



Sudong Qi^{1,2}, Bin Gao¹ and Shunyi Zhu^{1,*}

- ¹ Group of Peptide Biology and Evolution, State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China; qisudong@ioz.ac.cn (S.Q.); gaob@ioz.ac.cn (B.G.)
- ² University of Chinese Academy of Sciences, Beijing 100049, China

Correspondence: zhusy@ioz.ac.cn; Tel.: +86-010-64807112

Abstract: As a worldwide sanitary insect pest, the housefly *Musca domestica* can carry and transmit more than 100 human pathogens without suffering any illness itself, indicative of the high efficiency of its innate immune system. Antimicrobial peptides (AMPs) are the effectors of the innate immune system of multicellular organisms and establish the first line of defense to protect hosts from microbial infection. To explore the molecular diversity of the *M. domestica* AMPs and related evolutionary basis, we conducted a systematic survey of its full AMP components based on a combination of computational approaches. These components include the cysteine-containing peptides (MdDefensins, MdEppins, MdMuslins, MdSVWCs and MdCrustins), the linear α -helical peptides (MdCecropins) and the specific amino acid-rich peptides (MdDomesticins, MdDiptericins, MdEdins and MdAttacins). On this basis, we identified multiple genetic mechanisms that could have shaped the molecular and structural diversity of the *M. domestica* AMPs, including: (1) Gene duplication; (2) Exon duplication via shuffling; (3) Protein terminal variations; (4) Evolution of disulfide bridges via compensation. Our results not only enlarge the insect AMP family members, but also offer a basic platform for further studying the roles of such molecular diversity in contributing to the high efficiency of the housefly antimicrobial immune system.

Keywords: molecular diversity; evolution; defensin; gene duplication; homodimer; disulfide bridge

1. Introduction

Insects account for 90% of all extant animal organisms in the world [1,2] and coexist with a variety of microorganisms in different environments [3]. Therefore, they need to evolve a potent defense system for clearing potential invaders. Antimicrobial peptides (AMPs) are effectors of the innate immunity against bacteria, fungi, parasites and viruses [4], which exhibit some common properties (e.g., cationicity, hydrophobicity and amphipathcity) for their antimicrobial activity [5]. Many AMPs are induced from the insect immune organs (e.g., fat body) in response to microbial infections. They secret into hemolymph to reach a concentration between 0.1 and 100 μ M to inhibit the growth of exotic microorganisms [6].

Like the counterparts in non-insect organisms, insect AMPs are also classified into three distinct structural classes. They are the cysteine-rich peptides (e.g., drosomycins and insect defensins, two subfamilies of defensins in *Drosophila*) [7–9]; the peptides adopting an α -helical conformation (e.g., cecropins and moricins) [10]; and the peptides with an unusual bias in certain amino acids, such as proline-rich peptides (e.g., metchnikowins, apidaecins, drosocins, and lebocins) [11–13] and glycine-rich peptides/proteins (e.g., diptericins, attacins and gloverins) [6]. In *Drosophila*, their AMPs are initially divided into three functional classes based on their target specificity, which comprises antifungal AMPs (drosomysins and metchnikowins), anti-Gram positive bacterial AMPs (Defensin) and anti-Gram negative bacterial AMPs (cecropins, drosocin, attacins, diptericins and



Citation: Qi, S.; Gao, B.; Zhu, S. Molecular Diversity and Evolution of Antimicrobial Peptides in *Musca domestica*. *Diversity* **2021**, *13*, 107. https://doi.org/10.3390/d13030107

Academic Editor: Michael Wink

Received: 12 January 2021 Accepted: 21 February 2021 Published: 1 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). MPAC) [6]. However, subsequent studies demonstrated that some of them exhibit functional overlapping. For example, although the prototypical Drosomycin is a strictly antifungal defensin [6,9], homologs from *Drosphila takahashii* possess antibacterial activity [14], suggesting that functional diversification occurred in these drosomycin-type defensins. Similarly, the previously defined strictly antibacterial insect defensins and cecropins were later found to have antifungal activity [15,16].

Insects are an important resource for understanding the basic biology of the immune system and for searching for new peptides for anti-infective drugs. In recent years, some computational approaches have been applied to discover AMPs in a given species with whole genomes sequenced [17–20]. *Musca domestica* is a worldwide sanitary insect pest whose larvae often feed on microbe-rich, decaying organic materials and the adults are the major vectors of pathogens causing human or animal diseases. Thus, more AMPs might have been evolved [21]. Thanks to the release of its whole sequences [22], we have an opportunity to survey the AMPs in a vector insect for studying their evolution.

Here, we report the molecular diversity of the *M. dometica* AMPs based on a systematic database search, which could provide us with a special perspective to understand how the evolution of the antimicrobial immune system occurs in a species with the tenacious vitality. We found that different from *Drosophila* that has a limited number of AMPs [6], *M. domestica* has largely expanded its AMP number via multiple genetic mechanisms to create structural diversity of their AMPs, which would have commonly shaped its high-efficient antimicrobial immune system. These include: (1) Gene duplication; (2) Exon duplication via shuffling; (3) Protein terminal variations; (4) Evolution of disulfide bridges via compensation. A similar phenomenon was also observed previously in the evolution of AMPs in the parasitoid *Nasonia vitripennis* [20], suggesting that these two species of insects could suffer from a similar selective pressure to drive the evolution of their antimicrobial systems.

2. Materials and Methods

2.1. Database Searches

Strategies for gene discovery used in this study are provided in the supplementary information (Supplementary Information Figure S1). In brief, potential AMPs were firstly searched against the proteome of *M. domestica* that was downloaded from the Genome Database (up to September 2, 2020) (https://www.ncbi.nlm.nih.gov/) by filtering using a threshold < 100 amino acids with a signal peptide. The potential peptides were predicted in the Collection of Anti-microbial Peptides (CAMPR3) server (http://www.camp.bicnirrh. res.in/prediction.php), and then they were used as templates to search for more peptides from the non-redundant sequences database until no new peptides appeared by BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Secondly, these newly-discovered peptides were again used as queries to mine the whole genome shotgun and nucleotide collection databases (https://blast.ncbi.nlm.nih.gov/Blast.cgi) by TBLASTN. Gene runner (http: //www.generunner.net/) was employed to translate a complete open reading frame from a selected nucleotide sequence. Retrieved sequences with a signal peptide and a classical AMP signature were blasted again for new rounds of TBLASTN and BLASTP searches. The method would be continuously repeated until no new hit appeared. Thirdly, BLASTP and TBLASTN programs were used to characterize orthologues of known peptides of Drosophila melanogaster against the database of M. domestica. Finally, the protein pattern method by the PHI-BLAST algorithm program (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was conducted to search for *M. domestica* AMPs based on the cysteine arrangement pattern of defensins, i.e., the CXXC/CXC motif [23].

2.2. Characteristics Identification

All the potential AMP-like peptides were submitted to SignalP5.1 server (http://www.cbs.dtu.dk/services/SignalP/) for predicting a signal peptide. Pro-peptides were detected by ProP 1.0 Server (http://www.cbs.dtu.dk/services/ProP/). The net charge (NC)

(pH = 7), molecular weight (MW) and isoelectric point (PI) were predicted at PROTEIN CALCULATOR v3.4 server (http://protcalc.sourceforge.net/). Peptide properties were then analyzed by ESPript 3.0 server (ESPript 3.x /ENDscript 2.x (ibcp.fr) for secondary structure prediction [24].

2.3. Phylogenetic Tree Construction

Multiple sequences were aligned with MUSCLE (https://www.ebi.ac.uk/Tools/msa/ muscle/) and the alignments were used to build phylogenetic trees by iqtree-2.0-rc2 with substitution models BLOSUM62 and PMB. Phylogenetic testing included 1000 replicates of Ultrafast bootstrap (UFBoot) and 1000 replicates of SH-aLRT to provide support for tree branches [25]. In our study, both BLOSUM62 and PMB models generated very similar results with good agreement. The trees presented here were prepared by Evolview v2 (https://evolgenius.info/evolview-v2).

2.4. Structure Modeling and Analysis

The three-dimensional (3D) structures of *M. domestica* AMPs described here were predicted by I-TASSER server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) and evaluated by the Verify3D. Except homodimers of MdDefensin20, MdMuslin1 and Md-Muslin26 that are displayed by PyMol (https://pymol.org/2/), all structural images are displayed by MolMol (https://sourceforge.net/projects/molmol/). The wheel projection was performed online using the Helical Wheel Projections (http://rzlab.ucr.edu/scripts/wheel/wheel.cgi).

To detect whether cysteines in a specific *M. domestica* AMP would form one disulfide bridge, we first built its initial structure by I-TASSER and then refined the structure with the help of molecular dynamics simulations or energy minimization. For MdEppin35-1, its model was first obtained on I-TASSER with a position restraint to Cys2 (position 15) and Cys6 (position 45) and Cys3 (position 32) and Cys5 (position 42). The MdMuslin1 and MdMuslin26 were modeled with the potential disulfide bridges in MdMuslin2 and MdMuslin25, respectively, and their initial homodimer structures were assembled by Z-DOCK (http://zdock.umassmed.edu/) [26]. The MdMulsin1 homodimer was used to MD simulations (50 ns) with GROMACS 2020.1 with the OPLS (Optimized Potential for Liquid Simulations)-AA/L all-atom force field (2001 amino acid dihedrals) [27]. Sodium and chloride ions were added to neutralize the total system charge and simulated after the peptide was immersed in a cubic box from the surface at least 1 nm and solvated with SPC water. Solved structure was energy minimized for 5000 steps of steepest descent minimization termination with a maximum force less than 1000 KJ/mol/nm. The temperature at 300 K was maintained by velocity rescaling method, along with the pressure at 1 bar being kept by Parrinello-Rahman methods, followed after the system was equilibration phase of 100 ps number of particles, volume, and temperature (NVT) equilibration and 100 ps number of particles, pressure, and temperature (NPT) equilibration. The particle mesh Ewald method was used for long range electrostatic interactions and the linear constraint solver (LINCS) algorithm constrained all bonds. Trajectories were saved every 2 fs for analysis. A homodimer snapshot was extracted from the simulations for constructing the inter-monomer disulfide bridge by Swiss-PdbViewer (https://spdbv.vital-it.ch/). Energy minimizations were performed with the force fields [28] implemented in the MOE2019 software (OPLS-AA for MdEppin35-1, AMBER 10 for Md-Defensin20, MdMuslin1 and MdMuslin26). These homodimers were analyzed by PDBsum (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) to evaluate their quality.

2.5. Positive Selection Analysis

Codon-substitution models were selected to estimate the nonsynonymous-to-synonymous rate ration ($\omega = dN/dS$) using CODEML implemented in the PAML software package [29,30]. In these models, M0 assumes that all sites have a ω ratio and is used as a control. Two pairs of codon-based likelihood models (M1a/M2a, M7/M8) were chosen for making two likelihood ratio tests (LRTs). M1a (nearly neutral model) constraints a proportion p0 of conserved sites with $0 < \omega < 1$, while a proportion p1 = 1 – p0 of neutral sites with $\omega 1$ = 1; M2a (positive selection model) adds an extra class of sites with the proportion p2 = 1 – p0 – p1 and with ω estimated from the data. M7 (β distribution model) does not allow for positively selected sites and M8 (β and ω model) adds an extra class of sites to M7, allowing for $\omega > 1$, which means the presence of positively selected sites. The calculation of posterior probabilities was completed using the Bayes Empirical Bayes (BEB) method [31].

2.6. Co-Evolutionary Analysis

Multiple sequence alignments (MSA) of AMPs from the housefly and a representative structure were submitted to MISTIC2 for co-evolutionary analysis [32] (https://mistic2.leloir.org.ar). Four covariation methods: (1) corrected mutual information (MIp), (2) mean field direct coupling analysis (mfDCA), (3) pseudo-likelihood maximization DCA (plmDCA) and (4) multivariate Gaussian modeling DCA (gaussianDCA) were chosen for analyzing the inter relationship of residues in a protein sequence, which could identify the structurally or functionally important positions. At the same time, these sequences were also input into the Weblogo server [33] (http://weblogo.berkeley.edu/logo.cgi) for creating sequence logos using default parameters.

3. Results

Using a combination of computational approaches, we have largely enlarged the *Musca domestica* AMP repertoire to 186 AMP-like peptides/proteins (Table A1) [22], in which 148 are considered as newly discovered members. These components included the cysteine-containing peptides (MdDefensins, MdEppins, MdMuslins, MdSVWCs and MdCrustins); the linear α -helical peptides (i.e., MdCecropins); and the specific amino acid-rich AMPs (i.e., MdDomesticins, MdDiptericins, MdEdins and MdAttacins). Their characteristics including length, pI and net charges at pH = 7.0 are provided in Table A1. Our results indicate that most AMPs described here are smaller than 150 amino acids in length with some typical AMP features. Some putative AMPs are larger in length due to internal duplication. These peptides are described in details as follows:

3.1. Cysteine-Containing Peptides

3.1.1. MdDefensins

Defensins are approximately 4 kDa AMPs with three or four conserved disulfide bridges, which exist in nearly all multicellular organisms [34]. Based on their structural characteristics, defensins can be classified into two distinct superfamilies called *cis*- and *trans*-defensins. The former includes those with the cysteine-stabilized α -helix/ β -sheet (CS $\alpha\beta$) fold produced by plants, fungi and invertebrates; the latter includes α -defensins, β -defensins and θ -defensin from vertebrates as well as big defensins from invertebrates. In insects, Tian and colleagues found three different types of defensin from *N. vitripennis* including classical insect-type defensins (CITDs), nasonins and navitricins [20]. Insect defensins are composed of an n-terminal loop, an α -helix, followed by an antiparallel β sheet. Defensins in insects show antimicrobial activity on Gram-positive bacteria by forming voltage-dependent channels to disrupt the permeability barrier of the cytoplasmic membrane resulting in cytoplasmic potassium loss [35]. Besides, some insect defensins can kill the Gram-negative *Escherichia coli* and some fungi [15,36].

In the housefly, there are a total of 21 defensins-like AMPs (named MdDefensins) with 11 new members described here (Figure 1a and Figure S2). Their mature peptides are composed of 40–65 residues and contain three disulfide bridges with net charges ranging from 0.9 to + 4.2 (Table A1). Based on our phylogenetic tree analysis, these peptides can be divided into three groups (Figure 1b): Group I includes MdDefensin1-MdDefensin9 which all belong to the CITDs; Group II includes MdDefensin10-MdDefensin15 which all lack

a classical pro-peptide; and Group III includes MdDefensin16-MdDefensin21 which all contain a short n-loop without a pro-peptide (Figure 1b). Among them, some members (e.g., MdDefensin4, 6, MdDefensin13, 15 and MdDefensin16) have been identified to be transcriptionally active after body wall injury [37]. MdDefensin1-MdDefensin16, MdDefensin18, and MdDefensin19 share their precursor organization to CITDs that comprises a signal peptide, an acidic propeptide ending with an R/KXKR motif (X denoting F, Q or Y) or its variants followed by a mature peptide. This motif is lost in MdDefensin17, -20 and -21, giving rise to failure in propeptide processing and thus generating an extended Nterminus (Figure 1a and Figure S2). The 3D structures of four representative MdDefensins with different N-terminal lengths (Figure 1a) show they all adopt a typical fold of $CS\alpha\beta$ (Figure 1b). The α -helix spans residues L¹⁵-I²¹ in MdDefensin2, G² -L²⁸ in MdDefensin10, G^{24} -L³⁴ in MdDefensin¹⁷, and N¹⁸-I²⁶ in MdDefensin²⁰ and a hydrophobic cluster is present in MdDefensin2 (L¹⁵, A¹⁷ and A¹⁸), MdDefensin10 (W²⁷ and L²⁸), MdDefensin17 $(W^{26}, M^{29}, and L^{34})$ and MdDefensin20 (L^{20}, L^{23}, I^{26}) . The two β -strands constitute an antiparallel sheet linked by a loop that commonly forms a functional γ -core contributing to the antimicrobial activity in some defensins [38]. The N-terminally extended region in MdDefensin17 folds into a short two-stranded antiparallel sheet followed by an α -helix. This unique subdomain structure is firstly found in an insect defensin [39]. MdDefensin20 has evolved a free cysteine that is not involved in the intramolecular disulfide bridges (Figure 1). Compared with the CITDs, MdDefensin17 - MdDefensin21 have a shorter n-loop (Figure 1b and Figure S2), analogous to the antibacterial ancient invertebrate-type defensins (AITDs) [40]. In our analysis, M1a/M2a models identified Ser⁷ (numbered according to Mddefensin1) as a positively selected site (Table 1). The lacking of positive selection signals in M7/M8 would be due to the sparse sampling of species and low sequence divergence.

Model	1	LRT	Parameters	Positive Selected Sites
M0	-918.761		$\omega = 0.16294$	None
M1a	-891.66		$\rho 0 = 0.24997, \omega 0 = 0.00000$	Not allowed
			$\rho 1 = 0.75003, \omega 1 = 1.00000$	
M2a	-888.78	5.74	$\rho 0 = 0.24996, \omega 0 = 0.00000$	
			$\rho 1 = 0.66063, \omega 1 = 1.00000$	7 S *, 22 L
			$\rho 2 = 0.08941 \ \omega 2 = 3.89787$	
M7	-875.388		<i>p</i> = 0.39543, q = 1.05516	Not allowed
M8	-875.26	0.26	p0 = 0.97856, <i>p</i> = 0.41043, q = 1.16670	7 S
			$(p1 = 0.02144), \omega = 2.30984$	

Table 1. Maximum likelihood estimates of parameters and sites inferred to be under positive selection in the *M. domestica* defensins.

Note: *l* is the log likelihood; LRT is likelihood ratio test, which is twice the log likelihood difference (2 Δ l) between the null models (M1a and M7) and their alternative models (M2a and M8): M1a/M2a = 5.74 (df = 2, *p* = 0.05670); M7/M8 = 0.26 (*p* = 1). Positively selected sites identified by the BEB method under M2a with posterior probabilities *P* > 0.6 are shown and with *P* > 0.95 by "*".



Figure 1. Defensin-like AMPs. (a) Comparison of representative MdDefensins (For the full-set MdDefensins sequences, see Figure S2) and insect defensins from other insects. Previously known sequences are labeled by a red "#". Cysteines are shaded in yellow and the conserved glycines in grey. Basic residues (K, R, H) and acidic residues (E, D) are highlighted in

blue and red, respectively. The extended N-terminus in MdDefesin10, MdDefesin 17, and MdDefesin 20 is shadowed in cyan. Structural elements, including three loops (designated as n-loop, m-loop and c-loop), the α -helix, the two-stranded β -sheet, the γ -Core motif as well as the three conserved disulfides are indicated at the bottom. The free cysteine (Cys1) is underlined once. Dm: *Drosophila melanogaster* (GenBank: P36192.1), Lucifensin: *Lucilia sericata* (PDB:2LLD), Pt: *Protophormia terraenovae* (GenBank: P10891.2), Sb: *Simulium bannaense* (GenBank: AJP36711.1). (b) Phylogenetic tree and 3D structure representatives of the defensins from the housefly and other insects. This tree was inferred using iqtree-2.0-rc2. Significant bootstrap values are indicated by a black circle for SH-aLRT > 90 and white for UFBoot > 85. See GenBank IDs and other details for each AMP in Table A1. The structures are shown as ribbons by MolMol, with their N- and C-termini labeled. The n-loops for all the structures and the free cysteine (Cys1) in MdDefensin20 are denoted. The extra N-terminal domain in MdDefensin17 is circled in blue.

3.1.2. MdEppins

Serine protease inhibitors (SPIs) exist in all organisms that participate in many important metabolisms progresses, such as blood coagulation, fibrinolysis, inflammation and immunity. The inhibitors are categorized into four groups (Kazal, Kunitz, Serpin and α macroglobulin), all with a disulfide-rich α/β fold and a P1 site [41]. They show an inhibitory activity against a broad-spectrum of enzymes, e.g., trypsin, chymotrypsin, plasmin, elastase and microbial serine proteases. P1 site plays an important role in specificity and binding strength of serine protease inhibitors (KuSPIs) are extensively distributed in microbes, plants, insects and mammals. In recent years, many studies focused on their antimicrobial activity. For example, in human eppins comprising two potential protease inhibitory domains (a whey acid protein (WAP) or four disulfide core domain and a kunitz domain) has been reported to kill Gram-negative bacteria [43]. IPS1-3, a KuSPI isolated from the cell-free hemolymph of the *Galleria mellonella* larvae, can be induced to respond to the injected fungal elicitor zymosan [44]. In insect silk, KuSPIs can inhibit bacterial and fungal proteinases [45].

The Eppin family contains 35 members in the housefly (herein named MdEppin1–MdEppin35) and shares a conserved domain with the KuSPI family. Among them, MdEppin35 contains nine KuSPI domains (named MdEppin35-1 to MdEppin35-9) (Figure 2a and Figure S3). The pattern can be drawn as CX₈₋₁₄CX₁₅₋₁₇CX₆₋₇GCX₁₂₋₁₃CX₃C with three disulfide bridges (C1-C6, C2-C4 and C3-C5) (Figure 2a and Figure S3), in which C1-C6 and C3-C5 are essential for the maintenance of a native conformation but the third one (C2-C4) appears to be involved in stabilizing the binding domains in the loops containing the active site (P1) [46]. The phylogenetic tree reveals that Mdeppin1-Mdeppin8, Mdeppin 35-1 and 35-9 share high similarity to the kuntiz domain of eppin isolated from human whereas Mdeppin35-5 and Mdeppin35-6 are separated as single taxa (Figure 2b). In MdEppin35-1, an N-terminal deletion led to the loss of the first two cysteines. Alternatively, two C-terminal cysteines are evolved, which could compensate the loss via the formation of new bridge bridges to stabilize its structures, as verified by our structural modeling (Figure 2a,c).

Evolutionary analysis identified two positively selected sites (K¹² and L¹⁸, shown in MdEppin1) whose mutations might be relevant to their functional divergence (Table 2). Consistently, L¹⁸ was also identified as an essential site potentially related to the activity of MdEppin, as analyzed by MISTIC. In addition, this analysis suggests its possible connection with other amino acids, including K¹², R¹⁶ (active site P1), I¹⁹, P²⁰ and E³³ (Figure 2d,e).



Figure 2. Eppin-like AMPs. (**a**) Comparison of representative MdEppins (For the full-set MdEppin sequences, see Figure S3) and Kuntiz-domain-type AMPs from other species. Cysteines involved in the formation of disulfides are colored in yellow. Identical residues are shadowed in grey. Basic residues (K, R, H) and acidic residues (E, D) are highlighted in blue and red,

respectively. The highly conserved domain in eppins is boxed in green. The P1 amino acids are italicized and shadowed in cyan. Conserved disulfides, α -helix and two-stranded β -sheets are also indicated at the bottom, while the disulfides in MdEppin35-1 are displayed above MdEppin35-1, in which newly emerged ones are shown in dark red. Gm: *Galleria mellonella* (GenBank: AAK40037.1), Pp: *Pseudechis_porphyriacus* (GenBank: sp_B5G6G6.1), Hm: *Homo sapiens* (GenBank:AAG00547.1). The phase 1 intron or 2 intron is boxed in green or red, and phase 0 intron showed by black lines. [&] represents only signal peptide and kuntiz-domain in human eppin are displayed. (b) Phylogenetic tree constructed from the alignment of amino acid sequences present in Figure S3 by iqtree with a maximum-likelihood method. Branches with a significant bootstrap value are indicated by black circles for SH-aLRT > 90 and white for UFBoot > 85. (c) 3D models of MdEppin-1 and 35-1. The disulfides are shown as color sticks (blue for MdEppin-1 and red for MdEppin35-1) with the conserved one indicated by a blue arrow. (d) The cicro visualization of the coevolution of MdEppin. Amino acid names and the position are in the outer ring. Conservation (second ring) from light blue (lower) to red (higher); cScore (third ring) from yellow (lower) to violet (higher). pScore (inner ring) form green (lower) to red (higher). Inner lines are the top 5% covariation scores. (e) Weblogo of MdEppins, and the cola of positions on cScore from yellow (lower) to violet (higher) is shown on the top with the distance being displayed. The P1 site is arrowed in blue and the positively selected sites are arrowed in turquoise. The highly conserved domain is boxed in green.

Table 2. Maximum likelihood estimates of parameters and sites inferred to be under positive selection in the *M. domestica* eppins.

Model	l	LRT	Parameters	Positive Selected Sites
M0	-5115.94		$\omega = 0.10448$	None
M1a	-4880.09		$\rho 0 = 0.32678, \omega 0 = 0.05181$	Not allowed
			$\rho 1 = 0.67322, \omega 1 = 1.00000$	Not allowed
M2a	-4872.92	11.22	$\rho 0 = 0.32607, \ \omega 0 = 0.05760$	7R. 12K *. 14V. 18L *. 20P. 33E.
			$\rho 1 = 0.52462, \ \omega 1 = 1.00000$	47P, 57Q
M7	1765 64		$\rho 2 = 0.14931, \omega 2 = 2.14278$ $m = 0.52104, \alpha = 2.05121$	Not allowed
1017	-4703.04		p = 0.32104, q = 2.03131	Not allowed
M8	-4765.64	0.00	$p0 = 1.00000, p = 0.52104, q = 2.05131$ $(p1 = 0.00000), \omega = 1.00000$	None

Note: $2\Delta l$ between null models (M1a and M7) and their alternative models (M2a and M8): M1a/M2a = 11.22 (df = 2, p = 0.00366); M7/M8 = 0.00 (p = 1). Positively selected sites identified by the BEB method under M2a with posterior probabilities P > 0.6 are shown and those with P > 0.95 indicated by "*".

3.1.3. MdMuslins

Kazal-type serine proteinase inhibitors (KaSPI) were firstly isolated by Kazal and colleagues from pancrease [41]. KaSPIs have been identified in many insects such as mosquitos [47], *Drosophila* [48] and locusts [49]. They have a broad activity in various biological and physiological processes in many organisms, such as blood coagulation and innate immunity. Interestingly, KaSPIs exhibit an antibacterial activity in response to microbial infection. The recombinant CsKSPI inhibits the growth of Gram-positive and Gram-negative [50], and PSKP-1 and its variants reduce *E. coli* mobility and cell agglutination [51].

In the housefly, 24 muslins (named MdMuslin1-MdMuslin24) are identified to contain the typical kazal domain with three disulfide bridges and nine muslins (named MdMuslin25–MdMuslin33) contain four disulfide bridges (Figure 3a and Figure S4). Among them, four muslins (MdMuslin15, MdMuslin16, MdMuslin23 and MdMuslin24) contain two kazal domains and named -1 and -2. Our phylogenetic tree reveals four types of Mdmuslins (Figure 3b), each type clustering together, in favor of their monophyletic origin. Of them, type I to III contain three disulfide bridges (i.e., C1-C5, C2-C4, C3-C6) which are different from the peptides containing the kuntiz-domain. Their sequence pattern can be drawn as CX1-3CX5PVCX0-5GX6-9NX1-5CX3-6CX7-22C. In the tree, the two domains in the paralogous MdMuslin15 and MdMuslin16 are classified into two different types, indicating that their significant divergence occurred after domain repeats. Similarly, in other two kazal domain-containing members (MdMuslin23 and MdMuslin24) their domain-2 is categorized into type III and domain 1 into type IV. For the eight cysteinescontaining members except MdMuslin23-1 and MdMuslin24-1, their sequence pattern can be described as CX1-3CX5PVCX5-6GX5-6CX3NXCX6CX7-12CX2-5C (Figure 3c).



Figure 3. MdMuslins-like AMPs. (a) Multiple sequence alignment (MSA) of representative MdMuslins (For the full-set MdMuslins, see Figure S4) and kazal domains from other sepcies. Two serine residues mutated from the conserved cysteines

are circled. Cysteines involved in the formation of disulfides are colored in yellow. Conservation replacements are shadowed in grey. Basic residues (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. The P1 amino acids are italicized and shadowed in cyan. Residues split by phase 1 introns are shadowed in green. α -helix and β -sheets are also indicated at the bottom together with the potential disulfides being showed with black lines, and the fourth bridge by dotted line. Aae: *Aedes aegypti* (GenBank: ABF18209.1). Aal: *Aedes albopictus* (GenBank: JAC06964.1), Ac: *Apis cerana* (GenBank: AGW24880.1), Cs: *Channa striata*(GenBank: CDG86164.1), Cp: *Culex pipiens pallens* (GenBank: AFN41343.1), DaCOW: *Drosophila ananassae* (GenBank: XP_001953960.2:219-270), Df: *Drosophila ficusphila* (GenBank: XP_017047190.1), Ds: *Drosophila simulans* (GenBank: XP_002105007.1). (b) Phylogenetic tree of the sequences constructed from the alignment of amino acid sequences present in Figure S4 by iqtree with a maximum-likelihood method. Branches with a significant bootstrap value are indicated by black circle for SH-aLRT > 90 and white for UFBoot > 85. Muslins could be divided into four groups denoted by different colors. (c) 3D structures of MdMuslin1, MdMuslin2, MdMuslin25 and MdMuslin26 (Cys1) are displayed. C4-C8 in MdMuslin25 and MdMuslin26 is pointed out by a black arrow. (d) The coevolution of MdMuslins in *M. domestica*. (e) Weblogo of MdMuslins. The results indicate the residues in MdMuslins are changeable, and the potential site likely contributing to evolution of P1 (blue) are displayed with the distance (cola of the c-Score) and position (turquoise).

In MdMuslin1 the first cysteine is replaced by a serine and in Muslin26, the fifth cysteine is lost, leading to the initial first disulfide bridge disrupted. In these two molecules, the free cysteines are exposed to their molecular surface, as revealed by our structural models (Figure 3c), suggesting that they might participate in the formation of a homodimer. The conservation analysis indicates that the R¹⁰ (P1 site shown as MdMuslin2) has an inner relationship with P⁷, N¹¹ and P¹³ (Figure 3d,e).

3.1.4. SVWC Domain AMPs

Single domain von willebrand factor type C (SVWC) proteins mostly contain eight cysteines. Although the *Bombyx mori BmSVWC* gene was decreased in the cuticle when the insect was infected with fungi [52], granularin from the snails *Lymnaea stagnalis* was up-regulated during parasitation of the avian schistosome *Trichobilharzia ocellata*, in favor of a role of this class of proteins in the molluscan internal defense response [53].

In the housefly, the SVWC-type AMPs are expanded to a family comprising 29 small, single domain secreted proteins (named MdSVWC1-MdSVWC29) (Figure 4a and Figure S5). Among them, MdSVWC29 is unique in that it contains two classical fragments, named MdSVWC29-1 and MdSVWC29-2. The family displays a consensus pattern as CX18-23CX4CX10-12CX7-10CX11-14CCX1-5C. They are diverse in sequence, but the eight cysteines are conserved throughout the group to form four disulfide bridges (i.e., C1-C3, C2-C6, C4-C7, and C5-C8). Both conserved introns (phase 1 and phase 2) are located at the nearly identical position among all peptides whose gene structures are available, supporting their common origin (Figure S5). Based on our phylogenetic tree analysis, these MdSVWC peptides can be divided into two distinct groups (Figure 4b). The predicted 3D structures of MdSVWC1 and MdSVWC17 reveal a typical structure in their N-termini that contain a four β-stranded sheet (residues M³-F⁵, T¹²-E¹⁴, S¹⁸-E²⁰, R²⁷-T²⁹ in MdSVWC1 and C⁶-V⁸, K¹¹-V¹³, G¹⁶-H²¹, T²⁷-D³² in MdSVWC17). Residues P³⁷-L⁴¹ are folded into an α -helix in MdSVWC1, but G³⁶-E⁴¹ in MdSVWC17 are folded into a β -sheet. In addition, the C-terminus of MdSVWC1 forms a β -sheet (residues K⁵⁸-D⁶⁰ and F⁷⁷-C⁷⁹) and an α -helix $(Y^{87}-V^{91})$ (Figure 4c).



Figure 4. MdSVWC-domain peptides. (**a**) Comparison of representative MdSVWCs (For the full-set sequences, see Figure S5) and homologs from other species. Cysteines involved in the formation of disulfide bridges are colored in yellow. Basic residues (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. Disulfide bridges, α-helix and β-sheets are showed at the bottom. Residues split by phase 1 and 2 introns are shadowed in green and red, respectively. Bm: *Bombyx mandarina* (GenBank: XP_028031732.1), Dw: *Drosophila willistoni* (GenBank: XP_015032364.1), Granularin: *Lymnaea stagnalis* (GenBank: AAS20460.1), Lv: *Litopenaeus vannamei* (GenBank: HQ541159.1). (**b**) Phylogenetic tree constructed from the alignment of amino acid sequences present in Figure S5. Branches with a significant bootstrap value are indicated by black circle for SH-aLRT > 90 and white for UFBoot > 85. (**c**) The 3D models of MdSVWC1 and MdSVWC17. The four-β-stranded sheet and the α-helix are located on the N-terminus of MdSVWC1. For MdSVWC17, the α-helix in its N-terminus is replaced by a β-strand.

3.1.5. Crustin-like AMPs

Crustins are antibacterial proteins with a precursor organization including a signal peptide at the N-terminus and a whey acidic protein (WAP) at the C-terminus [54,55]. Crustins have been identified in diverse invertebrate animals, with a WAP domain of approximately 50 amino acids containing a conserved motif and a four disulfide bridge core (4-DSC) in the C-terminus. In previous studies crustins are divided into four families, crustin I contains a cys-rich domain, crustin II contains a glycine-rich domain in the N-terminus. Only one WAP domain exist in crustin III and thus are also named a single WAP domain-containing peptides (SWDs). Type IV contains a cysteine-rich, an aromatic-rich region and a WAP domain in the C-terminus. This type of crustins is exclusively present in ants. Crustins are widely regarded as antimicrobial molecules since they can kill Grampositive bacteria [56,57] and some fungi [58]. Moreover, LcSWD3 isolated from *L.vannamei* may contribute to antiviral immune response [59].

Even though the mechanism of action of crustins on pathogens remains unknown, it appears clear that the common WAP domain in all crustins plays a key role in their antibacterial effect. This has been evidenced by several previous observations: (1) the crustin in *Fenneropenaeus chinensis*, which only contains a glycine-rich region without a WAP domain, exhibits no antibacterial activity [60]. (2) SWAM1 and SWAM2 in mice which belong to the crustin III family can inhibit the growth of both *E. coli* and *S. aureus* [61]. Additionally, the reduction and alkylation of the cysteine residues in the WAP domain of a crustin-like peptide from a snake venom destroyed its antimicrobial activity [62].

The housefly crustins (named MdCrustins) are identified as types V and VI based on their sequence characteristics. In comparison with all other crustins, these peptides have developed an extra C-terminal region accompanying the loss of the cysteine-rich or glycine-rich region located between the signal peptide and the WAP region (Figure 5a; see sequences in Figure S6). In the phylogenetic tree, the housefly crustins are clustered together with type III crustins (Figure 5a), supporting their close evolutionary relationship. Structural modeling suggests that the WAP domain in both MdCrustin1 and MdCrustin3 may form four disulfide bridges with a connectivity pattern as C1-C6, C2-C7, C3-C5 and C4-C8. Compared with MdCrustin1, the C-terminal region of MdCrustin3 may be more rigid given that it folds into several β -strands stabilized by two disulfide bridges (C9-C10 and C11-C12) (Figure 5b).



Figure 5. The phylogenetic tree and structural models of Crustins. (**a**) Phylogenetic tree of crustins (details in Figure S6) inferred using iqtree-1.62. Branches with a significant bootstrap value are indicated by black circle for SH-aLRT > 90 and white for UFBoot > 85. Members belonging to the same subtype are clustered together and their branches are marked by the same color with six subgroups designated. The predicted domains are also showed. (**b**) Structure of MdCrustin1 (yellow) and MdCrustin3 (blue). Disulfides are shown as green sticks and the C-terminal cysteine-rich domain in MdCrustins3 is circled in red.

3.2. *Linear α-helical Peptides* Cecropins

Cecropins are a group of classical linear α -helical peptides containing 31–39 residues with a molecular weight of about 4 kDa. The first cecropin was firstly isolated from *Hyalophora cecropia* [63] and later found in many insects, such as Diptera [10,64–68], Lepidoptera [69] and Coleoptera [70] and so on, but not in Hemiptera [4]. Cecropins display a broad spectrum of activity against Gram-positive and Gram-negative bacteria, fungi and HIV virus [63,71]. In some cecropins, Trp² and Phe⁵ (e.g., cecropin A and papiliocin) was found to contribute to interactions with the negatively charged bacterial membrane [72] whereas Gly ¹, Trp², Lys⁴ and Lys⁵ in sarcotoxin-IA are important for binding with lipid A of LPS [73,74]. In cecropins, the hinge region disrupting the long helix is important for their structural flexibility [75].

In the housefly, 11 cecropins have been identified to contain a typical precursor organization [64]. Using our method, we found another five homologs that share similarity with those found in Diptera (Figure 6a). As reflected by MdCecropin2, in a hydrophobic environment, these peptides adopt an α -helical conformation, in which two α -helices (W²-Q²³ and G²⁶-G⁴⁰) are joined by a flexible hinge comprising G²⁴ and L²⁵ (Figure 6b). The N-terminal helix is strongly basic and the C-terminal one is hydrophobic (Figure 6b).



Figure 6. Cecropin-like peptides. (**a**) MSA of cecropins. Basic residues (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. Conserved residues are shadowed in cyan. # represent the peptides previously known. Dm: *Drosophila melanogaster* (NP_524588.1). The black line denotes the position a conserved phase 0 intron. (**b**) Helical-wheel and

spheres diagram of MdCecropin2. Left: The helical wheel projection showing the amphiphilic characteristics of the cecropin. By default, the output presents the hydrophilic residues as circles, hydrophobic residues as diamonds, potentially negatively charged as triangles, and potentially positively charged as pentagons. Hydrophobicity is color coded as well: the most hydrophobic residue is green, and the amount of green is decreasing proportionally to the hydrophobicity, with zero hydrophobicity coded as yellow. Hydrophilic residues are coded red with pure red being the most hydrophilic (uncharged) residue, and the amount of red decreasing proportionally to the hydrophilicity. The potentially charged residues are light blue. Middle: Carton model of MdCecropin2. The residues are colored as followed: the hydrophilic residues are in cyan, positive ones are in blue, negative are in red, and the hydrophobic ones are in gray. Glycine and Leucine residue at position 24, 25 serves to connect the two helices which are shown as stick as well. Right: The spheres diagram shows the structure of MdCecropin2 with the carton picture.

3.3. Specific Amino Acid-Rich AMPs

3.3.1. Domesticins, Diptericins and Edins

Domesticins are a class of proline-rich AMPs active against Gram-positive and Gramnegative bacteria and some fungi [76–78]. In our data mining, we found a new domesticin, named MdDomesticin2 (Figure 7 and Figure S7), with a proline content of 27.5%. Diptericins are about 9 kDa of glycine-rich peptides active on Gram-negative bacteria, which were initially isolated from the fly *Phormia terranovae* [79]. Some studies have shown that they are bacterially induced through the IMD signaling pathway [80]. In addition, *Diptericin* in mosquito larva is up-regulated after Sindbis virus infection [81]. For two housefly-sourced diptericin genes that were not named previously but have been found to be transcriptionally active when house fly larvae and pupae were injured [37], we named them *MdDiptericinD* and *MdDiptericinD1*. Their proteins share high similarity with other four diptericins previously named (Figure 7 and Figure S8). These diptericins can be clearly divided into two domains: a proline-rich domain and a glycine-rich domain. We found there are a phase 0 intron disrupting the proline rich domain of *MdDiptericinD* and *D1* (Figure S8).

Edins are a class of inducible insect AMPs [82,83] with a precursor comprising a signal peptide, a propertide ending with a RXXR motif and a mature glycine-rich domain. Interestingly, in the housefly MdEdin6-MdEdin10 have two Gly-rich domains, which is different from the only single glycine-rich domain in the orthologs from other insects and MdEdin1-MdEdin5 in the housefly (Figure 7 and Figure S9).

3.3.2. Attacins

Attacins are a class of 20–23 kDa proteins found in Lepidoptera and Diptera. Attacins have two types, the basic attacins (A-D) and acidic attacins (E-F) [6]. All attacins share a high similarity in their amino acid sequences, but more aspartic acids exist in the acidic attacins. The precursor organization of attacins contains a signal peptide, a propeptide ending with a conserved RXXR, an attacin-N domain, an attacin-C domain (also called G1 domain and G2 domain). In the housefly, 16 attacins can be divided into three types (named MdAttacinA, MdAttacinC and MdAttacinD), in which three MdAttacinA, one MdAttacinC (Herein named MdAttacinC2) and one MdAttacinD (MdAttacinD4) have been named. Not like attacinA and attacinB in Drosophila, MdAttacinAs in the housefly lacks a propeptide. Their glycine proportion in the N-domain is universally higher than that in the N-domain of the fruit fly counterparts (Figure S10). In our study, we found two new MdAttacinCs (MdAttacinC6 and MdAttacin7) (Figure S11). In addition, although some MdAttacinCs have been reported [37], we found that the original MdAttacinC3 contains two whole attacin sequences and thus named C3-1 and C3-2 (Figure 8). MdAttacinCs contain a longer propeptide ending with an RXXR motif. The glycine residues in the attacin-N domain are much less than those in the domain of the fruitfly counterparts, but higher in attacin-G1 and G2 domain (Figure S11). MdAttacinDs lack a signal peptide (Figure S12). However, like *MdAttacinA3*, *MdAttacinD3* (herein named) was also expressed in larval tissues after injury [37]. Compared with the fruit fly attacinDs, we found that the glycine percentages largely varied due to sequence divergence.

а

DsIp18 MdDomesticin1# MdDomesticin2 DmDiptericin MdDiptericin1# MdDiptericinD1# DmEdin MdEdin1 MdEdin2 Mdedin6 DmAttacinC MdAttacinC1# MdAttacinC2# DmAttacinA MdAttacinA1# MdAttacinA2

b

DmAttacinA MdAttacinA1# MdAttacinA2 DmAttacinC MdAttacinC1# MdAttacinC2# DmAttacinD MdAttacinD1 MdAttacinD3#

DmAttacinA MdAttacinA1# MdAttacinA2 DmAttacinC MdAttacinC1# MdAttacinC2# DmAttacinD MdAttacinD1 MdAttacinD3#

DmDiptericin

DmEdin MdEdin1 MdEdin2 MdEdin6 MdEdin6-2

DmAttacinA MdAttacinA1# MdAttacinA2 DmAttacinC MdAttacinC1# MdAttacinC2# DmAttacinD

MdAttacinD1

MdAttacinD3#

С



Figure 7. MSA of specific amino acid-rich AMPs. (a) Proline-rich AMPs. Prolines are shadowed in pink and the RXXR motifs are boxed in red. (b) and (c) Glycine-rich AMPs (G1 and G2 domains). Glycines are highlighted in dark red. Basic

residues (K, R, H) and acidic residues (E, D) are highlighted in blue and red, respectively. # represent the peptides previously known. The phase-2 introns are shadowed in red. The details of these peptides including Domesticin, Diptericin, Edin, AttacinA and AttacinC, and AttacinD are provided in Supplementary Information Figures S7–S12, respectively. DsIp18: *Drosophila serrata* (GenBank: XP_020804442.1), DmDiptericin: *Drosophila melanogaster* (GenBank: AAB82521.1), DmEdin: *Drosophila melanogaster* (GenBank: NP_730278.1), DmAttacinA: *Drosophila melanogaster* (GenBank: ABS52579.1), DmAttacinC: *Drosophila melanogaster* (GenBank: NP_523729.3), DmAttacinD: *Drosophila melanogaster* (GenBank: NP_524391.2).



Figure 8. Phylogenetic tree of proline-rich AMPs in insects. The tree was constructed based on the MSA of proline-rich peptides in *M. domestica* and other insects. Branches with a significant bootstrap value are indicated by black circle for SH-aLRT > 90 and white for UFBoot > 85. For the sequence, prolines are shadowed in pink with percentage of proline residues being calculated. The identity percentages compared with Formaecin-1 were calculated as well. The RxRR motifs are boxed in red. # represent the peptides previously known and "a" represents known mature peptides. The taxonomy is displayed on the right. BmLebocin: *Bombyx mori* (GenBank: AAB35218.1), DmAttacinA: *Drosophila melanogaster* (GenBank: ABS52579.1), DmAttacinB: *Drosophila melanogaster* (GenBank: Q9V751.2), DmAttacinC: *Drosophila melanogaster* (GenBank: NP_523729.3), DmDiptericin: *Drosophila melanogaster* (GenBank: AAB82521.1), DmDrosocin: *Drosophila melanogaster* (GenBank: CAA79936.1), DmMetchnikowin: *Drosophila melanogaster* (GenBank: NP_523752.1),HvHeliocin: *Heliothis virescens* (GenBank: P83427.1), Mg: *Myrmecia gulosa* (GenBank: P81438), PaPyrrhocoricin: *Pyrrhocoris apterus* (GenBank: P37362), PpMetalnikowin: *Palomena prasine* (GenBank: P80408).

Our phylogenetic analysis based on the MSA of diptericin, domesticin and attacinC, a group of proline-rich AMPs described above, reveals a close relationship between MdAttacinC and the *D. melanogaster* attacinC and a paralogous relationship between diptericins and domesticins (Figure 9). In a similar manner, we analyzed edins, diptericins and attacins, a group of glycine-rich AMPs described above. These can be divided into eight groups based on the evolutionary tree (Figure 9) where the Attacin-N domains belonging to the type attacinA are grouped together (named AttacinA-N group) with the exogenous Attacin-N in the attacinA-D from the fruit fly (Figure 9). The N-domain of attacinD except attacinD4 in the housefly are clustered with the AttacinA-N group. The Attacin-N peptides in type attacinC and attacinD4 are clustered together (named AttacinD1. AttacinC-G1 group contains the G1 domain of attacinC and attacinD4. AttacinA-G2 group contains the G2 domain of attacinC and attacinD4. The diptericin group contains only one Gly-rich domain. All Edins are grouped together (Figure 9). These results reveal that edins and the G2 domain of attacinC have the closest evolutionary relationship and diptericins are closer to the G2 domain of attacinA. Even though MdEdin10 has been named Attacin, we found that it has a closer relationship with the Edin family according to the evolutionary tree (Figure 9). The domain architecture of different proline-rich and glycine-rich AMPs are presented in Figure 10.



Figure 9. Phylogenetic tree of the glycine-rich domain in the MdAttacins, MdEdins, MdDiptericins families. The G1-domain of attacinA-D from the fruit fly was used as an outgroup. The different group are shadowed in different colors. The N-domain, G1 and G2 domain are abbreviated as N, G1 and G2. Braches with a significant bootstrap value are indicated by black circle for SH-aLRT > 90 and white for UFBoot > 85. # represent the peptides previously known.



Figure 10. Structural domains of the specific amino acid-rich AMPs. Different domains are presented in different colors. Notes are shown in the right bottom. # represent the peptides previously known.

4. Discussion

Although some AMPs have been identified previously in *M. domestica* [77,84,85], more potential AMPs may be discovered by surveying its whole genome and related databases. Here, we identified 148 potential AMPs in *M. domestica* with different structural types. Compared with *D. melanogaster*, *M. domestica* has evolved a set of more complex antimicrobial components with diversification in number, protein size, and structure. For instance, the housefly has 21 defensins with a variable n-loop in length, but *D. melanogaster* has only one. Besides the increase in number, structural alteration by extension or truncation also contributes to the diversity. For example, some members belonging to MdEppins, MdMuslins and MdEdins have extended their C-termini compared with those in *D. melanogaster*. What attracts us most is which evolutionary or genetic mechanisms have shaped the diversity in the housefly?

4.1. Gene Duplication

Gene duplication has been found in nearly all organisms. In the *M. domestica* genome, there are at least seven AMPs families exhibit an adjacent chromosome location with gene structure conservation, which can be considered as a consequence of tandem duplication (Figure 11). Gene duplication followed by positive selection is important for the creation of new biological functions, which has been observed in the evolution of insect multigene of AMPs [20]. In the housefly, we have detected some positive selection signals in the MdDefensin and MdEppin families (Tables 1 and 2) but not in other families. In addition,

as mentioned above, some structural variations are also observed among different paralogs in the housefly defensin family via dynamic insertion/deletions (indels) in their n-loop (Figure S2). Such change could have an impact on their antimicrobial activity [20].



Figure 11. Gene duplication in the *M. domestica* AMP genes. (a) The clusters of MdDefensins. The two clusters are respectively located on Genbank ID AQPM01000006.1 and AQPM01000010.1. (b) The clusters of MdEppins. The big cluster is located on GenBank ID AQPM01069148.1; and two small clusters are respectively located on GenBank ID NDYK01230132.1 and AQPM01069146.1. (c) The cluster of MdMuslins (AQPM01019378.1 and AQPM0100603.1). (d) The clusters of MdSVWCs (AQPM01095428.1 and AQPM01095425.1). (e) The clusters of MdCecropins (AQPM01058001.1, AQPM01058004.1 and AQPM01100428.1). (f) The clusters of MdEdins (AQPM01067938.1 and AQPM01067936.1). (g) The MdAttacinC cluster (AQPM01059487.1). Chromosome fragments are shown in lines and genes coding AMPs are represented by boxes. Introns are indicated in triangles with different colors.

4.2. Exon Duplication via Shuffling

Exon duplication-mediated internal repeats plays an important role in the evolution of proteins as it can create an obvious complexity increase and likely contributes to the emergence of new functions. From the house fly genome, we identified an unusual eppin protein (MdEppin35) that carries nine repeats of kuntiz domain without a protease processing signal. In the MdMuslin family, MdMuslin15, MdMuslin16, MdMuslin23 and MdMuslin24 all share two kazal domains. Moreover, in the MdEdin family, MdEdin6 - MdEdin10 contain two glycine-rich domains (Figure 12a–c). Since there are two same phase introns (i.e., phase 0) corresponding to the boundaries of the MdEppin35-1 domain, we speculate that the evolution of multiple Kuntiz-domains might be a result of exon-shuffling, as illustrated in the schematic diagram (Figure 12d). In this case, the lack of introns at other domain boundaries could be explained by intron loss during evolution [86].



Figure 12. Domain repeats in different *M. domestica* AMPs. (a) MdEppin35. (b) MdMuslin16. (c) MdEdin6. Each repeated domain structure is also shown here. The disulfides are shown as blue sticks. Introns are indicated in triangles with phase 1 in green, phase 2 in red, and phase 0 in blue. (d) Schematic diagram showing a potential exon-shuffling creating the multiple-domain MdEppin35. In this process, the two same phase introns (here phase 0) located on the boundaries of the domain could contribute to its insertion into an ancestor via intron-mediated shuffling.

4.3. Protein Terminal Variations

In comparison with their paralogs some members belonging to the three housefly AMP families (i.e., MdCrustins, MdDefensins and MdEppins) have changed their terminal length through truncation or extension. For example, compared with the ancestral state present in the insect lineage (e.g., MdCrustin3, MdCrustin4 and ArWaprinThr1), MdCrustin1 and MdCrustin2 have evolved to form a truncated C-terminus (Figure 5) through the deletion of a disulfide-bonded sub-domain structure (Figure 5 and Figure S6). In the MdDefensins, several members extended their N-termini via the loss of a pro-peptide processing signal and the extended fragment in MdDefensin17 forms an isolated sub-domain structure. In the MdEppin members (e.g., MdEppin4, MdEppin6, MdEppin25, MdEppin33 and MdEppin34), all have an extended C-terminus of 20–103 amino acids (Figure S3). These observations indicate a clear structural diversification occurring among different paralogs of housefly AMPs and could hint at their functional divergence, an open question to be answered in the future.

4.4. Evolution of New Disulfide Bridges

Disulfide bridges are important to both protein structure and function [87]. In this work, we found that several *M. domestica* AMPs have odd cysteines. For example, Md-Defensin20 has evolved one additional cysteine in its N-terminus (Figure 1) whereas Md-Muslin1 and MdMuslin26 have a cysteine mutation to a non-cysteine residue (Figure 3c). To a secreted protein, the presence of a free cysteine often is detrimental to its structural stability due to air oxidization, especially when this residue is exposed to the molecular surface. We thus speculated this free cysteine might be involved in the formations of a homodimer structure, as previously observed in scorpion venom lipolysis activating peptides [88]. This speculation is supported by our structural modeling, in which one intermolecular disulfide bridge is clearly formed between two monomer MdDefensin20 (Figure 13a). A Ramachandran plot indicates that in this homodimer almost all φ/ψ torsion angles are found in the favored or additionally allowed regions except several residues (Figure 13b). In MdMuslin1, the fifth cysteine in monomer lost the opportunity to form

the intramolecular disulfide bridge, but alternatively an intermolecular disulfide bridge could be found by ZDOCK method (Figure 13c), Ramachandran plot indicates that in this homodimer almost all φ/ψ torsion angles are found in the favored or additionally allowed regions except a serine in position 5 (Figure 13d). A similar observation is also made in MdMuslin26, in which the loss of the sixth cysteine leads to the first cysteine residue being not paired and thus one predicted intermolecular disulfide bridge links the two monomers into a homodimer (Figure 13e,f).



Figure 13. The homodimers of MdDefensin20, MdMuslin1 and MdMuslin26. (**a**,**c**,**e**) The 3D structures with the interchain disulfide bridges built between two monomers denoted by blue cyan arrows. (**b**,**d**,**f**) Ramachandran plot analysis with PROCHECK. In this plot, almost all φ/ψ torsion angles are found in the favored or additionally allowed regions and only few residues are in a disallowed region.

In the two types of MdMuslins, the type I peptides are shared with the orthologs from a diversity of species whereas the type II peptides are only shared with the orthologs from a Diptera species. This observation suggests that the former might represent the ancestor from which the latter emerged via evolutionary gain of one additional disulfide bridge in the common ancestor of Diptera. Since members with this disulfide bridge all have a long C-terminus whereas members without this bridge exhibit more divergence in their C-terminal length, it appears that gaining a disulfide bridge during evolution could have an impact on the stabilization of the size of a protein. In Mdeppin35-1, its two unusual cysteines are located at the C-terminus. Due to a deletion in its N-terminus, this peptide lose the first two cysteine residues compared with other paralogs (Figure 2), leading to the disruption of the original disulfide bridge pattern, which probably makes it become a pseudogene in the case of the loss of its structural stability. However, our structural modeling indicates that the two C-terminal cysteine residues can provide new pairings for disulfide bridge formation (Figure 2) to save the life of this gene via restoring its structural stability. These observations suggest that when the structurally important cysteines are substituted or deleted in protein evolution, compensation might be a choice.

5. Conclusions

In prior studies, some insect-derived AMPs have been found in the housefly based on biochemical purification combined with functional identification as well as the analysis of genomic sequences. In this work, we used a combination of computational approaches to establish a relative complete *M. domestica* peptidome associated with its antimicrobial immunity, which largely expands the repertoire of AMP-like molecules in a sanitary insect pest. These molecules exhibit considerable diversity in their gene numbers and structural type with some new architectures that may be assembled as a homodimer structure or display repeats by multiple homologous domains or even a novel fold different from its ancestral state. It is clear that such diversity can attribute to several evolutionary scenarios involving gene and exon duplication, terminal variations and disulfide bridge reconstruction. Although the housefly has a closer phylogenetic relationship with Drosophila (both belonging to the Order Diptera), their evolution in developing their antimicrobial immune system remarkably differs. Compared with Drosophila, the housefly seems to have increased the complexity of the system, a similar case also previously reported in the parasitoid Naso*nia vitripennis*. This might be a result of evolutionary convergence in facing the selective pressure towards parasitism and a dirty microbe-rich environment. Our work will offer a basic platform to further study the immune and evolutionary significance of these newly discovered AMPs and the role of the molecular and structural diversity in contributing to the immune response of houseflies.

Supplementary Materials: The following are available online at https://www.mdpi.com/1424-281 8/13/3/107/s1, Figure S1: The strategy for database searches for putative *M. domestica* antimicrobial peptides, Figure S2: Multiple sequence alignment (MSA) of defensins, Figure S3: MSA of MdEppins, Figure S4: MSA of MdMuslins, Figure S5: MSA of SVWC AMPs, Figure S6: MSA of Crustins, Figure S7: Domesticins, Figure S8: MSA of Diptericins, Figure S9: MSA of Edins, Figure S10: MSA of AttacinA, Figure S11: MSA of AttacinC, Figure S12: MSA of AttacinD.

Author Contributions: S.Z. conceived and designed research. S.Q. performed sequence, structural and evolutionary analyses. B.G. performed energy minimization analyses of peptide structures. S.Q. wrote the manuscript with assistance from all other authors. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 31870766) to S.Z.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data supporting the reported results can be found at https://www.ncbi.nlm.nih.gov/.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Characteristics o	AMPs in Musca d	omestica.
-----------------------------	-----------------	-----------

			-	Size				
Name	WGS/EST	GenBank No.	Туре	ORF	MP	MW(Da)	Net Charge	pl
MdDofoncin1#		A ID08287 1	CoxB	02	41	4159 74	2.2	8 20
MdDefensin2#	AV260152 1	AII 90307.1 A A P33451 1	Csap	93	41	3005 56	3.2	8 31
MdDefensin3 #	K1867444 1	AII 24687 1	Csap	92	40	3995.56	3.2	8 31
MdDefensin4 #	AOPM0100006 1:16485-16782	XP_005174767.1	Csab	93	40	3995 56	3.2	8 31
MdDefensin5#	FF175879 1	ABM66377 1	Csab	93	40	3995 56	3.2	8 31
MdDefensin6 [#]	AOPM0100006 1:14681-14958	XP_005174766.1	Csab	92	40	3996 56	3.2	8 31
MdDefensin7#	AOPM01000006 1:10815-10994	AGS57597 1	Csab	91	40	4009 59	3.2	8 31
MdDefensin8	AOPM01000006 1·19117-19389	XP_005174768.1	Csab	91	40	4228.88	4.2	8.55
MdDefensin9	KM047667.1		Csαβ	97	41	4158.74	3.2	8.29
MdDefensin10	AQPM01000006.1:27384-27647		Csαβ	88	44	4642.39	3.9	8.29
MdDefensin11	NDYK01163469.1:1738-2002		Csαβ	72	44	4642.39	3.9	8.29
MdDefensin12	NDYK01012409.1:563-805		Csαβ	81	41	4454.16	2.7	8.29
MdDefensin13 #	AQPM01000009.1:1990-2309	XP_005174769.1	Csαβ	83	42	4383.17	2.0	8.03
MdDefensin14	NDYK01067664.1:1236-24		Csαβ	83	42	4383.17	2.0	8.03
MdDefensin15 #	AQPM01000008.1:929-1186	XP_011291282.1	Csαβ	86	41	4232.99	2.2	8.03
MdDefensin16 #	AQPM01000010.1:7421-7708	XP_011292449.1	Csαβ	96	43	4707.76	6.0	8.99
MdDefensin17		XP_011290793.1	Csαβ	75	52	5492.29	1.0	7.66
MdDefensin18	AQPM01000010.1:5050-5322	XP_019893215.1	Csαβ	91	41	4495.13	3.0	8.29
MdDefensin19	AQPM01000010.1:909-1677		Csαβ	84	38	4360.19	2.7	8.29
MdDefensin20	NDYK01134510.1:970-1187		Csαβ	64	45	5075.93	0.9	7.61
MdDefensin21	AQPM01000007.1:876-1139		Csαβ	88	65	7013.82	1.0	7.66
MdEppin1		XP_005182439.1	kuntiz domain	120	90	10213.55	-1.2	5.55
MdEppin2	NDYK010/3931.1:19/5-2334		kuntiz domain	103	73	8437.57	1.7	7.98
MdEppin3	AQPM01068620.1:8089-8448	VD 00E100(0E1	kuntiz domain	103	73	8437.57	1.7	7.98
MdEppin4	AQPM01069148.1:21728-22230	XP_005182685.1	kuntiz domain	108	149	10083,41	-19	4.55
MdEppin5		XP_005192098.1 XP_005102000.1	kuntiz domain	108	80 172	9321.00	4./	8./1 5.72
MdEppino MdEppin7	NDVK01008682·1427-1791	XF_003192099.1	kuntiz domain	201	75	8489 55	-1.2	7.66
MdEppin/ MdEppin8	ND1R01000002.1427-1791	YP 005182689.1	kuntiz domain	99	73	8636 78	7.0	0.00
MdEppin9	AOPM01082327 1.250-573	XI_000102009.1	kuntiz domain	108	91	10121 48	-1.1	5.97
MdEppin10	AQ1 W01002527.11.250 575	XP 0198930361	kuntiz domain	83	66	7318 19	1.0	7.66
MdEppin11	AOPM01068767.1:4133-4939	01)0000001	kuntiz domain	83	60	7255.20	7.9	9.13
MdEppin12	AOPM01058660.1:4502-4813		kuntiz domain	104	83	9255.26	-2.1	5.27
MdEppin13	2	XP 011291014.1	kuntiz domain	84	63	6886.53	-6.0	4.30
MdEppin14	AQPM01069148.1:1494-1791	XP_005182706.1	kuntiz domain	85	63	6862.55	-3.0	4.87
MdEppin15	AQPM01069148.1:7578-7914	XP_005182682.2	kuntiz domain	91	63	6966.53	-8.0	4.26
MdEppin16	AQPM01069148.1:12552-12854	XP_005182683.1	kuntiz domain	82	63	6946.74	-2.0	5.27
MdEppin17	AQPM01069148.1:3783-4100	XP_005182681.1	kuntiz domain	84	64	7101.81	-2.8	5.36
MdEppin18	AQPM01069148.1:15772-16090		kuntiz domain	86	66	7424.27	-1.0	5.97
MdEppin19		XP_005182684.1	kuntiz domain	83	63	7036.78	-3.0	5.05
MdEppin20	NDYK01230132:2648-2965		kuntiz domain	84	64	7098.93	-1.8	5.78
MdEppin21	NDYK01230132:3440-3753	20 6 005100 (00 0	kuntiz domain	87	67	7507.45	-1.6	6.25
MdEppin22		XM_005182623.3	kuntiz domain	83	64	7162.96	-2.5	5.69
MdEppin23	A ODM 010(014(1.10504 10800	XP_005182678.1	kuntiz domain	79	59	6444.22	-3.0	4.94
MdEppin24	AQPM01069146.1:10524-10829	XP_011292137.1 XP_010802012.1	kuntiz domain	78	59	0001.00	2.7	8.27 4.75
MdEppin25	AQF101009140.1:7608-6195	XF_019692012.1 XP_010802010.1	kuntiz domain	107	00 59	6468 20	-3.2	4.75
MdEppin20		XP_005190040.1	kuntiz domain	88	68	7497 46	-3.0	9.41
MdEppin27	NDYK01190170·236-565	XI_000170040.1	kuntiz domain	82	62	7018 89	87	9.70
MdEppin29	AOPM01002184.1:2478-2770	XP 005192097.1	kuntiz domain	77	57	6485.14	-1.0	5.97
MdEppin30		XP 019895340.1	kuntiz domain	88	68	7688.60	4.2	8.50
MdEppin31	AQPM01011896.1:10591-10914	-	kuntiz domain	85	65	7374.26	4.2	8.50
MdEppin32	NDYK01025868:2318-2646		kuntiz domain	85	65	7445.39	6.2	8.95
MdEppin33		XP_005182687.1	kuntiz domain	145	120	12984.71	7.7	9.13
MdEppin34	AQPM01069148.1:24856-25351	XP_019892019.1	kuntiz domain	146	121	13055.78	7.7	9.13
MdEppin35-1	NDYK01174947.1:14053-17900		kuntiz domain	638	46	5130.92	0.4	7.23
MdEppin35-2			kuntiz domain		74	8622.47	-5.0	4.80
MdEppin35-3			kuntiz domain		86	9705.43	-9.0	4.46
MdEppin35-4			kuntiz domain		84	9442.15	-14.2	4.05
MdEppin35-5			kuntiz domain		58	6481.97	-7.0	4.23
MdEppin35-6			kuntiz domain		65	7352.07	-4.8	4.77
MdEppin35-7			kuntiz domain		0U 71	0020.10 8226.17	-4.0	4.70
MdEppin25.0			kuntiz domain		/1 Q1	0220.17	-0.5	7.44
MdMuelin1	AOPM01019378 1.40008-41220		kazal domain	81	61	7210.20 6923 73	_27	7.01 5.36
MdMuelin?	AOPM01019378 1.43260-41220		kazal domain	75	56	6374 40	-2.7	6.95
MdMuslin3	AOPM01019378 1.47740-48051		kazal domain	81	62	6880.61	-4.3	4.50
MdMuslin4	AOPM01019378.1:52156-52447		kazal domain	76	57	6384.33	-1.0	5.88
MdMuslin5	AQPM01019378.1:56578-56881	XP_005175924.1	kazal domain	78	59	6608.54	-3.0	4.98

		GenBank No.	Туре	Size		>		Ţ
Name	WGS/EST			ORF	MP	MW(Da)	Net Charge	pI
MdMuslin6	AOPM01019378 1:57384-57671	XP 0051759231	kazal domain	73	53	5905 82	5.0	8 73
MdMuslin7	NDYK01036986.1:1550-1765	/4_0001/0/20.1	kazal domain	70	51	5592.32	-1.3	5.12
MdMuslin8	NDYK01185191.1:567-854	XP_011294170.1	kazal domain	75	56	6360.27	4.2	8.52
MdMuslin9		XP_005175922.1	kazal domain	75	56	6165.88	-1.3	5.27
MdMuslin10	AQPM01042957.1:886-1101	XP_005175386.1	kazal domain	72	48	5340.00	0.7	7.61
MdMuslin11	NDYK01172454.1:1382-1630	XP_005190386.1	kazal domain	83	63	6751.57	-1.0	5.97
MdMuslin12 MdMuslin13	ND1K01221455.1:787-1002	XP_005188808.1 XP_005178654.1	kazal domain	82 136	49 117	12354 50	3.7	8.50 8.55
MdMuslin14		XP_005188224.1	kazal domain	79	59	6456.06	-53	4 49
MdMuslin15-1		XP 005188061.1	kazal domain	154	76	8566.70	-3.3	4.77
MdMuslin15-2		-	kazal domain		57	6219.32	1.7	7.98
MdMuslin16-1	AQPM01094907.1:8864-9376		kazal domain	148	77	8667.80	-3.3	4.77
MdMuslin16-2			kazal domain		50	5537.53	1.7	7.98
MdMuslin17	NDYK01122947.1:2962-3420	XP_005190966.1	kazal domain	88	67	7092.01	0.7	7.61
MdMuslin18	NDVV01014762-1020 2222	XP_005190967.1	kazal domain	98	70	7304.27	4.7	8.80
MdMuslin20	ND1R01014703.1939-2223	XP 0198936341	kazal domain	83	62	6943.64	-8.0	9.30 4.16
MdMuslin20		XP_011296278.1	kazal domain	98	77	8811 17	4 2	8.52
MdMuslin22		XP 005188497.1	kazal domain	91	72	7752.70	2.0	7.98
MdMuslin23-1		XP_005190965.2	kazal domain	136	57	6848.86	3.0	8.27
MdMuslin23-2			kazal domain		52	5968.77	0.7	7.61
MdMuslin24-1		XP_019894976.1	kazal domain	135	61	7210.20	-1.3	5.41
MdMuslin24-2			kazal domain		52	6100.76	-0.5	6.72
MdMuslin25	AQPM01000599.1:6069-6748		kazal domain	87	64	7410.43	10.4	10.09
MdMuslin26	NDYK011901/7.1:7-310	VD 00E1000021	kazal domain	//	56	6388.29	5.2	8.65
MdMuslin27	AOPM01000604 1.11632-11959	XP_011295627.1	kazal domain	89	66	7208.50	9.7 7 9	930
MdMuslin29	AOPM01000605 1:5737-6069	XI_0112/3027.1	kazal domain	87	64	7267 41	89	9.30
MdMuslin30	11011101000000110707 0007	XP 005190994.1	kazal domain	90	67	7619.84	9.9	9.51
MdMuslin31	NDYK01044765.1:1300-1610		kazal domain	83	61	7115.23	10.9	10.61
MdMuslin32	AQPM01000603.1:4008-4322		kazal domain	84	59	6658.82	5.9	8.66
MdMuslin33	AQPM01000603.1:639-963		kazal domain	89	66	7007.07	7.1	8.90
MdSVWC1		XP_005183514.1	svwc	149	130	14382.45	-4.3	5.08
MdSVWC2	NIDV/K01201922 1.12/ E90	XP_011295325.1	svwc	102	81	8733.00	-1.1	5.88
MdSVWC3	NDYK01201822.1:136-580	XP_005175282.1 XD_005180656.1	svwc	102	81	8/41.00	-1.1	5.88 E 01
MdSVWC4	AOPM01015706 1.698-1200	AF_003189636.1	SVWC	134	94	10684 30	-5.5	6.25
MdSVWC6	AQI 101013/00.1.078-1200	XP 0051754081	SVWC	120	100	11347 19	0.4	7 19
MdSVWC7	AOPM01092559.1:3374-5544	XP 005187657.1	svwc	182	163	19163.26	1.9	7.87
MdSVWC8	AQPM01024958.1:41-530	XP_005190002.1	SVWC	113	94	10879.12	-2.6	5.78
MdSVWC9	AQPM01015051.1: 3997-5502	XP_005175283.1	svwc	111	87	10386.73	-2.1	6.34
MdSVWC10	AQPM01015049.1:461-992		SVWC	100	80	9437.82	7.4	8.88
MdSVWC11		XP_011295271.1	svwc	104	84	9971.32	0.6	7.26
MdSVWC12	AQPM01015050.1:115-2228	XP_005175281.1	SVWC	104	79	9215.33	-1.6	6.25
MdSVWC13	A ODM01081212 1.14881 15505	XP_005175284.1	svwc	88	68 70	7510.65	3.9	8.29
WIUSV WC14	NW 004765359 1.113654-		SVWC	102	19	0904.37	0.9	7.52
MdSVWC15	114278	JZ121963.1	svwc	95	72	8095.36	-0.1	6.91
MdSVWC16	1112/0	XP 005184317.1	svwc	124	96	10843.56	0.2	7.09
MdSVWC17	AQPM01092395.1:1108-1569	XP_005187600.1	svwc	121	102	11254.72	-0.4	6.86
MdSVWC18		XP_005180761.1	svwc	107	88	9682.02	4.2	8.31
MdSVWC19	AQPM01056437.1:2396-2894	XP_011290794.1	SVWC	113	89	9482.15	-0.8	6.44
MdSVWC20	AQPM01095428.1:3680-4100	XP_005188179.1	SVWC	102	84	9247.47	-5.1	4.75
MdSVWC21	AQPM01060615.1:2747-3172	XP_005180301.1	svwc	103	85	9126.49	-1.6	6.25
MdSVWC22 MdSVWC23	AQPM01095425.1:2644-3102	XP_005188177.1 XP_011294423.1	SVWC	113	95 102	10296.72	-4.6	5.10
MdSVWC23	AOPM01095428 1 13247-13749	XP_011294424.1	SVWC	105	85	9320 74	4 2	8.31
MdSVWC25	AOPM01095423.1:787-1232	XP 005188176.1	svwc	106	84	8913.27	3.2	8.12
MdSVWC26	AQPM01019310.1:3498-3948	XP_005175918.1	svwc	106	84	9408.71	-0.6	6.72
MdSVWC27	AQPM01095427.1:815-1458	XP_005188181.1	svwc	115	95	10876.26	-6.9	4.77
MdSVWC28	AQPM01095425.1:13938-14587		SVWC	115	95	10891.33	-7.8	4.67
MdSVWC29-1		XP_005188180.2	svwc	137	117	13288.15	-7.9	4.60
MdSVWC29-2			svwc	101	83	9053.29	-0.8	6.48
MdCrustin1	AQPM01030484.1:548-1269		domain	100	67	7493.40	2.9	8.10
MdCrustin2		XP_011295532.1	domain	92	67	7493.40	2.9	8.10
MdCrustin3		XP_005190815.1	domain	115	94	9819.93	5.8	8.37
MdCrustin4		XP_011295531.1	domain	120	95	10541.92	6.8	8.48
MdCecropin1 "	A OD M01050001 1 555 1000	ABB17292.1	α -helix	63	40	42/1.97	5.1	10.56
MdCecropin2 *	AQPM01058001.1:775-1030	XP_005179713.1 XP_005170700.1	α -helix	64	41	4342.06	6.1	10.66
MdCorropin3 "	AQPIVIU1030001.1:8963-9222	AF_0051/9/00.1	α-nelix	62	41	4386.11	0.1 E 1	10.66
MdCorronin ^E	AUTW01030004.1:2369-2661	XF_019890986.1	a-nelix	03 44	41U 41	4257.94	5.1 4.2	10.56
MdCecropin6 [#]	AOPM01058004 1.5352-5771	XP 0051797171	α-nelix α-helix	64	41	4370 11	+.2 61	10.94
MdCecropin7 [#]	AOPM01058000.1:20546-20804	XP 0051797121	α-helix	63	41	4356.05	6.1	11.12
MdCecropin8 #	AQPM01058001.1:4386-4681	AXG50148.1	α -helix	64	41	4356.09	6.1	10.66

Table A1. Cont.

		GenBank No.	Туре	Si	ze	- MW(Da)	Net Charge	pI
Name	WGS/EST			ORF	MP			
MdCecropin9 #	AQPM01058006.1:1083-1338	XP_005179718.1	α-helix	64	41	4461.10	4.2	10.94
MdCecropin10 #	AQPM01058008.1:3578-4516	XP_005179719.1	α-helix	62	41	4464.06	3.2	10.28
MdCecropin11 #	AOPM01058004.1:928-1210	AIW52264.1	α-helix	64	41	4356.09	6.1	10.66
MdCecropin12	IZ121081.1		α-helix	63	40	4227.91	5.1	10.56
MdCecropin13	ES608288 1		α -helix	64	41	4341 12	61	10.66
MdCecropin14 #	AOPM01100428 1·3910-4122	XP 011294761 1	α-helix	69	44	4546 24	29	9 34
MdCecropin15	AOPM01100427 1.1792-2047	XP_019894290.1	α-helix	69	44	4518 23	3.9	9.72
MdCecropin16	AOPM01100428 1.8445-8711	, <u>1</u> _01)0)12)0.1	a-helix	69	44	4560.27	29	934
Maccelopinio	MQ1 1001100420.1.0445 0/11		Proline and	07		4500.27	2.)	7.54
MdDiptericin1 #	FJ748596.1:148-344	ACN61637.1	Glycine rich	99	79	8721.39	1.4	8.50
MdDiptericin2 #	FJ794602.1:25-321	ACO35257.1	Glycine rich	99	79	8721.39	1.4	8.50
MdDiptericin3 #	KM205631.151-347		Proline and Glycine rich	99	79	8596.58	3.5	8.69
			Proline and					
MdDiptericin4 #	FJ795370.1:65-364	ACN93798.1	Glycine rich	99	79	8725.34	1.4	8.50
			Proline and					
MdDiptericinD *	AQPM01092243.1:221-1086	NP_001295957.2	Glycine rich	99	79	8770.47	1.4	8.41
			Proline and		-			
MdDiptericinD1 *	AQPM01092241.1:4245-4608	XP_005187575.1	Glycine rich	99	79	8/11.36	1.4	8.50
MdDomesticin1 #		AHA56721 1	Proline rich	65	40	4583 33	69	11 41
MdDomesticin?	AOPM01056449 1.10	440-12589	Proline rich	65	40	4525 20	5.9	10.98
MdEdin1	AOPM01067038 1.2033_2341	12303	Clycino rich	102	62	4923.20	_0.9	6.67
MdEdin2	AOPM01067938 1.5150-5500		Clycine rich	116	65	7321 74	24	9.20
MdEdin3	AOPM01067936 1.1742-2049		Clycine rich	101	61	6827.11	_19	6.34
MdEdin4	AOPM01067026 1:4517 4852		Clycine rich	120	65	7221 70	26	0.54
MdEdin5	I7121804 1		Clycine rich	116	65	7350.70	2.4	0.25
MdEding	JZ121094.1		Glycine rich	110	107	12760 56	2.4	9.23
MdEdin6	AQPM0106/939.1:3916-4450		Glycine rich	177	127	13760.36	1.1	7.00
MaEdin/	ND1K01101123.1:1387-1910		Glycine rich	174	127	13721.52	1.9	8.50
MaEdin8	AQPM0106/938.1:9198-9/31		Glycine rich	177	127	13/52.54	0.7	7.47
MdEdin9	AQPM01067938.1:15033-15560		Glycine rich	174	125	13612.44	3.6	9.34
MdEdin10 *		AFP64086.1	Glycine rich	175	127	13740.57	1.7	8.50
MdAttacinA1 #		XP_011296530.1	Glycine rich	208	188	20002.01	6.9	9.66
MdAttacinA2	AQPM01013309.1:3192-3887	XP_019890218.1	Glycine rich	208	188	19613.34	6.9	9.81
MdAttacinA3 #		AAY59540.1	Glycine rich	208	188	19688.41	6.9	9.81
MdAttacinA4 #		AAR23786.1	Glycine rich	208	188	19672.41	6.9	9.81
MdAttacinA5		XP_019890219.1	Glycine rich	208	188	19718.44	6.9	9.81
MdAttacinC1 #	AQPM01059487.1:2038-2826		Proline and Glycine rich	241	192	20420.06	2.9	8.92
			Proline and					
MdAttacinC2 #		ACO35258.1	Clycine rich	241	192	20334.91	1.9	8.41
			Brolino and					
MdAttacinC3-1		XP_005180079.2	Glycine rich	250	201	21385.12	5.1	9.34
MdAttacinC3-2			Proline and Glycine rich	241	192	20449.11	4.1	9.23
MdAttacinC4 #	AQPM01059487.1:7828-8614	XP_005180076.1	Proline and Glycine rich	241	192	20383.93	3.1	8.92
MdAttacinC5 #	NDYK01054543.1:3986-4696		Proline and Glycine rich	241	192	19973.48	5.1	9.34
MdAttacinC6	AQPM01059487.1:3997-5074		Proline and Glycine rich	241	192	20349.91	3.1	8.92
MdAttacinC7	JZ121354.1:31-753		Proline and Glycine rich	241	192	20384.91	1.7	8.41
MdAttacinD1		XP 0112965381	Glycine rich	181	181	19122.94	8.1	9.91
MdAttacinD2		XP_0051785161	Glycine rich	189	189	19412 19	5.3	9 41
MdAttacinD3 #	NDYK01109436 1·1208-1899	XP 005178550 1	Glycine rich	191	191	19712 79	78	9 70
MdAttacinD4 #	112 11(0110) 100.111200 1077	AFP64340.1	Glycine rich	197	197	21110.83	6.1	9.65

Table A1. Cont.

Note: Previously known sequences are labeled by "#".

References

- 1. Hoffmann, J.A. Innate immunity of insects. Curr. Opin. Immunol. 1995, 7, 4–10. [CrossRef]
- Stork, N.E.; McBroom, J.; Gely, C.; Hamilton, A.J. New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proc. Natl. Acad. Sci. USA* 2015, 112, 7519–7523. [CrossRef]
- Lemaitre, B.; Hoffmann, J. The host defense of Drosophila melanogaster. Annu. Rev. Immunol. 2007, 25, 697–743. [CrossRef] [PubMed]
- Flores-Villegas, A.L.; Salazar-Schettino, P.M.; Cordoba-Aguilar, A.; Gutierrez-Cabrera, A.E.; Rojas-Wastavino, G.E.; Bucio-Torres, M.I.; Cabrera-Bravo, M. Immune defence mechanisms of triatomines against bacteria, viruses, fungi and parasites. *Bull. Entomol. Res.* 2015, 105, 523–532. [CrossRef] [PubMed]
- 5. Park, Y.; Hahm, K.S. Antimicrobial peptides (AMPs): Peptide structure and mode of action. *J. Biochem. Mol. Biol.* **2005**, *38*, 507–516. [CrossRef] [PubMed]

- Imler, J.L.; Bulet, P. Antimicrobial peptides in *Drosophila*: Structures, activities and gene regulation. *Chem. Immunol. Allergy* 2005, 86, 1–21. [PubMed]
- Zhang, Z.T.; Zhu, S.Y. Drosomycin, an essential component of antifungal defence in *Drosophila*. *Insect. Mol. Biol.* 2009, 18, 549–556. [CrossRef]
- 8. Hanson, M.A.; Lemaitre, B. New insights on *Drosophila* antimicrobial peptide function in host defense and beyond. *Curr. Opin. Immunol.* **2020**, *62*, 22–30. [CrossRef]
- 9. Bulet, P.; Hetru, C.; Dimarcq, J.L.; Hoffmann, D. Antimicrobial peptides in insects; structure and function. *Dev. Comp. Immunol.* **1999**, *23*, 329–344. [CrossRef]
- 10. Kaushal, A.; Gupta, K.; Shah, R.; van Hoek, M.L. Antimicrobial activity of mosquito cecropin peptides against Francisella. *Dev Comp. Immunol.* **2016**, *63*, 171–180. [CrossRef]
- 11. Casteels, P.; Ampe, C.; Jacobs, F.; Vaeck, M.; Tempst, P. Apidaecins: Antibacterial peptides from honeybees. *EMBO J.* **1989**, *8*, 2387–2391. [CrossRef]
- 12. Charlet, M.; Lagueux, M.; Reichhart, J.M.; Hoffmann, D.; Braun, A.; Meister, M. Cloning of the gene encoding the antibacterial peptide drosocin involved in *Drosophila* immunity. *Eur. J. Biochem.* **1996**, 241, 699–706. [CrossRef] [PubMed]
- Chowdhury, S.; Taniai, K.; Hara, S.; Kadonookuda, K.; Kato, Y.; Yamamoto, M.; Xu, J.; Choi, S.K.; Debnath, N.C.; Choi, H.K.; et al. cDNA cloning and gene expression of lebocin, a novel member of antibacterial peptides from the silkworm, *Bombyx mori. Biochem. Biophys. Res. Commun.* 1995, 214, 271–278. [CrossRef]
- 14. Gao, B.; Zhu, S. The drosomycin multigene family: Three-disulfide variants from *Drosophila takahashii* possess antibacterial activity. *Sci. Rep.* **2016**, *6*, 32175–32186. [CrossRef]
- Thevissen, K.; Kristensen, H.H.; Thomma, B.P.; Cammue, B.P.; François, I.E. Therapeutic potential of antifungal plant and insect defensins. *Drug Discov. Today* 2007, 12, 966–971. [CrossRef] [PubMed]
- 16. Mylonakis, E.; Podsiadlowski, L.; Muhammed, M.; Vilcinskas, A. Diversity, evolution and medical applications of insect antimicrobial peptides. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2016**, *371*, 1695–1705. [CrossRef] [PubMed]
- 17. Christophides, G.K.; Zdobnov, E.; Barillas-Mury, C.; Birney, E.; Blandin, S.; Blass, C.; Brey, P.T.; Collins, F.H.; Danielli, A.; Dimopoulos, G.; et al. Immunity-related genes and gene families in Anopheles gambiae. *Science* **2002**, *298*, 159–165. [CrossRef]
- Tanaka, H.; Ishibashi, J.; Fujita, K.; Nakajima, Y.; Sagisaka, A.; Tomimoto, K.; Suzuki, N.; Yoshiyama, M.; Kaneko, Y.; Iwasaki, T.; et al. A genome-wide analysis of genes and gene families involved in innate immunity of Bombyx mori. *Insect. Biochem. Mol. Biol.* 2008, *38*, 1087–1110. [CrossRef]
- Waterhouse, R.M.; Kriventseva, E.V.; Meister, S.; Xi, Z.; Alvarez, K.S.; Bartholomay, L.C.; Barillas-Mury, C.; Bian, G.; Blandin, S.; Christensen, B.M.; et al. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* 2007, 316, 1738–1743. [CrossRef] [PubMed]
- Tian, C.; Gao, B.; Fang, Q.; Ye, G.; Zhu, S. Antimicrobial peptide-like genes in *Nasonia vitripennis*: A genomic perspective. *BMC Genomics* 2010, *11*, 187. [CrossRef] [PubMed]
- 21. Niu, Y.; Zheng, D.; Yao, B.; Cai, Z.; Zhao, Z.; Wu, S.; Cong, P.; Yang, D. A novel bioconversion for value-added products from food waste using Musca domestica. *Waste Manag.* **2017**, *61*, 455–460. [CrossRef]
- Scott, J.G.; Warren, W.C.; Beukeboom, L.W.; Bopp, D.; Clark, A.G.; Giers, S.D.; Hediger, M.; Jones, A.K.; Kasai, S.; Leichter, C.A.; et al. Genome of the house fly, Musca domestica L. a global vector of diseases with adaptations to a septic environment. *Genome Biol.* 2014, 15, 466–482. [CrossRef] [PubMed]
- Zhu, S.; Gao, B.; Tytgat, J. Phylogenetic distribution, functional epitopes and evolution of the CSαβ superfamily. *Cell Mol. Life Sci.* 2005, *62*, 2257–2269. [CrossRef] [PubMed]
- 24. Robert, X.; Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **2014**, 42, W320–W324. [CrossRef] [PubMed]
- 25. Zhu, S.; Gao, B.; Peigneur, S.; Tytgat, J. How a scorpion toxin selectively captures a prey sodium channel: The molecular and evolutionary basis uncovered. *Mol. Biol. Evol.* **2020**, *37*, 3149–3164. [CrossRef]
- Pierce, B.G.; Wiehe, K.; Hwang, H.; Kim, B.H.; Vreven, T.; Weng, Z. ZDOCK server: Interactive docking prediction of proteinprotein complexes and symmetric multimers. *Bioinformatics* 2014, *30*, 1771–1773. [CrossRef]
- 27. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. GROMACS: Fast, flexible, free. J. Comput. Chem. 2005, 26, 1701–1718. [CrossRef]
- Kaminski, G.A.; Friesner, R.A.; Tirado-Rives, J.; Jorgensen, W.L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J. Phys. Chem. B 2001, 105, 6474–6487. [CrossRef]
- 29. Zhu, S.; Gao, B. Positive selection in cathelicidin host defense peptides: Adaptation to exogenous pathogens or endogenous receptors? *Heredity* **2017**, *118*, 453–465. [CrossRef]
- Yang, Z.; Nielsen, R.; Goldman, N.; Pedersen, A.M.K. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 2000, 155, 431–449.
- Yang, Z.; Wong, W.S.W.; Nielsen, R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 2005, 22, 1107–1118. [CrossRef] [PubMed]
- 32. Colell, E.A.; Iserte, J.A.; Simonetti, F.L.; Marino-Buslje, C. MISTIC2: Comprehensive server to study coevolution in protein families. *Nucleic Acids Res.* 2018, *46*, W323–W328. [CrossRef] [PubMed]

- Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* 2004, 14, 1188–1190. [CrossRef] [PubMed]
- 34. Zhou, W.; Gao, B.; Zhu, S. Did cis- and trans-defensins derive from a common ancestor? Immunogenetics 2019, 71, 61–69. [CrossRef]
- 35. Cociancich, S.; Ghazi, A.; Hetru, C.; Hoffmann, J.A.; Letellier, L. Insect defensin, an inducible antibacterial peptide, forms voltage-dependent channels in Micrococcus luteus. *J. Biol. Chem.* **1993**, *268*, 19239–19245. [CrossRef]
- 36. Lee, Y.S.; Yun, E.K.; Jang, W.S.; Kim, I.; Lee, J.H.; Park, S.Y.; Ryu, K.S.; Seo, S.J.; Kim, C.H.; Lee, I.H. Purification, cDNA cloning and expression of an insect defensin from the great wax moth, *Galleria mellonella*. *Insect. Mol. Biol.* **2004**, *13*, 65–72. [CrossRef]
- 37. Andoh, M.; Ueno, T.; Kawasaki, K. Tissue-dependent induction of antimicrobial peptide genes after body wall injury in house fly (Musca domestica) larvae. *Drug Discov. Ther.* **2018**, *12*, 355–362. [CrossRef]
- Yount, N.Y.; Yeaman, M.R. Multidimensional signatures in antimicrobial peptides. Proc. Natl. Acad. Sci. USA 2004, 101, 7363–7368.
 [CrossRef]
- 39. Koehbach, J. Structure-Activity Relationships of Insect. Defensins. Front. Chem. 2007, 5, 45–54. [CrossRef]
- 40. Zhu, S. Discovery of six families of fungal defensin-like peptides provides insights into origin and evolution of the CSαβ defensins. *Mol. Immunol.* **2008**, 45, 828–838. [CrossRef]
- 41. Laskowski, M., Jr.; Kato, I. Protein inhibitors of proteinases. Annu. Rev. Biochem. 1980, 49, 593–626. [CrossRef] [PubMed]
- 42. Kanost, M.R. Serine proteinase inhibitors in arthropod immunity. Dev. Comp. Immunol. 1999, 23, 291–301. [CrossRef]
- McCrudden, M.T.; Dafforn, T.R.; Houston, D.F.; Turkington, P.T.; Timson, D.J. Functional domains of the human epididymal protease inhibitor, eppin. FEBS J. 2008, 275, 1742–1750. [CrossRef]
- 44. Fröbius, A.C.; Kanost, M.R.; Götz, P.; Vilcinskas, A. Isolation and characterization of novel inducible serine protease inhibitors from larval hemolymph of the greater wax moth Galleria mellonella. *Eur. J. Biochem.* **2000**, *267*, 2046–2053. [CrossRef] [PubMed]
- 45. Nirmala, X.; Kodrík, D.; Žurovec, M.; Sehnal, F. Insect silk contains both a Kunitz-type and a unique Kazal-typeproteinase inhibitor. *Eur. J. Biochem.* **2001**, *268*, 2064–2073. [CrossRef]
- 46. de Magalhaes, M.T.Q.; Mambelli, F.S.; Santos, B.P.O.; Morais, S.B.; Oliveira, S.C. Serine protease inhibitors containing a Kunitz domain: Their role in modulation of host inflammatory responses and parasite survival. *Microbes Infect.* **2018**, *20*, 606–609. [CrossRef]
- Watanabe, R.M.; Soares, T.S.; Morais-Zani, K.; Tanaka-Azevedo, A.M.; Maciel, C.; Capurro, M.L.; Torquato, R.J.; Tanaka, A.S. A novel trypsin Kazal-type inhibitor from Aedes aegypti with thrombin coagulant inhibitory activity. *Biochimie* 2010, 92, 933–939. [CrossRef]
- 48. Niimi, T.; Yokoyama, H.; Goto, A.; Beck, K.; Kitagawa, Y. A Drosophila gene encoding multiple splice variants of Kazal-type serine protease inhibitor-like proteins with potential destinations of mitochondria, cytosol and the secretory pathway. *Eur. J. Biochem.* **1999**, 266, 282–292. [CrossRef]
- 49. Brillard-Bourdet, M.; Hamdaoui, A.; Hajjar, E.; Boudier, C.; Reuter, N.; Ehret-Sabatier, L.; Bieth, J.G.; Gauthier, F. A novel locust (*Schistocerca gregaria*) serine protease inhibitor with a high affinity for neutrophil elastase. *Biochem. J.* **2006**, 400, 467–476. [CrossRef]
- 50. Kumaresan, V.; Harikrishnan, R.; Arockiaraj, J. A potential Kazal-type serine protease inhibitor involves in kinetics of protease inhibition and bacteriostatic activity. *Fish Shellfish Immunol.* **2015**, *42*, 430–438. [CrossRef]
- 51. Gebhard, L.G.; Carrizo, F.U.; Stern, A.L.; Burgardt, N.I.; Faivovich, J.; Lavilla, E.; Ermacora, M.R. A Kazal prolyl endopeptidase inhibitor isolated from the skin of Phyllomedusa sauvagii. *Eur. J. Biochem.* **2004**, *271*, 2117–2126. [CrossRef]
- Han, F.; Lu, A.; Yuan, Y.; Huang, W.; Beerntsen, B.T.; Huang, J.; Ling, E. Characterization of an entomopathogenic fungi target integument protein, Bombyx mori single domain von Willebrand factor type C, in the silkworm, *Bombyx mori. Insect. Mol. Biol.* 2017, 26, 308–316. [CrossRef] [PubMed]
- Smit, A.B.; de Jong-Brink, M.; Li, K.W.; Sassen, M.M.J.; Spijker, S.; van Elk, R.; Buijs, S.P.; van Minnen, J.; van Kesteren, R.E. Granularin, a novel molluscan opsonin comprising a single vWF type C domain is up-regulated during parasitation. *FEBS J.* 2004, 18, 845–847. [CrossRef] [PubMed]
- 54. Smith, V.J.; Fernandes, J.M.; Kemp, G.D.; Hauton, C. Crustins: Enigmatic WAP domain-containing antibacterial proteins from crustaceans. *Dev. Comp. Immunol.* 2008, 32, 758–772. [CrossRef]
- 55. Vargas-Albores, F.; Martinez-Porchas, M. Crustins are distinctive members of the WAP-containing protein superfamily: An improved classification approach. *Dev. Comp. Immunol.* **2017**, *76*, 9–17. [CrossRef]
- 56. Afsal, V.V.; Antony, S.P.; Sathyan, N.; Philip, R. Molecular characterization and phylogenetic analysis of two antimicrobial peptides: Anti-lipopolysaccharide factor and crustin from the brown mud crab, *Scylla serrata. Results Immunol.* 2011, 1, 6–10. [CrossRef] [PubMed]
- 57. Brockton, V.; Hammond, J.A.; Smith, V.J. Gene characterisation, isoforms and recombinant expression of carcinin, an antibacterial protein from the shore crab, Carcinus maenas. *Mol. Immunol.* **2007**, *44*, 943–949. [CrossRef]
- 58. Antony, S.P.; Singh, I.S.; Sudheer, N.S.; Vrinda, S.; Priyaja, P.; Philip, R. Molecular characterization of a crustin-like antimicrobial peptide in the giant tiger shrimp, Penaeus monodon, and its expression profile in response to various immunostimulants and challenge with WSSV. *Immunobiology* **2011**, *216*, 184–194. [CrossRef]
- 59. Yang, L.; Niu, S.; Gao, J.; Zuo, H.; Yuan, J.; Weng, S.; He, J.; Xu, X. A single WAP domain (SWD)-containing protein with antiviral activity from Pacific white shrimp Litopenaeus vannamei. *Fish Shellfish Immunol.* **2018**, *73*, 167–174. [CrossRef]
- 60. Zhang, J.; Li, F.; Wang, Z.; Xiang, J. Cloning and recombinant expression of a crustin-like gene from Chinese shrimp, *Fenneropenaeus chinensis*. J. Biotechnol. 2007, 127, 605–614. [CrossRef]
- 61. Hagiwara, K.; Kikuchi, T.; Endo, Y.; Usui, K.; Takahashi, M.; Shibata, N.; Kusakabe, T.; Xin, H.; Hoshi, S.; Miki, M.; et al. Mouse SWAM1 and SWAM2 are antibacterial proteins composed of a single whey acidic protein motif. *J. Immunol.* **2003**, *170*, 1973–1979. [CrossRef] [PubMed]

- 62. Nair, D.G.; Fry, B.G.; Alewood, P.; Kumar, P.P.; Kini, R.M. Antimicrobial activity of omwaprin, a new member of the waprin family of snake venom proteins. *Biochem. J.* **2007**, *402*, 93–104. [CrossRef]
- 63. Steiner, H.; Hultmark, D.; Engström, A.; Bennich, H.; Boman, H.G. Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature* **1981**, *182*, 246–248. [CrossRef]
- 64. Peng, J.; Wu, Z.; Liu, W.; Long, H.; Zhu, G.; Guo, G.; Wu, J. Antimicrobial functional divergence of the cecropin antibacterial peptide gene family in Musca domestica. *Parasit. Vectors* **2019**, *12*, 537–546. [CrossRef] [PubMed]
- 65. Boulanger, N.; Munks, R.J.; Hamilton, J.V.; Vovelle, F.; Brun, R.; Lehane, M.J.; Bulet, P. Epithelial innate immunity. A novel antimicrobial peptide with antiparasitic activity in the blood-sucking insect *Stomoxys calcitrans*. *J. Biol. Chem.* **2002**, 277, 49921–49926. [CrossRef] [PubMed]
- Vizioli, J.; Bulet, P.; Charlet, M.; Lowenberger, C.; Blass, C.; Muller, H.M.; Dimopoulos, G.; Hoffmann, J.; Kafatos, F.C.; Richman, A. Cloning and analysis of a cecropin gene from the malaria vector mosquito, Anopheles gambiae. *Insect. Mol. Biol.* 2000, *9*, 75–84. [CrossRef]
- 67. Ekengren, S.; Hultmark, D. Drosophila cecropin as an antifungal agent. Insect. Biochem. Mol. Biol. 1999, 29, 965–972. [CrossRef]
- 68. Okada, M.; Natori, S. Primary structure of sarcotoxin I, an antibacterial protein induced in the hemolymph of *Sarcophaga peregrina* (flesh fly) larvae. *J. Biol. Chem.* **1985**, 260, 7174–7177. [CrossRef]
- 69. Ouyang, L.; Xu, X.; Freed, S.; Gao, Y.; Yu, J.; Wang, S.; Ju, W.; Zhang, Y.; Jin, F. Cecropins from *Plutella xylostella* and their interaction with *Metarhizium anisopliae*. *PLoS ONE* **2015**, *10*, e0142451. [CrossRef] [PubMed]
- Saito, A.; Ueda, K.; Imamura, M.; Atsumi, S.; Tabunoki, H.; Miura, N.; Watanabe, A.; Kitami, M.; Sato, R. Purification and cDNA cloning of a cecropin from the longicorn beetle, Acalolepta luxuriosa. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 2005, 142, 317–323. [CrossRef]
- Kim, J.K.; Lee, E.; Shin, S.; Jeong, K.W.; Lee, J.Y.; Bae, S.Y.; Kim, S.H.; Lee, J.; Kim, S.R.; Lee, D.G.; et al. Structure and function of papiliocin with antimicrobial and anti-inflammatory activities isolated from the swallowtail butterfly, *Papilio xuthus. J. Biol. Chem.* 2011, 286, 41296–41311. [CrossRef] [PubMed]
- 72. Lee, E.; Jeong, K.W.; Lee, J.; Shin, A.; Kim, J.-K.; Lee, J.; Lee, D.G.; Kim, Y. Structure-activity relationships of cecropin-like peptides and their interactions with phospholipid membrane. *BMB Rep.* **2013**, *46*, 282–287. [CrossRef]
- Yagi-Utsumi, M.; Yamaguchi, Y.; Boonsri, P.; Iguchi, T.; Okemoto, K.; Natori, S.; Kato, K. Stable isotope-assisted NMR characterization of interaction between lipid A and sarcotoxin IA, a cecropin-type antibacterial peptide. *Biochem. Biophys. Res. Commun.* 2013, 431, 136–140. [CrossRef] [PubMed]
- Okemoto, K.; Nakajima, Y.; Fujioka, T.; Natori, S. Participation of two N-terminal residues in LPS neutralizing activity of sarcotoxin IA. J. Biochem. 2002, 131, 277–281. [CrossRef] [PubMed]
- 75. Oh, D.; Shin, S.Y.; Lee, S.; Kang, J.H.; Kim, S.D.; Ryu, P.D.; Hahm, K.S.; Kim, Y. Role of the hinge region and the tryptophan residue in the synthetic antimicrobial peptides, cecropin A(1–8)-magainin 2(1–12) and its analogues, on their antibiotic activities and structures. *Biochemistry* 2000, *39*, 11855–11864. [CrossRef] [PubMed]
- 76. 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. [CrossRef]
- 77. Tang, T.; Li, X.; Yang, X.; Yu, X.; Wang, J.; Liu, F.; Huang, D. Transcriptional response of *Musca domestica* larvae to bacterial infection. *PLoS ONE* **2014**, *9*, e104867. [CrossRef]
- Scocchi, M.; Tossi, A.; Gennaro, R. Proline-rich antimicrobial peptides: Converging to a non-lytic mechanism of action. *Cell Mol. Life Sci.* 2011, 68, 2317–2330. [CrossRef] [PubMed]
- 79. Dimarcq, J.L.; Zachary, D.; Hoffmann, J.A.; Hoffmann, D.; Reichhart, J.M. Insect immunity: Expression of the two major inducible antibacterial peptides, defensin and diptericin, in *Phormia terranovae*. *EMBO J.* **1990**, *9*, 2507–2515. [CrossRef] [PubMed]
- 80. Lee, J.H.; Cho, K.S.; Lee, J.; Yoo, J.; Lee, J.; Chung, J. Diptericin-like protein: An immune response gene regulated by the anti-bacterial gene induction pathway in *Drosophila*. *Gene* **2001**, *271*, 233–238. [CrossRef]
- Kim, C.H.; Muturi, E.J. Effect of larval density and Sindbis virus infection on immune responses in *Aedes aegypti*. J. Insect. Physiol. 2013, 59, 604–610. [CrossRef]
- 82. Vanha-Aho, L.M.; Anderl, I.; Vesala, L.; Hultmark, D.; Valanne, S.; Ramet, M. Edin expression in the fat body is required in the defense against parasitic wasps in *Drosophila* melanogaster. *PLoS Pathog.* **2015**, *11*, e1004895. [CrossRef] [PubMed]
- 83. Verleyen, P.; Baggerman, G.; D'Hertog, W.; Vierstraete, E.; Husson, S.J.; Schoofs, L. Identification of new immune induced molecules in the haemolymph of *Drosophila melanogaster* by 2D-nanoLC MS/MS. *J. Insect. Physiol.* **2006**, *52*, 379–388. [CrossRef] [PubMed]
- 84. Mura, M.E.; Ruiu, L. Brevibacillus laterosporus pathogenesis and local immune response regulation in the house fly midgut. *J. Invertebr. Pathol.* **2017**, 145, 55–61. [CrossRef]
- 85. Kawasaki, K.; Andoh, M. Properties of induced antimicrobial activity in *Musca domestica* larvae. *Drug Discov. Ther.* **2017**, *11*, 156–160. [CrossRef] [PubMed]
- 86. Mourier, T.; Jeffares, D.C. Eukaryotic intron loss. Science 2003, 300, 1393. [CrossRef]
- 87. Hogg, P.J. Disulfide bonds as switches for protein function. Trends Biochem. Sci. 2003, 28, 210–214. [CrossRef]
- 88. Zhu, S.; Gao, B. Molecular characterization of a new scorpion venom lipolysis activating peptide: Evidence for disulfide bridge-mediated functional switch of peptides. *FEBS Lett.* **2006**, *580*, *6825–6836*. [CrossRef] [PubMed]