Computational and Structural Biotechnology Journal 21 (2023) 5818–5828

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

The evolutionary diversification and antimicrobial potential of MPEG1 in Metazoa

Yuan Chen [a,b,c,1], Zihao Yuan [a,b,c,*,1], Li Sun [a,b,c,*]

[a] CAS and Shandong Province Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Center for Ocean Mega-Science, Chinese Academy of Sciences (CAS), Qingdao, China
[b] Laboratory for Marine Biology and Biotechnology, Laoshan Laboratory, Qingdao 266237, China
[c] College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Keywords:
Macrophage-expressed gene 1/Perforin 2
Data mining
Evolution
Metazoa
Antimicrobial

ABSTRACT

Macrophage-expressed gene 1 (MPEG1) is an ancient immune effector known to exist in Cnidaria, Mollusca, Actinopterygii, and Mammalia. In this study, we examined the evolution and antibacterial potential of MPEG1 across Metazoa. By unbiased data-mining, MPEG1 orthologs were found in 11 of 34 screened phyla. In invertebrates, MPEG1 is present in the major phyla and exhibits intensive duplication. In vertebrates, class-based clades were formed by the major, generic MPEG1 (gMPEG1) in each class. However, there is a minority of unique MPEG1 (uMPEG1) from 71 species of 4 classes that clustered into a separate clade detached from all major class-based clades. gMPEG1 and uMPEG1 exhibit strong genomic collinearity and are surrounded by high-density transposons. gMPEG1 and uMPEG1 transcript expressions were most abundant in immune organs, but differed markedly in tissue specificity. Systematic analysis identified an antimicrobial peptide (AMP)-like segment in the C-terminal (CT) tail of MPEG1. Peptides based on the AMP-like regions of 35 representative MPEG1 were synthesized. Bactericidal activities were displayed by all peptides. Together these results suggest transposon-propelled evolutionary diversification of MPEG1 in Metazoa that has likely led to functional specialisation. This study also reveals a possible antimicrobial mechanism mediated directly and solely by the CT tail of MPEG1.

1. Introduction

MPEG1 (macrophage-expressed gene 1), or perforin-2, is a member of the Membrane Attack Complex, Perforin / Cholesterol-Dependent Cytolysin (MACPF/CDC) superfamily. Like other members of this family, such as complement C9 and perforin-1, MPEG1 is endowed with pore-forming capacity [1–3]. MPEG1 was first identified through differential cDNA analysis in mouse mature and immature macrophages in 1990 s [4], and subsequently identified in diverse species of Metazoa. MPEG1 is a Type I transmembrane protein with conserved structural organization, which includes an N-terminal (NT) MACPF domain characteristic of all MACPF proteins, a P2 domain that is unique to MPEG1, a transmembrane region, and a short C-terminal (CT) tail of ~ 40 residues. However, a human MPEG1 isoform was reported to be secreted into the extracellular space [5]. There is evidence that MPEG1 is localized in the endoplasmic reticulum, Golgi apparatus, endosome, phagosome, and the plasma membrane [5,6]. In orientation, the MACPF and P2 domains form the ectodomain, while the CT tail faces towards the cytoplasm [7]. The MACPF domain, together with the P2 domain, functions in pore formation. The CT tail is associated with MPEG1 trafficking and interferon signalling [7–9].

MPEG1 has been shown to act as an immune effector against bacterial infection. MPEG1 can restrict the transition of bacterial pathogens from vacuole to cytosol, and retard the survival and replication of intracellular bacteria [6,10–12]. As a result, mice with MPEG1 knockout became susceptible to the infection of methicillin-resistant *Staphylococcus aureus* and *Salmonella typhimurium* [6]. However, it is more likely that MPEG1 does not kill bacteria directly, but rather it promotes the bacteria-damaging effect of other antimicrobial agents by breaching the envelope of phagocytosed bacteria [13]. It has been proposed that microbial phagocytosis stimulates the intracellular trafficking of MPEG1 to the phagosome, where the low pH enables MPEG1 to oligomerize and

* Corresponding authors at: CAS and Shandong Province Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Center for Ocean Mega-Science, Chinese Academy of Sciences (CAS), Qingdao, China.
E-mail addresses: yuanzihao@qdio.ac.cn (Z. Yuan), lsun@qdio.ac.cn (L. Sun).
[1] These two authors contribute equally to this work.

https://doi.org/10.1016/j.csbj.2023.11.032
Received 25 May 2023; Received in revised form 17 November 2023; Accepted 17 November 2023
Available online 19 November 2023

form pores on the bacterial membrane, which facilitates the entry of various phagosome proteases that cause further damage to the bacteria by proteolytic digestion of bacterial proteins [7]. The intracellular trafficking of MPEG1 requires ubiquitination of conserved lysine residues in the CT tail by the cullin-RING E3 ubiquitin ligase [8]. Once in the phagosome, the CT tail is thought to be cleaved off [14]. In addition to its bactericidal activity, MPEG1 was reported to directly associate with IFN-α/β receptors and STAT2, whereby playing an essential role in Type I IFN Signaling [9,15]. Two recent studies revealed an implication of MPEG1 in dendritic cell antigen presentation [14,16].

MPEG1 is an ancient gene known to exist in Porifera, Cnidaria, Brachiopoda, Mollusca, and Chordata [7,17–19]. In the sponge *Suberites domuncula*, the MPEG1 gene was cloned, and its encoded recombinant protein was shown to exhibit antibacterial activity *in vitro* [20]. In Cnidaria, there is a diversity of MPEG1 homologues, but their functions remain to be studied. Interestingly, Cnidaria MPEG1 seems to exist only in Anthozoa, such as coral and sea anemone, and is absent in Hydra [17,18]. In Mollusca and Actinopterygii, MPEG1 expression in a number of species was found to be responsive to pathogen treatment [21–26]. *In vitro* antimicrobial activity has been observed with truncated recombinant MPEG1, consisting of the MACPF domain or the ectodomain, of *Crassostrea gigas*, *Haliotis discus discus*, *Charonia tritonis*, *Platichthys stellatus*, and *Epinephelus coioides* [21,23–25,27]. These findings suggest a conserved function of MPEG1 in the immune defense against pathogen infection.

In the present study, in order to gain a comprehensive understanding of MPEG1 evolution, we analyzed MPEG1 orthologs across metazoan taxa and examined their evolutionary relationships. We also investigated the bactericidal ability of the MPEG1 CT tail in a systematic manner. Our work provides new insights into the evolution and immune function of MPEG1 in Metazoa.

2. Results

2.1. MPEG1 orthologs are widely distributed in major invertebrate phyla

To explore the origin and evolution of MPEG1 in Metazoa, a data-driven approach was used to screen 34 metazoan phyla with available genomic sequences. MPEG1 sequences were found in 11 phyla (Fig. 1A), including 7 phyla, i.e., Platyhelminthes, Nematoda, Arthropoda, Rotifera, Annelida, Brachiopoda, and Echinodermata. Except for Brachiopoda [19], no MPEG1 has been reported previously in these phyla. Detailed phylogenetic analyses revealed that invertebrate and vertebrate MPEG1 formed distinctly separated clades (Fig. 1B, Fig S1). In invertebrate, the MPEG1 from Brachiopod, Platyhelminthes, Nematoda, Arthropoda, and Rotifera formed highly supported monophyletic clades (Fig. 1C, Fig S2). The copy number of MPEG1 per species in invertebrate is significantly (p = 9.8e-5) higher than that in vertebrate. Fifteen invertebrate species across 5 phyla possess more than 5 copies of MPEG1, e.g., Platyhelminthes *Macrostomum lignano* has 13 MPEG1, whereas only 4 vertebrate species, all belonging to tetraploid Actinopterygii, have more than 5 copies of MPEG1 (Fig. 1D). It is worth noting that all the 357 Aves species analyzed have only one MPEG1.

2.2. Vertebrate MPEG1 form the major class-based clades and a minor nonclass-based clade

Within vertebrate, phylogenetic analysis indicated that the MPEG1 from Chondrichthyes, Actinopterygii, Amphibia, Reptilia, Aves, and Mammalia were generally separated into different class-based clades. Besides these major clades, a small group of MPEG1 from 71 species belonging to the classes of Osteichthyes, Amphibia, Reptilia, and Mammalia deviated from their respective class clades and clustered into a distinct clade (Fig. 2A, Fig S3, Table S1). We named these mixed-class MPEG1 “unique MPEG1” (uMPEG1) to distinguish them from the rest “generic MPEG1” (gMPEG1) that formed the main clade characteristic of each class. The phylogenetic result of Fig. 2A supported a monophyletic origin of gMPEG1 and uMPEG1. For all the 71 species that possess uMPEG1, they also contain gMPEG1. A large-scale pairwise sequence alignment of 72 uMPEG1 and 69 gMPEG1 revealed that the similarity among the uMPEG1 (average 63.38%) was significantly ($p < 2.2e-16$) lower than the similarity among the gMPEG1 (average 81.01%) (Fig. 2B). Similar differences in sequence variation and site conservation between uMPEG1 and gMPEG1 were observed via Weblogo analysis (Fig S4). It is notable that the similarity between uMPEG1 and gMPEG1 decreases from Actinopterygii to Mammalia (Fig. 2B).

2.3. gMPEG1 and uMPEG1 exhibit genomic collinearity and are surrounded by transposons

Synteny analysis was performed to further investigate the evolutionary relationship between uMPEG1 and gMPEG1. Except for Actinopterygii *Lepisosteus oculatus*, uMPEG1 and gMPEG1 were found to locate on the same chromosome in all examined species. In these species, a markedly conserved genomic synteny was observed proximal to u/gMPEG1. For example, DTX4 and ZFP91 are highly conserved around the u/gMPEG1 of Mammalia (*Homo sapiens*, *Ovis aries*, and *Tachyglossus aculeatus*), Aves (*Gallus gallus*), Reptilia (*Chelonia mydas*), Amphibia (*Rhinatrema bivittatum*), and Actinopterygii (*L. oculatus*) (Fig. 3A). Unlike the majority of vertebrate species, Actinopterygii *L. oculatus* harbors uMPEG1 (one copy) and gMPEG1 (two copies) on different chromosomes (LG15 and LG9, respectively). Conserved neighbor genes were found between *L. oculatus* gMPEG1 and the gMPEG1 of other species, such as human, sheep, and chicken. In contrast, no conserved neighborhood synteny was found between *L. oculatus* uMPEG1 and the uMPEG1 of other species (Fig. 3B). Abundant repetitive elements, most of which are transposons, were found surrounding the two gMPEG1 of *L. oculatus* and the adjacent gMPEG1 and uMPEG1 of *O. aries*, *R. bivittatum*, and *C. mydas* (Tables S2-S7). The densities of the repetitive elements around MPEG1 (over 90%, 50%, 90%, and 90% in the hot spots in *O. aries*, *L. oculatus*, *R. bivittatum*, and *C. mydas*, respectively) are much higher than that in the genome background (~43%, 21%, 54%, and 43% in *O. aries, L. oculatus*, *R. bivittatum*, and *C. mydas*, respectively) (Figs. 3C, 3D, Fig S5), suggesting a role of transposable elements in MPEG1 gene duplication.

2.4. gMPEG1 and uMPEG1 differ in tissue-specific expression patterns

To better understand the biological role of MPEG1, the transcript expression profiles of g/uMPEG1 were examined in Actinopterygii and Mammalia represented by *L. oculatus* and *O. aries*, respectively. In *L. oculatus,* g/uMPEG1 transcripts were detected in eight tissues, with the highest expression levels of gMPEG1 copy 1, gMPEG1 copy 2, and uMPEG1 occurring in the spleen, blood and gill, respectively, and the lowest expression levels of these genes all occurring in the muscle (Fig. 4A). The differences between the highest and lowest expression levels of gMPEG1 copy 1 and copy 2 were ~150- and ~20-fold, respectively, while the difference between the highest and lowest expression levels of uMPEG1 was ~1500-fold. In *O. aries*, the expression levels of g/uMPEG1 in 55 tissues were examined based on RNA-seq. gMPEG1 mRNA was relatively abundantly distributed in most tissues, whereas uMPEG1 mRNA was almost exclusively expressed in placenta (Fig. 4B).

2.5. The C-terminal cationic segment of MPEG1 exhibits antibacterial activity

MPEG1 possess conserved domain structures, including a short CT tail of ~40 residues that is thought to orient toward the cytoplasm. Systematic analysis of over 1000 metazoan MPEG1 showed that the cytosolic tail is in general constituted by two biochemically distinct segments rich in positively and negatively charged residues, respectively

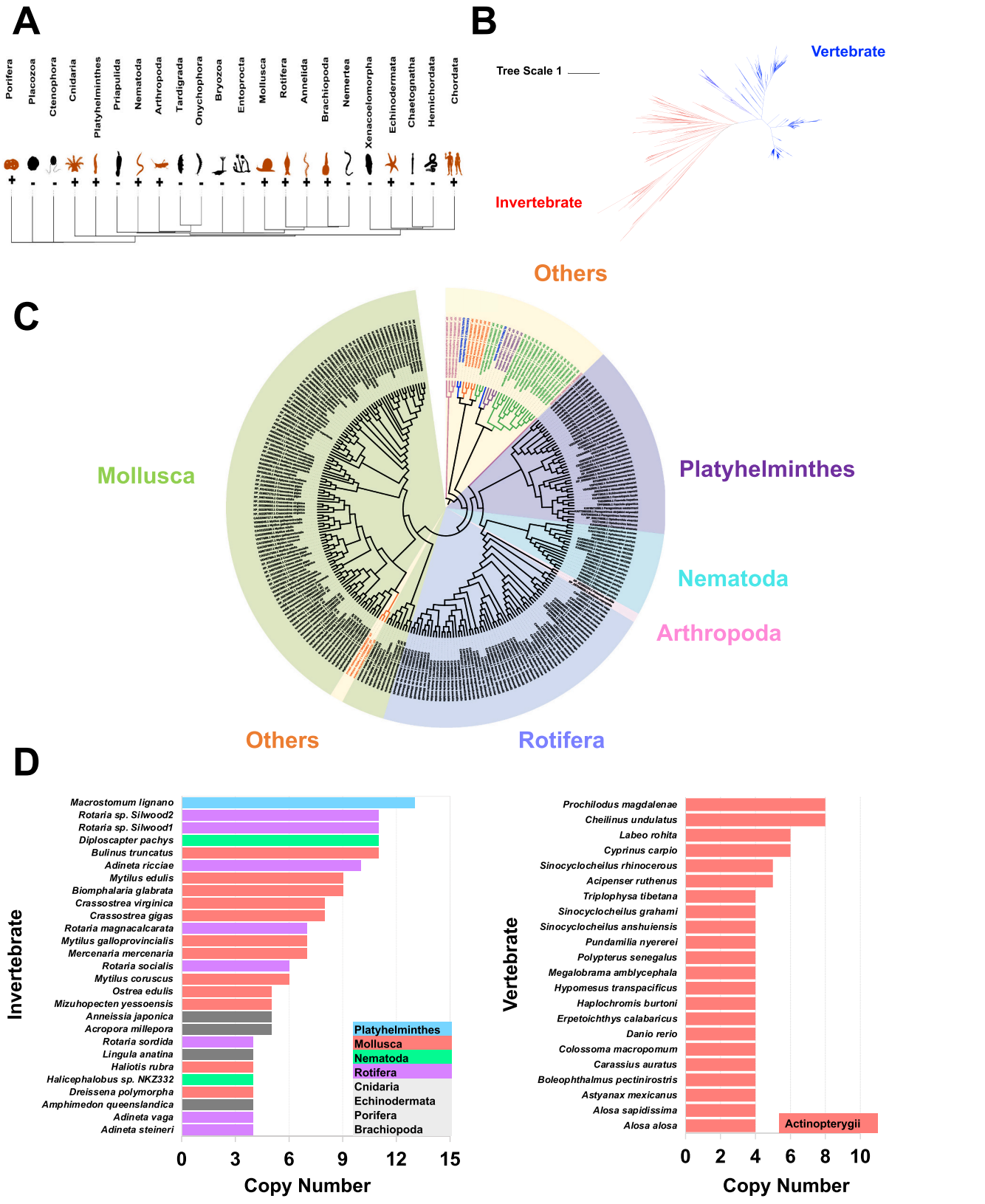


Fig. 1. Phylogenetic analysis of MPEG1 in Metazoa. (A) The phylogenetic tree of the major metazoan phyla with MPEG1 distribution. Orange and black icons indicate MPEG1-positive (+) and negative (-) phyla, respectively. (B) The unrooted phylogenetic tree indicating the genetic distance of vertebrate (blue) and invertebrate (red) MPEG1. The phylogenetic tree was constructed via maximum likelihood analysis with the JTT+R10 substitution model implemented in IQ-TREE 2. (C) The phylogenetic analysis of the MPEG1 in invertebrate. The phylogenetic tree was constructed via maximum likelihood analysis with the WAG+F+R8 substitution model implemented in IQ-TREE 2. Porifera was used as an outgroup for rooting. Different phyla are presented in different colors. In "Others", Porifera, Cnidaria, Annelida, Brachiopoda, and Echinodermata MPEG1 are colored in magenta, green, blue, purple, and red, respectively. (D) The invertebrate and vertebrate species that contain at least four copies of MPEG1.

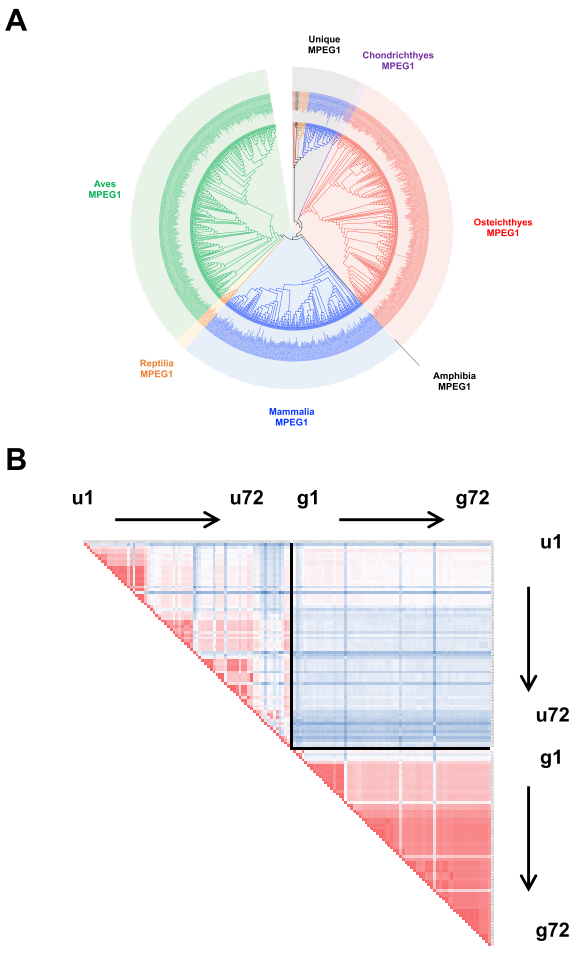


Fig. 2. Phylogenetic and sequence analyses of vertebrate MPEG1. (A) The phylogenetic tree of vertebrate MPEG1. The phylogenetic tree was constructed via maximum likelihood analysis with the JTT+R9 substitution model implemented in IQ-TREE 2. The tree was rooted at midpoint. Different classes are presented in different colors. In the "Unique MPEG1" clade, the MPEG1 of Osteichthyes, Amphibia, Reptilia and Mammalia are colored in red, black, orange, and blue, respectively. (B) Pairwise sequence comparison of 72 uMPEG1 (u1 to 72) and 69 gMPEG1 (g1 to 69). The color from blue to red indicates increasing sequence similarity. The full names of abbreviations are listed in Table S1.

(Fig. 5A). In most cases, a conserved tyrosine residue demarcates the two segments. The cationic segment ranges mainly from 10 to 20 amino acid residues in length and exhibits a high pI value (mostly over 10). These biochemical features resemble that of antimicrobial peptides (AMPs). AMPs are short, positively charged peptides widely distributed in all kingdoms of life as one of the most ancient innate immune effectors that can directly kill microbial pathogens by damaging the cellular membrane of the microbes [28,29]. To examine whether the cationic CT region possessed antimicrobial activity, peptides based on the basic segments of 35 invertebrate and vertebrate MPEG1 were synthesized. These peptides ranged from 7 to 22 aa in size and 10.58–13.20 in pI. The potential bactericidal activity of the peptides was tested against Gram-positive (*Streptococcus iniae* and *Micrococcus luteus*) and Gram-negative (*Escherichia coli*, *Edwardsiella tarda*, *Pseudomonas fluorescens* and *Vibrio harveyi*) bacteria. Apparent bactericidal activity against one or more bacteria was observed with all peptides, in particular, the peptides of *Bradysia coprophila*, *Amphimedon queenslandica*, *Danio rerio*, and *Siniperca chuatsi* exhibited relatively strong and broad-spectrum antibacterial effects (Figs. 5B and 5C; Figs S6-S17). Compared with other vertebrate peptides, Actinopterygii peptides appeared to target a broader range of bacteria. It is interesting that some peptides, such as those from *Anneissia japonica* (Aj-2) and *Nematostella vectensis* (Nv), appeared to slightly enhance the propagation of *P. fluorescens*, suggesting that these peptides might be utilized by the bacteria as a nutrient that promoted bacterial growth.

2.6. The antibacterial activity of the MPEG1 peptides depends on certain key residues

Four of the peptides, i.e., Aq-2, Bc, Dr-2, and Sc from *A. queenslandica*, *B. coprophila*, *D. rerio*, and *S. chuatsi*, respectively, that exhibited relatively strong antibacterial effects were further analyzed. The results showed that the antibacterial activities of these peptides were apparently dose-dependent (Fig. 6A). The minimum inhibitory concentrations (MICs) of Aq-2, Bc, Dr-2, and Sc were 4, 6, 4, and 8 μM, respectively, and the minimum bactericidal concentrations (MBCs) of these peptides were 40, 20, 10, and 20 μM, respectively. To identify the key loci in the peptides that were essential to antibacterial activity, three consecutive Arg residues in Aq-2, Bc, Dr-2, and Sc were substituted by Glu (Fig. 6B). The resulting mutant peptides, i.e., Aq-2M, BcM, Dr-2M, and ScM, exhibited much reduced pI (Fig. 6B). Unlike the wild type peptides, these mutant peptides displayed no detectable bactericidal activity (Fig. 6C). Consistently, the mutant peptides failed to bind to the bacteria (Fig. 6D-G).

3. Discussion

Previous reports showed that MPEG1 is present in Porifera, Cnidaria, Mollusca, Actinopterygii, Brachiopoda, and Mammalia [7,17–19]. With the advance of genome sequencing and data mining technology, MPEG1-like genes are expected to be found in more taxa. In this study, we conducted an unbiased data-mining of MPEG1 across 34 major metazoan phyla. MPEG1 was absent in 23 phyla and present in 11 phyla, most of the latter were reported for the first time to harbor MPEG1. The most primitive MPEG1 sequences were identified in Porifera *S. domuncula* and *A. queenslandica*, suggesting an ancient origin of MPEG1 in invertebrate and the existence of a common ancestor for all metazoan MPEG1. It is possible that over the course of evolution, MPEG1 had duplicated and expanded in some phyla and entirely lost in some other phyla, in the latter case, MPEG1 functional alternatives may have come into existence. Similar discontinuous distribution patterns were observed in other immune genes such as big defensins and gasdermin [30–33]. Phylogenetically, the MPEG1 in invertebrate are distant from that in vertebrates, implying diversified evolution of MPEG1 in invertebrate and vertebrate. Previous works have shown that the innate immune genes, such as TLR, NLR, and the pore-forming toxin actinoporin, are prone to expansion and duplication in invertebrate, as a compensation for the lack of adaptive immunity [34–36]. For MPEG1, its duplication in Euteleostomi led to the emergence of perforin-1 [37]. In this study, we observed significant duplications of MPEG1 in invertebrate species, notably *M. lignano* and *Rotaria*. Given the pore-forming ability of MPEG1, this observation suggests an important role of MPEG1 in immune defense that may be played to a better effect by increasing copy number.

In vertebrate, detailed phylogenetic analysis of MPEG1 identified distinct clades of Chondrichthyes, Actinopterygii, Amphibia, Reptilia, Aves and Mammalia formed by the major gMPEG1 in each of these classes. Unexpectedly, an additional clade was also identified, which was constituted by a group of unique uMPEG1 from 71 species belonging to the classes of Osteichthyes, Amphibia, Reptilia and Mammalia. The similarities between gMPEG1 and uMPEG1 are low and decrease from Actinopterygii to Mammalia, implying an increased divergence of MPEG1 in Mammalia, which might lead to functional difference in the MPEG1 of fish and mammals. Synteny analysis revealed marked conservations between gMPEG1 and uMPEG1, which promoted us to explore the genomic mechanism of MPEG1 evolution. It has been reported that repetitive elements, especially the transposable elements,

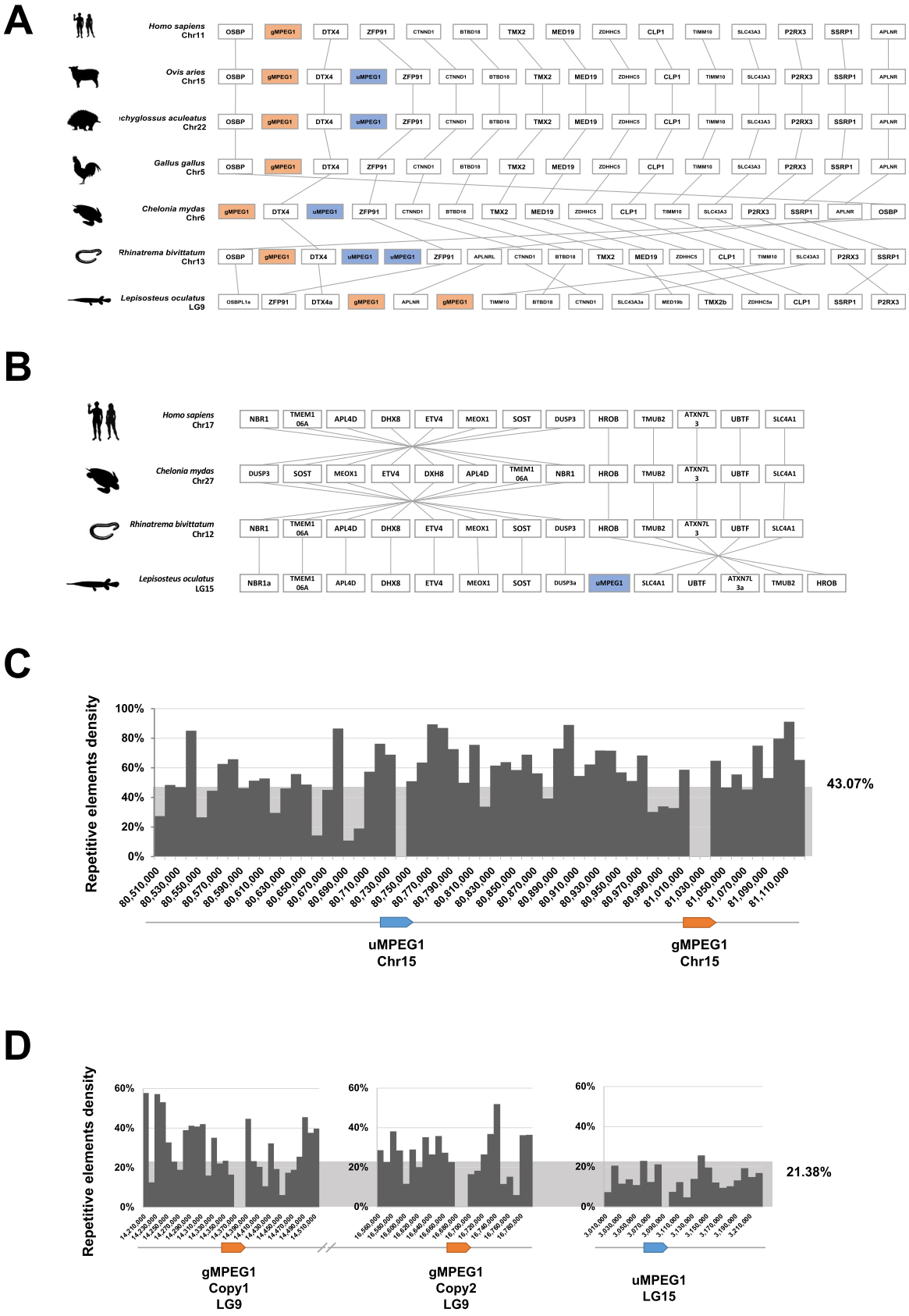


Fig. 3. Genomic synteny and transposon density surrounding the g/uMPEG1 loci. (A-B) Schematic diagrams showing conservation of the neighbor genes of gMPEG1 (orange) and uMPEG1 (blue) in vertebrate. (C-D) The MPEG1 locations and the adjacent transposon densities in *Ovis aries* (C) and *Lepisosteus oculatus* (D). The Y-axis represents the percentage of repetitive elements in a 10,000 bp bin; the height of the grey shade is the percentage of repetitive elements to the whole genome.

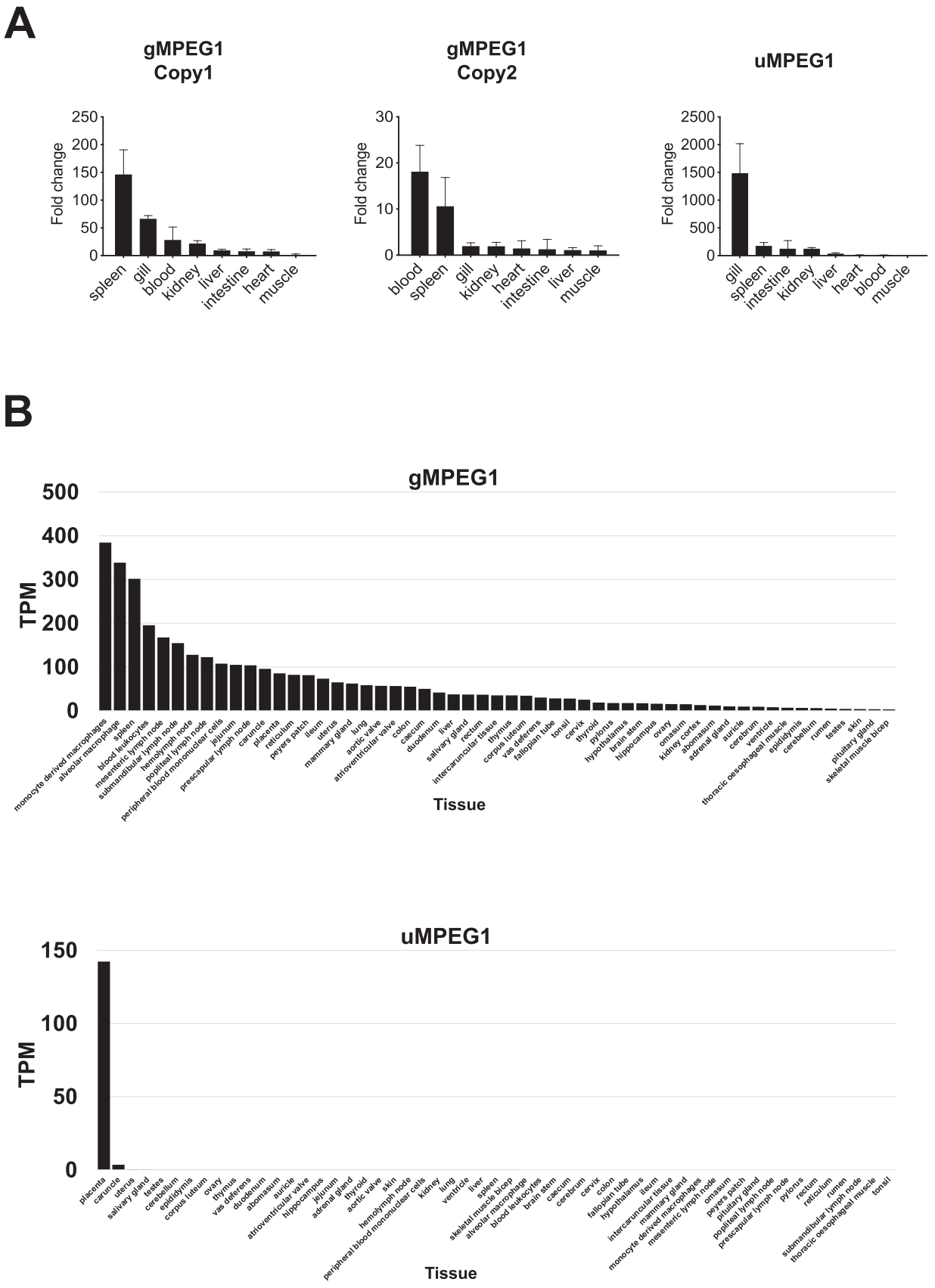


Fig. 4. The tissue-specific expression profiles of g/uMPEG1. (A) g/uMPEG1 expressions in eight tissues of *Lepisosteus oculatus* were determined by quantitative real time RT-PCR. For convenience of comparison, the expression level in muscle (the lowest) was set as 1. Data are the means of triplicate experiments and shown as means ± SD. (B) The expression profiles of gMPEG1 (upper panel) and uMPEG1 (lower panel) in 55 *Ovis aries* tissues based on transcriptome analysis. Gene expression is presented as TPM (transcripts per million) in the Y-axis.

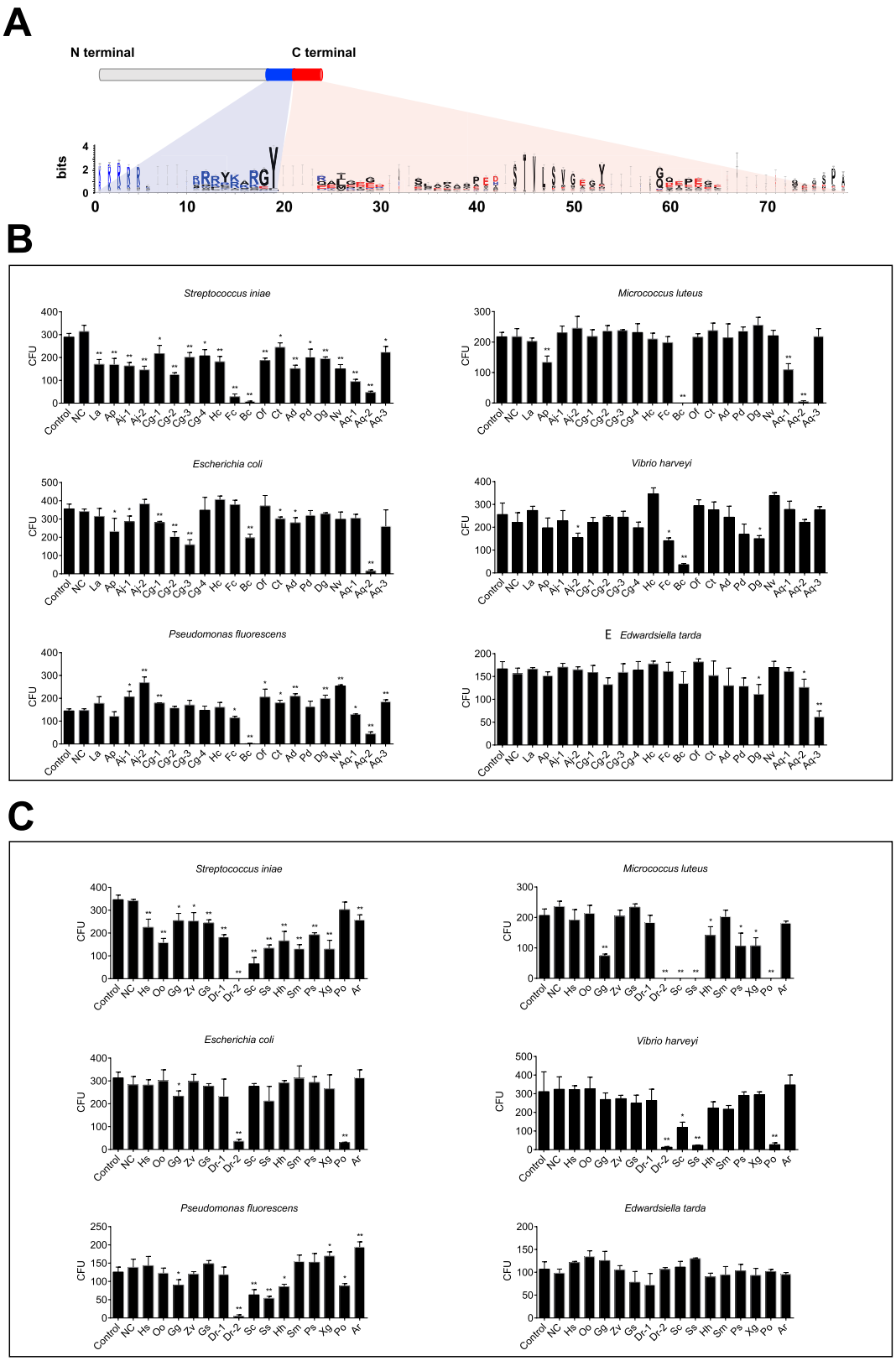
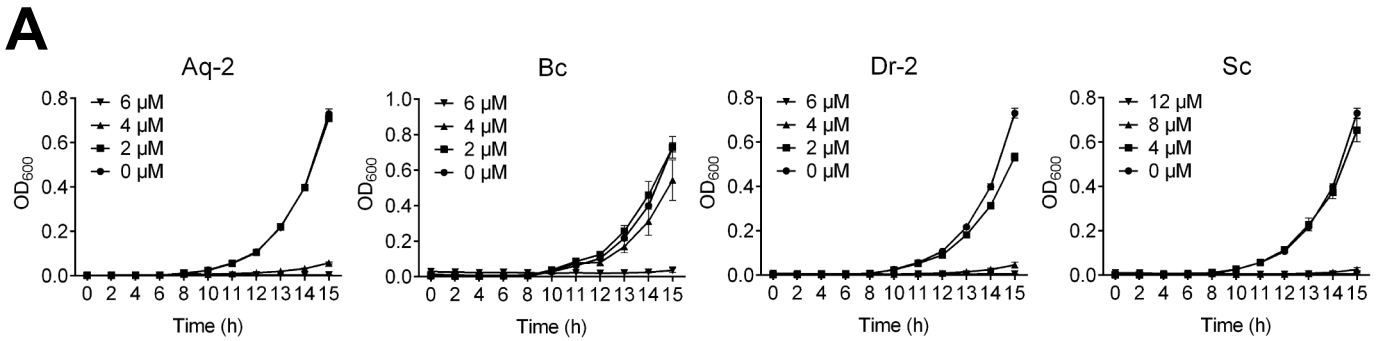
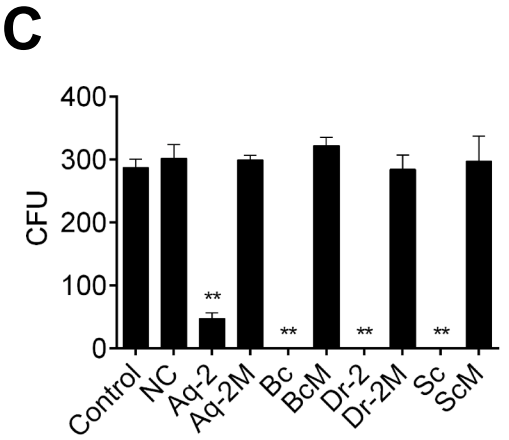


Fig. 5. The antibacterial effects of MPEG1 peptides. (A) Domain schematic of MPEG1. The conserved basic (blue) and acidic (red) regions, with their constitutive cationic/anionic residues, of the C-terminal tail are indicated. (B-C) The bactericidal activity of invertebrate (B) and vertebrate (C) MPEG1 peptides. The bacteria were treated with or without (control) the MPEG1 peptides of different species or with the negative control peptide (NC) for 4 h. Bacterial survival was then determined by plate count and shown as CFU (Colony Forming Unit). The full names of the species abbreviations are listed in Table S8. The Y-axis is the CFU. Values are the means of triplicate experiments and shown as means ± SD. $^{**}p < 0.01$; $^{*}p < 0.05$.



B

Peptide	Sequence	pI
Aq-2	RKWRSKRKQKSKKSSQRYM	12.55
Aq-2M	RKWESKEKQKSKKSSQEYM	10.58
Bc	KCFLRRRRSRA	12.68
BcM	KCFLEEERSRA	6.43
Dr-2	MKRRRKFSHFSFKRGY	12.53
Dr-2M	MKEEEKFSHFSFKRGY	9.22
Sc	VKRRRRLAGFRTVRGY	12.80
ScM	VKEEERLAGFRTVRGY	9.30



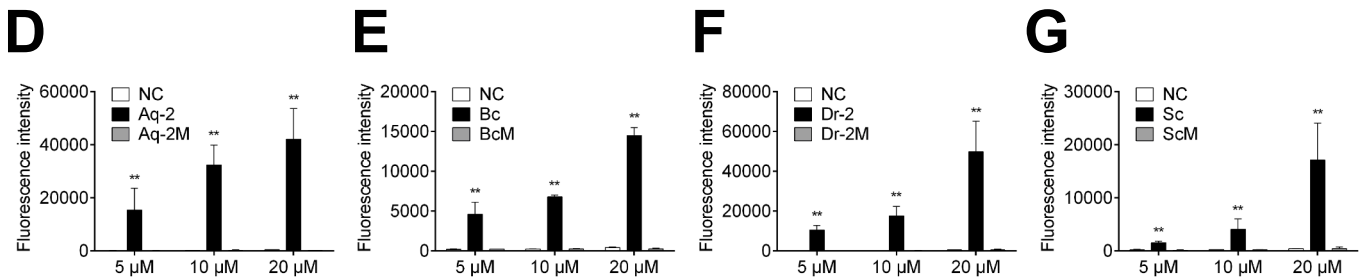


Fig. 6. The antibacterial characteristics of Aq-2, Bc, Dr-2, and Sc. (A) Aq-2, Bc, Dr-2, and Sc in different concentrations (0–12 μM) were incubated with *Micrococcus luteus*, and bacterial growth was determined at different hours by measuring absorbance at OD_{600}. (B) The sequences of the wild type and mutant Aq-2, Bc, Dr-2, and Sc. The mutated residues are shown in red. (C) *M. luteus* was incubated with 20 μM wild type or mutant Aq-2, Bc, Dr-2, and Sc, or with the negative control peptide (NC) for 4 h. Bacterial survival was determined by plate count and shown as CFU (Colony Forming Unit). (D-G) *M. luteus* was incubated with different concentrations of wild type or mutant Aq-2, Bc, Dr-2, and Sc, or with the negative control peptide (NC) for 2 h. Bacteria-bound peptides were determined with ELISA. For panels (A) and (C-G), values are the means of triplicate experiments and shown as means ± SD. $^{**}p < 0.01$.

are the propelling force of genetic variation and new gene generation [38–41]. In the reference genomes of the well-studied *O. aries* (representing terrestrial Mammalia) and *L. oculatus* (representing aquatic Actinopterygii), as well as *R. bivittatum* and *C. mydas*, we found high densities of transposable elements, especially RNA transposons, surrounding the gMPEG1 loci. Given their ability to facilitate evolution, these transposable elements likely play a role in the generation of duplicated gMPEG1 that may have possibly evolved into uMPEG1.

To examine the activeness of the MPEG1 genes, the expression profiles of gMPEG1 and uMPEG1 in fish and mammal represented by *L. oculatus* and *O. aries*, respectively, were analyzed. In *L. oculatus*, the two copies of gMPEG1 expressed most abundantly in the internal immune organs of spleen and blood, while the uMPEG1 expressed most abundantly in gill, the immune organ of fish that comes into contact directly with pathogens in the external environment. These results support an immune associated function of gMPEG1 and uMPEG1. A previous report showed that in zebrafish, which possesses only gMPEG1, MPEG1 expression was associated with, besides macrophages, a subpopulation of B-lymphocytes in most adult fish organs, especially in irf8^{null} myeloid-defective mutant [42], suggesting a role of fish MPEG1 in not only innate immunity but also adaptive immunity. It is interesting that in our study, although both gMPEG1 and uMPEG1 expressions were detected in multiple organs, the fold difference of uMPEG1 expression between the highest- and the lowest-expression tissues was ~10 times more than that of the gMPEG1 (1500-fold vs 20–150-fold). Hence, compared with gMPEG1, uMPEG1 exhibited much stronger tissue specificity in expression and is predominately expressed in a single immune organ (gill). Similarly, in *O. aries*, gMPEG1 expressed in a relatively ubiquitous manner in multiple tissues, whereas uMPEG1 expressed highly specifically in one tissue, the placenta. Like the fish gill, the mammalian placenta, in delivering nutrients and oxygen to the fetus, serves as the first immune barrier to various external pathogens. The almost exclusive expression of uMPEG1 in the gill of fish and the placenta of mammal suggests a specific and evolutionary conserved role of uMPEG1 in immune defense associated with host-pathogen interaction.

As a Type I transmembrane protein, MPEG1 orients its CT tail in the cytoplasm. The function of this cytoplasmic tail remains to be investigated. A previous report showed that the CT tail was involved in the intracellular trafficking of MPEG1 in a manner that depended on ubiquitylation on certain lysine residues, which could be blocked by pathogenic *E. coli* and *Yersinia pseudotuberculosis*, thus aborting the bactericidal activity of MPEG1 [8]. In our study, a large-scale analysis of metazoan MPEG1 revealed that the CT tail is generally divided into two different regions with opposite electrostatic properties. In structure, the positively charged region is akin to classical AMPs. In line with this observation, a test of 35 CT tail-based peptides indicated that all peptides exhibited apparent bactericidal activity against one or more bacterial species, including the well-known aquatic pathogens of *S. iniae* and *V. harveyi*. It is notable that relatively strong and broad-spectrum bactericidal effects were observed with sponge, gnat, and fish derived peptides in a manner that depended on the highly positively charged residues, suggesting an AMP-like working mechanism of these peptides. Considering that sponge, gnat, and fish live in microbe-rich water-y/moisture environments whole life or at a certain life stage, the antibacterial effects of the CT tail might be a beneficial trait for these animals to adapt to the environment. It has been proposed that in order to exert its pore forming activity, the MPEG1 N-terminal ectodomain containing the MACPF and P2 regions is likely cleaved, via some unknown mechanism, off the anchoring membrane [7]. It is possible that during bacterial infection, the cleaved ectodomain monomers form oligomers that perforate the cellular membrane of the bacteria in the phagosome, while the released CT tail interacts with and kills free bacteria in the cytoplasm via the cationic segment. The CT tail may also be further cleaved to release the cationic segment (Fig S18). Taken together, the results of our study suggest a possibility that, in addition to the proposed role of mediating MPEG1 trafficking, the CT tail may also have an immune function of its own by directly binding and killing the pathogens that have invaded into the cytoplasm. All these results support a role of MPEG1 in antimicrobial immunity, in particular in invertebrate and teleost that lack or have relatively primitive adaptive immune systems.

4. Materials and methods

4.1. Sequence collection

A total of 238 MPEG1 reference sequences (Table S9) were collected from NCBI Orthologs and used as queries to search against the non-redundant database via TBLASTN with E-value set as 1e-5 to ensure, in part, accuracy. The sequences were validated using the conserved domains database (https://www.ncbi.nlm.nih.gov/cdd/) [43]. The protein sequences of the putative MPEG1 homologs were further aligned using Clustal Omega [44] and genomic locations to remove duplicates (Table S9).

4.2. Phylogenetic and syntenic analysis

The phylogenetic tree of life representing the major metazoan phyla was fetched from the public knowledge-based TimeTree (http://timetree.org/) [45]. The phyla icons used in the tree were downloaded from PhyloPic (http://www.phylopic.org/), with the detailed credentials provided in the Table S10. Sequence alignments were conducted with Clustal Omega [44]. For the phylogenetic analysis of MPEG1, a maximum likelihood tree was generated using IQ-TREE 2 v.2.1.2 with 1000 bootstrap [46]. The JTT+R10, WAG+F+R8, and JTT+R9 substitution models were used based on the BIC criterion for the metazoan, invertebrate, and vertebrate trees, respectively. The final presented tree was visualized with iTOL (https://itol.embl.de/) [47]. The sequence similarities were calculated with needleall.

4.3. Repetitive element analysis

The repetitive elements in *L. oculatus*, *O. aries*, *R. bivittatum*, and *C. mydas* were identified using RepeatModeler 1.0.8 containing RECON and RepeatScout with default parameters [48,49]. The derived repetitive sequences were searched against Repbase [50]. The total repetitive elements on the chromosomes / scaffolds were subtotaled and collected in a 10,000 bp bin (for *L. oculatus* and *O. aries*) or 20,000 bp bin (for *R. bivittatum* and *C. mydas*).

4.4. Quantitative real-time reverse transcription-PCR (qRT-PCR) and RNA-sequencing (RNA-seq)

L. oculatus (average weight of 40 g) were purchased from a commercial company in Guangdong province, China, and kept at 28 ± 1 °C in tanks with aerated water for at least one week before the experiments. For experiments involving tissue collection, the fish were euthanized with tricaine methanesulfonate (Sigma, St. Louis, MO, USA). qRT-PCR was performed as described previously [51], with the primers listed in Table S11. The expression levels of MPEG1 were calculated using the comparative threshold cycle method ($2^{-\Delta\Delta CT}$) with elongation factor-1-α (Ef1α) as an internal reference [52]. The RNA-seq was based on the gene expression atlas in domestic sheep [53].

4.5. Antibacterial assays

The peptides (Table S8) derived from the C-terminal tail of 35 invertebrate and vertebrate MPEG1 were synthesized by Sangon (Shanghai, China). To determine the antibacterial effects of the peptides, *V. harveyi*, *E. coli*, *P. fluorescens*, *M. luteus*, *E. tarda*, and *S. iniae* were cultured in LB or TSB medium as reported previously to logarithmic

phase and resuspended in PBS to 5×10^3 CFU/ml [54]. Each of the bacterial suspension was incubated with the MPEG1 peptides or the negative control peptide P86P15 [55] (final concentration of 20 μM), or PBS (control) for 4 h at room temperature. The number of survived bacteria was determined by plate count as reported previously [56]. The minimum inhibitory concentrations (MICs) and the minimum bactericidal concentrations (MBCs) of Aq-2, Bc, Dr-2, and Sc against were determined as reported previously [54,57].

4.6. Binding of peptides to bacteria

Peptide-bacteria interaction was determined with enzyme-linked immunosorbent assay (ELISA) as reported previously [56] with slight adjustments. Briefly, *M. luteus* was cultured as above to logarithmic phase and resuspended in coating buffer (15 mM Na_2CO_3, 35 mM $NaHCO_3$, pH 9.6) to a final concentration of 2×10^8 CFU/ml. The bacterial suspension was added to a 96-well plate (100 μl/well), and the plate was incubated at 4 °C for overnight. The plate was blocked with 5% skim milk and incubated at room temperature for 2 h. The plate was washed three times with PBST (PBS containing 0.1% Tween-20). Then, fluorescein isothiocyanate (FITC) labeled-peptides at various concentrations were added to the plate. The plate was incubated at room temperature for 2 h and washed five times with PBST. The fluorescence intensity was detected with a multifunctional microplate reader (TECAN Infinite M200 PRO, Switzerland).

4.7. Statistical analysis

The comparison of similarity and copy number between vertebrate and invertebrate MPEG1 were conducted via Wilcoxon rank test function in R language [58]. The qRT-PCR, antibacterial assays, and bacteria-binding assay were performed in triplicate. Statistical analyses were carried out using GraphPad Prism version 6.01 (GraphPad Software Inc., San Diego, CA, USA). Data were analyzed with student's t-test and considered statistically significant when $p < 0.05$.

Funding

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA22050402), the Science & Technology Innovation Project of Laoshan Laboratory (LSKJ202203000), and the Innovation Research Group Project of the National Natural Science Foundation of China (42221005).

CRediT authorship contribution statement

Conceptualization: LS, ZHY. Funding acquisition: LS. Methodology & Investigation: YC, ZHY. Writing - original draft: YC, ZHY. Writing - review & editing: LS.

Declaration of Competing Interest

none.

Data availability

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article / Table S9.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.11.032.

References

[1] Bayly-Jones C, Pang SS, Spicer BA, Whisstock JC, Dunstone MA. Ancient but not forgotten: new insights into MPEG1, a macrophage perforin-like immune effector. Front Immunol 2020;11:581906.
[2] Ni T, Jiao F, Yu X, Aden S, Ginger L, Williams SI, et al. Structure and mechanism of bactericidal mammalian perforin-2, an ancient agent of innate immunity. Sci Adv 2020;6(5):eaax8286.
[3] Jiao F, Dehez F, Ni T, Yu X, Dittman JS, Gilbert R, et al. Perforin-2 clockwise hand-over-hand pre-pore to pore transition mechanism. Nat Commun 2022;13(1):5039.
[4] Spilsbury K, O'Mara M-A, Wu WM, Rowe PB, Symonds G, Takayama Y. Isolation of a novel macrophage-specific gene by differential cDNA analysis. Blood 1995;85(6):10.
[5] Xiong P, Shiratsuchi M, Matsushima T, Liao J, Tanaka E, Nakashima Y, et al. Regulation of expression and trafficking of perforin-2 by LPS and TNF-α. Cell Immunol 2017;320:1–10.
[6] McCormack RM, de Armas LR, Shiratsuchi M, Fiorentino DG, Olsson ML, Lichtenheld MG, et al. Perforin-2 is essential for intracellular defense of parenchymal cells and phagocytes against pathogenic bacteria. Elife 2015;4:e06508.
[7] Merselis LC, Rivas ZP, Munson GP. Breaching the bacterial envelope: the pivotal role of perforin-2 (MPEG1) within phagocytes. Front Immunol 2021;12:597951.
[8] McCormack RM, Lyapichev K, Olsson ML, Podack ER, Munson GP. Enteric pathogens deploy cell cycle inhibiting factors to block the bactericidal activity of Perforin-2. Elife 2015;4:e06505.
[9] McCormack R, Hunte R, Podack ER, Plano GV, Shembade N. An essential role for perforin-2 in type I IFN signaling. J Immunol 2020;204(8):2242–56.
[10] McCormack R, De Armas LR, Shiratsuchi M, Ramos JE, Podack ER. Inhibition of intracellular bacterial replication in fibroblasts is dependent on the perforin-like protein (perforin-2) encoded by macrophage-expressed gene 1. J innate Immun 2013;5(2):185–94.
[11] Fields K, McCormack R, De Armas L, Podack E. Perforin-2 restricts growth of Chlamydia trachomatis in macrophages. Infect Immun 2013;81(8):3045–54.
[12] McCormack R, Bahnan W, Shrestha N, Boucher J, Barreto M, Barrera CM, et al. Perforin-2 protects host cells and mice by restricting the vacuole to cytosol transitioning of a bacterial pathogen. Infect Immun 2016;84(4):1083–91.
[13] Bai F, McCormack RM, Hower S, Plano GV, Lichtenheld MG, Munson GP. Perforin-2 breaches the envelope of phagocytosed bacteria allowing antimicrobial effectors access to intracellular targets. J Immunol 2018;201(9):2710–20.
[14] Rodríguez-Silvestre P, Laub M, Krawczyk PA, Davies AK, Schessner JP, Parveen R, et al. Perforin-2 is a pore-forming effector of endocytic escape in cross-presenting dendritic cells. Science 2023;380(6651):1258–65.
[15] Plano GV, Shembade N. Role of Perforin-2 in regulating type i interferon signaling. J Cell Immunol 2021;3(1):8.
[16] Ebrahimnezhaddarzi S, Bird CH, Allison CC, Tuipulotu DE, Kostoulias X, Macri C, et al. Mpeg1 is not essential for antibacterial or antiviral immunity, but is implicated in antigen presentation. Immunol Cell Biol 2022;100(7):529–46.
[17] Walters BM, Connelly MT, Young B, Traylor-Knowles N. The complicated evolutionary diversification of the Mpeg-1/Perforin-2 family in cnidarians. Front Immunol 2020;11:1690.
[18] Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, Funayama N, et al. The innate immune repertoire in Cnidaria-ancestral complexity and stochastic gene loss. Genome Biol 2007;8:1–13.
[19] Gerdol M, Luo Y-J, Satoh N, Pallavicini A. Genetic and molecular basis of the immune system in the brachiopod Lingula anatina. Dev Comp Immunol 2018;82:7–30.
[20] Wiens M, Korzhev M, Krasko A, Thakur NL, Perovic-Ottstadt S, Breter HJ, et al. Innate immune defense of the sponge Suberites domuncula against bacteria involves a MyD88-dependent signaling pathway: induction of a perforin-like molecule. J Biol Chem 2005;280(30):27949–59.
[21] He X, Zhang Y, Yu Z. An Mpeg (macrophage expressed gene) from the Pacific oyster Crassostrea gigas: molecular characterization and gene expression. Fish Shellfish Immunol 2011;30(3):870–6.
[22] Kemp IK, Coyne VE. Identification and characterisation of the Mpeg1 homologue in the South African abalone, Haliotis midae. Fish Shellfish Immunol 2011;31(6):754–64.
[23] Bathige S, Umasuthan N, Whang I, Lim B-S, Won SH, Lee J. Antibacterial activity and immune responses of a molluscan macrophage expressed gene-1 from disk abalone, Haliotis discus discus. Fish Shellfish Immunol 2014;39(2):263–72.
[24] Ni L-Y, Han Q, Chen H-P, Luo X-C, Li A-X, Dan X-M, et al. Grouper (Epinephelus coioides) Mpeg1s: molecular identification, expression analysis, and antimicrobial activity. Fish Shellfish Immunol 2019;92:690–7.
[25] Choi K-M, Cho D-H, Joo M-S, Choi H-S, Kim MS, Han H-J, et al. Functional characterization and gene expression profile of perforin-2 in starry flounder (Platichthys stellatus). Fish Shellfish Immunol 2020;107:511–8.
[26] Wang G-D, Zhang K-F, Zhang Z-P, Zou Z-H, Jia X-W, Wang S-H, et al. Molecular cloning and responsive expression of macrophage expressed gene from small abalone Haliotis diversicolor supertexta. Fish Shellfish Immunol 2008;24(3):346–59.
[27] Liu W, Liu B, Zhang G, Yao G, Zhang Y, Cen X, et al. Giant triton snail charonia tritonis macrophage-expressed gene 1 Protein Ct-Mpeg1: molecular identification, expression analysis, and antimicrobial activity. Int J Mol Sci 2022;23(21):13415.
[28] Pasupuleti M, Schmidtchen A, Malmsten M. Antimicrobial peptides: key components of the innate immune system. Crit Rev Biotechnol 2012;32(2):143–71.
[29] Wimley WC. Describing the mechanism of antimicrobial peptide action with the interfacial activity model. ACS Chem Biol 2010;5(10):905–17.

[30] Gerdol M, Schmitt P, Venier P, Rocha G, Rosa RD, Destoumieux-Garzón D. Functional insights from the evolutionary diversification of big defensins. Front Immunol 2020;11:758.
[31] Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, et al. Massive gene presence-absence variation shapes an open pan-genome in the Mediterranean mussel. Genome Biol 2020;21:1–21.
[32] Yuan Z, Jiang S, Qin K, Sun L. New insights into the evolutionary dynamic and lineage divergence of gasdermin E in metazoa. Front Cell Dev Biol 2022;10:952015.
[33] Angosto-Bazarra D, Alarcón-Vila C, Hurtado-Navarro L, Baños MC, Rivers-Auty J, Pelegrín P. Evolutionary analyses of the gasdermin family suggest conserved roles in infection response despite loss of pore-forming functionality. BMC Biol 2022;20(1):9.
[34] Zhang L, Li L, Guo X, Litman GW, Dishaw LJ, Zhang G. Massive expansion and functional divergence of innate immune genes in a protostome. Sci Rep 2015;5(1):8693.
[35] Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. Cell 2006;124(4):783–801.
[36] Koritnik N, Gerdol M, Šolinc G, Švigelj T, Caserman S, Merzel F, et al. Expansion and neofunctionalization of actinoporin-like genes in Mediterranean mussel (Mytilus galloprovincialis). Genome Biol Evol 2022;14(11):evac151.
[37] D'Angelo ME, Dunstone MA, Whisstock JC, Trapani JA, Bird PI. Perforin evolved from a gene duplication of MPEG1, followed by a complex pattern of gene gain and loss within Euteleostomi. BMC Evolut Biol 2012;12:1–12.
[38] Cerbin S, Jiang N. Duplication of host genes by transposable elements. Curr Opin Genet Dev 2018;49:63–9.
[39] Shao F, Han M, Peng Z. Evolution and diversity of transposable elements in fish genomes. Sci Rep 2019;9(1):15399.
[40] Cosby RL, Judd J, Zhang R, Zhong A, Garry N, Pritham EJ, et al. Recurrent evolution of vertebrate transcription factors by transposase capture. Science 2021;371(6531):eabc6405.
[41] Oliver KR, Greene WK. Transposable elements: powerful facilitators of evolution. Bioessays 2009;31(7):703–14.
[42] Ferrero G, Gomez E, Lyer S, Rovira M, Miserocchi M, Langenau DM, et al. The macrophage-expressed gene (mpeg) 1 identifies a subpopulation of B cells in the adult zebrafish. J Leukoc Biol 2020;107(3):431–43.
[43] Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, et al. The conserved domain database in 2023. Nucleic Acids Res 2023;51(D1):D384–8.
[44] Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 2019;47(W1):W636–41.
[45] Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol 2017;34(7):1812–9.
[46] Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams M, Haeseler AV, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era (vol 37, pg 1530, 2020). Mol Biol Evol 2020;(8):37.
[47] Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 2021;49(W1):W293–6.
[48] Bao Z, Eddy SR. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res 2002;12(8):1269–76.
[49] Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics 2005;21(suppl_1):i351–8.
[50] Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob Dna 2015;6:1–6.
[51] Chen Y, Yu C, Jiang S, Sun L. Japanese flounder HMGB1: a DAMP Molecule that promotes antimicrobial immunity by interacting with immune cells and bacterial pathogen. Genes 2022;13(9):1509.
[52] Venkatachalam AB, Fontenot Q, Farrara A, Wright JM. Fatty acid-binding protein genes of the ancient, air-breathing, ray-finned fish, spotted gar (Lepisosteus oculatus). Comp Biochem Physiol Part D: Genom Proteom 2018;25:19–25.
[53] Clark E, Bush S, McCulloch E, Farquhar I, Young R, Lefevre L, et al. A high resolution atlas of gene expression in the domestic sheep (Ovis aries). PLoS genetics 2017;13(9):e1006997.
[54] Xu H, Yuan Z, Sun L, Non-Canonical A. Teleost NK-Lysin: antimicrobial activity via multiple mechanisms. Int J Mol Sci 2022;23(21):12722.
[55] Zhang M, Hu Y-H, Xiao Z-Z, Sun Y, Sun L. Construction and analysis of experimental DNA vaccines against megalocytivirus. Fish Shellfish Immunol 2012;33(5):1192–8.
[56] Li H, Sun Y, Sun L. A teleost CXCL10 is both an immunoregulator and an antimicrobial. Front Immunol 2022;13.
[57] Catherine, Cesa-Luna, Jesús, Muoz-Rojas, Gloria, Saab-Rincon, et al. Structural characterization of scorpion peptides and their bactericidal activity against clinical isolates of multidrug-resistant bacteria. PloS One 2019;14(11):e0222438.
[58] R. R Core Team, R: A language and environment for statistical computing, (2013).