Identification of new insect Host Defense Peptides (HDP) from Dung Beetles (Coleoptera: Scarabaeidae) transcriptomes

New insect Host Defense Peptides from Dung Beetles

Author's names and affiliations:

Germán Alberto Téllez Ramirez^{1*}

<https://orcid.org/0000-0002-3057-9290>

Center of Biomedical Research. Group of Molecular Immunology. Universidad del Quindío. Cra. 15 calle 12 norte. Armenia, Quindío – Colombia.

Author Contributions: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing.

Juan Felipe Osorio-Méndez¹

<https://orcid.org/0000-0002-7326-8871>

Center of Biomedical Research. Group of Molecular Immunology. Universidad del Quindío. Cra. 15 calle 12 norte. Armenia, Quindío – Colombia.

Author Contributions: Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Diana Carolina Henao Arias¹,

<https://orcid.org/0000-0001-8136-1059>

Center of Biomedical Research. Group of Molecular Immunology. Universidad del Quindío. Cra. 15 calle 12 norte. Armenia, Quindío – Colombia.

Author Contributions: Investigation, Methodology, Visualization, Writing – Original Draft Preparation.

26 **Lily Johanna Toro S¹,**

27 <http://orcid.org/0000-0001-7906-1997>

28 Center of Biomedical Research. Group of Molecular Immunology. Universidad del

29 Quindío. Cra. 15 calle 12 norte. Armenia, Quindío – Colombia.

30 Author Contributions: Conceptualization, Investigation, Methodology, Project

31 Administration, Writing – Original Draft Preparation, Writing – Review & Editing.

32 **Juliana Franco Castrillón¹**

33 <https://orcid.org/0000-0003-4942-5096>

34 Center of Biomedical Research. Group of Molecular Immunology. Universidad del

35 Quindío. Cra. 15 calle 12 norte. Armenia, Quindío – Colombia.

36 Author Contributions: Writing – Original Draft Preparation.

37 **Maribel Rojas-Montoya¹,**

38 <https://orcid.org/0000-0002-0966-7165>

39 Center of Biomedical Research. Group of Molecular Immunology. Universidad del

40 Quindío. Cra. 15 calle 12 norte. Armenia, Quindío – Colombia.

41 Author Contributions: Visualization, Writing – Original Draft Preparation.

42 **Jhon Carlos Castaño Osorio¹**

43 <https://orcid.org/0000-0002-7639-3053>

44 Center of Biomedical Research. Group of Molecular Immunology. Universidad del

45 Quindío. Cra. 15 calle 12 norte. Armenia, Quindío – Colombia.

46 Author Contributions: Conceptualization, Project Administration, Writing – Original

47 Draft Preparation, Writing – Review & Editing.

48 **E-mail address:** gymol@uniquindio.edu.co

49 ***corresponding author: Email address:** gatellez@uniquindio.edu.co

50

Abstract

Introduction: The Coleoptera Scarabaeidae family is one of the most diverse groups of insects on the planet, living in complex microbiological environments. Their immune systems have evolved with the generation of host defense peptides but only a small number of these peptides have been characterized.

Methods: In this work two sources of information were retrieved: 1) *De novo* transcriptomic data from two species of neotropical Scarabidae (*Dichotomius satanas* and *Ontophagus curvicornis*); 2) Sequence data deposited in available databases. A BLAST-based search against the transcriptomes with a subset of sequences representative of the Host Defense Peptides (HDP). The HDP were described with the cecropin, defensin, attacin, and coleopteracin families; their physical/chemical and structural properties were described.

Results: This work reports 155 novel sequences of HDP identified in 9 transcriptomes from seven species from the Coleoptera order: *D. satanas* (n= 76; 49.03%), *O. curvicornis* (n= 23; 14.83%), *T. dichotomus* (n= 18; 11.61%), *O. nigriventris* (n= 10; 6.45%), *Heterochelus sp* (n= 6; 3.87%), *O. conspicillatum* (n= 18; 11.61%) and *P. 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 coleopteracins (n= 38; 24.52%), with their physicochemical, structural characteristics, and sequence relationship to other insect HDP were analyzed.

Conclusions:

155 new HDP could be identified, based on similarity to known HDP insect families on nine transcriptome sequences of seven beetle species.

Keywords:

Antimicrobial cationic peptides, insect proteins, computational biology, RNA-Seq, cecropins, defensins, attacin, coleopteracin.

Highlights

- This work identified 155 novel sequences of HDP found in nine transcriptomes from seven Coleoptera species.
- *De novo* transcriptomic data from two species of neotropical Scarabaeidae (*Dichotomius satanas* and *Ontophagus curvicornis*).
- *In silico* prediction of physicochemical properties, structural features, sequence similarity, and antimicrobial activity of Scarabaeidae HDP.

Introduction

One of the main effectors of the insect 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 [1–3]. Host Defense Peptides are cationic, amphipathic polypeptides, produced in all known genera of living organisms, and represent an ancient innate defense mechanism [4–6]. Once activated by post-translational proteolysis [1,7–9], they act as effector molecules against pathogens in a broad spectrum of antimicrobial activity against Gram-positive and Gram-negative bacteria, protozoa, fungi, and viruses. They also have low propensity

for developing resistance. This efficiency is thought to be one of the biological attributes that would explain the evolutionary success of insects [10,11]. Therefore, they have received attention for the development of new antimicrobials with clinical applications in recent years [12,13].

Insect HDP have been classified according to their sequence, physicochemical and structural properties, in cecropin, defensin, attacin, and coleopteracin families. Other families, such as moricin and gloverin have been identified only in Lepidoptera [2,14,15]. Cecropins are a family of 3–4 kDa, cationic, alpha helix, amphipathic peptides devoid of cysteine residues [16,17]. Mature active cecropins are generated after removal of the signal peptide and form two α -helices connected by a hinge. A long hydrophobic C-terminal and a strongly basic N-terminal domain is presumptively required for the biological activity and they are most active against Gram-negative bacteria [18–22].

Defensins are the largest family of HDP and are ubiquitous in almost all forms of life, including animals, fungi, and plants [23]. They have broad-spectrum antimicrobial activity against bacteria, fungi, and viruses [24]. The majority of mature defensins are cationic peptides composed of 24 to 42 amino acid residues, characterized by six cysteines [25]. The structure of insect defensins is composed of an α -helix, followed by an antiparallel β -sheet linked by three intramolecular disulfide bonds, forming a “Cysteine-Stabilized alpha Beta (CS $\alpha\beta$)” or “loop-helix-beta-sheet” structure [26,27].

Attacins are larger peptides with basic (~ 8.3) or acidic (~ 5.7) isoelectric points, and molecular weights of 20–23 kDa; we can find two isoforms, basic and acids attacins

[28]. Their secondary structure is composed of a hydrophobic alpha helix similar to glycine-rich peptides [29]. They can inhibit growth of Gram-negative bacteria, and the synthesis of bacterial proteins like OmpC, OmpF, OmpA and LamB [30,31].

Coleoptericsins contain approximately 70 amino acid residues and are characterized as glycine and proline-rich antimicrobial peptides that are bactericidal against Gram-positive and Gram-negative bacteria [32]. There are two subgroups, one positively and other negatively charged, and their C-terminus has a basic nature [32,33].

With approximately one-million characterized species, insects represent the largest class within the animal kingdom and Coleoptera are the most diverse order and is the most diverse in insects [34–36]. Nevertheless, only 305 of the 3070 HDP sequences deposited in the Antimicrobial Peptide Database (APD) are derived from insects [37]. To date some HDP have been described in the Coleoptera order, as is the case of *Allomyrina dichotoma* [38], *Octodonta nipae* [39], *Hylobius abietis* [40], *Nicrophorus vespilloides* [41], *Tenebrio molitor* [42], *Calomera littoralis* [43], *Protaetia brevitarsis seulensis* [44], *Tribolium castaneum* [45–47], *Holotrichia diomphalia* [48], *Zophobas atratus* [33], *Allomyrina dichotoma* [32], *Acalolepta luxuriosa* [49], and *Sitophilus oryzae* [50]. 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 those of the Coprisin, Oxysterlin, and Scarabaeicin families [51–57]. Therefore, we propose 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 [58–61].

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 m.a.s.l. The beetles captured were identified as *Dichotomius satanas* and *Ontophagus curvicornis* with the Cultid-Medina, 2012 taxonomic key [58]. Once collected, they were maintained in a terrarium with organic soil and human feces bait for 12 h. Then, they were separated into two groups of five individuals each. One group was inoculated in the ventral lateral abdomen with 10 µL of a pool of 1X10⁶ UFC/mL formalin-fixed bacteria (*E.coli* and *Staphylococcus aureus*) and fungi (*Candida albicans*). The other group was used as an untreated control. Finally, the fat body and part of the hindgut were dissected 12 h post-inoculation.

Total RNA extraction, transcriptome sequencing, and *de novo* assembly

Total RNA was extracted by using Ambion total RNA extraction kit (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 transcriptome was sequenced with Illumina Hiseq and pair-end read with 150bp length. The FASTAQ files were checked by FastQC [62], the trimming was done by

Trimmomatic V0.36 [63], the 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 (<https://galaxy.ncgas-trinity.indiana.edu/>) [64].

Transcriptome shotgun assemblies

Transcriptome shotgun assemblies (TSA) 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/>). The fasta files of the assembled transcriptomes were converted into BLAST databases with CLC main workbench software 7.9.1.

Homology identification of HDP

The queries from the Cecropin family were constructed from Oxysterlins [53] and different sequences related to the HDP InterPro families (Cecropin: IPR020400; Defensin: IPR017982; Coleopteracin: IPR009382, and Attacin: IPR005520 IPR005521) and a multi-TBLASTn search was constructed and the result sequences were filtered according to the E-score ≤ 0.01 and the open reading frame (ORF) was identified for each sequence [65]. The list of sequence codes identified in the tBLASTn was extracted from the different transcriptomes and the related amino-acid sequence

was identified by using the ORF finder tool (<https://www.ncbi.nlm.nih.gov/orffinder/>) [66].

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/>) [67]. The physicochemical 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>) calculation and prediction tool [37]. Prediction of the antimicrobial function was conducted with the SVMC, RFC, and DAC algorithms available in Campr3 [68], Classamp [69], and iAmpred tools [37].

Structural analysis

The secondary structure prediction was done with the Psipred (<http://bioinf.cs.ucl.ac.uk/psipred/>) server [70]. Prediction of functional domains was done with Interpro (<https://www.ebi.ac.uk/interpro/>) [71]. The tertiary structure was predicted with RAPTOR X (<http://raptorx.uchicago.edu/StructurePrediction/predict/>) [72], and structural alignments were done with the 3Dcomb V1.18 tool [73]. All the models were visualized in UCSF Chimera V1.13.1. [74]. The protein-protein interactions of the sequences modeled were constructed by using the Cluspro server [75]. Illustration of the general structural properties for each family was made by DOG V1.0 [76].

218

219 **Similarity dendrogram**

220 The sequences of InterPro families (cecropin: IPR020400; insect Defensin:
221 IPR017982; Coleopteracin: IPR009382 Attacin: IPR005520 IPR005521) and the
222 taxonomic key corresponding to each sequence were downloaded from the PIR batch
223 server (<https://pir.georgetown.edu/pirwww/search/batch.shtml>). The signal and
224 propeptide were identified and removed and the mature peptides were aligned with
225 the HDP from the Scarabaeidae with Clustal Omega
226 (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) or MUSCLE [77]. Dendrograms were
227 generated by Neighbor-joining with Jukes-Cantor protein distance and bootstrap with
228 10000 replicates in CLC main workbench V7.9.1.

229

230 **Ethics statement**

231 This work was approved by the Universidad del Quindío bioethics committee under
232 act number 8 of 06 May 2016. The contract of access to genetic resources was drawn
233 through resolution N° 120 of 22 October 2015 with the Colombian Ministry of
234 Environment and Sustainable Development.

235

236 **Results and Discussion**

237 **Transcriptome shotgun assemblies (TSA)**

238 Traditional strategies for HPD identification and characterization involve biochemical
239 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 with RACE-PCR (rapid amplification of cDNA ends) [78]. Artificial neural network algorithms have also been trained with information about structural and physicochemical characteristics to identify novel HDP sequences with proteomic and genomic methodologies [79]. Recent developments in high-throughput sequencing technologies have represented a novel and efficient method for gene identification [80]. Transcriptome-based approaches using next-generation sequencing are particularly useful because they focus on the expressed (*i.e.* exomic) portion of the genome. This information can be exploited by using bioinformatic tools to search for target genes/proteins using a step-by-step selection strategy [81]. In this work, the transcriptomes of two Scarabaeidae species (*O. curvicornis* and *D. satanas*) were sequenced, assembled and submitted to the DDB/EMBL/GenBank database under the accession codes GHMD000000000 - GHMA000000000, bio-sample: SAMN10614998 - SAMN10614917 and bio-project: PRJNA510790 - PRJNA510790, respectively. The transcriptome characteristics used for this work are shown in **Table 1**.

Table 1. TSA from Scarabaeidae.

Prefix (TSA code)	Organism	Bio project code	Biosample code	Source	Contigs number
-------------------	----------	------------------	----------------	--------	----------------

GAQV01	<i>Trypoxylus dichotomus</i>	PRJNA231720	SAMN02444008	3rd instar larval and prepupal	34455
			SAMN02444009		
			SAMN02444010		
			SAMN02444011		
			SAMN02444012		
			SAMN02444013		
			SAMN02444014		
			SAMN02444015		
			SAMN02444016		
			SAMN02444017		
			SAMN02444018		
			SAMN02444019		
IABQ01	<i>Trypoxylus dichotomus tunobosonis</i>	PRJDB4830	SAMD00051587	3rd larvae	30157
GAQW01	<i>Onthophagus nigriventris</i>	PRJNA231725	SAMN02444020	3rd instar larval and prepupal	59302
			SAMN02444021		
			SAMN02444022		
			SAMN02444023		
GARJ01	<i>Popillia japonica</i>	PRJNA233626	SAMN02569976	Antenna	698
GARK01	<i>Popillia japonica</i>	PRJNA198730	SAMN02055564	Grub	1916
GDNJ01	<i>Heterochelus</i> sp. AD-2015	PRJNA286531	SAMN03799575	Tissue	50435
GEXM01	<i>Oxysternon conspicillatum</i>	PRJNA339294	SAMN05589108	Adult fat body	27567
GHMA01	<i>Dichotomius satanas</i>	PRJNA510790	SAMN10614917	Adult fat body and intestine	463430

	<i>Onthophagus</i>			Adult fat body	
GHMD01	<i>curvicornis</i>	PRJNA510790	SAMN10614998	and intestine	172518

259

260 A tBLASTn search on the nine assembled transcriptomes was done with the
 261 constructed HDP sequence queries, the result contigs were retrieved and only those
 262 containing complete ORF were considered for further analysis. In total, 155 contigs
 263 encoding for potential HDP were identified. The number of HDP identified correlated
 264 with the size of the transcriptome and the condition of the sample (**Fig 1A**). These
 265 three species, *O. curvicornis*, *D. satanas*, and *O. conspiciellatum*, have most of the
 266 HDP. This may be explained by the fact that these three specimens had been obtained
 267 from the fat body of adult specimens previously subjected to stimulation to boost the
 268 insect immune response; these results agree with the different expression patterns of
 269 HDP with basal and inducible forms in different tissues and conditions [2,82].

270

271 **Fig 1.** Percentage of contigs in nine Scarabaeidae TSA and percentage of HDP
 272 identified by sequence homology. **A.** Percentage of contigs and HDP per species. **B.**
 273 Percentage of Scarabaeidae HDP families by dung beetle species.

274 Four HDP family distribution also varied according to species (**Fig 1B**) with members
 275 from these families identified in all species, with two exceptions *Popillia japonica* with
 276 only defensin peptides and *Onthophagus nigriventris* where the cecropin family was
 277 missing. *Heterochelus* sp and *O. nigriventris* have less number of defensins in contrast
 278 with *O. conspiciellatum*, *O. curvicornis*, *T. dichotomus*, and *D. satanas* wich have more
 279 attacins. The defensin HDP family has the majority of peptides representing 37.42 %
 280 (**Fig 1B**). The difference in the number of HDP found among species and type of HDP

may be explained by the different tissues sampled, size of the transcriptome, and life cycle stage of the individual sample, according to these differences, 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 [82].

This article, Identified 155 new putative HDP sequences from nine transcriptomes (**S1 appendix**), while another work with a similar approach using tBLASTn and BLASTp searches in newly sequenced arthropod genomes and expressed sequence tags (EST) derived from the red flour beetle (*Tribolium castaneum*), monarch butterfly (*Danaus plexippus*) and human body louse (*Pediculus humanus humanus*) identified six HDP [83]. One difference in the results may be explained by the use of the TSA database where the sequences may be more refined in the context of genome complexity with different introns and gene arrangements; another difference was the query construction, where sequences reported in APD were used and our work used a greater query search according to the InterPro related sequences.

Cecropin family description

Cecropins are the most abundant family of linear α -helical HDP in insects [84]. They have been identified in Hexapoda orders like Coleoptera, Diptera, and Lepidoptera [26]. Insect and non-insect cecropins are encoded by non-homologous genes. However, in holometabolous orders they have evolved only once [85]. Genes from this family frequently encode for tryptophans at the first or second position of the mature

peptide and lack cysteines. They are basic amphipathic peptides rich in cationic amino acids, with an N-terminal hydrophilic domain and a C-terminal hydrophobic domain. Structurally, they are characterized by two distinctive alpha-helical segments linked by a short hinge. In the post-translational level, they have a signal peptide and a frequent C-terminal amidation [16,17,19,26,86].

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 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 2**). 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 3**) [73,87,88]. We 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.

Fig 2. 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.

Fig 3. Structure of Scarabaeidae mature cecropins. A. Structural superposition constructed in DeepAlign. TMScore: 0.517 with Alpha helical conserved tertiary conformation. **B.** Ramachandran plot of the superimposed structures.

329

330 The insect cecropin dendrogram shows that the majority of sequences are from
331 Diptera and Lepidoptera orders as already described [85]. The dendrogram structure
332 has four main clades that are well related to the phylogenetic orders, a highly
333 conserved Diptera clade representing flies, Lepidoptera, Ascaridida, and a distant
334 share clade of Coleoptera and mosquitoes (**Fig 4**). Compared to other orders like
335 Diptera and Lepidoptera, cecropins from Scarabaeidae show more variation and
336 divergence among the sequences within the same order.

337

338 **Fig 4.** Neighbor-joining similarity dendrogram of invertebrate cecropins. Sequences
339 from the mature peptides were aligned by MUSCLE. The taxonomic distribution of the
340 sequences is indicated in the presented color code. Confidence values of the branches
341 were calculated with 10.000 bootstrap replicates. Thicker lines show branches with
342 bootstrap threshold value > 70.

343

344 According to the previously described cecropins in Coleoptera, there are few
345 representatives with Oxysterlins (1, 2, and 3) (*O. conspicillatum*), Cec (*Acalolepta*
346 *luxuriosa*), and Sarcotoxin Pd (*Paederus dermatitis*) [53,89,90]. Interestingly, the
347 Interpro data set, also has low representation of Coleopteran cecropins, with only two
348 representative sequences, *Acalolepta luxuriosa* (Q5W8G6) and *Agrilus planipennis*
349 (A0A1W4XF49). Additionally, the InterPro search fails to classify the Scarabaeidae
350 cecropin sequences as insect cecropins, indicating that, these sequences are different
351 from the InterPro signatures; nevertheless, as described, these new peptides had the
352 same physicochemical and structural characteristics of the insect cecropins [84].

353

Defensin family description

Insect defensins are inducible antibacterial peptides with a spectrum of activity focused mainly against Gram-positive bacteria, although activity against fungi and parasites has also been reported [91,92]. They are extensively distributed in nearly all life forms [24]. In insects, they have been found in Diptera, Hymenoptera, Coleoptera, Trichoptera, Hemiptera, and Odonata orders [14,91]. Defensins are 18 - 45 amino acids long with 6 - 8 conserved cysteine residues stabilized by three or four disulfide bonds [92,93]. Typically, insect defensins have the same cysteine pairing: Cys1-Cys4, Cys2-Cys5, and Cys3-Cys6. Two disulfide bonds connect the C-terminal β -sheet and the α -helix; and the third connects the N-terminal loop with the second β -sheet. This structural topology is known as cysteine-stabilized $\alpha\beta$ motif (CS $\alpha\beta$) and is common among defensin peptides across different organisms [34,94–96]. Once synthesized, pre-pro-defensins are proteolytically processed. First is the removal of the signal peptide to produce an inactive prodefensin; then, it suffers an additional cut of the propeptide by a furin-like enzyme in an R-X-[RK]-R conserved site to produce an active mature peptide [14,97].

This work identified 58 new defensin sequences in Scarabaeidae. As described for other insect defensins, those encoded by Scarabaeidae encode for a signal peptide followed by a propeptide (position 51-80 **Fig 5**) characterized by acid residues and an R-X-[RK]-R furin-like cleavage site (**Fig 5**). These sequences were classified into three groups (Group A, B, and C), according to their sequence, structural, and physicochemical properties.

Fig 5. Multiple-sequence alignment of defensin HDP found in Scarabaeidae. The

Scarabaeidae defensin groups are annotated in the left keys. The predicted signal peptide cleavage site is in position 50. Position 72-75 indicates the propeptide furin-like cleavage site.

Defensin group A with 31 sequences show 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. There is only one sequence (GHMA01_94621) with 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 6**).

Fig 6. 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).

Defensins group B is represented by 14 sequences. One key feature in this subgroup is the left-handed helix 6 to 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 (GHMA01_21200; GHMD01_63103) have one pair Cys1-Cys6 and loss predicted helix of the tertiary structure prediction. A hydrophobic random coil region was found in the C-terminal

end of the sequences (**Fig 6**). 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 [99,100]. 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 6**). These acid residues explain the negative charge of this group (**Fig 7**). 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 [98]. The anionic antimicrobial peptides, although rarely documented appear to complement the cationic antimicrobial peptides, offering a complete spectrum of antimicrobial peptides [99,100].

Fig 7. The net charge of the Scarabaeidae defensins predicted by APD. Statistical

differences were evaluated by ANOVA test with Tukey's multiple comparisons (p-value: ****<0.0001).

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 8**). 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 to 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 to be related to a single defensin encoded by Hymenoptera (VespidaeXP_014602557) and other Coleoptera species. These results are compatible with the idea that groups B and C are Scarabaeidae-specific groups of defensins with novel structural and physicochemical properties.

Fig 8. Neighbor-joining similarity dendrogram of invertebrate defensins. Sequences from the 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.

Attacin family description

Attacins are peptides with molecular masses of 20-23 kDa with a wide spectrum of antimicrobial activities, including Gram-positive and Gram-negative bacteria, fungi, and protozoan parasites. As described for defensins, attacins are synthesized as pre-proteins containing a signal peptide, and a conserved R-X-[RK]-R motif, which can be recognized by furin-like enzymes. The mature peptide is composed of an N-terminal attacin domain, followed by a glycine-rich segment. Attacins may be further divided according to basic and acid attacins [28,101–103].

In this study, we identified 41 new Scarabaeidae attacin sequences that fulfill the characteristics described (**Fig 9**). They are also recognized as sequences from this family by the Pfam signature Attacin-C (PF03769) [104,105]. The predicted mature sequences of the identified attacins can be further divided into two groups (named group A or B) according to the net charge (**Fig 10**). Compared to group B attacins, those belonging to group A are more cationic as they are highly enriched in positively charged residues. Additionally, group A attacins contain significantly more GNTS polar residues.

Fig 9. 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 to 60. Position 80 indicates the propeptide furin-like cleavage site.

Fig 10. Physicochemical characteristics of the Scarabaeidae attacins. **A.** Net charge and **B.** GNTS composition. Statistical differences were evaluated by ANOVA test with Tukey's multiple comparisons (p-value: ****<0.0001).

The secondary and tertiary structures predicted for Scarabaeidae attacins were characterized by a predominant antiparallel eight string beta-sheet configuration (**Fig 11**). 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 translocates protein substrates across bacterial membranes [106].

Fig 11. Structural superposition (left) and Ramachandran plot (right) of the mature attacins. **A.** Analysis of group A, TMscore=0.7366. **B.** Analysis of group B, TMscore=0.5844. The analysis was conducted in DeepAlign.

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 12**). 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 such as sodium and potassium [106]. 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 [101]. However, the distinctive feature of α -helical proteins in CD are negative bands at 208 and 222 nm [107], so the structure of this protein family remains an open question.

Fig 12. 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/(mole^e)). **B.** and **D.** Dimer protein-protein interaction prediction in ribbons chain coloring.

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 13**).

Fig 13. 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

526 presented. The confidence values of the branches were calculated with 10.000
527 bootstrap replicates. Thicker lines show branches with bootstrap threshold value > 70.
528
529

Coleopteracin family description

Coleopteracins are a group of antimicrobial peptides found exclusively in the Coleoptera order. The first coleopteracin was described in the tenebrionid beetle, *Zophobas atratus* [33]. This family is characterized by a signal peptide and a furin-like cleavage site with a mature peptide of 71 to 75 residues [108]. This group of HDP shows antimicrobial activity to Gram-positive and Gram-negative bacteria, with bacteriostatic and bactericidal activity at low and high concentrations respectively. 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 the bacteria [32]. A coleopteracin has recently been shown to play a central role in host-symbiont interactions in the weevil *S. zeamais* [50].

Coleopteracins encoded by Scarabaeidae (37 sequences) were identified as pre-pro-proteins with a furin-like cleavage site, they share the motif G-P-[GNS]-[KR]-[GSA]-K-P from position 97 to 103 (**Fig 14**). 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) (**Figs 15-16**); 2. Group B is a small group in Scarabaeidae (three sequences), but shows high similarity with already reported coleopteracins (**Fig 17**). They 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 16**); 3. Group C (17 sequences) is 57 to 59 amino acids in length and is enriched in acidic residues that

provide an mean positive net charge of 1.41. They have a relatively disordered structure with a coil-beta sheet-helix-coil configuration (**Figs 14-17**). In general, the conserved coleopteracin motif is structurally preceded by an alpha helix or related to the third beta loop in the structural conformation.

Fig 14. Multiple sequence alignment of Coleopteracin HDP found in Scarabaeidae. Position 23 in the alignment represents the cleavage site of the signal peptide predicted with SignalP. Positions 94 to 97 furin-like cleavage site.

Fig 15. The net charge of the Scarabaeidae Coleopteracins. Statistical differences were evaluated by ANOVA test with Tukey's multiple comparisons (p-value: ****<0.0001).

Fig 16. Structural superposition (left) and Ramachandran plot (right) of the mature coleopteracins. **A.** Analysis of group A, TMscore=0.2753. **B.** Analysis of group B, TMscore=0.6248, **C.** Analysis of group C, TMscore=0.3925. The analysis was conducted in DeepAlign.

Fig 17. Neighbor-joining similarity dendrogram of Coleopteracins. The taxonomic distribution of the sequences and the corresponding coleopteracin 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.

Coleopterics A, B, and C identified form three separate groups exclusive of the Scarabaeidae family. This work, reports two new groups, A and C coleopterics, which are a group in a different clade and are exclusive of Scarabaeidae (**Fig 17**).

Distinct physicochemical properties characterize the HDP families of Scarabaeidae

Families of insect antimicrobial peptides have distinct physicochemical properties, amino acid composition, and structural features. This has allowed the classification of HDP according to them [13]. 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) [37,109]. With this in mind, those features were analyzed in the HDP identified from Scarabaeidae. Attacins and coleopterics were the families with higher molecular mass (Mean: 12931Da SD: 1075Da and Mean: 7097Da SD: 736.6Da, respectively) and lowest hydrophobic nature (0.27 and 0.28 ratios respectively) (**Fig 18**). Defensins and cecropins showed similar hydrophobic ratios (0.37 and 0.47, respectively). As described for other members of cecropins [26], 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) (**Fig 18**). These characteristics are reflected in the amino acid composition of the peptide families (**Fig 19**). Cecropins show a higher proportion of positively charged amino acids, and attacins and coleopterics a higher content of polar residues. Prolines and glycines are structurally important amino acids, hence, differences were sought in the content of these amino acids in the HDP. In contrast to the other two families, more than 10 % of the amino

acids encoded by attacins and coleopterins correspond to these two residues (**Fig 19**).

Fig 18. Physicochemical properties of Scarabaeidae HDP families predicted with APD3. **A.** Molar mass. **B.** Isoelectric point. **C.** Net charge. **D.** Hydrophobic ratio

Fig 19. Representation of amino acid groups in the families of Scarabaeidae HDP. **A.** Positively charged amino acids (RK). **B.** Proline and glycine amino acids **C.** Polar amino acids (GNST), statistical difference was observed among all groups (****). Statistical differences were evaluated by ANOVA test with Tukey's multiple comparisons (p-value: ***<0.001; ****<0.0001).

The amino-acid distribution represents a key characteristic to classify HDP families, like proline, lysine, and cysteine-rich antimicrobial peptides [13,37,109,110]. According to this, in Scarabaeidae HDP it was found that positively charged amino-acid proportion only differentiates the cecropin family (**Fig 19A**). The percentage of glycine and proline obtained showed a statistical difference among all groups except between the attacin and coleopterins. The polar amino acid (GNST) distribution differentiates the HDP families with a statistically significant difference (**Fig 19B, C**).

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 the accuracy, capacity of prediction, and training set for the construction of the program [13,68,69].

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 coleopterics. ClassAmp2 predicted a high probability for all the families involved (**Fig 20**). 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 wide range of organisms and thus have a higher diversity and representation in the data bases and training libraries, whereas, cecropins had been found majorly in insects (limited taxon) with a low level of representation for some groups, like Coleoptera [111].

Fig 20. 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 in their respective groups: Attacin A and B; Coleopterics 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.

In general, CAMPr3 has a higher probability of prediction 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 iAMPpred; nevertheless, iAMPpred includes a wide diversity of physicochemical 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 4**). This also may explain the low prediction probabilities for the coleopterics in general.

Conclusion

Using HDP has been suggested as a complement to traditional antibiotics, and insects are considered attractive sources of novel HDP, given that their immune response relies – in part – on the expression of diverse families of these peptides showing strong antimicrobial activity at low concentrations and against a wide range of pathogens. So far, identification of insect HDP has focused on the Hemiptera, Hymenoptera, Lepidoptera, and Diptera orders [85]. In Coleoptera, the most diverse order of insects, only a few HDP have been reported [2,24,53,85]. This work identified 155 novel sequences of HDP found in nine transcriptomes from seven Coleoptera species: *D. satanas* (n= 76; 49.03%), *O. curvicornis* (n= 23; 14.83%), *T. dichotomus* (n= 18; 11.61%), *O. nigriventris* (n= 10; 6.45%), *Heterochelus* sp. (n= 6; 3.87%), *O. conspicillatum* (n= 18; 11.61%) and *P. japonica* (n= 4; 2.58%). Two information sources were retrieved: 1) *De novo* transcriptomic data from two neotropical

Scarabaeidae species (*Dichotomius satanas* and *Ontophagus curvicornis*); 2) Sequence data deposited in public sequence databases. The HDP sequences from Coleoptera 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 coleoptericsins (n = 38; 24.52%) were detected. These families were described based on their physicochemical properties, structural features, and sequence similarity. Both novel and previously described attributes were observed in different families. These features are summarized in **Fig 21**. Most HDP sequences show predicted antibacterial, antiviral, and antifungal activities (**Fig 20**), so they may be promising candidates for experimental characterization in anti-microbial and cytotoxicity screening assays. Additionally, this work would also help distinguish key attributes associated with different peptide activities, allowing identification and description of new anti-microbial peptides.

Fig 21. Scarabaeidae HDP general sequence and structural features **A.** Cecropins. **B.** Defensins. **C.** Attacins **D.** Coleoptericsins.

Acknowledgments

Students of the research seedbed 2017, Universidad del Quindío, Jean Carlos Sanchez Rojas, Luz Dahiana Erazo Urquijo; Alejandra Vélez García ; Helen Vanessa Pinchao Salgar and Angélica Maria Madrid Ortiz.

Funding

699 This work was funded by Colombia's Ministry of Science, Technology and Innovation
700 (MINCIENCIAS) call, through grant number 744-2016 and project number 732; grant
701 number 727-2015 for doctorate study in Colombia (to LJT) and project number:
702 111356933173, grant number 784-2017 for postdoctoral stay in Colombia (to GAT)
703 and project number: FP44842-250-2018; grant number 844-2019 and project number:
704 111384466847; grant number 848-2020 and project number: 80740-145-2020.

References

1. Hoffmann JA. Innate immunity of insects. *Curr Opin Immunol.* 1995;7: 4–10.
doi:10.1016/0952-7915(95)80022-0
2. Cociancich S, Bulet P, Hetru C, Hoffmann JA. The inducible antibacterial peptides of insects. *Parasitol Today Pers Ed.* 1994;10: 132–139. doi:10.1016/0169-4758(94)90260-7
3. Hoffmann JA, Reichhart JM, Hetru C. Innate immunity in higher insects. *Curr Opin Immunol.* 1996;8: 8–13. doi:10.1016/s0952-7915(96)80098-7
4. Faruck MO, Yusof F, Chowdhury S. An overview of antifungal peptides derived from insect. *Peptides.* 2016;80: 80–88. doi:10.1016/j.peptides.2015.06.001
5. Saikia K, Belwal VK, Datta D, Chaudhary N. Aromatic-rich C-terminal region of LCI is a potent antimicrobial peptide in itself. *Biochem Biophys Res Commun.* 2019;519: 372–377. doi:10.1016/j.bbrc.2019.09.013
6. Alencar-Silva T, Braga MC, Santana GOS, Saldanha-Araujo F, Pogue R, Dias SC, et al. Breaking the frontiers of cosmetology with antimicrobial peptides. *Biotechnol Adv.* 2018;36: 2019–2031. doi:10.1016/j.biotechadv.2018.08.005
7. Boman HG. Peptide antibiotics and their role in innate immunity. *Annu Rev Immunol.* 1995;13: 61–92. doi:10.1146/annurev.iy.13.040195.000425
8. Zhang L, Gallo RL. Antimicrobial peptides. *Curr Biol.* 2016;26: R14–R19.
doi:10.1016/j.cub.2015.11.017
9. Gómez EA, Giraldo P, Orduz S. InverPep: A database of invertebrate antimicrobial peptides. *J Glob Antimicrob Resist.* 2017;8: 13–17. doi:10.1016/j.jgar.2016.10.003
10. Pasupuleti M, Schmidtchen A, Malmsten M. Antimicrobial peptides: key components of the innate immune system. *Crit Rev Biotechnol.* 2012;32: 143–171.
doi:10.3109/07388551.2011.594423

11. Imler J-L, Bulet P. Antimicrobial peptides in *Drosophila*: structures, activities and gene regulation. *Chem Immunol Allergy*. 2005;86: 1–21. doi:10.1159/000086648
12. Xiao X, Wang P, Lin W-Z, Jia J-H, Chou K-C. iAMP-2L: A two-level multi-label classifier for identifying antimicrobial peptides and their functional types. *Anal Biochem*. 2013;436: 168–177. doi:10.1016/j.ab.2013.01.019
13. Meher PK, Sahu TK, Saini V, Rao AR. Predicting antimicrobial peptides with improved accuracy by incorporating the compositional, physico-chemical and structural features into Chou's general PseAAC. *Sci Rep*. 2017;7: 42362. doi:10.1038/srep42362
14. Bulet P, Hetru C, Dimarcq JL, Hoffmann D. Antimicrobial peptides in insects; structure and function. *Dev Comp Immunol*. 1999;23: 329–344. doi:10.1016/s0145-305x(99)00015-4
15. Wiesner J, Vilcinskas A. Antimicrobial peptides: The ancient arm of the human immune system. *Virulence*. 2010;1: 440–464. doi:10.4161/viru.1.5.12983
16. Hultmark D, Steiner H, Rasmuson T, Boman HG. Insect Immunity. Purification and Properties of Three Inducible Bactericidal Proteins from Hemolymph of Immunized Pupae of *Hyalophora cecropia*. *Eur J Biochem*. 2005;106: 7–16. doi:10.1111/j.1432-1033.1980.tb05991.x
17. Boman. Gene-Encoded Peptide Antibiotics and the Concept of Innate Immunity: An Update Review. *Scand J Immunol*. 1998;48: 15–25. doi:10.1046/j.1365-3083.1998.00343.x
18. Wang L, Li Z, Du C, Chen W, Pang Y. Characterization and expression of a cecropin-like gene from *Helicoverpa armigera*. *Comp Biochem Physiol B Biochem Mol Biol*. 2007;148: 417–425. doi:10.1016/j.cbpb.2007.07.010
19. Marshall SH, Arenas G. Antimicrobial peptides: A natural alternative to chemical antibiotics and a potential for applied biotechnology. *Electron J Biotechnol*. 2003;6: 0–0. doi:10.2225/vol6-issue3-fulltext-1
20. Cavallarin L, Andreu D, San Segundo B. Cecropin A—Derived Peptides Are Potent

- Inhibitors of Fungal Plant Pathogens. *Mol Plant Microbe Interact.* 1998;11: 218–227.
doi:10.1094/MPMI.1998.11.3.218
21. DeLucca AJ, Bland JM, Jacks TJ, Grimm C, Cleveland TE, Walsh TJ. Fungicidal activity of cecropin A. *Antimicrob Agents Chemother.* 1997;41: 481–483.
22. Hultmark D, Engstrom A, Bennich H, Kapur R, Boman HG. Insect Immunity: Isolation and Structure of Cecropin D and Four Minor Antibacterial Components from *Cecropia* Pupae. *Eur J Biochem.* 1982;127: 207–217. doi:10.1111/j.1432-1033.1982.tb06857.x
23. Volkoff A-N, Rocher J, d’Alençon E, Bouton M, Landais I, Quesada-Moraga E, et al. Characterization and transcriptional profiles of three *Spodoptera frugiperda* genes encoding cysteine-rich peptides. A new class of defensin-like genes from lepidopteran insects? *Gene.* 2003;319: 43–53. doi:10.1016/S0378-1119(03)00789-3
24. Bulet P, Stocklin R, Menin L. Anti-microbial peptides: from invertebrates to vertebrates. *Immunol Rev.* 2004;198: 169–184. doi:10.1111/j.0105-2896.2004.0124.x
25. Araújo CAC, Waniek PJ, Stock P, Mayer C, Jansen AM, Schaub GA. Sequence characterization and expression patterns of defensin and lysozyme encoding genes from the gut of the reduviid bug *Triatoma brasiliensis*. *Insect Biochem Mol Biol.* 2006;36: 547–560. doi:10.1016/j.ibmb.2006.04.003
26. Yi H-Y, Chowdhury M, Huang Y-D, Yu X-Q. Insect antimicrobial peptides and their applications. *Appl Microbiol Biotechnol.* 2014;98: 5807–5822. doi:10.1007/s00253-014-5792-6
27. Cornet B, Bonmatin J-M, Hetru C, Hoffmann JA, Ptak M, Vovelle F. Refined three-dimensional solution structure of insect defensin A. *Structure.* 1995;3: 435–448. doi:10.1016/S0969-2126(01)00177-0
28. Hultmark D, Engström A, Andersson K, Steiner H, Bennich H, Boman HG. Insect immunity. Attacins, a family of antibacterial proteins from *Hyalophora cecropia*. *EMBO J.* 1983;2: 571–576.
29. Wang J, Hu C, Wu Y, Stuart A, Amemiya C, Berriman M, et al. Characterization of the

- antimicrobial peptide attacin loci from *Glossina morsitans*. *Insect Mol Biol.* 2008;17: 293–302. doi:10.1111/j.1365-2583.2008.00805.x
30. Engström A, Engström P, Tao ZJ, Carlsson A, Bennich H. Insect immunity. The primary structure of the antibacterial protein attacin F and its relation to two native attacins from *Hyalophora cecropia*. *EMBO J.* 1984;3: 2065–2070.
31. Carlsson A, Engström P, Palva ET, Bennich H. Attacin, an antibacterial protein from *Hyalophora cecropia*, inhibits synthesis of outer membrane proteins in *Escherichia coli* by interfering with omp gene transcription. *Infect Immun.* 1991;59: 3040–3045.
32. Sagisaka A, Miyanoshita A, Ishibashi J, Yamakawa M. 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.* 2001;10: 293–302. doi:10.1046/j.0962-1075.2001.00261.x
33. Bulet P, Cociancich S, Dimarcq JL, Lambert J, Reichhart JM, Hoffmann D, et al. 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.* 1991;266: 24520–24525.
34. Hwang J-S, Lee J, Kim Y-J, Bang H-S, Yun E-Y, Kim S-R, et al. Isolation and Characterization of a Defensin-Like Peptide (Coprinsin) from the Dung Beetle, *Copris tripartitus*. *Int J Pept.* 2009;2009: 1–5. doi:10.1155/2009/136284
35. Purvis A, Hector A. Getting the measure of biodiversity. *Nature.* 2000;405: 212–219. doi:10.1038/35012221
36. Hunt T, Bergsten J, Levkanicova Z, Papadopoulou A, John OSt, Wild R, et al. A Comprehensive Phylogeny of Beetles Reveals the Evolutionary Origins of a Superradiation. *Science.* 2007;318: 1913–1916. doi:10.1126/science.1146954
37. Wang G, Li X, Wang Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* 2016;44: D1087–D1093. doi:10.1093/nar/gkv1278
38. Lee JH, Seo M, Lee HJ, Baek M, Kim IW, Kim SY, et al. Anti-Inflammatory Activity of

- Antimicrobial Peptide Allomyrinasin Derived from the Dynastid Beetle, *Allomyrina*
dichotoma. J Microbiol Biotechnol. 2019;29: 687–695. doi:10.4014/jmb.1809.09031
39. Meng E, Li J, Tang B, Hu Y, Qiao T, Hou Y, et al. 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. 2019;109: 248–256. doi:10.1017/S0007485318000780
40. Namara LM, Griffin CT, Fitzpatrick D, Kavanagh K, Carolan JC. 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. 2018;101:
1–13. doi:10.1016/j.ibmb.2018.07.001
41. Vogel H, Badapanda C, Vilcinskis A. Identification of immunity-related genes in the
burying beetle *Nicrophorus vespilloides* by suppression subtractive hybridization:
Burying beetle immunity. Insect Mol Biol. 2011;20: 787–800. doi:10.1111/j.1365-
2583.2011.01109.x
42. Jacobs CGC, Gallagher JD, Evison SEF, Heckel DG, Vilcinskis A, Vogel H.
Endogenous egg immune defenses in the yellow mealworm beetle (*Tenebrio molitor*).
Dev Comp Immunol. 2017;70: 1–8. doi:10.1016/j.dci.2016.12.007
43. Rodríguez-García MJ, García-Reina A, Machado V, Galián J. Identification, structural
characterisation and expression analysis of a defensin gene from the tiger beetle
Calomera littoralis (Coleoptera: Cicindelidae). Gene. 2016;589: 56–62.
doi:10.1016/j.gene.2016.05.030
44. Lee J, Bang K, Hwang S, Cho S. cDNA cloning and molecular characterization of a
defensin-like antimicrobial peptide from larvae of *Protaetia brevitarsis seoulensis*
(Kolbe). Mol Biol Rep. 2016;43: 371–379. doi:10.1007/s11033-016-3967-1
45. Altincicek B, Vilcinskis A. Identification of immune-related genes from an apterygote
insect, the firebrat *Thermobia domestica*. Insect Biochem Mol Biol. 2007;37: 726–731.
doi:10.1016/j.ibmb.2007.03.012

46. Zou Z, Evans JD, Lu Z, Zhao P, Williams M, Sumathipala N, et al. Comparative genomic analysis of the *Tribolium* immune system. *Genome Biol.* 2007;8: R177. doi:10.1186/gb-2007-8-8-r177
47. Altincicek B, Elashry A, Guz N, Grundler FMW, Vilcinskas A, Dehne H-W. Next Generation Sequencing Based Transcriptome Analysis of Septic-Injury Responsive Genes in the Beetle *Tribolium castaneum*. Mariño-Ramírez L, editor. *PLoS ONE.* 2013;8: e52004. doi:10.1371/journal.pone.0052004
48. Lee SY, Moon HJ, Kurata S, Kurama T, Natori S, Lee BL. Purification and Molecular Cloning of cDNA for an Inducible Antibacterial Protein of Larvae of a Coleopteran Insect, *Holotrichia diomphalia*1. *J Biochem (Tokyo).* 1994;115: 82–86. doi:10.1093/oxfordjournals.jbchem.a124309
49. Imamura M, Wada S, Ueda K, Saito A, Koizumi N, Iwahana H, et al. Multi-peptide precursor structure of acaloleptin A isoforms, antibacterial peptides from the Udo longicorn beetle, *Acalolepta luxuriosa*. *Dev Comp Immunol.* 2009;33: 1120–1127. doi:10.1016/j.dci.2009.06.004
50. Login FH, Balmand S, Vallier A, Vincent-Monegat C, Vigneron A, Weiss-Gayet M, et al. Antimicrobial Peptides Keep Insect Endosymbionts Under Control. *Science.* 2011;334: 362–365. doi:10.1126/science.1209728
51. Lee J, Lee D, Choi H, Kim HH, Kim H, Hwang JS, et al. Structure-activity relationships of the intramolecular disulfide bonds in coprisin, a defensin from the dung beetle. *BMB Rep.* 2014;47: 625–630. doi:10.5483/BMBRep.2014.47.11.262
52. Hull R, Alaouna M, Khanyile L, Byrne M, Ntwasa M. Lifestyle and Host Defense Mechanisms of the Dung Beetle, *Euoniticellus intermedius* : The Toll Signaling Pathway. *J Insect Sci.* 2013;13: 1–25. doi:10.1673/031.013.10801
53. Toro Segovia LJ, Téllez Ramírez GA, Henao Arias DC, Rivera Duran JD, Bedoya JP, Castaño Osorio JC. Identification and characterization of novel cecropins from the *Oxytetrone conspicillatum* neotropic dung beetle. van Raaij MJ, editor. *PLOS ONE.*

- 2017;12: e0187914. doi:10.1371/journal.pone.0187914
54. Yoon IN, Hwang JS, Lee JH, Kim H. The Antimicrobial Peptide CopA3 Inhibits Clostridium difficile Toxin A-Induced Viability Loss and Apoptosis in Neural Cells. J Microbiol Biotechnol. 2019;29: 30–36. doi:10.4014/jmb.1809.08065
55. Kim DH, Hwang JS, Lee IH, Nam ST, Hong J, Zhang P, et al. The Insect Peptide CopA3 Increases Colonic Epithelial Cell Proliferation and Mucosal Barrier Function to Prevent Inflammatory Responses in the Gut. J Biol Chem. 2016;291: 3209–3223. doi:10.1074/jbc.M115.682856
56. Yoon IN, Hong J, Zhang P, Hwang JS, Kim H. An Analog of the Antimicrobial Peptide CopA5 Inhibits Lipopolysaccharide-Induced Macrophage Activation. J Microbiol Biotechnol. 2017;27: 350–356. doi:10.4014/jmb.1607.07065
57. Tomie T, Ishibashi J, Furukawa S, Kobayashi S, Sawahata R, Asaoka A, et al. Scarabaecin, a novel cysteine-containing antifungal peptide from the rhinoceros beetle, Oryctes rhinoceros. Biochem Biophys Res Commun. 2003;307: 261–266. doi:10.1016/S0006-291X(03)01162-8
58. Cultid C, Medina C, Constantino LM, Bedir Martinez, Escobar AF. Escarabajos coprófagos (Scarabaeinae) del eje cafetero: Guia para el estudio ecológico (Field Guide). Cenicafé, Federación Nacional de Cafeteros de Colombia and Wild life Conservation Society; 2012. doi:10.13140/RG.2.1.1013.9049
59. CULTID-MEDINA CA, LOBO JM, MEDINA CA, GONZÁLEZ FA, ESCOBAR F, CHACÓN DE ULLOA P. Completitud del inventario de escarabajos coprófagos (Coleoptera: Scarabaeinae) en la Ecorregión del Eje Cafetero, Colombia. Rev Colomb Entomol. 2014;40: 111–119.
60. Villada-Bedoya S, Cultid-Medina CA. ESTRUCTURA POBLACIONAL DE DOS ESPECIES DE Dichotomius HOPE, 1838 (COLEOPTERA: SCARABAEINAE) EN UN PAISAJE CAFETERO DE LOS ANDES OCCIDENTALES DE COLOMBIA, RISARALDA. Boletín Científico de la Facultad de Ciencias Exactas y Naturales. 2017;21: 188–198.

61. Bouchard P, Bousquet Y, Davies A, Alonso-Zarazaga M, Lawrence J, Lyal C, et al. Family-Group Names In Coleoptera (Insecta). ZooKeys. 2011;88: 1–972. doi:10.3897/zookeys.88.807
62. FastQC. 2015. Available: <https://qubeshub.org/resources/fastqc>
63. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170
64. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 2018;46: W537–W544. doi:10.1093/nar/gky379
65. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215: 403–410. doi:10.1016/S0022-2836(05)80360-2
66. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2011;39: D38–D51. doi:10.1093/nar/gkq1172
67. Nielsen H. Predicting Secretory Proteins with SignalP. Methods Mol Biol Clifton NJ. 2017;1611: 59–73. doi:10.1007/978-1-4939-7015-5_6
68. Waghu FH, Barai RS, Gurung P, Idicula-Thomas S. CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. Nucleic Acids Res. 2016;44: D1094–1097. doi:10.1093/nar/gkv1051
69. Joseph S, Karnik S, Nilawe P, Jayaraman VK, Idicula-Thomas S. ClassAMP: A Prediction Tool for Classification of Antimicrobial Peptides. IEEE/ACM Trans Comput Biol Bioinform. 2012;9: 1535–1538. doi:10.1109/TCBB.2012.89
70. Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 2013;41: W349–W357. doi:10.1093/nar/gkt381
71. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30: 1236–1240.

doi:10.1093/bioinformatics/btu031

72. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, et al. Template-based protein structure modeling using the RaptorX web server. *Nat Protoc.* 2012;7: 1511–1522. doi:10.1038/nprot.2012.085
73. Wang S, Peng J, Xu J. Alignment of distantly related protein structures: algorithm, bound and implications to homology modeling. *Bioinformatics.* 2011;27: 2537–2545. doi:10.1093/bioinformatics/btr432
74. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera A visualization system for exploratory research and analysis. *J Comput Chem.* 2004;25: 1605–1612. doi:10.1002/jcc.20084
75. Vajda S, Yueh C, Beglov D, Bohnuud T, Mottarella SE, Xia B, et al. New additions to the ClusPro server motivated by CAPRI. *Proteins.* 2017;85: 435–444. doi:10.1002/prot.25219
76. Ren J, Wen L, Gao X, Jin C, Xue Y, Yao X. DOG 1.0: illustrator of protein domain structures. *Cell Res.* 2009;19: 271–273. doi:10.1038/cr.2009.6
77. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011;7: 539. doi:10.1038/msb.2011.75
78. Pei Z, Sun X, Tang Y, Wang K, Gao Y, Ma H. Cloning, expression, and purification of a new antimicrobial peptide gene from *Musca domestica* larva. *Gene.* 2014;549: 41–45. doi:10.1016/j.gene.2014.07.028
79. Wang G, editor. *Antimicrobial peptides: discovery, design and novel therapeutic strategies.* Wallingford, Oxfordshire, UK ; Cambridge, MA: CABI; 2010.
80. Pane K, Durante L, Crescenzi O, Cafaro V, Pizzo E, Varcamonti M, et al. 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.* 2017;419: 254–265. doi:10.1016/j.jtbi.2017.02.012

81. Yoo WG, Lee JH, Shin Y, Shim J-Y, Jung M, Kang B-C, et al. Antimicrobial peptides in the centipede *Scolopendra subspinipes mutilans*. *Funct Integr Genomics*. 2014;14: 275–283. doi:10.1007/s10142-014-0366-3
82. Jacobs CGC, Steiger S, Heckel DG, Wielsch N, Vilcinskas A, Vogel H. Sex, offspring and carcass determine antimicrobial peptide expression in the burying beetle. *Sci Rep*. 2016;6: 25409. doi:10.1038/srep25409
83. Duwadi D, Shrestha A, Yilma B, Kozlovski I, Sa-eed M, Dahal N, et al. Identification and screening of potent antimicrobial peptides in arthropod genomes. *Peptides*. 2018;103: 26–30. doi:10.1016/j.peptides.2018.01.017
84. Brady D, Grapputo A, Romoli O, Sandrelli F. Insect Cecropins, Antimicrobial Peptides with Potential Therapeutic Applications. *Int J Mol Sci*. 2019;20: 5862. doi:10.3390/ijms20235862
85. Mylonakis E, Podsiadlowski L, Muhammed M, Vilcinskas A. Diversity, evolution and medical applications of insect antimicrobial peptides. *Philos Trans R Soc B Biol Sci*. 2016;371: 20150290. doi:10.1098/rstb.2015.0290
86. Wang Y-P, Lai R. Insect Antimicrobial Peptides: Structures, Properties and Gene Regulation: Insect Antimicrobial Peptides: Structures, Properties and Gene Regulation. *Zool Res*. 2010;31: 27–34. doi:10.3724/SP.J.1141.2010.01027
87. Xu J, Zhang Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinforma Oxf Engl*. 2010;26: 889–895. doi:10.1093/bioinformatics/btq066
88. Wang S, Ma J, Peng J, Xu J. Protein structure alignment beyond spatial proximity. *Sci Rep*. 2013;3: 1448. doi:10.1038/srep01448
89. Memarpour-Yazdi M, Zare-Zardini H, Asoodeh A. A Novel Antimicrobial Peptide Derived from the Insect *Paederus dermatitis*. *Int J Pept Res Ther*. 2013;19: 99–108. doi:10.1007/s10989-012-9320-1
90. Saito A, Ueda K, Imamura M, Atsumi S, Tabunoki H, Miura N, et al. Purification and cDNA cloning of a cecropin from the longicorn beetle, *Acalolepta luxuriosa*. *Comp*

974 Biochem Physiol B Biochem Mol Biol. 2005;142: 317–323.
 975 doi:10.1016/j.cbpb.2005.08.001

976 91. Hoffmann JA, Hetru C. Insect defensins: inducible antibacterial peptides. Immunol
 977 Today. 1992;13: 411–415. doi:10.1016/0167-5699(92)90092-L

978 92. Wu Q, Patočka J, Kuča K. Insect Antimicrobial Peptides, a Mini Review. Toxins.
 979 2018;10: 461. doi:10.3390/toxins10110461

980 93. Zhu S, Gao B. Evolutionary origin of β -defensins. Dev Comp Immunol. 2013;39: 79–84.
 981 doi:10.1016/j.dci.2012.02.011

982 94. Dias R de O, Franco OL. Cysteine-stabilized $\alpha\beta$ defensins: From a common fold to
 983 antibacterial activity. Peptides. 2015;72: 64–72. doi:10.1016/j.peptides.2015.04.017

984 95. Tarr DEK. Establishing a reference array for the CS- $\alpha\beta$ superfamily of defensive
 985 peptides. BMC Res Notes. 2016;9: 490. doi:10.1186/s13104-016-2291-0

986 96. Shafee TMA, Lay FT, Phan TK, Anderson MA, Hulett MD. Convergent evolution of
 987 defensin sequence, structure and function. Cell Mol Life Sci. 2017;74: 663–682.
 988 doi:10.1007/s00018-016-2344-5

989 97. Lowenberger CA, Smartt CT, Bulet P, Ferdig MT, Severson DW, Hoffmann JA, et al.
 990 Insect immunity: molecular cloning, expression, and characterization of cDNAs and
 991 genomic DNA encoding three isoforms of insect defensin in *Aedes aegypti*. Insect Mol
 992 Biol. 1999;8: 107–118. doi:10.1046/j.1365-2583.1999.810107.x

993 98. Peschel A, Jack RW, Otto M, Collins LV, Staubitz P, Nicholson G, et al.
 994 *Staphylococcus aureus* Resistance to Human Defensins and Evasion of Neutrophil
 995 Killing via the Novel Virulence Factor Mprf Is Based on Modification of Membrane
 996 Lipids with L-Lysine. J Exp Med. 2001;193: 1067–1076. doi:10.1084/jem.193.9.1067

997 99. Brogden KA, De Lucca AJ, Bland J, Elliott S. Isolation of an ovine pulmonary
 998 surfactant-associated anionic peptide bactericidal for *Pasteurella haemolytica*. Proc
 999 Natl Acad Sci. 1996;93: 412–416. doi:10.1073/pnas.93.1.412

1000 100. Lai R, Liu H, Hui Lee W, Zhang Y. An anionic antimicrobial peptide from toad *Bombina*

maxima. Biochem Biophys Res Commun. 2002;295: 796–799. doi:10.1016/S0006-291X(02)00762-3

101. Gunne H, Hellers M, Steiner H. Structure of preproattacin and its processing in insect cells infected with a recombinant baculovirus. Eur J Biochem. 1990;187: 699–703. doi:10.1111/j.1432-1033.1990.tb15356.x

102. Kockum K, Faye I, Hofsten PV, Lee JY, Xanthopoulos KG, Boman HG. Insect immunity. Isolation and sequence of two cDNA clones corresponding to acidic and basic attacins from *Hyalophora cecropia*. EMBO J. 1984;3: 2071–2075.

103. Shin HS, Park S-I. Novel attacin from *Hermetia illucens*: cDNA cloning, characterization, and antibacterial properties. Prep Biochem Biotechnol. 2019;49: 279–285. doi:10.1080/10826068.2018.1541807

104. Ando K, Natori S. Molecular cloning, sequencing, and characterization of cDNA for sarcotoxin IIA, an inducible antibacterial protein of *Sarcophaga peregrina* (flesh fly). Biochemistry. 1988;27: 1715–1721. doi:10.1021/bi00405a050

105. Sun S-C, Lindstrom I, Lee J-Y, Faye I. Structure and expression of the attacin genes in *Hyalophora cecropia*. Eur J Biochem. 1991;196: 247–254. doi:10.1111/j.1432-1033.1991.tb15811.x

106. Engström P, Carlsson A, Engström A, Tao ZJ, Bennich H. The antibacterial effect of attacins from the silk moth *Hyalophora cecropia* is directed against the outer membrane of *Escherichia coli*. EMBO J. 1984;3: 3347–3351. doi:10.1002/j.1460-2075.1984.tb02302.x

107. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc. 2006;1: 2876–2890. doi:10.1038/nprot.2006.202

108. Vilcinskas A, Mukherjee K, Vogel H. Expansion of the antimicrobial peptide repertoire in the invasive ladybird *Harmonia axyridis*. Proc R Soc B Biol Sci. 2013;280: 20122113. doi:10.1098/rspb.2012.2113

109. Zhou X, Peng L-Y, Wang Z-C, Wang W, Zhu Z, Huang X-H, et al. Identification of novel

1028 antimicrobial peptides from rice planthopper, *Nilaparvata lugens*. *Insect Biochem Mol*
1029 *Biol.* 2019;113: 103215. doi:10.1016/j.ibmb.2019.103215
1030 110. Lu W. Antimicrobial peptides. *Semin Cell Dev Biol.* 2019;88: 105–106.
1031 doi:10.1016/j.semcdb.2018.05.026
1032 111. Waghu FH, Barai RS, Idicula-Thomas S. Leveraging family-specific signatures for AMP
1033 discovery and high-throughput annotation. *Sci Rep.* 2016;6: 24684.
1034 doi:10.1038/srep24684

A.

HDPs and contigs per species

B.

Families of HDPs per species

Figure 1

Figure 2

Figure 3

bioRxiv preprint doi: <https://doi.org/10.1101/2020.06.15.152363>; this version posted June 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 4

A. **B.** **C.**

Figure 6

Defensin groups net charge

bioRxiv preprint doi: <https://doi.org/10.1101/2020.06.15.152363>; this version posted June 15, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 7

Figure 8

Figure 9

A.**Attacins net charge** **B.****Attacins GNTS ratio**

Figure 10

Figure 11

Inner face

Outer face

A.

B.

C.

D.

Figure 12

Figure 13

Figure 15

Figure 16

Figure 17

A.

B.

C.

D.

Figure 18

Figure 19

Figure 20

Figure 21