peptide, and a propeptide that is cleaved through a furin-like site to produce a mature peptide of 71–75 residues (Vilcinskas et al. 2013). They are glycine- and proline-rich antimicrobial peptides with bacteriostatic and bactericidal activity. Their action mechanism is unclear, but liposome-leaking experiments suggest that it does not involve the formation of pores. Instead, they elicit the formation of an elongated and chain formation morphology in bacteria (Sagisaka et al. 2001). Two subgroups exist, one positively and the other negatively charged, and their C-terminus has a basic nature (Bulet et al. 1991, Sagisaka et al. 2001).

The HDP reported in the Scarabaeidae family are scarce compared with their wide diversity of species, consisting of over 30,000 globally. The sequence and function of only a few of these peptides have been characterized, including some defensin and cecropin families, like Coprisin and Oxysterlin (Tomie et al. 2003, Lee et al. 2014, Toro Segovia et al. 2017). Therefore, this work sought to identify and describe new putative HDP in publicly available assembled transcript sequences from the NCBI Transcriptome Shotgun Assembly (TSA) database of seven different species of Scarabaeidae and two new transcriptomes from the neotropical beetles *Dichotomius satanas* and *Ontophagus curvicornis*, both species widely distributed inhabiting the Andean region of Colombia (Bouchard et al. 2011, Cultid-Medina et al. 2012).

Materials and Methods

Collection and Maintenance of Beetles

Neotropical dung beetles used in this research were obtained in the municipality of Filandia, Quindío-Colombia (4.686998°N and −75.614500°W; datum = WGS84) 1.923 masl. The beetles captured were identified as *Dichotomius satanas* and *Ontophagus curvicornis* with the Cultid-Medina et al. 2014 taxonomic key (Cultid-Medina et al. 2014). Once collected, they were maintained in a terrarium with organic soil and human feces bait for 12 hr. Then, they were inoculated in the ventral lateral abdomen with 10 µl of a pool of 1×10^6 UFC/ml formalin-fixed bacteria (*Escherichia coli* and *Staphylococcus aureus*) and fungi (*Candida albicans*). Finally, the fat body and part of the hindgut were dissected 12 hr post-inoculation.

Total RNA Extraction, Transcriptome Sequencing, and De Novo Assembly

Total RNA was extracted by using an Ambion total RNA extraction kit with in column DNAse treatment (Invitrogen cat PureLink RNA Mini Kit, Life Technologies 12183025). Total RNA was prepared by using bead clean-up and library preparation with Illumina RNA poly-A selection. The RNA quality and quantity were assessed via spectrophotometry (Nanodrop2000), fluorometry (Qubit), and electrophoretic profile (Labchip) for further processing. The transcriptome was sequenced with Illumina Hiseq and pair-end read with 150 pb length. The FASTAq files were checked through FastQC (Thrash et al. 2018); trimming was done by Trimmomatic V0.36 (Bolger et al. 2014); data from each species were merged and the transcriptome was de-novo assembled by using Trinity V2.5 on Indiana University National center for genome analysis support (Afgan et al. 2018).

Transcriptome Shotgun Assemblies

TSAs from Scarabaeidae species: *Trypoxylus dichotomus*, *Onthophagus nigriventris*, *Onthophagus curvicornis*, *Popillia japonica*, *Heterochelus* sp, *Dichotomius satanas*, and *Oxysternon conspicillatum* were downloaded from the sequence set browser (https://www.ncbi.nlm.nih.gov/Traces/wgs/; Table 1). Fasta files of the assembled transcriptomes were converted into BLAST databases for each TSA with CLC main workbench software 7.9.1.

Homology Identification of HDP

Queries lists from different HDP InterPro families were constructed (Cecropin: IPR020400; Defensin: IPR017982; Coleoptericin: IPR009382, and Attacin: IPR005520 IPR005521; Supp Table S1 [online only]); the Cecropin family was complemented with Oxysterlins (Toro Segovia et al. 2017). With the TSA BLAST databases and the HDP queries, a multi-TBLASTn search was constructed and the resulting sequences were filtered according to the E-score ≤ 0.01 (Altschul et al. 1990). The list of sequence codes identified in the tBLASTn were extracted from the different transcriptomes and, according to the high-scoring segment pair frame, the related amino-acid sequence was identified

Table 1. TSA from Scarabaeidae

Prefix (TSA code)	Organism	Bio-project code	Bio-sample code	Source	Contigs number
GAQV01	*Trypoxylus dichotomus*	PRJNA231720	SAMN02444008 SAMN02444009 SAMN02444010 SAMN02444011 SAMN02444012 SAMN02444013 SAMN02444014 SAMN02444015 SAMN02444016 SAMN02444017 SAMN02444018 SAMN02444019	3rd-instar larval and prepupal	34,455
IABQ01	*Trypoxylus dichotomus* tunobosonis	PRJDB4830	SAMD00051587	3rd larvae	30,157
GAQW01	*Onthophagus nigriventris*	PRJNA231725	SAMN02444020 SAMN02444021 SAMN02444022 SAMN02444023	3rd-instar larval and prepupal	59,302
GARJ01	*Popillia japonica*	PRJNA233626	SAMN02569976	Antenna	698
GARK01	*Popillia japonica*	PRJNA198730	SAMN02055564	Grub	1,916
GDNJ01	*Heterochelus sp*. AD-2015	PRJNA286531	SAMN03799575	Tissue	50,435
GEXM01	*Oxysternon conspicillatum*	PRJNA339294	SAMN05589108	Adult fat body	27,567
GHMA01	*Dichotomius satanas*	PRJNA510790	SAMN10614917	Adult fat body and intestine	4,63,430
GHMD01	*Onthophagus curvicornis*	PRJNA510790	SAMN10614998	Adult fat body and intestine	1,72,518

by using the ORF finder tool (https://www.ncbi.nlm.nih.gov/orffinder/; Sayers et al. 2011).

Workflow to Analyze Putative HDP Sequences

The presence and location of signal peptides in the deduced HDP amino acid sequences were predicted with SignalP (http://www.cbs.dtu.dk/services/SignalP/; Nielsen 2017). The physico-chemical characteristics of the peptides (molecular mass, isoelectric point, and Kyte–Doolittle hydrophobic profile) were calculated with the protein report tool from the CLC main workbench V7.9.1. The total net charge was calculated with the APD3 (http://aps.unmc.edu/AP/main.php; Wang et al. 2016). Prediction of the antimicrobial function was conducted with the SVMC, RFC, and DAC algorithms available in Campr3 (Waghu, Barai, Gurung, et al. 2016a), Classamp (Joseph et al. 2012), and iAmppred tools (Bhadra et al. 2018).

Structural Analysis

The secondary structure prediction was conducted with the Psipred (http://bioinf.cs.ucl.ac.uk/psipred/) server (Buchan et al. 2013). Prediction of functional domains was carried out with Interpro (https://www.ebi.ac.uk/interpro/; Jones et al. 2014). The tertiary structure was predicted with RAPTOR X (http://raptorx.uchicago.edu/StructurePrediction/predict/; Källberg et al. 2012), and structural alignments were conducted with the 3Dcomb V1.18 tool (Wang et al. 2011). All the models were visualized in UCSF Chimera V1.13.1 (Pettersen et al. 2004). Protein-protein interactions of the sequences modeled were constructed by using the Cluspro server (Vajda et al. 2017).

Similarity Dendrogram

The sequences of InterPro families (cecropin: IPR020400; insect Defensin: IPR017982; Coleoptericin: IPR009382 Attacin: IPR005520 IPR005521) and the taxonomic key corresponding to each sequence were downloaded from the PIR batch server (https://pir.georgetown.edu/pirwww/search/batch.shtml). The signal and propeptide were identified and removed and mature peptides were aligned with the HDP from the Scarabaeidae with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) or MUSCLE (Sievers et al. 2011). Dendrograms were generated by Neighbor-joining using the Jukes-Cantor model to calculate protein distance and bootstrap with 10,000 replicates in CLC main workbench V7.9.1.

Ethics Statement

This work was approved by the bioethics committee at Universidad del Quindío under act number 8 of 6 May 2016. The contract of access to genetic resources was drawn through resolution No. 120 of 22 October 2015 with the Colombian Ministry of Environment and Sustainable Development.

Results and Discussion

Transcriptome Shotgun Assemblies

In this work, the transcriptomes of two Scarabaeidae species (*O. curvicornis* [Reads: 78,643,148] and *D. satanas* [Reads: 99,661,453]) were sequenced, assembled, and submitted to the DDB/EMBL/GenBank database under access codes GHMD00000000–GHMA00000000, bio-sample: SAMN10614998–SAMN10614917 and bio-project: PRJNA510790–PRJNA510790, respectively. The RNA quantity and quality scores used for these can be found in Supp Table S2 (online only). Another seven Scarabaeidae transcriptomes were used from the DDB/EMBL/GenBank database. The TSA characteristics used for this work are shown in Table 1.

A tBLAStn search on the nine assembled transcriptomes was undertaken with the HDP sequence queries constructed; the resulting contigs were retrieved and only those containing complete ORF were considered for further analysis. In total, 155 contigs encoding for potential HDP were identified. The number of ORFs

encoding for potential HDP is proportional to the number of assembled contigs (Supp Fig. 1A [online only]).

Sequences from four HDP families (Attacin, Coleoptericin, Defensin, and Cecropin) were identified in the TSAs. Members from all families were found in five of seven species, the exception being *Onthophagus nigriventris* and *Popillia japonica* (Supp Fig. 1B [online only]). The most abundant family was Defensin, followed by Attacin, Coleoptericin, and Cecropin. However, the size of each family varies depending on the species. It has been described that the expression of HDP is influenced by context-specific characteristics, where sex, presence of offspring, and carcass affect their expression in a complex system of transcriptional reprogramming, reflecting adaptations to specific ecological niches (Jacobs et al. 2016). For these reasons, the differences observed may be explained by the sample origin of the RNA (tissue and life-cycle stage) and the size of each transcriptome.

Traditional strategies for HPD identification and characterization involve biochemical purification methods with RP-HPLC (reverse-phase high-pressure liquid chromatography) coupled with mass spectrometry and functional assays. Other strategies use highly conserved positions of some HDP families to identify potential HDP by similarity searches or molecular biology approaches by RACE-PCR (rapid amplification of cDNA ends; Pei et al. 2014). Artificial neural network algorithms have also been trained with structural

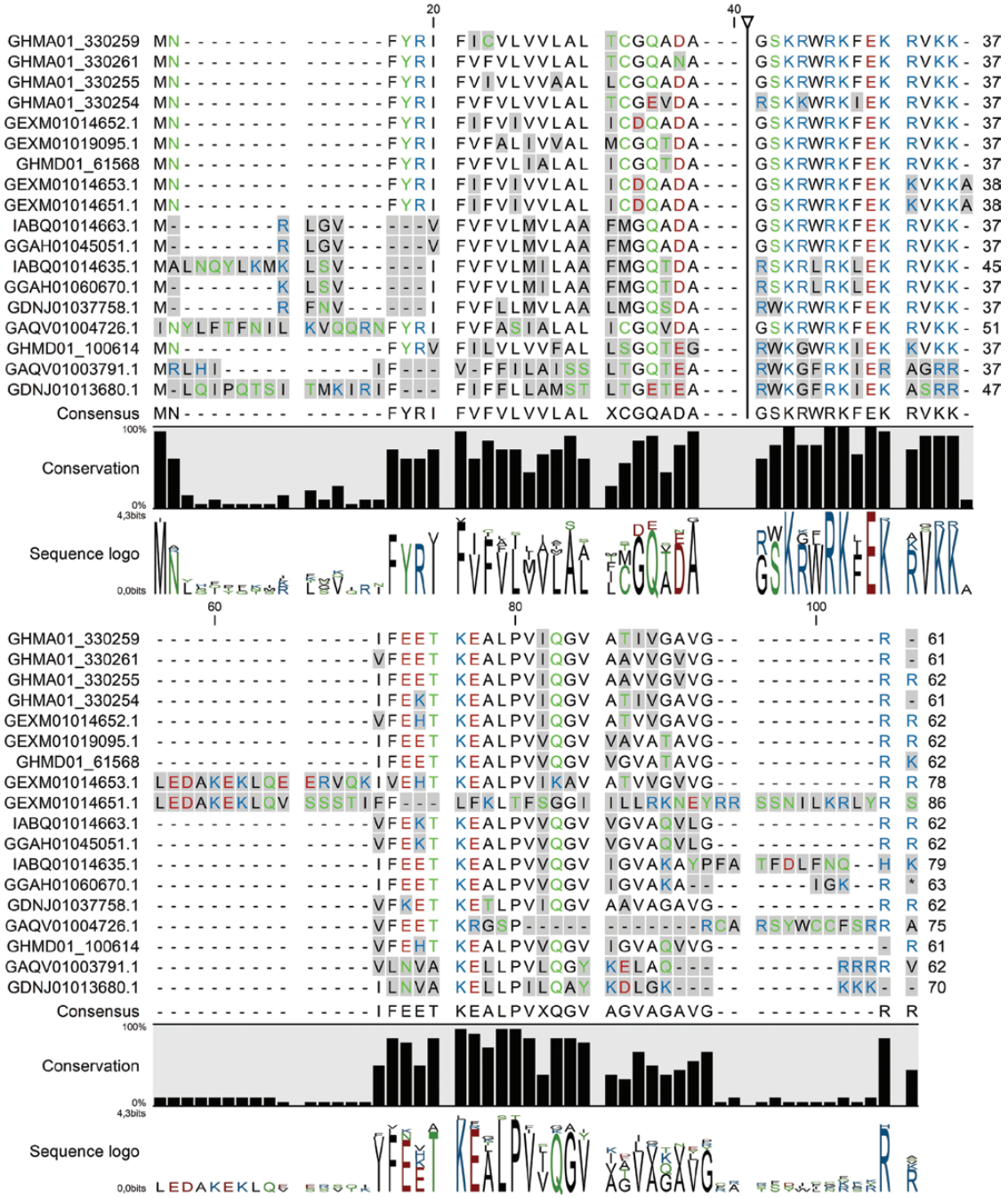

Fig. 1. Multiple-sequence alignment of cecropin HDP found in Scarabaeidae. Position 40 in the alignment represents the cleavage site of the signal peptide predicted with SignalP.

and physicochemical information to identify novel HDP sequences (Wang 2010). Recent developments in high-throughput sequencing technologies have represented a novel and efficient method for gene identification (Pane et al. 2017). Transcriptome-based approaches, using next-generation sequencing, are particularly useful because they focus on the expressed (i.e., exomic) portion of the genome.

This strategy proved successful in this work, given that it identified 155 new putative HDP sequences from nine transcriptomes (Supp Table S3 [online only]). In contrast, an approach using similarity searches on deduced sequences derived genomic and expressed sequences tags (ESTs) from three insect species only identified six putative HDPs (Duwadi et al. 2018). One difference in these works is the presence of elements, such as introns and regulatory elements within the genomic sequences as they may limit the efficacy of gene identification by gene prediction methods. Additionally, a greater number of sequences were used as queries for similarity searches.

With approximately one-million species, insects represent the largest class within the animal kingdom. Within insects, Coleoptera is the most diverse order (Purvis and Hector 2000, Hunt et al. 2007, Hwang et al. 2009). Nevertheless, only 305 of the 3070 HDP sequences deposited in the Antimicrobial Peptide Database (APD) are derived from insects (Wang et al. 2016). The HDP sequences and their biological activities have been described for several Coleoptera species, including: (*Allomyrina dichotoma*) (Lee et al. 2019), (*Octodonta nipae*) (Meng et al. 2019), *Hylobius abietis* (Namara et al. 2018), (*Nicrophorus vespilloides*) (Vogel et al. 2011), (*Tenebrio molitor*) (Jacobs et al. 2017), (*Calomera littoralis*) (Rodríguez-García et al. 2016), (*Protaetia brevitarsis seulensis*) (Lee et al. 2016), (*Tribolium castaneum*) (Altincicek and Vilcinskas 2007, Zou et al. 2007, Altincicek et al. 2013), (*Holotrichia diomphalia*) (Lee et al. 1994), *Zophobas atratus* (Bulet et al. 1991), *Allomyrina dichotoma* (Sagisaka et al. 2001), (*Acalolepta luxuriosa*) (Imamura et al. 2009), and (*Sitophilus oryzae*) (Login et al. 2011). Our work further extends the list of HDP sequences to other species of Coleoptera. Future work would involve the characterization of the biological activity of some of these peptides.

Cecropin Family Description

In Scarabaeidae, 18 cecropin sequences were found with mature peptide lengths ranging from 37 to 55 residues, with molecular weights around 4 kDa. The sequence alignment shows a highly conserved signal peptide, the mature peptide has an N-terminal cationic domain [GR]-[SW]-K-[RKG]-[WLF]-R-K-[FIL]-E-[KR]-[RKA]-[VSG]-[KR]-[KR] with a high frequency of K-R residues and hydrophobic angle from 120° to 180°. The C-terminal domain has a higher degree of variability with a region rich in acid residues and an aliphatic hydrophobic region (Fig. 1). The predicted secondary and tertiary structures show highly conserved alpha-helix with a TM score of 0.51 in the structural alignment, suggesting that they can be classified as a single structural family (Fig. 2; Xu and Zhang 2010; Wang et al. 2011, 2013). It was found that only GEXM01014653.1 and GEXM01014651.1 sequences have an additional segment in the N-terminal domain, probably because of alternative splicing or a transcriptome assembly error.

The insect cecropin dendrogram shows that the majority of sequences are from Diptera and Lepidoptera orders, as described (Mylonakis et al. 2016). The dendrogram structure has four main clades that are well related to the phylogenetic orders, a highly conserved Diptera clade representing flies, Lepidoptera, Ascaridida, and a distant share clade of Coleoptera and mosquitoes (Fig. 3). Compared to other orders, like Diptera and Lepidoptera, cecropins from Scarabaeidae show more variation and divergence among the sequences within the same order.

According to the previously described cecropins in Coleoptera, few representatives exist with Oxysterlins (1, 2, and 3; *O. conspicillatum*), Cec (*Acalolepta luxuriosa*), and Sarcotoxin Pd (*Paederus dermatitis*); Saito et al. 2005, Memarpoor-Yazdi et al. 2013, Toro Segovia et al. 2017). Interestingly, the InterPro data

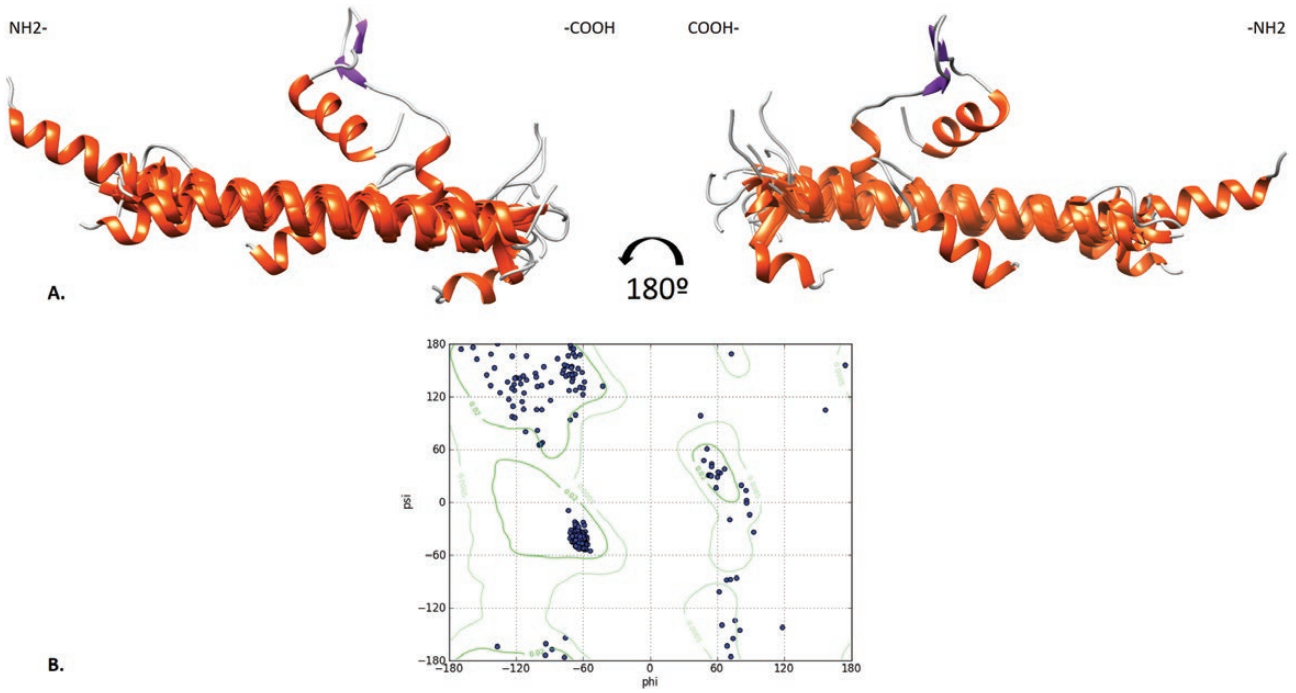

Fig. 2. Structure of Scarabaeidae mature cecropins. A. Structural superposition constructed in DeepAlign. TM score: 0.517 with alpha helical conserved tertiary conformation. B. Ramachandran plot of the superimposed structures.

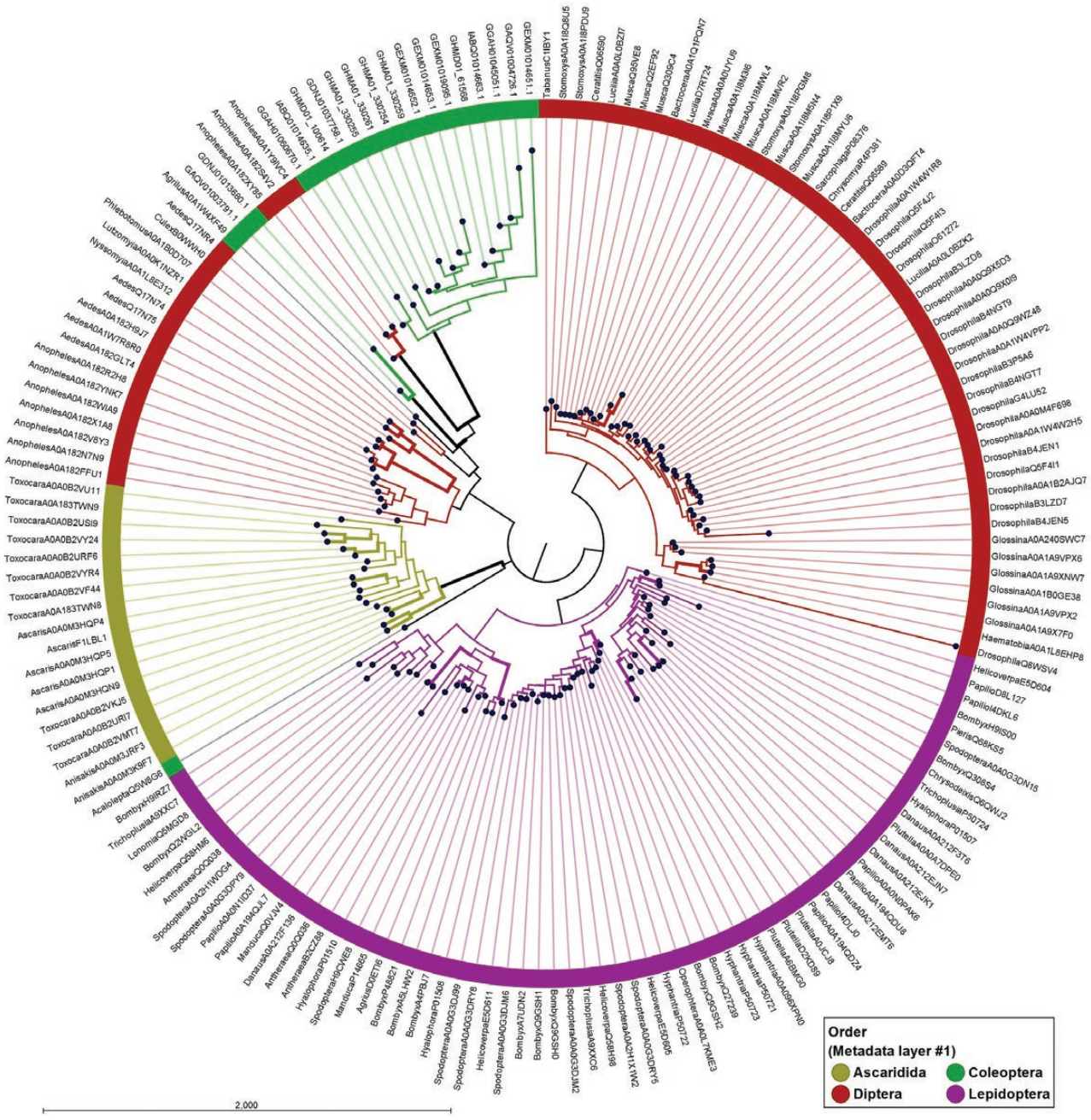

Fig. 3. Neighbor-joining similarity dendrogram of invertebrate cecropins. Sequences from the mature peptides were aligned by MUSCLE. The taxonomic distribution of the sequences is indicated in the color code presented. Confidence values of the branches were calculated with 10,000 bootstrap replicates. Thicker lines show branches with bootstrap threshold value > 70.

set also has a low representation of Coleopteran cecropins, with only two representative sequences, *Acalolepta luxuriosa* (Q5W8G6) and (*Agrilus planipennis*) (A0A1W4XF49). Additionally, the InterPro search fails to classify the Scarabaeidae cecropin sequences as insect cecropins, indicating that these sequences are different from the InterPro signatures; nevertheless, as described, these new peptides had similar physico-chemical and structural characteristics of the insect cecropins (Brady et al. 2019).

Defensin Family Description

This work identified 58 new putative defensin sequences in Scarabaeidae transcriptomes. As described for other insect defensins, those encoded by Scarabaeidae encode for a signal peptide followed by a propeptide (position 5–80) characterized by acid residues and an R-X-[RK]-R furin-like cleavage site (Fig. 4). These sequences were classified into three groups (group A, B, and C), according to their sequence and structural and physico-chemical properties.

Defensin group A, with 31 sequences, shows a classical defensin pattern of helix beta-sheet structure with three disulfide bridges between cysteine pairs Cys1–Cys4, Cys2–Cys5, and Cys3–Cys6 (20 sequences). For 10 sequences, the first cysteine pairing seems to be lost, but they keep the Cys2–Cys5 and Cys3–Cys6 binding pattern. Only one sequence (GHMA01_94621) has no predicted disulfide bridges. Structurally, there is one subgroup within Group A that can be distinguished by the absence of the N-terminal hydrophobic loop (six sequences; Fig. 5).

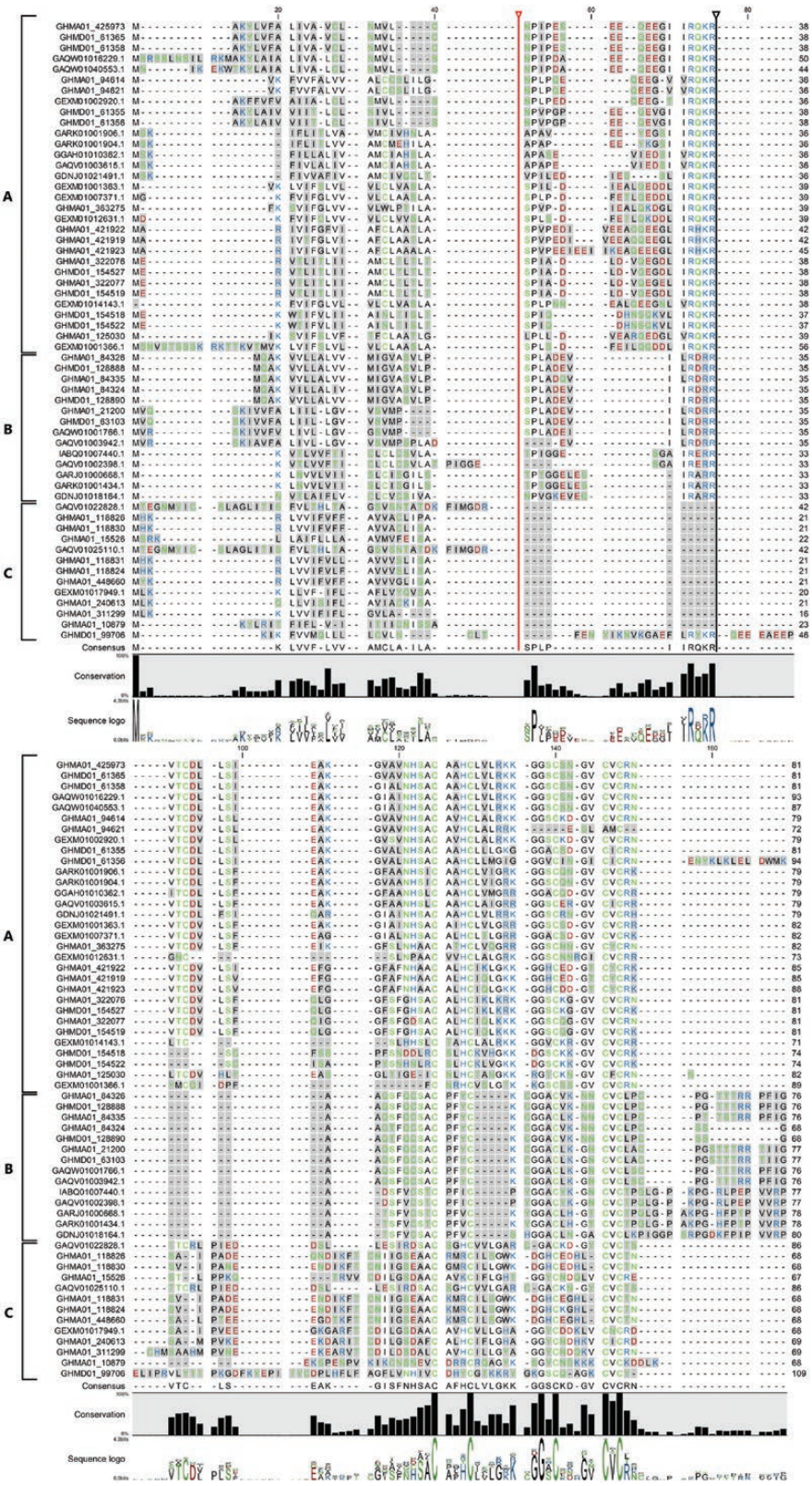

Fig. 4. Multiple-sequence alignment of defensin HDP found in Scarabaeidae. A, B, and C groups are annotated. The signal peptide cleavage site predicted is in position 50. Positions 72–75 indicate the propeptide furin-like cleavage site.

Defensins group B is represented by 14 sequences. One key feature in this subgroup is the left-handed helix 6–8 residues long, related to a P-F-[YVI] motif in position 11. They encode for 6 or 8 cysteine residues (nine sequences) that form two predicted disulfide bridges between pairs Cys2–Cys4 and Cys3–Cys5 (six sequences), or Cys2–Cys5 and Cys3–Cys6 (six sequences). Two sequences

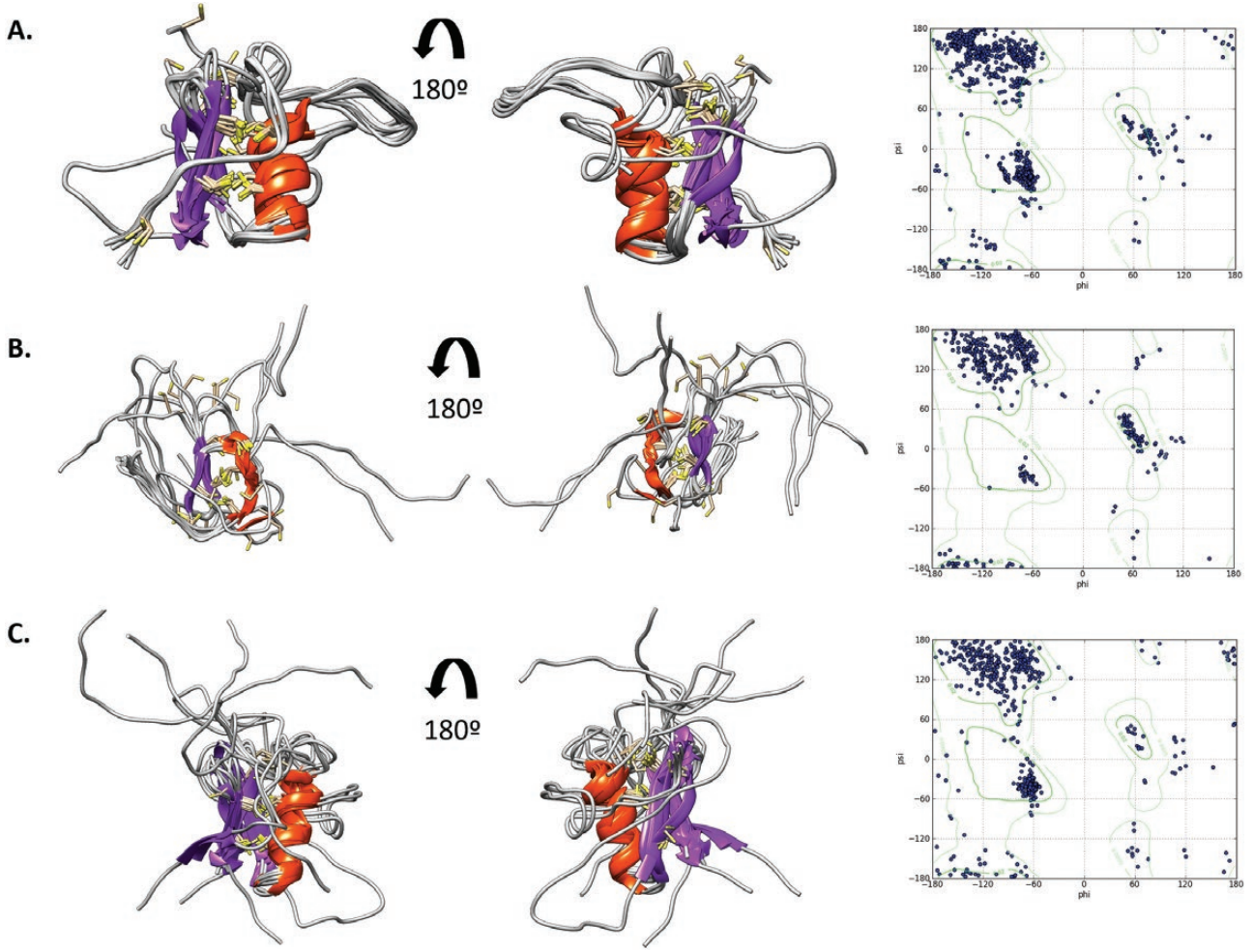

Fig. 5. Structural superposition of the mature Scarabaeidae Defensins and Ramachandran plot, constructed in DeepAlign. A. Group A with a helix beta-sheet structure (TM score: 0.569). B. Group B with left-handed helix. and high frequency of residues in the left-handed helix region lying between 30° and 130° in the Φ angle and –50° and 100° in the ψ angle in the Ramachandran plot (TM score: 0.571). C. Group C with helix beta-sheet structure (TM score: 0.425).

(GHMA01_21200; GHMD01_63103) have one pair Cys1–Cys6 and loss of predicted helix of the tertiary structure prediction. A hydrophobic random coil region was found in the C-terminal end of the sequences (Fig. 5). The left-handed helix is a rare structural motif found in peptides and proteins. It has been found in regions related to protein stability, ligand binding, or as part of an enzyme's active site. Thus, a significant structural or functional role for this secondary structure element has been suggested. The motif related to this particular structure agrees with the described propensity of amino acids to form such structure, preferring aromatic and large aliphatic amino acids (Brogden et al. 1996, Lai et al. 2002). To our knowledge, this kind of structure has not been described in the defensin family but its appearance may indicate functional importance due to its unique structural parameters.

Defensins group C, with 13 sequences, contains the classic three disulfide bridges of the insect defensins between cysteine pairs Cys1–Cys4, Cys2–Cys5, and Cys3–Cys6. Interestingly, they lost the R-X-[RK]-R cleavage site conserved for the other defensins, thus, adding 15 N-terminal negatively charged residues to the mature peptide. In addition, there are two highly conserved acidic residues [DE] in 45 and 64 positions located at the beginning of the alpha-helix and beta-sheet loop (Fig. 5). These acid residues explain the negative charge of this group (Supp Fig. 2 [online only]). These types of anionic antimicrobial peptides have been shown to kill the human B-defensin-resistant Gram-positive bacterium (*Staphylococcus aureus*), which escapes attacks from cationic peptides probably by incorporating positive charges on the membrane surface by adding Lys to lipids (Peschel et al. 2001). The anionic antimicrobial peptides, although rarely documented, appear to complement the cationic antimicrobial peptides, offering a complete spectrum of antimicrobial peptides (Brogden et al. 1996, Lai et al. 2002).

To evaluate the relationships of the Scarabaeidae defensins identified, a dendrogram was constructed with the retrieved insect defensins reported under the InterPro IPR017982. The sequences analyzed are distributed in the six major orders of insects (Fig. 6). The distribution shows four main clades, two related to the Hymenoptera order, and one distinctive clade for Diptera and Hemiptera. Orders, like Coleoptera, (Phthiraptera), and (Archaeogastropoda) were not grouped in a single clade, representing a more diverse distribution throughout the diversity of sequences. The defensins group A from Scarabaeidae was related with the Coleopteran defensin described from the InterPro. Group B of Scarabaeidae defensins were exclusively found in Scarabaeidae in a closer relationship with the clade corresponding to Hymenoptera. Group C seems related to a single defensin encoded by Hymenoptera (VespidaeXP_014602557) and other Coleoptera species. These results are compatible with the idea

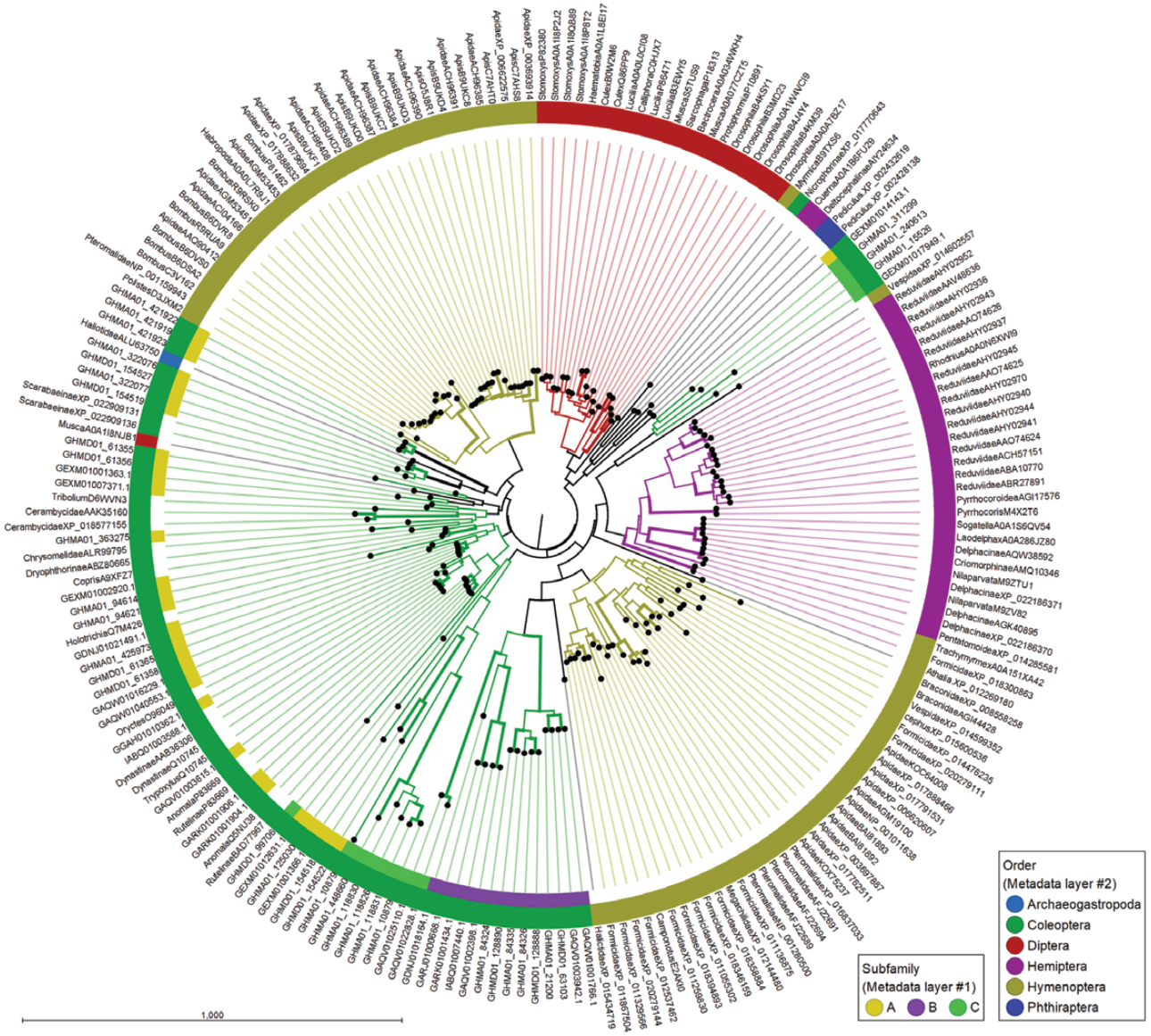

Fig. 6. Neighbor-joining similarity dendrogram of invertebrate defensins. Sequences from mature peptides were aligned by MUSCLE. The taxonomic distribution of the sequences and the corresponding defensin group is indicated in the color code presented. The confidence values of the branches were calculated with 10,000 bootstrap replicates. Thicker lines show branches with bootstrap threshold value > 70.

that groups B and C are Scarabaeidae-specific groups of defensins with novel structural and physicochemical properties.

Attacin Family Description

This study identified 41 new Scarabaeidae attacin sequences that fulfill the characteristics described (Fig. 7). They are also recognized as sequences from this family by the Pfam signature Attacin-C (PF03769; Ando and Natori 1988, Sun et al. 1991). The predicted mature sequences of the attacins identified can be further divided into two groups (named group A or B) according to the net charge (Supp Fig. 3 [online only]). Compared to group B attacins, those belonging to group A are more cationic given that they are highly enriched in positively charged residues. Additionally, group A attacins contain significantly more GNTS polar residues.

The secondary and tertiary structures predicted for Scarabaeidae attacins were characterized by a predominant antiparallel eight-string beta-sheet configuration (Fig. 8). The structural similarity was higher for group A attacins, as evidenced by higher TM scores (Group A = 0.73 and Group B = 0.58). The tertiary structure partially resembles those adopted by barrel channels. Interestingly, their structure was modeled by using the *E. coli* TamA barrel domain (PBD:4N74) as template by the automatic server predictor RaptorX. TamA forms a barrel channel with 16 transmembrane beta-sheets that translocate protein substrates across bacterial membranes (Engström et al. 1984a).

Based on these observations, we hypothesized that attacin antimicrobial activity may be related to the formation of similar configurations in the bacterial membranes. To evaluate this possibility further, a protein-protein interaction modeling was conducted. This analysis predicts homodimers of attacins with barrel-like structures containing an outward hydrophobic face, conserved acidic residues facing inward, and glycine-rich regions corresponding to the beta loops (Fig. 9). These findings are compatible with the experimental results of attacin E of (*Hyalophora cecropia*), which is targeted to the outer membrane of (*E. coli*) and facilitates the penetration of cations,

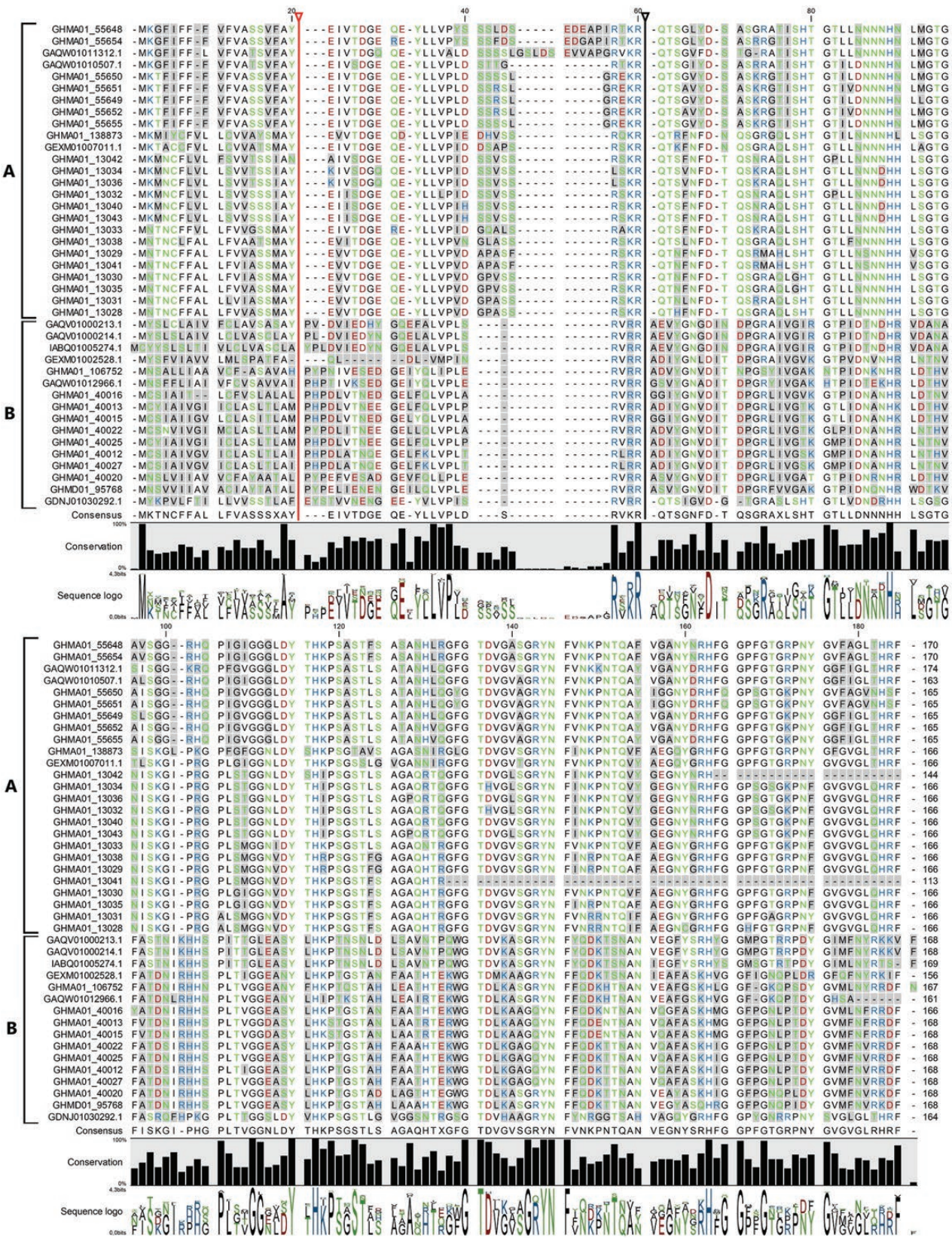

Fig. 7. Multiple-sequence alignment of attacin HDP found in Scarabaeidae. Position 19 in the alignment represents the cleavage site of the signal peptide predicted with SignalP. The propeptide comprises positions 20–60. Position 80 indicates the propeptide furin-like cleavage site.

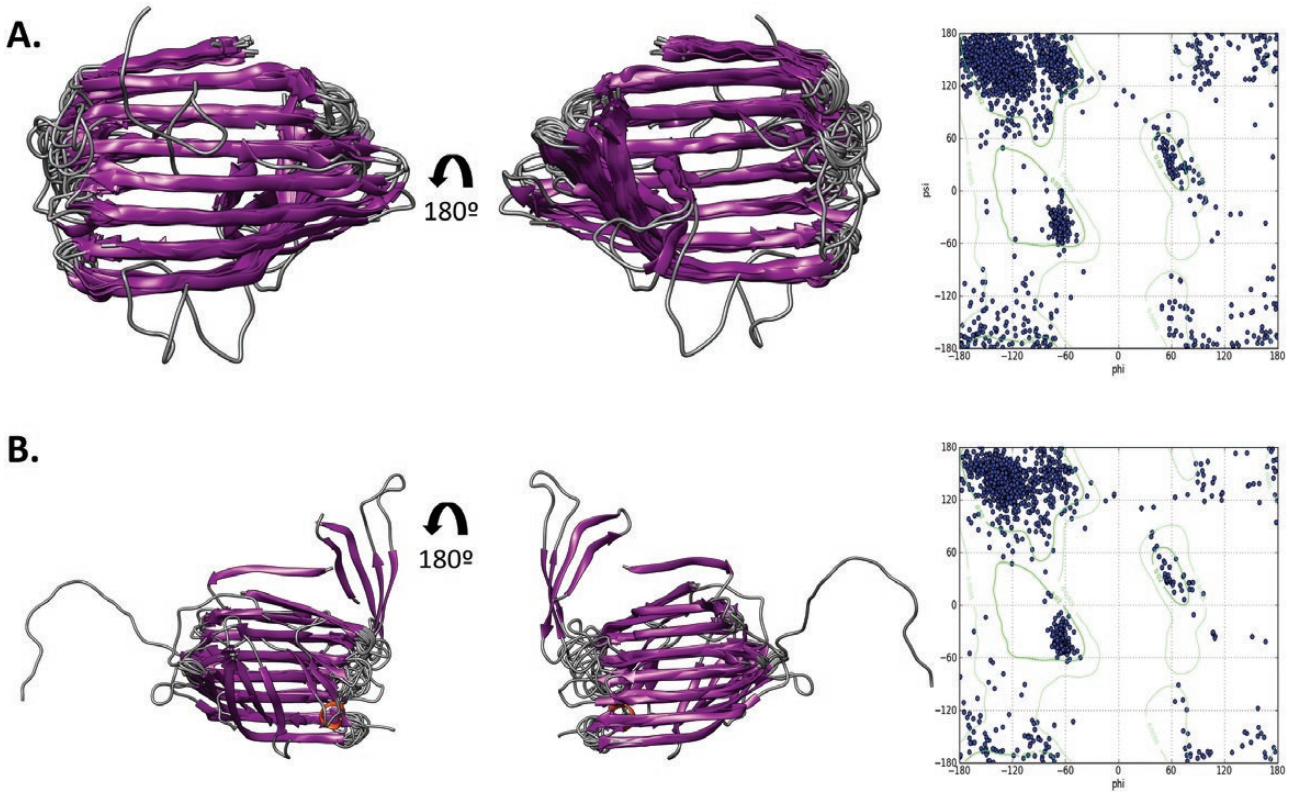

Fig. 8. Structural superposition (left) and Ramachandran plot (right) of the mature attacins. A. Analysis of group A, TM score = 0.7366. B. Analysis of group B, TM score = 0.5844. The analysis was conducted in DeepAlign.

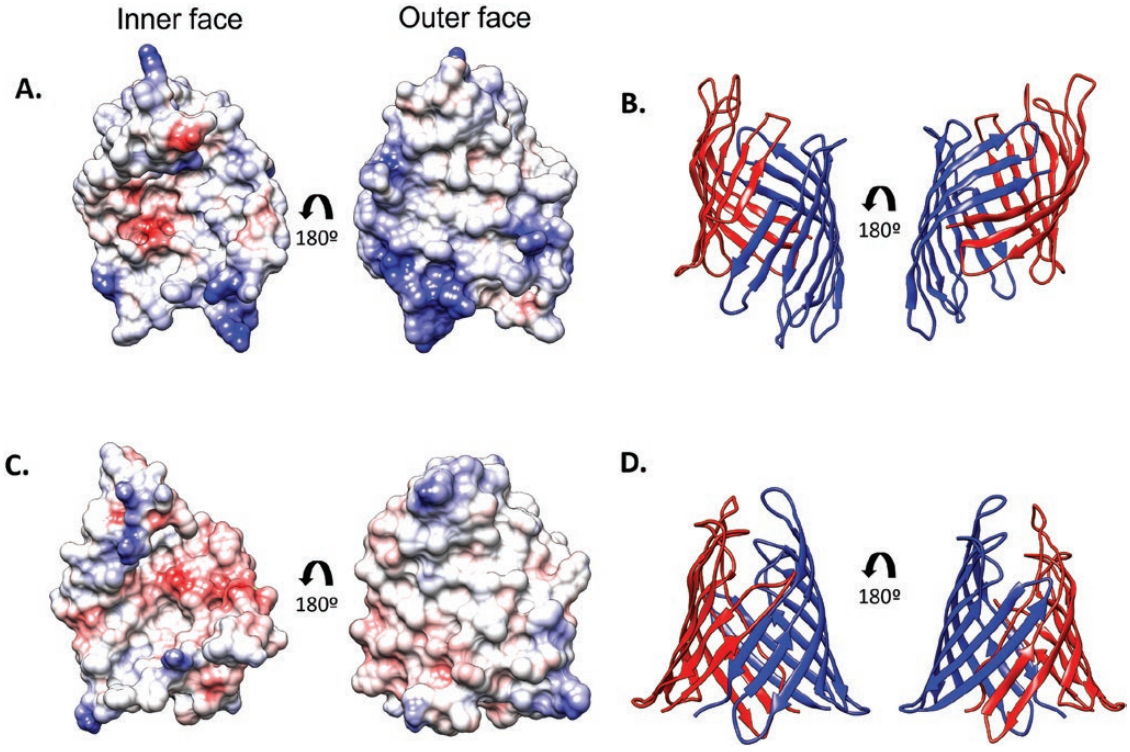

Fig. 9. Attacin structural protein-protein interaction prediction. A and B. Attacin A (GHMA01_13033). C and D. Attacin B (GHMA01_106752). A and C. Monomer electrostatic potential, according to Coulomb's law surface coloring (red –10, white 0, blue 10 kcal/(mol·e)). B. and D. Dimer protein-protein interaction prediction in ribbon chain coloring.

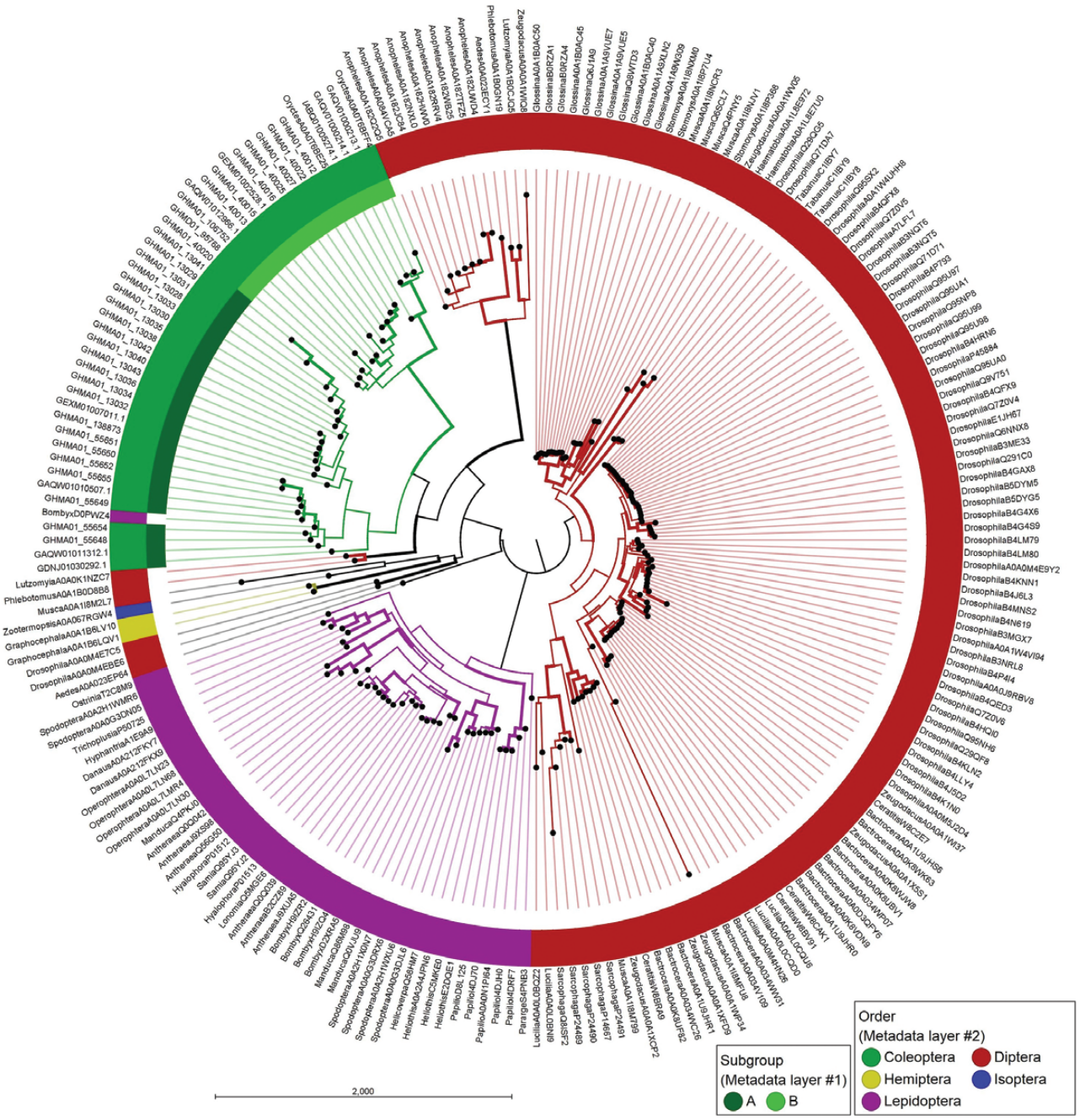

Fig. 10. Neighbor-joining similarity dendrogram of insect attacins. Sequences from the mature peptides were aligned by MUSCLE. The taxonomic distribution of the sequences and the corresponding attacin group is indicated in the color code presented. The confidence values of the branches were calculated with 10,000 bootstrap replicates. Thicker lines show branches with bootstrap threshold value > 70.

such as sodium and potassium (Engström et al. 1984a). The tertiary structure of attacins has not been experimentally determined in the PDB. The structural information available was obtained by circular dichroism (CD) on a recombinant attacin encoded by *Hyalophora cecropia*. An α-helical structure for this protein was deduced based on the presence of a single peak at 222 nm in CD (Gunne et al. 1990). However, the distinctive feature of α-helical proteins in CD are negative bands at 208 and 222 nm (Greenfield 2006), so the structure of this protein family remains an open question.

The dendrogram for insect attacins shows a clear overrepresentation of sequences from Diptera, especially flies representing 59% of the sequences. This group of diptera sequences, representing flies, forms a single clade, as well as Lepidoptera, for Coleoptera attacins; the group in a distant clade in close proximity to a second diptera group representing mosquitoes. The Coleoptera attacins had a clear differentiation in the A and B groups described. Only two sequences of Coleoptera attacins were previously annotated in the InterPro (Oryctes A0A0T6BE25; Oryctes A0A0T6BFF4) and were grouped into the B Scarabaeidae attacins (Fig. 10).

Coleoptericin Family Description

Coleoptericins encoded by Scarabaeidae (37 sequences) were identified as pre-pro-proteins with a furin-like cleavage site, they share

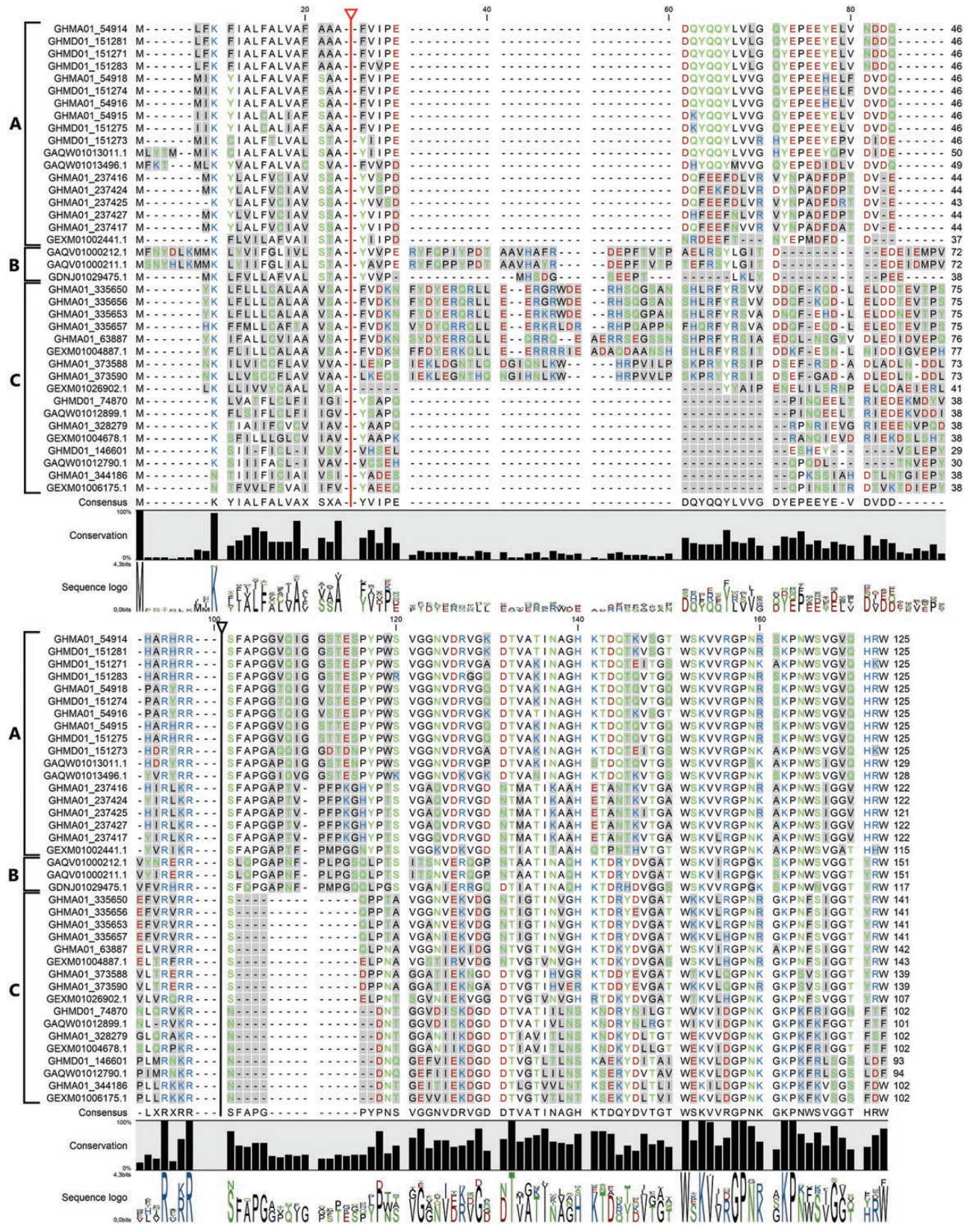

Fig. 11. Multiple sequence alignment of Coleoptericin HDP found in Scarabaeidae. Position 23 in the alignment represents the cleavage site of the signal peptide predicted with SignalP. Positions 94–97 are furin-like cleavage sites.

the motif G-P-[GNS]-[KR]-[GSA]-K-P from position 97–103 (Fig. 11). They were classified according to their amino acid sequence conservation into three major groups: 1) Group A (18 sequences) with approximately 73 amino acid residues in length, a cationic character with a mean net charge of 5.63, and a relative disorder structure with a random coil-beta sheet structure (TM score: 0.27; Fig. 12 and Supp Fig. 4 [online only]); 2) Group B is a small group in Scarabaeidae (three sequences), but shows high similarity with already reported coleoptericins (Fig. 13). These are typically 72 residues in length, with a mean net charge of 5 and a coil-beta sheet-helix-coil configuration in the tertiary structure (Fig. 12); and 3) Group C (17 sequences) is 57–59 amino acids in length and is enriched in acidic residues that provide a mean positive net charge of 1.41. They have a relatively disordered structure with a coil-beta sheet-helix-coil configuration (Figs. 11–13). In general, the conserved coleoptericin motif is structurally preceded by an alpha helix or related to the third beta loop in the structural conformation.

Coleoptericins A, B, and C were identified from three separate groups exclusive of the Scarabaeidae family. This work reports two new groups, A and C coleoptericins, which are a group in a different clade and are exclusive of Scarabaeidae (Fig. 13).

Distinct Physico-Chemical Properties Characterize the HDP Families of Scarabaeidae

Families of insect antimicrobial peptides have distinct physico-chemical properties, amino acid composition, and structural features. This has allowed classifying the HDP according to them (Meher et al. 2017). Insect antimicrobial peptides can be divided into four classes: α-helical peptides (cecropin and moricin), cysteine-rich peptides (insect defensin and drosomycin), proline-rich peptides (apidaecin, drosocin, and lebocin), and glycine-rich peptides (attacin and gloverin; Wang et al. 2016, Zhou et al. 2019). With this in mind, said features were analyzed in the HDP identified from Scarabaeidae (Table 2). Attacins and coleoptericins were the families with higher

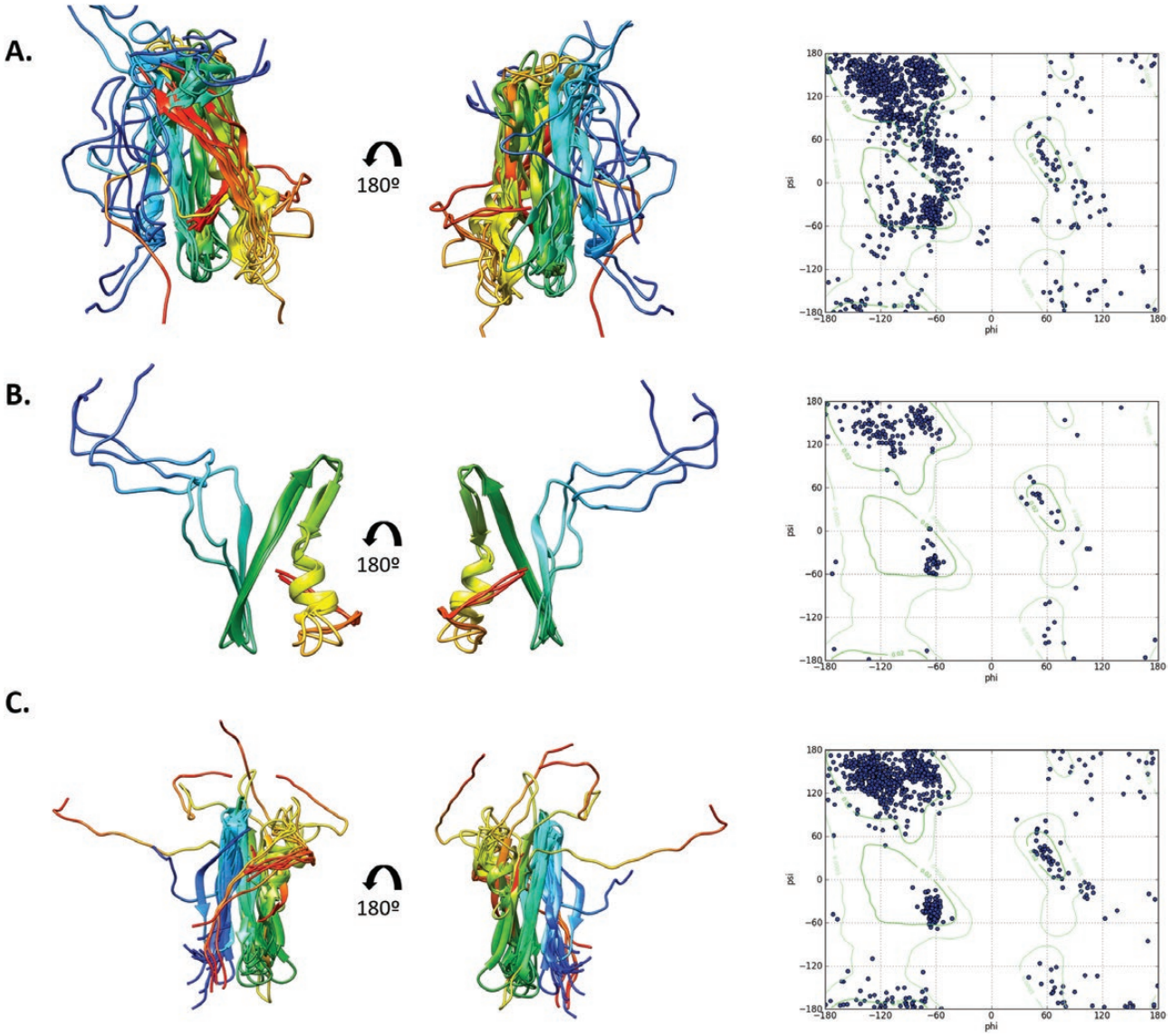

Fig. 12. Structural superposition (left) and Ramachandran plot (right) of the mature coleoptericins. A. Analysis of group A, TM score = 0.2753. B. Analysis of group B, TM score = 0.6248, C. Analysis of group C, TM score = 0.3925. The analysis was conducted in DeepAlign.

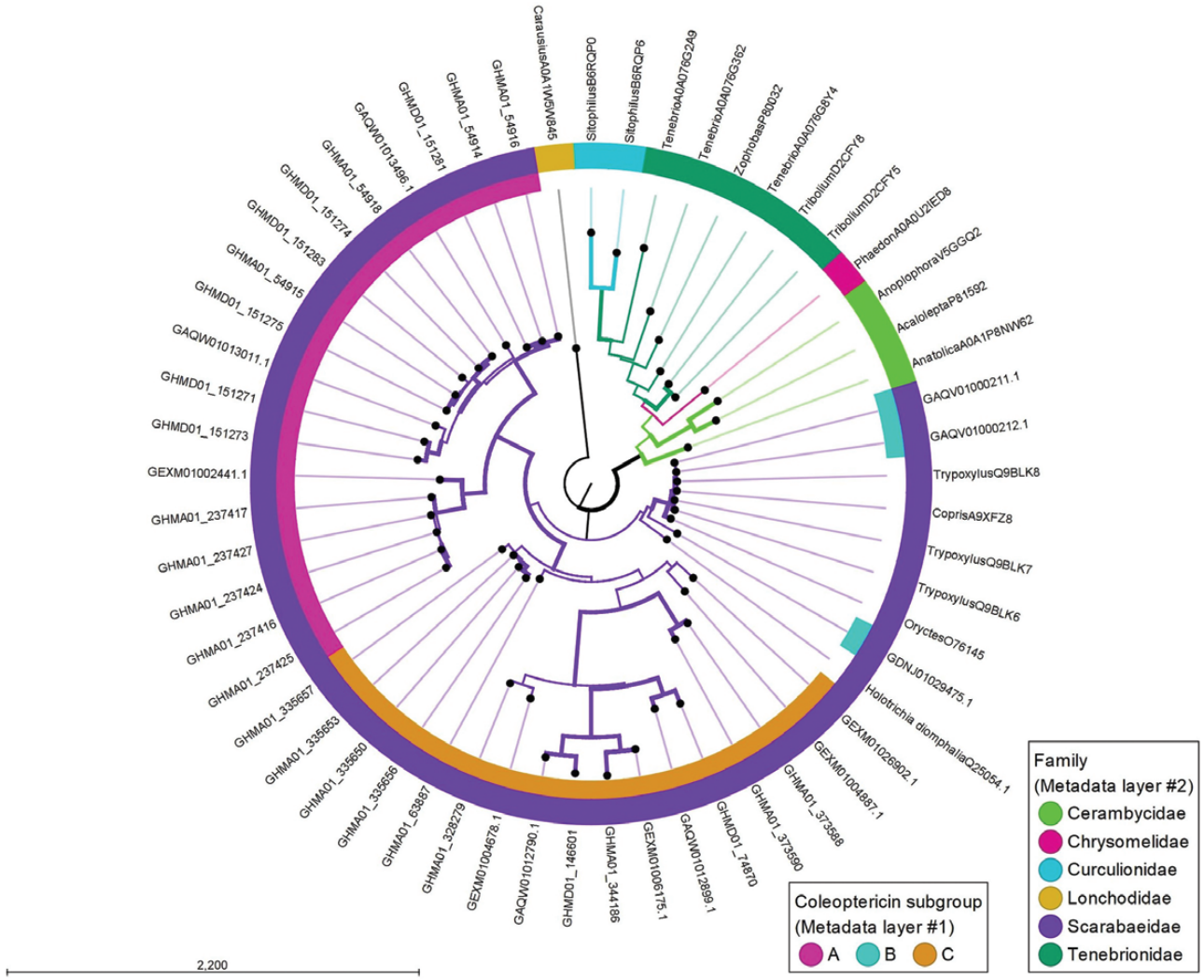

Fig. 13. Neighbor-joining similarity dendrogram of Coleoptericins. The taxonomic distribution of the sequences and the corresponding coleoptericin group is indicated in the color code presented. A sequence from Carausius (Phasmatodea order) was used as the root node. The remaining sequences belong to the Coleoptera order. The confidence values of the branches were calculated with 10,000 bootstrap replicates. Thicker lines show branches with bootstrap threshold value > 70.

Table 2. Physicochemical and structural properties of antimicrobial peptide families

Family property	Cecropin		Defensin		Attacin		Coleoptericin	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mass	4,806	850.7	4537	693	12,931	1,075	7,097	736.6
pI	11.38	0.3633	7.958	1.471	9.573	1.551	9.596	2.075
Net charge pH 7	8.528	1.835	2.483	2.57	6.11	3.329	3.697	3.116
Hydrophobic ratio	0.3844	0.02706	0.4609	0.05614	0.2641	0.03654	0.2813	0.02642
% positively charged	30.48	3.029	14.86	5.15	14.23	1.563	14.87	1.387
% proline and glycine	10.64	2.527	13.81	5.53	20.22	3.955	20.54	2.866
% polar amino acids	15.16	3.393	25.92	4.015	41.26	6.008	36.58	3.011

molecular mass (mean: 12931 Da; SD: 1075 Da and mean: 7097 Da; SD: 736.6 Da, respectively) and lowest hydrophobic nature (0.27 and 0.28 ratios, respectively; Supp Fig. 5 [online only]). Defensins and cecropins showed similar hydrophobic ratios (0.37 and 0.47, respectively). As described for other members of cecropins (Yi et al. 2014), those identified in Scarabaeidae are characterized by high net charges (mean: 9.25; SD: 1.71) and pI values (mean: 11.3; SD: 0.41; Supp Fig. 5 [online only]). These characteristics are reflected in the amino acid composition of the peptide families (Supp Fig. 6 [online only]). Cecropins show a higher proportion of positively charged amino acids, and attacins and coleoptericins show a higher content of polar residues. Prolines and glycines are structurally important

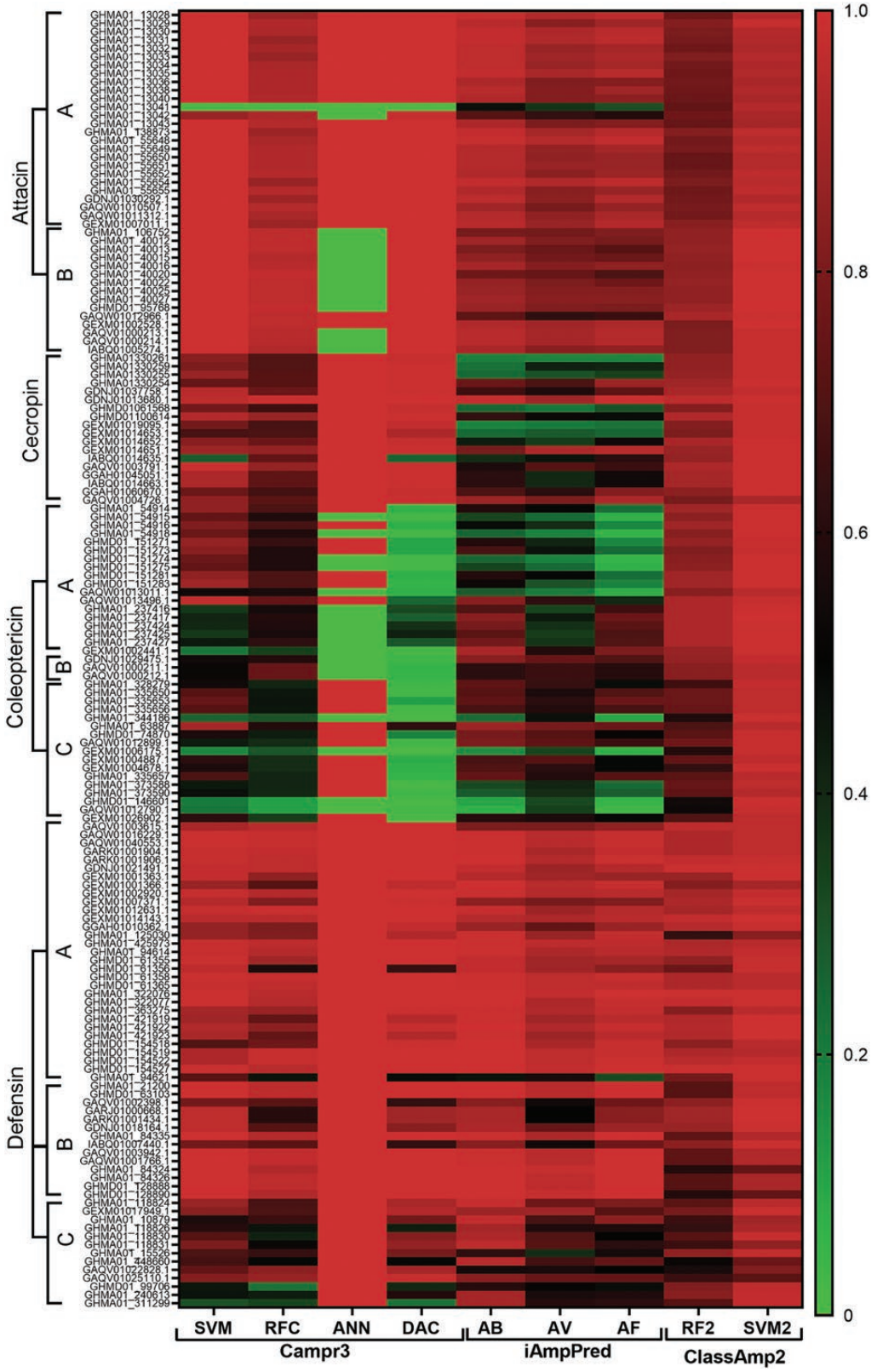

Fig. 14. Heat map for the antimicrobial prediction of Scarabaeidae HDP. The gradient represents the probability score for each server from inactive (green), medium activity (black), and active (red). The HDP families were separated into their respective groups: Attacin A and B; Coleoptericins A, B, and C; Defensins A, B, and C. SVM, support vector machine; RFC, random forest; ANN, artificial neural network; DAC, discriminant analysis classification; AB, antibacterial; AV, antiviral; AF, antifungal.

amino acids, hence, differences were sought in the content of these amino acids in the HDP. In contrast to the other two families, >10% of the amino acids encoded by attacins and coleoptericins correspond to these two residues (Supp Fig. 6 [online only]).

The amino-acid distribution represents a key characteristic to classify HDP families, like proline, lysine, and cysteine-rich antimicrobial peptides (Wang et al. 2016, Meher et al. 2017, Lu 2019, Zhou et al. 2019). According to this, in Scarabaeidae HDP it was found that positively charged amino-acid proportion only differentiates the cecropin family (Supp Fig. 6A [online only]). The percentage of glycine and proline obtained showed a statistical difference among all groups except between the attacin and coleoptericin. The polar amino acid (GNST) distribution differentiates the HDP families with a statistically significant difference (Supp Fig. 6B and C [online only]).

Antimicrobial Activity Prediction of Dung Beetle HDP In Silico

Development of sequence-based computational tools can be helpful in identifying candidate HDP for experimental characterization. Different tools classified the sequences according to different parameters, like physicochemical characteristics, compositional amino acids, family signatures, and hidden Markov models (HMM), with complementary approaches regarding accuracy, capacity of prediction, and training set to construct the program (Joseph et al. 2012, Waghu et al. 2016a, Meher et al. 2017).

The cecropin family in CAMPr3 had 17 of 18 sequences with a positive prediction with four different tools (Random forest, support vector machine, discriminant classifier, artificial neural network) compared to iAmPred, which shows lower scores. iAmPred and Campr3 predicted high antimicrobial probability to defensins (groups A and B) and attacins; and low antimicrobial probability to cecropins and coleoptericins. ClassAmp2 predicted a high probability for all the families involved (Fig. 14). This high probability prediction for defensins and low activity prediction for cecropins may be related with the fact that β-defensins had been reported in a vast range of organisms and, thus, have a higher diversity and representation in the databases and training libraries, whereas, cecropins had been found mostly in insects (limited taxon) with a low level of representation for some groups, like Coleoptera (Waghu et al. 2016b).

In general, CAMPr3 has a higher prediction probability compared with iAMPpred. These results may be due to the greater data set used to construct the prediction in CAMPr3, including family signatures, compared with iAMPred; nevertheless, iAMPpred includes a wide diversity of physico-chemical parameters and compositional AA parameters. The cecropin family has a lower probability in iAMPpred, compared with CAMPr3; this result may be explained by the underrepresentation of the cecropin family of Coleoptera in the databases and the relative divergence of the Scarabaeidae cecropins compared with other taxa (Fig. 3). This also may explain the low prediction probabilities for the coleoptericins in general.

Conclusion and Perspectives

The Scarabaeidae family is a complex, and diverse group of insects that have evolved in a wide range of habitats and environmental conditions, these adaptations make them a rich source of new and diverse antimicrobial peptides that can be identified by transcriptomic approaches. This new diversity of antimicrobial peptides is a generous source to explore their antimicrobial activities, as well as different biological processes like wound healing and immunomodulation.

Supplementary Data

Supplementary data are available at *Journal of Insect Science* online.

Acknowledgments

We thank the students from the 2017 seedbed of research at Universidad del Quindío Jean Carlos Sanchez Rojas, Luz Dahiana Erazo Urquijo, Alejandra Vélez García, Helen Vanessa Pinchao Salgar, and Angélica María Madrid Ortiz. This work was funded by Colombia's Ministry of Science, Technology, and Innovation (MINCIENCIAS), through grant number 111374455390, call 744-2016, project number 732; and, grant number 111384466847, call 844-2019 and project 859-2019. General benefits system BPIN code 2020000100077, JFOM was funded by MINCIENCIAS postdoctoral fellowship call 784-2017, contract number FP44842-250-2018.

Author Contributions

G.A.T.R.: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Visualization, Writing—Original Draft Preparation, Writing—Review and Editing. J.F.O.-M.: Visualization, Writing—Original Draft Preparation, Writing—Review and Editing. L.J.T.S.: Conceptualization, Investigation, Methodology, Project Administration, Writing—Original Draft Preparation, Writing—Review and Editing. J.F.C.: Writing—Original Draft Preparation. M. R.-M.: Visualization, Writing—Original Draft Preparation. J.C.C.O.: Conceptualization, Project Administration, Writing—Original Draft Preparation, Writing—Review and Editing.

References Cited

Afgan, E., D. Baker, B. Batut, M. van den Beek, D. Bouvier, M. Cech, J. Chilton, D. Clements, N. Coraor, B. A. Grüning, et al. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 46: W537–W544.

Alencar-Silva, T., M. C. Braga, G. O. S. Santana, F. Saldanha-Araujo, R. Pogue, S. C. Dias, O. L. Franco, and J. L. Carvalho. 2018. Breaking the frontiers of cosmetology with antimicrobial peptides. Biotechnol. Adv. 36: 2019–2031.

Altincicek, B., and A. Vilcinskas. 2007. Identification of immune-related genes from an apterygote insect, the firebrat *Thermobia domestica*. Insect Biochem. Mol. Biol. 37: 726–731.

Altincicek, B., A. Elashry, N. Guz, F. M. Grundler, A. Vilcinskas, and H. W. Dehne. 2013. Next generation sequencing based transcriptome analysis of septic-injury responsive genes in the beetle *Tribolium castaneum*. PLoS One 8: e52004.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403–410.

Ando, K., and S. Natori. 1988. Molecular cloning, sequencing, and characterization of cDNA for sarcotoxin IIA, an inducible ntibacterial protein of *Sarcophaga peregrina* (flesh fly). Biochemistry. 27: 1715–1721.

Araújo, C. A., P. J. Waniek, P. Stock, C. Mayer, A. M. Jansen, and G. A. Schaub. 2006. Sequence characterization and expression patterns of defensin and lysozyme encoding genes from the gut of the reduviid bug *Triatoma brasiliensis*. Insect Biochem. Mol. Biol. 36: 547–560.

Bhadra, P., J. Yan, J. Li, S. Fong, and S. W. I. Siu. 2018. AmPEP: sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest. Sci. Rep. 8: 1697.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120.

Boman. 1998. Gene-encoded peptide antibiotics and the concept of innate immunity: an update review. Scand. J. Immunol. 48: 15–25.

Boman, H. G. 1995. Peptide antibiotics and their role in innate immunity. Annu. Rev. Immunol. 13: 61–92.

Bouchard, P., Y. Bousquet, A. Davies, M. Alonso-Zarazaga, J. Lawrence, C. Lyal, A. Newton, C. Reid, M. Schmitt, A. Slipinski, et al. 2011. Family-group names in coleoptera (Insecta). ZooKeys. 88: 1–972.

Brady, D., A. Grapputo, O. Romoli, and F. Sandrelli. 2019. Insect cecropins, antimicrobial peptides with potential therapeutic applications. Int. J. Mol. Sci. 20: 5862.

Brogden, K. A., A. J. De Lucca, J. Bland, and S. Elliott. 1996. Isolation of an ovine pulmonary surfactant-associated anionic peptide bactericidal for *Pasteurella haemolytica*. Proc. Natl. Acad. Sci. USA 93: 412–416.

Buchan, D. W., F. Minneci, T. C. Nugent, K. Bryson, and D. T. Jones. 2013. Scalable web services for the PSIPRED protein analysis workbench. Nucleic Acids Res. 41: W349–W357.

Bulet, P., S. Cociancich, J. L. Dimarcq, J. Lambert, J. M. Reichhart, D. Hoffmann, C. Hetru, and J. A. Hoffmann. 1991. Insect immunity. Isolation from a coleopteran insect of a novel inducible antibacterial peptide and of new members of the insect defensin family. J. Biol. Chem. 266: 24520–24525.

Bulet, P., C. Hetru, J. L. Dimarcq, and D. Hoffmann. 1999. Antimicrobial peptides in insects; structure and function. Dev. Comp. Immunol. 23: 329–344.

Bulet, P., R. Stöcklin, and L. Menin. 2004. Anti-microbial peptides: from invertebrates to vertebrates. Immunol. Rev. 198: 169–184.

Carlsson, A., P. Engström, E. T. Palva, and H. Bennich. 1991. Attacin, an antibacterial protein from *Hyalophora cecropia*, inhibits synthesis of outer membrane proteins in Escherichia coli by interfering with omp gene transcription. Infect. Immun. 59: 3040–3045.

Cavallarin, L., D. Andreu, and B. San Segundo. 1998. Cecropin A—derived peptides are potent inhibitors of fungal plant pathogens. Mol. Plant-Microbe Interact. 11: 218–227.

Cociancich, S., P. Bulet, C. Hetru, and J. A. Hoffmann. 1994. The inducible antibacterial peptides of insects. Parasitol. Today 10: 132–139.

Cornet, B., J. M. Bonmatin, C. Hetru, J. A. Hoffmann, M. Ptak, and F. Vovelle. 1995. Refined three-dimensional solution structure of insect defensin A. Structure 3: 435–448.

Cultid-Medina, C. A., C. Medina, B. Martínez, F. Escobar, L. Constantino, and N. J. Betancur. 2012. Escarabajos coprófagos (Scarabaeinae) del eje cafetero: Guía para el estudio ecológico. Wildlife Conservation Society Colombia-CENICAFÉ y Federación Nacional de cafeteros. Villa María, Colombia. 196p.

Cultid-Medina, C. A., J. M. Lobo, C. Medina, F. González, F. Escobar, and P. Chacón-Ulloa. 2014. Completitud del inventario de escarabajos coprófagos (Coleoptera: Scarabaeinae) en la Ecorregión del Eje Cafetero, Colombia. Revista Colombiana de Entomología. 40(1).

DeLucca, A. J., J. M. Bland, T. J. Jacks, C. Grimm, T. E. Cleveland, and T. J. Walsh. 1997. Fungicidal activity of cecropin A. Antimicrob. Agents Chemother. 41: 481–483.

Dias, R. D. E. O., and O. L. Franco. 2015. Cysteine-stabilized αβ defensins: from a common fold to antibacterial activity. Peptides 72: 64–72.

Duwadi, D., A. Shrestha, B. Yilma, I. Kozlovski, M. Sa-Eed, N. Dahal, and J. Jukosky. 2018. Identification and screening of potent antimicrobial peptides in arthropod genomes. Peptides 103: 26–30.

Engström, P., A. Carlsson, A. Engström, Z. J. Tao, and H. Bennich. 1984a. The antibacterial effect of attacins from the silk moth Hyalophora cecropia is directed against the outer membrane of *Escherichia coli*. EMBO J. 3: 3347–3351.

Engström, A., P. Engström, Z. J. Tao, A. Carlsson, and H. Bennich. 1984b. Insect immunity. The primary structure of the antibacterial protein attacin F and its relation to two native attacins from *Hyalophora cecropia*. EMBO J. 3: 2065–2070.

Faruck, M. O., F. Yusof, and S. Chowdhury. 2016. An overview of antifungal peptides derived from insect. Peptides 80: 80–88.

Gómez, E. A., P. Giraldo, and S. Orduz. 2017. InverPep: a database of invertebrate antimicrobial peptides. J. Glob. Antimicrob. Resist. 8: 13–17.

Greenfield, N. J. 2006. Using circular dichroism spectra to estimate protein secondary structure. Nat. Protoc. 1: 2876–2890.

Gunne, H., M. Hellers, and H. Steiner. 1990. Structure of preproattacin and its processing in insect cells infected with a recombinant baculovirus. Eur. J. Biochem. 187: 699–703.

Hoffmann, J. A. 1995. Innate immunity of insects. Curr. Opin. Immunol. 7: 4–10.

Hoffmann, J. A., and C. Hetru. 1992. Insect defensins: inducible antibacterial peptides. Immunol. Today 13: 411–415.

Hoffmann, J. A., J. M. Reichhart, and C. Hetru. 1996. Innate immunity in higher insects. Curr. Opin. Immunol. 8: 8–13.

Hultmark, D., H. Steiner, T. Rasmuson, & H. G. Boman. 2005. Insect immunity. Purification and Properties of three inducible bactericidal proteins from hemolymph of immunized pupae of hyalophora cecropia. Eur. J. Biochem. 106: 7–16.

Hunt, T., J. Bergsten, Z. Levkanicova, A. Papadopoulou, O. S. John, R. Wild, P. M. Hammond, D. Ahrens, M. Balke, M. S. Caterino, et al. 2007. A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. Science 318: 1913–1916.

Hwang, J.-S., J. Lee, Y.-J. Kim, H.-S. Bang, E.-Y. Yun, S.-R. Kim, H.-J. Suh, B.-R. Kang, S.-H. Nam, J.-P. Jeon, et al. 2009. Isolation and characterization of a defensin-like peptide (coprisin) from the dung beetle, *Copris tripartitus*. Int. J. Pept. 2009: 1–5.

Imamura, M., S. Wada, K. Ueda, A. Saito, N. Koizumi, H. Iwahana, and R. Sato. 2009. Multipeptide precursor structure of acaloleptin A isoforms, antibacterial peptides from the Udo longicorn beetle, *Acalolepta luxuriosa*. Dev. Comp. Immunol. 33: 1120–1127.

Imler, J.-L., and P. Bulet. 2005. Antimicrobial peptides in drosophila: structures,activities and gene regulation, pp. 1–21. *In* D. Kabelitz, and J.-M. Schröder (eds.), Chemical immunology and allergy. Karger, Basel, Switzerland.

Jacobs, C. G. C., J. D. Gallagher, S. E. F. Evison, D. G. Heckel, A. Vilcinskas, and H. Vogel. 2017. Endogenous egg immune defenses in the yellow mealworm beetle (*Tenebrio molitor*). Dev. Comp. Immunol. 70: 1–8.

Jacobs, C. G., S. Steiger, D. G. Heckel, N. Wielsch, A. Vilcinskas, and H. Vogel. 2016. Sex, offspring and carcass determine antimicrobial peptide expression in the burying beetle. Sci. Rep. 6: 25409.

Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30: 1236–1240.

Joseph, S., S. Karnik, P. Nilawe, V. K. Jayaraman, and S. Idicula-Thomas. 2012. ClassAMP: a prediction tool for classification of antimicrobial peptides. IEEE/ACM Trans. Comput. Biol. Bioinform. 9: 1535–1538.

Källberg, M., H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu, and J. Xu. 2012. Template-based protein structure modeling using the RaptorX web server. Nat. Protoc. 7: 1511–1522.

Kockum, K., I. Faye, P. V. Hofsten, J. Y. Lee, K. G. Xanthopoulos, and H. G. Boman. 1984. Insect immunity. Isolation and sequence of two cDNA clones corresponding to acidic and basic attacins from *Hyalophora cecropia*. EMBO J. 3: 2071–2075.

Lai, R., H. Liu, W. Hui Lee, and Y. Zhang. 2002. An anionic antimicrobial peptide from toad *Bombina maxima*. Biochem. Biophys. Res. Commun. 295: 796–799.

Lee, S. Y., H. J. Moon, S. Kurata, T. Kurama, S. Natori, and B. L. Lee. 1994. Purification and molecular cloning of cDNA for an inducible antibacterial protein of larvae of a coleopteran insect, *Holotrichia diomphalia*. J. Biochem. (Tokyo). 115: 82–86.

Lee, J., D. Lee, H. Choi, H. H. Kim, H. Kim, J. S. Hwang, D. G. Lee, and J. I. Kim. 2014. Structure-activity relationships of the intramolecular disulfide bonds in coprisin, a defensin from the dung beetle. BMB Rep. 47: 625–630.

Lee, J., K. Bang, S. Hwang, and S. Cho. 2016. cDNA cloning and molecular characterization of a defensin-like antimicrobial peptide from larvae of *Protaetia brevitarsis seulensis* (Kolbe). Mol. Biol. Rep. 43: 371–379.

Lee, J. H., M. Seo, H. J. Lee, M. Baek, I. W. Kim, S. Y. Kim, M. A. Kim, S. H. Kim, and J. S. Hwang. 2019. Anti-inflammatory activity of antimicrobial peptide allomyrinasin derived from the dynastid beetle, *Allomyrina dichotoma*. J. Microbiol. Biotechnol. 29: 687–695.

Login, F. H., S. Balmand, A. Vallier, C. Vincent-Monégat, A. Vigneron, M. Weiss-Gayet, D. Rochat, and A. Heddi. 2011. Antimicrobial peptides keep insect endosymbionts under control. Science. 334: 362–365.

Lowenberger, C. A., C. T. Smartt, P. Bulet, M. T. Ferdig, D. W. Severson, J. A. Hoffmann, and B. M. Christensen. 1999. Insect immunity: molecular cloning, expression, and characterization of cDNAs and genomic DNA encoding three isoforms of insect defensin in *Aedes aegypti*. Insect Mol. Biol. 8: 107–118.

Lu, W. 2019. Antimicrobial peptides. Semin. Cell Dev. Biol. 88: 105–106.

Marshall, S. H., and G. Arenas. 2003. Antimicrobial peptides: a natural alternative to chemical antibiotics and a potential for applied biotechnology. Electron. J. Biotechnol. 6(3).

Meher, P. K., T. K. Sahu, V. Saini, and A. R. Rao. 2017. Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC. Sci. Rep. 7: 42362.

Memarpoor-Yazdi, M., H. Zare-Zardini, and A. Asoodeh. 2013. A novel antimicrobial peptide derived from the insect paederus dermatitis. Int. J. Pept. Res. Ther. 19: 99–108.

Meng, E., J. Li, B. Tang, Y. Hu, T. Qiao, Y. Hou, Y. Lin, and Z. Chen. 2019. Alteration of the phagocytosis and antimicrobial defense of *Octodonta nipae* (Coleoptera: Chrysomelidae) pupae to *Escherichia coli* following parasitism by *Tetrastichus brontispae* (Hymenoptera: Eulophidae). Bull. Entomol. Res. 109: 248–256.

Mylonakis, E., L. Podsiadlowski, M. Muhammed, and A. Vilcinskas. 2016. Diversity, evolution and medical applications of insect antimicrobial peptides. Philos. Trans. R. Soc. B Biol. Sci. 371: 20150290.

Namara, L. M., C. T. Griffin, D. Fitzpatrick, K. Kavanagh, and J. C. Carolan. 2018. The effect of entomopathogenic fungal culture filtrate on the immune response and haemolymph proteome of the large pine weevil, *Hylobius abietis*. Insect Biochem. Mol. Biol. 101: 1–13.

Nielsen, H. 2017. Predicting secretory proteins with SignalP, pp. 59–73. *In* D. Kihara (ed.), Protein Funct. Predict., methods in molecular biology. Springer, New York, NY.

Pane, K., L. Durante, O. Crescenzi, V. Cafaro, E. Pizzo, M. Varcamonti, A. Zanfardino, V. Izzo, A. Di Donato, and E. Notomista. 2017. Antimicrobial potency of cationic antimicrobial peptides can be predicted from their amino acid composition: application to the detection of "cryptic" antimicrobial peptides. J. Theor. Biol. 419: 254–265.

Pasupuleti, M., A. Schmidtchen, and M. Malmsten. 2012. Antimicrobial peptides: key components of the innate immune system. Crit. Rev. Biotechnol. 32: 143–171.

Pei, Z., X. Sun, Y. Tang, K. Wang, Y. Gao, and H. Ma. 2014. Cloning, expression, and purification of a new antimicrobial peptide gene from Musca domestica larva. Gene. 549: 41–45.

Peschel, A., R. W. Jack, M. Otto, L. V. Collins, P. Staubitz, G. Nicholson, H. Kalbacher, W. F. Nieuwenhuizen, G. Jung, A. Tarkowski, et al. **2001**. Staphylococcus aureus resistance to human defensins and evasion of neutrophil killing via the novel virulence factor MprF is based on modification of membrane lipids with l-lysine. J. Exp. Med. 193: 1067–1076.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. UCSF Chimera? A visualization system for exploratory research and analysis. J. Comput. Chem. 25: 1605–1612.

Purvis, A., and A. Hector. 2000. Getting the measure of biodiversity. Nature 405: 212–219.

Rodríguez-García, M. J., A. García-Reina, V. Machado, and J. Galián. 2016. Identification, structural characterisation and expression analysis of a defensin gene from the tiger beetle *Calomera littoralis* (Coleoptera: Cicindelidae). Gene 589: 56–62.

Sagisaka, A., A. Miyanoshita, J. Ishibashi, and M. Yamakawa. 2001. Purification, characterization and gene expression of a glycine and proline-rich antibacterial protein family from larvae of a beetle, *Allomyrina dichotoma*: antibacterial proteins from *A. dichotoma*. Insect Mol. Biol. 10: 293–302.

Saikia, K., V. K. Belwal, D. Datta, and N. Chaudhary. 2019. Aromatic-rich C-terminal region of LCI is a potent antimicrobial peptide in itself. Biochem. Biophys. Res. Commun. 519: 372–377.

Saito, A., K. Ueda, M. Imamura, S. Atsumi, H. Tabunoki, N. Miura, A. Watanabe, M. Kitami, and R. Sato. 2005. Purification and cDNA cloning of a cecropin from the longicorn beetle, *Acalolepta luxuriosa*. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 142: 317–323.

Sayers, E. W., T. Barrett, D. A. Benson, E. Bolton, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, S. Federhen, et al. 2011. Database resources of the national center for biotechnology information. Nucleic Acids Res. 39: D38–D51.

Shafee, T. M., F. T. Lay, T. K. Phan, M. A. Anderson, and M. D. Hulett. 2017. Convergent evolution of defensin sequence, structure and function. Cell. Mol. Life Sci. 74: 663–682.

Shin, H. S., and S.-I. Park. 2019. Novel attacin from *Hermetia illucens* : cDNA cloning, characterization, and antibacterial properties. Prep. Biochem. Biotechnol. 49: 279–285.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. Mol. Syst. Biol. 7: 539.

Sun, S. C., I. Lindström, J. Y. Lee, and I. Faye. 1991. Structure and expression of the attacin genes in *Hyalophora cecropia*. Eur. J. Biochem. 196: 247–254.

Tarr, D. E. 2016. Establishing a reference array for the CS-αβ superfamily of defensive peptides. BMC Res. Notes 9: 490.

Thrash, A., M. Arick, 2nd, and D. G. Peterson. 2018. Quack: a quality assurance tool for high throughput sequence data. Anal. Biochem. 548: 38–43.

Tomie, T., J. Ishibashi, S. Furukawa, S. Kobayashi, R. Sawahata, A. Asaoka, M. Tagawa, and M. Yamakawa. 2003. Scarabaecin, a novel cysteine-containing antifungal peptide from the rhinoceros beetle, *Oryctes rhinoceros*. Biochem. Biophys. Res. Commun. 307: 261–266.

Toro Segovia, L. J., G. A. Téllez Ramírez, D. C. Henao Arias, J. D. Rivera Duran, J. P. Bedoya, and J. C. Castaño Osorio. 2017. Identification and characterization of novel cecropins from the *Oxysternon conspicillatum* neotropic dung beetle. PLoS One 12: e0187914.

Vajda, S., C. Yueh, D. Beglov, T. Bohnuud, S. E. Mottarella, B. Xia, D. R. Hall, and D. Kozakov. 2017. New additions to the ClusPro server motivated by CAPRI. Proteins 85: 435–444.

Vilcinskas, A., K. Mukherjee, and H. Vogel. 2013. Expansion of the antimicrobial peptide repertoire in the invasive ladybird *Harmonia axyridis*. Proc. Biol. Sci. 280: 20122113.

Vogel, H., C. Badapanda, and A. Vilcinskas. 2011. Identification of immunity-related genes in the burying beetle *Nicrophorus vespilloides* by suppression subtractive hybridization: burying beetle immunity. Insect Mol. Biol. 20: 787–800.

Volkoff, A. N., J. Rocher, E. d'Alençon, M. Bouton, I. Landais, E. Quesada-Moraga, A. Vey, P. Fournier, K. Mita, and G. Devauchelle. 2003. Characterization and transcriptional profiles of three *Spodoptera frugiperda* genes encoding cysteine-rich peptides. A new class of defensin-like genes from lepidopteran insects? Gene 319: 43–53.

Waghu, F. H., R. S. Barai, P. Gurung, and S. Idicula-Thomas. 2016a. CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. Nucleic Acids Res. 44: D1094–D1097.

Waghu, F. H., R. S. Barai, and S. Idicula-Thomas. 2016b. Leveraging family-specific signatures for AMP discovery and high-throughput annotation. Sci. Rep. 6: 24684.

Wang, G., (Ed.). 2010. Antimicrobial peptides: discovery, design and novel therapeutic strategies. CABI, Wallingford, United Kingdom.

Wang, L., Z. Li, C. Du, W. Chen, and Y. Pang. 2007. Characterization and expression of a cecropin-like gene from *Helicoverpa armigera*. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 148: 417–425.

Wang, J., C. Hu, Y. Wu, A. Stuart, C. Amemiya, M. Berriman, A. Toyoda, M. Hattori, and S. Aksoy. 2008. Characterization of the antimicrobial peptide attacin loci from *Glossina morsitans*. Insect Mol. Biol. 17: 293–302.

Wang, S., J. Peng, and J. Xu. 2011. Alignment of distantly related protein structures: algorithm, bound and implications to homology modeling. Bioinformatics 27: 2537–2545.

Wang, S., J. Ma, J. Peng, and J. Xu. 2013. Protein structure alignment beyond spatial proximity. Sci. Rep. 3: 1448.

Wang, G., X. Li, and Z. Wang. 2016. APD3: the antimicrobial peptide database as a tool for research and education. Nucleic Acids Res. 44: D1087–D1093.

Wiesner, J., and A. Vilcinskas. 2010. Antimicrobial peptides: the ancient arm of the human immune system. Virulence 1: 440–464.

Wu, Q., J. Patočka, and K. Kuča. 2018. Insect antimicrobial peptides, a mini review. Toxins 10: 461.

Xiao, X., P. Wang, W. Z. Lin, J. H. Jia, and K. C. Chou. 2013. iAMP-2L: a two-level multi-label classifier for identifying antimicrobial peptides and their functional types. Anal. Biochem. 436: 168–177.

Xu, J., and Y. Zhang. 2010. How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics 26: 889–895.

Yi, H. Y., M. Chowdhury, Y. D. Huang, and X. Q. Yu. 2014. Insect antimicrobial peptides and their applications. Appl. Microbiol. Biotechnol. 98: 5807–5822.

Zhang, L. J., and R. L. Gallo. 2016. Antimicrobial peptides. Curr. Biol. 26: R14–R19.

Zhou, X., L. Y. Peng, Z. C. Wang, W. Wang, Z. Zhu, X. H. Huang, L. B. Chen, Q. S. Song, and Y. Y. Bao. 2019. Identification of novel antimicrobial peptides from rice planthopper, *Nilaparvata lugens*. Insect Biochem. Mol. Biol. 113: 103215.

Zhu, S., and B. Gao. 2013. Evolutionary origin of β-defensins. Dev. Comp. Immunol. 39: 79–84.

Zou, Z., J. D. Evans, Z. Lu, P. Zhao, M. Williams, N. Sumathipala, C. Hetru, D. Hultmark, and H. Jiang. 2007. Comparative genomic analysis of the *Tribolium* immune system. Genome Biol. 8: R177.