ISJ 16: 173-183, 2019 ISSN 1824-307X

RESEARCH REPORT

HC-SPH: A conserved serine protease homolog of S1 superfamily in the triangle-shell pearl mussel (*Hyriopsis cumingii*)

Q Liu[1,2], B Xu[1,2*], T Xiao[1,2*]

[1]Hunan Engineering Technology Research Center of Featured Aquatic Resources Utilization, Hunan Agricultural University, Changsha 410128, China
[2]Collaborative Innovation Center for Efficient and Health Production of Fisheries in Hunan Province, Changde 415000, China

Accepted September 12, 2019

Abstract

Serine proteases play central roles in immune defense in invertebrates through innate immunity, and are particularly important complement system in molluscs because their susceptibility to infection due to lack of an adaptive immune ability. A gene encoding the serine protease homolog from the triangle - shell pearl mussel (*Hyriopsis cumingii*) was identified and designated as HC-SPH in this study. Protein sequence analysis revealed that HC-SPH consists of a typical Tryp_SPc functional domain of serine protease of S1 family lead by a signal peptide, and the molecule shares a highly conserved sequence and structural organization with other members, including a cleavage site, 3 enzymatic active sites and 3 substrate binding sites, so that it was clustered into a trypsin-like serine protease subfamily of the S1 superfamily. Semi - quantitative analysis of the amplicons separated on agarose gel by comparing to the β-actin products revealed that the digestive gland had a strong expression while the gonads were seen as weak expression sites. Infected by *Aeromonas hydrophila*, the gene expression was significantly up - regulated in the kidney at the 6 hours post challenge (hpc), stomach at 12 hpc and gills at 24 hpc while the expression maintained steadily unchanged in the digestive gland. However, up to 48 hpc, the expression levels in all four tissues reached significantly high, and also joined by a high level of expression in intestine that was down - regulated before 24 hpc, to build up an enhanced immune defense. The complementary up - regulation of the gene expression in these tissues suggested a temporal and spatial reinforce model for HC-SPH in immune response.

Key Words: Tryp_SPc domain; Serine protease homolog; S1 superfamily; immune response; mollusk

Introduction

Serine proteases (SPs), commonly recognized by their serine activity, are hydrolyse protein family of enzymes with wide distribution and important biological actions (Ross *et al.*, 2003; Liu *et al.*, 2019; Wei *et al.*, 2019). SPs are molecularly characterized by a conserved Tryp_SPc domain together with the catalytic triad histidine (His, H), aspartic acid (Asp, D) and the serine (Ser, S) (Li *et al.*, 2012b; Wang *et al.*, 2014). Inactive SPs with key residue(s) of catalytic triads replaced by other amino acid are defined as SP homologs (SPHs), which are similar to SPs in amino acid sequence (Liu *et al.*, 2010) and classified as protease precursors. SPHs assist SPs as cofactors to exert their functions (Wang and Jiang, 2004; Lee *et al.*, 2018) by activating/inhibiting the original protease activities in variety of biological processes such as digestion, cell differentiation, tissue remodeling, angiogenesis and embryonic development as well as their involvement in cellular and humoral immune responses (Rawlings and Barrett, 1993; Krem and Di Cera, 2001, 2002).

Up to date, SPH was identified in many invertebrates and some of which were shown to play roles in different biological processes, including antimicrobial activity from horseshoe crab *Tachypleus tridentatus* (Kawabata *et al.*, 1996), defense of pathogen challenge in Chinese white shrimp *Fenneropenaeus chinensis* (Ren *et al.*, 2009,2011), cell adhesion in crayfish *Pacifastacus leniusculus* (Huang *et al.*, 2000) and the tiger shrimp *Penaeus monodon* (Liu *et al.*, 2006), and activation of the prophenoloxidase (proPO) system (Kwon *et al.*, 2000; Charoensapsri *et al.*, 2009; Cui *et al.*,

Corresponding authors:
Baohong Xu
Tiaoyi Xiao
Hunan Engineering Technology Research Center of Featured Aquatic Resources Utilization
Hunan Agricultural University
Changsha 410128, China
E-mail: xbht568@126.com; tyxiao1128@163.com

2010), which represent an important host natural immune system that recognizes foreign invasive substances by molecular pattern of proteins. The molecular characterization, gene cloning and expression profile of the SPH were also abundantly studied in insect and other arthropods (Jitvaropas *et al.*, 2009; Ren *et al.*, 2011) while little is known in mollusks.

The freshwater triangle - shell pearl mussel *Hyriopsis cumingii* is a mollusk species widely cultured in China for pearl production (Ren *et al.*, 2012). However, mussel aquaculture has been declining in the recent years because of the diseases by bacterial and viral infection. Like other mollusks, the mussel lacks acquired immune system and mainly relies on innate immunity to defense pathogens invasion (Wang *et al.*, 2013). In this study, a gene (termed HC-SPH) coding for SPH (termed HC-SPH for encoded proteins) was cloned and characterized from the *H. cumingii*. The tissue - specific gene expression profiles of the gene before and after infection with the common infectious bacteria *Aeromonas hydrophila* were analyzed to provide evidence of its potential role in innate immune responses as suggested from other species. The results reported here summarized the molecular identification and characterization of HC-SPH, which contribute to general knowledge of SPH from a mollusk and particularly provide important molecular data for further investigation of HC-SPH towards a better understanding for its role in immune defense against infections.

Materials and methods

Experimental animal and immune challenge

Mussels (*Hyriopsis cumingii*), 285.33 ± 55.18 g and 14.37 ± 1.28cm, were obtained from a commercial farm in Changde, Hunan Province, China and cultured in freshwater at 24 - 28 °C. They were acclimatized for a week before being used for experiments.

The healthy mussels were injected in axe feet with 10^9 cfu/mL (≈ 0.5 mL) of *Aeromonas hydrophila* isolated and purified from dying mussels for challenging, and with an equal volume of phosphate buffer saline (PBS) as controls. Three challenged and three control mussels were chosen for sample collection from the digestive gland, stomach, intestine, gill, heart, mantle, axe foot, adductor muscle, kidney and the gonad at the time points of 0 h, 3 h, 6 h, 12 h, 24 h, 48 h after injection. All samples were snap - frozen in liquid nitrogen and immediately stored at - 80 °C under RNase - free conditions.

RNA extraction and cDNA synthesis

Total RNA samples were isolated using the RNAprep Pure tissue kit (Tiangen Biotech, China) according to the manufacturer's instruction, and treated with DNase I. RNA samples were kept in DEPC - treated water, qualified by gel electrophoresis and quantified using a biophotometer (Eppendorf, Germany).

About 5 μg total RNA from the digestive gland of healthy mussel was converted into 5′-RACE - Ready cDNA and 3'-RACE-Ready cDNA by reverse transcription using the SMARTer™ RACE (Rapid amplification of cDNA ends) cDNA Amplification Kit (Clontech, USA) for full - length cDNA cloning. About 2 μg total RNA from each tissue sample was reverse transcribed by RevertAid™ First Strand cDNA Synthesis Kit (Fermentas, Canada) for gene expression study.

RACE and sequencing analysis

The designing of gene specific primers used for this study was based on the expressed sequence tag (EST, GenBank accession number: FE968619) from a cDNA library generated by suppression subtractive hybridization (Xiao *et al.*, 2009). The 5′-RACE - Ready cDNA was synthesized using the SMART II™ A Oligonucleotide and the 5′-RACE CDS Primer A (Table 1); The 5′-RACE was performed using gene specific primer pairs UPM/ HC-SPH - GSPR - out and UPM/ HC-SPH - GSPR - in (Table1) in nested PCRs. 3'-RACE-Ready cDNA was synthesized using the 3′ - RACE CDS Primer A (Table 1). The 3′-RACE was performed using gene specific primer pairs UPM/HC-SPH-GSPF - out and UPM/ HC-SPH - GSPF - in (Table 1) in nested PCRs. The PCR reactions were catalyzed by Ex Taq polymerase (Takara, Japan) and the amplified fragments were cloned into the pUCm - T vector (Biobasic, Canada) for sequencing analysis (Sangon, China). The full - length cDNA coding for the SP homolog of *H. cumingii* was amplified and identified by its overlapping sequence to the EST.

DNA and amino acid sequence analysis

The cDNA sequence of HC-SPH was submitted and translated into amino acid sequence (http://www.vivo.colostate.edu/molkit/translate/), termed as HC-SPH, which was further characterized for its molecular property and homology prediction. The protein sequence was submitted to BLASTP searching against various databases through UniProtKB/SwissProt (http://www.uniprot.org/blast/) for a primary homology result and collected close related amino acid sequence for alignment. The sequence was used for a prediction of its functional domains and motifs by the online service of the Eukaryotic Linear Motif (ELM) resource for functional sites in proteins. The SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/) was used to define the signal peptide and the cleavage site. The amino acid sequences related to HC-SPH found by UniProtKB were retrieved for annotating the place of HC-SPH, sequences clustered directly with HC-SPH and the sequences with well-defined Tryp_SPc domains used for functional motif alignment were listed in Table S1. The sequences were aligned and edited by CLUSTALX 1.83 (Thompson *et al.*, 1997). Phylogenetic analysis was performed with MEGA 6.0 using maximum likehood method with a bootstrap valve of 1000.

The HC-SPH *gene expression by RT-PCR and RT-qPCR*

The expression profiling of HC-SPH transcripts in the digestive gland, stomach, intestine, gill, heart, mantle, axe foot, adductor muscle, kidney and gonad from healthy groups were measured using RT-PCR. The quantity of HC-SPH transcripts in the

Table 1 Primers designed for cloning and expression analysis of HC-SPH

Primer name	Sequence
SMART II™ A Oligonucleotide	5'-AAGCAGTGGTATCAACGCAGAGTACGCGGG-3'
3′-RACE CDS Primer A(3′-CDS)	5'-AAGCAGTGGTATCAACGCAGAGTAC$(T)_{30}$V N-3' *
5′-RACE CDS Primer A (5′-CDS)	5'-$(T)_{25}$V N-3' *
10×Universal Primer A Mix (UPM)	5'-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3′
Long Primer	5'-CTAATACGACTCACTATAGGGC-3'
Short PrimerRandom Primer	5'-$(dN)_{9}$-3' *
Oligo $(dT)_{20}$	5'-$(dT)_{20}$-3'
HCSHP-GSPR-out	5'-GGCGATGTCGTCCTTGAGACTGGTG-3'
HCSPH-GSPR-in	5'-CGCTGAATCCTGGACCCTCAAAATG-3'
HCSPH-GSPF-out	5'-TTGACCTTGACCTCAGACATTGGCAG-3'
HCSPH-GSPF-in	5'-ATGCCTCTCATCACTACCGTGTTGGA-3'
HC-SPH-pf	5'-TGAGGGTCCAGGATTCAGCGTAT-3'
HC-SPH-pr	5'-CCCCAACCAGGGAGGTAGCAAAC-3'
β-actin -pf	5'-ACTCACACCGTCCCCATCTAT-3'
β-actin -pr	5'-CGATTTCTCTTTCAGCAGTGG-3'

* N = A, C, G, or T; V = A, G, or C in primer sequences

digestive gland, stomach, intestine, gill and kidney from both challenged and control groups were measured using RT-qPCR. The HC-SPH gene - specific primer pairs, HC-SPH - pf and HC-SPH - pr (Table 1), was designated to amplify a fragment of 279 bp located between the position 326 and 603 of the gene. The house - keeping gene β-actin was used as an internal control amplified by the gene specific primer pair β-actinpf/β-actin-pr (Table 1) for an expected fragment of 148 bp. The RT-qPCR reaction was conducted using the TransStart™ Top Green qPCR SuperMix (TransGen Biotech, China) on an ABI 7300 Real-Time Detection System (Applied Biosystems, USA) as instructed by the manufacturer. Briefly, the final PCR reaction mixture (25 µL) in H_2O consisted of 12.5 µL of 2 x TransStart™ Top Green qPCR SuperMix, 0.5 µL of 50 × Passive Reference Dye1, 1 µL cDNA and 0.5 µL of each primer, was incubated at the denaturing temperature of 94 °C for 30 s first followed by 40 cycles of 5 s at 94 °C, 31 s at 61 °C. All PCR products were dissociated and confirmed a single amplicon from each sample by the system. A series of 10 - fold diluted mixture of first - strand cDNAs from all selected tissues were used to construct two standard curves (Chang *et al.*, 2007) for HC-SPH and β-actin respectively against the no cDNA negative controls. Each sample was run in triplicates. Data collected were analyzed using the Sequence Detection System (SDS version1.4, Applied Biosystems, USA). The relative expression of HC-SPH was determined using the comparative threshold cycle (Ct) method ($2^{-(\Delta\Delta CT)}$ method) as described previously (Livak and Schmittgen, 2001; Jiang et al., 2019). The arithmetic mean ± standard deviation was calculated and used for determining the significance of differences by *t* - test using the Statistics Package for Social Science (SPSS) 17.0. $p < 0.05$ means significant differences and $p < 0.01$ for most significant differences.

Results

Isolation of HC-SPH*, a full - length cDNA sequence coding for* HC-SPH

A cDNA fragment of 478 bp (EST tag of SP from the triangle shell mussel, *Genbank* accession number: FE 968644) was retrieved from the database, based on which a pair of gene specific primers were designed (GSPF and GSPR listed in Table 1) for amplifying the full - length cDNA sequence (Fig. 1A). A 344 bps 5′-RACE fragment and a 495 bps 3′-RACE fragment (Fig. 1B) overlapped to the EST fragment to give the rise of a full - length cDNA sequence of 1046 bp (Fig. 1C). The HC-SPH contains a 5′ - terminal untranslated region (UTR) of 73 bps, an 864 bps open reading frame (ORF) encoding 287 amino acid residues and a 3′-UTR of 109 bps with a canonical polyadenylation signal sequence AATAAA and a labile motif ATTTA (both underlined) followed by a poly (A) tail (Fig. 1C).

Bioinformatic analysis of amino acid sequence of HC-SPH *protein*

Full-length HC-SPH cDNA and its encoded SP in *H. cumingii* (HC-SPH) were submitted to the GenBank Accession number: GU222695. The amino acid sequence from ORF (303 aa) of the HC-SPH was analyzed using UniProtKB and ELM in conjunction with biological databases and found a typical Tryp_SPc (Trypsin-like SP) domain of the SP family between R_{52} and I_{290} (Fig. 2A). The N - terminus (K_1 - I_{70}) was recognized as a signal peptide (Fig. 2A) for proteolytic cleavage around R_{52}. Highly conserved motifs important for proteolytic cleavage, typical enzymatic active sites, and substrate binding sites for the enzyme are shown along the molecule (Fig. 2A). Multiple sequence alignment of HC-SPH with six other well-defined SPs showed those conserved motifs (Fig. 2B) around the

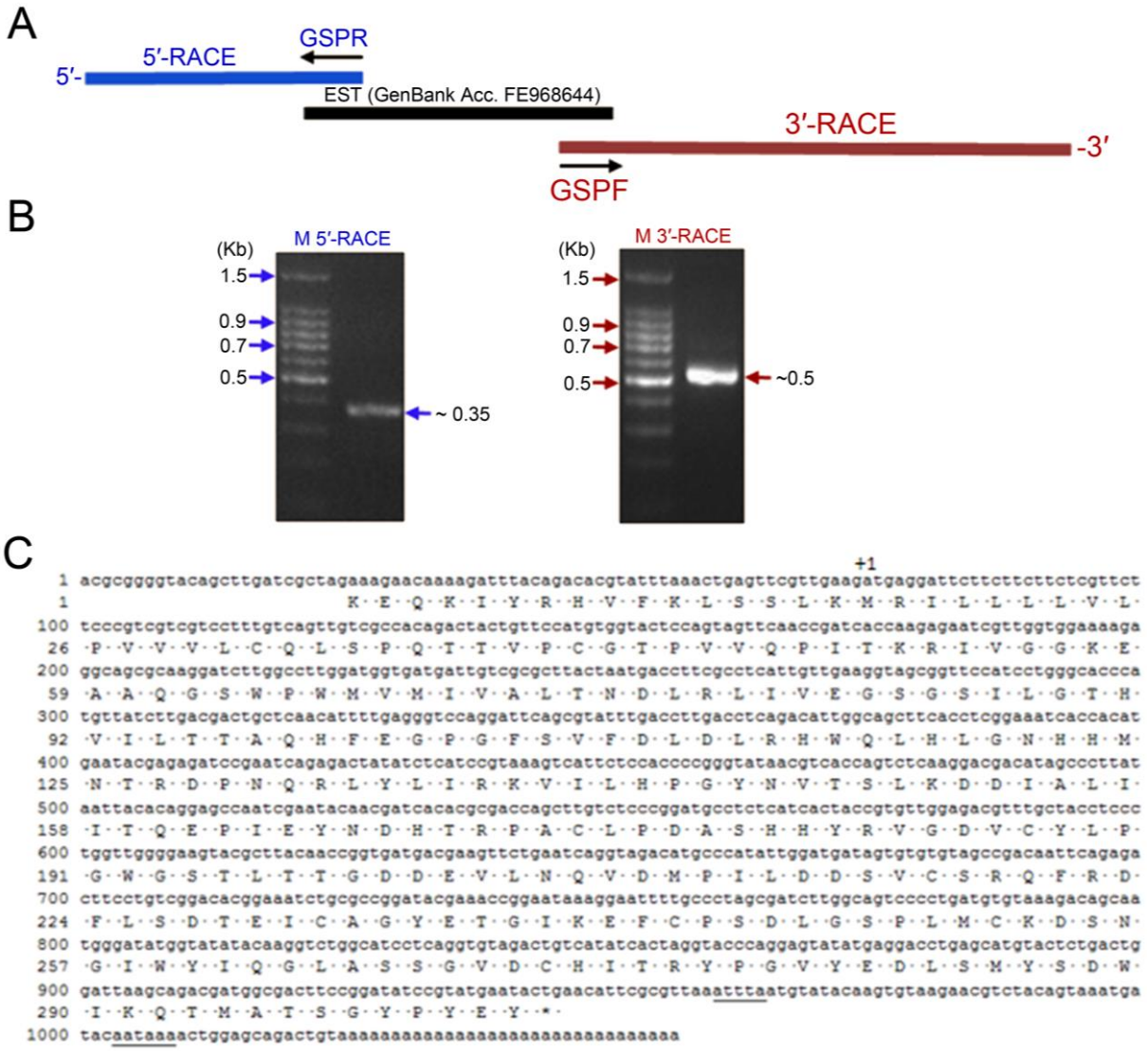

Fig 1 The full - length cDNA cloning of HC-SPH gene and its encoded amino acid sequence (HC-SPH). The HC-SPH EST - tag sequence from GenBank was used for designing of the gene specific primers for RACE PCRs (A), which amplified specific products as single bands (sizes of molecular weight maker and the products were given as indicated by arrows) for 5' - and 3' - RCAE respectively (B). The determined nucleotide sequence, contained a canonical polyadenylation signal sequence AATAAA and a labile motif ATTTA (both underlined) followed by a poly (A) tail, and the deduced amino acid sequence were given in single letter format (C). The translation initiation methionine (M) is labelled by +1 on top

enzymatic important sites with the ranges numbered at the bottom. The numbers of non - conserved amino acid residues before, between and after the motifs are given for each species respectively in brackets (Fig. 2B).

The UniProtKB BLAST search against databases revealed a high similarity of HC-SPH to the Tryp_SPc superfamily, amongst which protein sequences with defined biological properties were retrieved for similarity analysis (Fig. 3). A phylogenetic tree was generated to show the position and similarity relationship of HC-SPH in the unrooted tree (Fig. 3). HC-SPH was clearly clustered with a group of proteins that have strong trypsin activities.

The HC-SPH *expression in tissues and regulation in response to bacterial infection*

The presence of HC-SPH gene transcripts in different tissues of healthy *H. cumingii* was detected by RT-PCR as strategized (Fig. 4A) using a pair of gene - specific primers, HC-SPH-pf and HC-SPH-pr (Table 1). A fragment with the expected size (279 bp) was amplified from the tissues listed in all replicates in comparison with the relative amount of cDNA template reflected by PCR amplification of the house

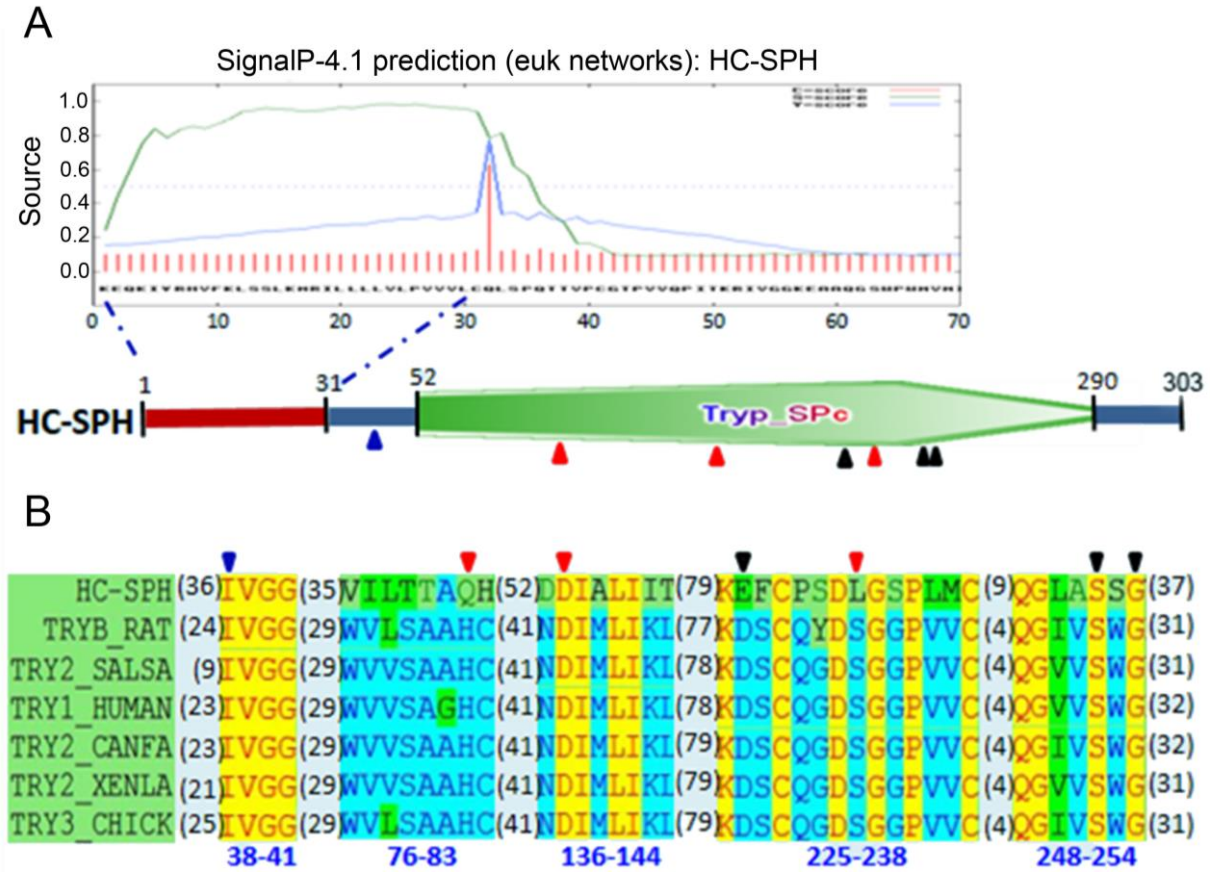

Fig. 2 Characterization of the molecular properties of HC-SPH. A signal peptide sequence was identified at the N - terminus with a proteolytic cleavage site, presented by SignalP 4.1 (A). Using UniProtKB BLAST conjunct with ELM searches, a typical Tryp_SPc functional domain was detected (A), including a cleavage site (indicated by blue ▲**),** 3 enzymatic active sites (indicated by red ▲) and 3 substrate binding sites (indicated by black ▲). The conserved motifs containing the sites were aligned with the corresponding residues numbered at the bottom and the numbers of non - conserved residues were given in brackets (B)

keeping gene β-actin (Fig. 4B). RT-PCR amplification result showed HC-SPH gene expression in tissues of 2 years old mussels examined, including the digestive gland, stomach, intestine, gills, heart, mantle, axe foot, adductor muscle, kidney and gonads, by the gene-specific primers (Fig. 4B). Semi-quantitative analysis of the amplicons separated on agarose gel by comparing to the β-actin products revealed that the digestive gland had a strong expression while the gonads were seen as weak expression sites.

In the time course RT-qPCR analysis of gene expression in five selected tissues at hours post challenge (hpc), the HC-SPH expression in digestive gland was slightly increased at 6 hpc followed by a decrease at 12 hpc detected, then the expression kept rising to a highest ($p < 0.01$) level till 48 hpc (Fig. 5A). The HC-SPH gene expression in kidney (Fig. 5E), stomach (Fig. 5B) and the gills (Fig. 5D) rather fluctuated with a highest expression detected at 6 hpc for kidney, 12 hpc for stomach and 24 hpc for the gills. While kidney expressed basal level of HC-SPH after a surge at 6 hpc, the expression in the stomach and gills showed complementary expression pattern (Fig. 5). The intestine showed a constant low levels of HC-SPH expression (although expression was significantly high at 3 hpc ($p < 0.01$) than that at 0 hpc), particularly between 6 hpc and 24 hpc at a significantly lower level ($p < 0.01$) than 0 hpc, before an expression surge starting from 24 hpc and reach a peak ($p < 0.01$) at 48 hpc (Fig. 5C).

Discussion

HC-SPH*, a secreted trypsin-like SP in* H. cumingii

SPs are enzymes widely conserved across species not only in their molecular structures but also in their biological activities (Ross *et al.*, 2003). A cDNA library was established using suppression subtraction hybridization achieves an EST tag for SP gene of the *H. cumingii* in our previous study (Xiao *et al.*, 2009), which is similar with the genes encoding the factors responded to immunogenic stimulation (Lund and Olafsen, 1999; Gerwick *et al.*, 2000,2002). In the present study, the information of the EST tag was used for identify and clone the full-length cDNA, which contends an ORF coding for HC-SPH. The UniProtKB BLAST analysis against databases

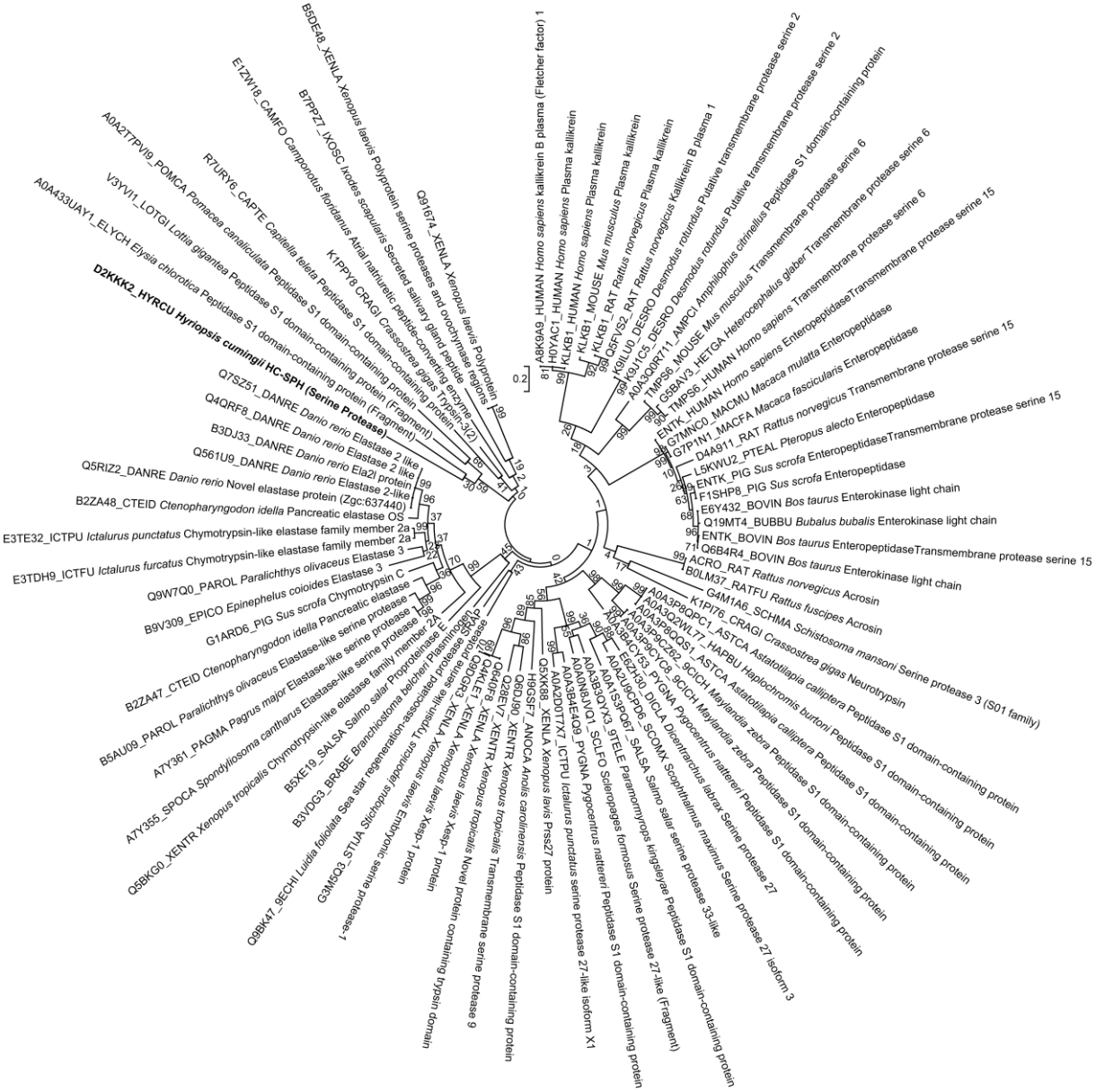

Fig. 3 The phylogenetic analysis of HC-SPH. The phylogenetic analysis showing the position of HC-SPH as a member of serine proteases of S1 superfamily (brunch was highlighted with **bold**) and protein was clustered into the group of with trypsin-like proteins

revealed HC-SPH exhibited a typical molecular characteristic of a SP.

A well-conserved Tryp_SPc domain of the SP family proteins was found in HC-SPH by functional search, including enzymatic active sites for its processing and activities along the molecule (Fig. 2A). The molecule was recognized as an inactive trypsin-like SP precursor. A cleavage site around I_{37} (Fig. 2B) was found in the precursor that are the characteristics of enzymatic activities and also known as catalytic triad for SPs (Perona and Craik, 1995), and 3 substrate binding sites at around E_{226}, S_{252} and 3 S_{254} (Fig. 2B) of the molecule, which bind to the substrates for the function. These conserved motifs are ensuring the actions (Li *et al.*, 2012a). However, key residues of catalytic triads were replaced by other amino acids (Q_{82} for H_{82}, L_{232} for S_{232}), similar to SPs in amino acid sequence, are classified as protease precursors (Liu *et al.*, 2010). Furthermore, the numbers of the amino acid residuals in non - conserved regions were very close between species (Fig. 2B), which were significant in ensuring similar molecular conformations essential for their function. SPHs assist SPs as cofactors to exert their functions (Wang and Jiang, 2004) by activating/inhibiting the original protease activities, including cell differentiation, tissue remodeling, embryonic development and humoral immune responses in the cell (Rawlings and Barrett, 1993; Krem and Di Cera, 2001, 2002).

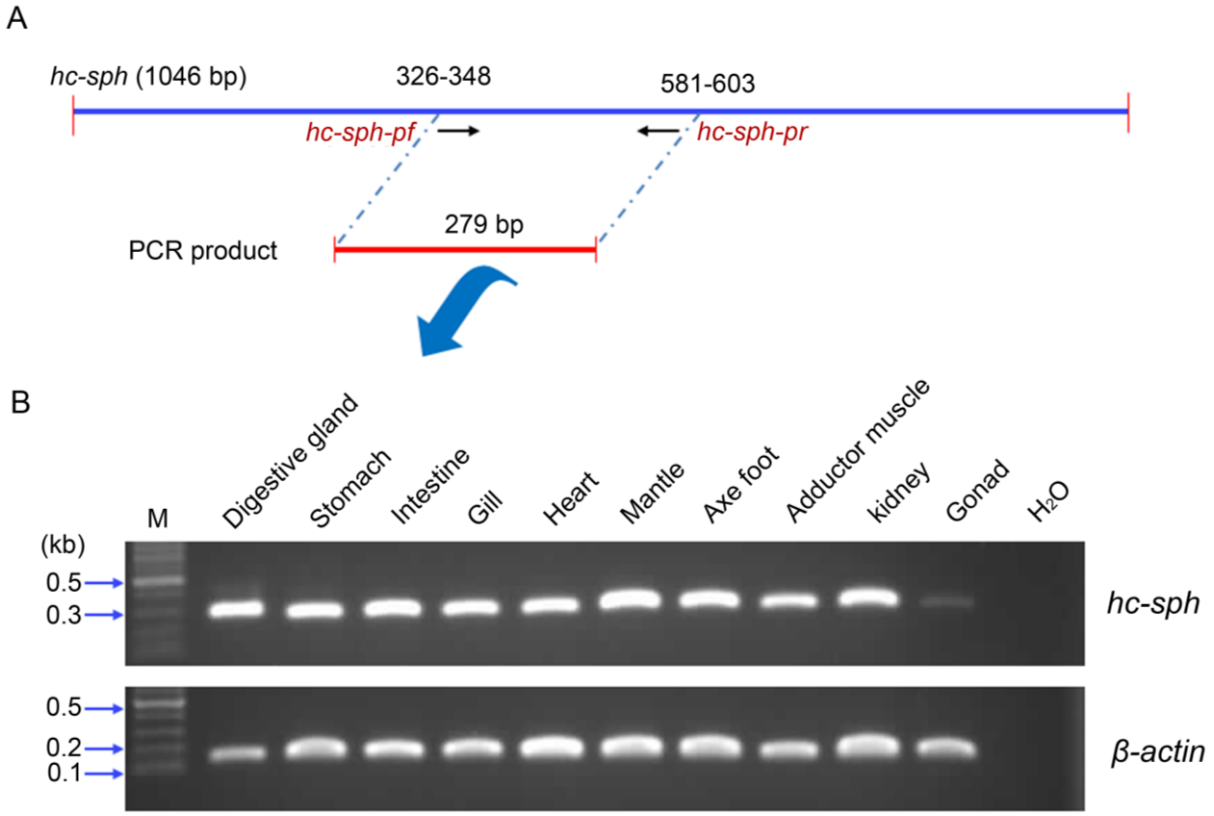

Fig. 4 The HC-SPH gene expression in tissues of healthy *Hyriopsis cumingii*. A pair of gene specific primers were designed to amplify a fragment of 279 bp between the nucleotide position 326 and 603 (A). The RT-PCR products were confirmed by agarose gel electrophoresis showing the right size (top, B) and the relative template concentrations were judged by the house - keeping gene β-actin (bottom, B)

Signal peptide at the amino terminus of a nascent protein, mediating protein targeting to the membrane of the endoplasmic reticulum (Blobel, 1980, 2000), directs the protein transportation and the secretory pathway, which is recognized by the signal recognition particle (SRP) and cleaved by the signal peptidase (Duffaud *et al.*, 1985). These residuals are comprising a characteristic tripartite structure of a hydrophilic, usually positively charged n - region, a central hydrophobic h - region and a cleavage site (von Heijne, 1985), which are represented by C-score (raw cleavage site score), S-score (signal peptide score) and Y-score (combined cleavage site score) in the SignalP prediction (Petersen *et al.*, 2011). The HC-SPH protein sequence was analyzed by software and a clear signal peptide region of 31 amino acids (the translation initiation starts at the M_{17}, see Fig. 1C) with a cleavage site (D valve = 0.829 by setting - up D-cutoff 0.450) between C_{31} and Q_{32} of the short stretch of VLC-QL recognized (Fig. 2A). The high raw C-score recognized this region as a signal peptide and enhanced by the Y-score of combined cleavage site, which compares the C-score peak(s) against the steep slope of the signal peptide score (S-score). Signal peptides are rather heterogeneous and interchangeable between different species, but the signal peptide specifies the protein secretion (von Heijne, 1985; Kober *et al.*, 2013). HC-SPH was therefore regarded as a secreted trypsin-like SP.

In addition, the HC-SPH was very similar to the complement factor D (CFAD_HUMAN, P00746), which is also single - domain SP. The factor D is a component of the alternative complement pathway. Recent studies show that an ancient proto - complement complex (C3, SP factor B and complement receptor CR) exists in bivalves (Gorbushin, 2018, 2019). A complement -like activity is shown in bivalves *Mytilus edulis*, *Mya Arenaria*, and *Sinonovacula constricta* (Klimovich and Gorbushin, 2017; Niu *et al.*, 2018). However, homolog of the factor D cleaving the factor B to Bb and Ba is not described until now (Gerdol *et al.*, 2018). Our results implied that the HC-SPH was a potential candidate for this role, although it was still need to further confirmation.

HC-SPH *is a member of the peptidases S1 family*

Hundreds of sequences producing significant alignments with low expect values (E values) were found by both BLASTP (Altschul *et al.*, 2005) and UniProtKB (Apweiler *et al.*, 2012) searches. They share similar enzymatic functions and classified as proteins of the peptidases S1 family (Goldman *et al.*, 2006), the largest of all of the peptidase families, by both the number of sequenced proteins and the

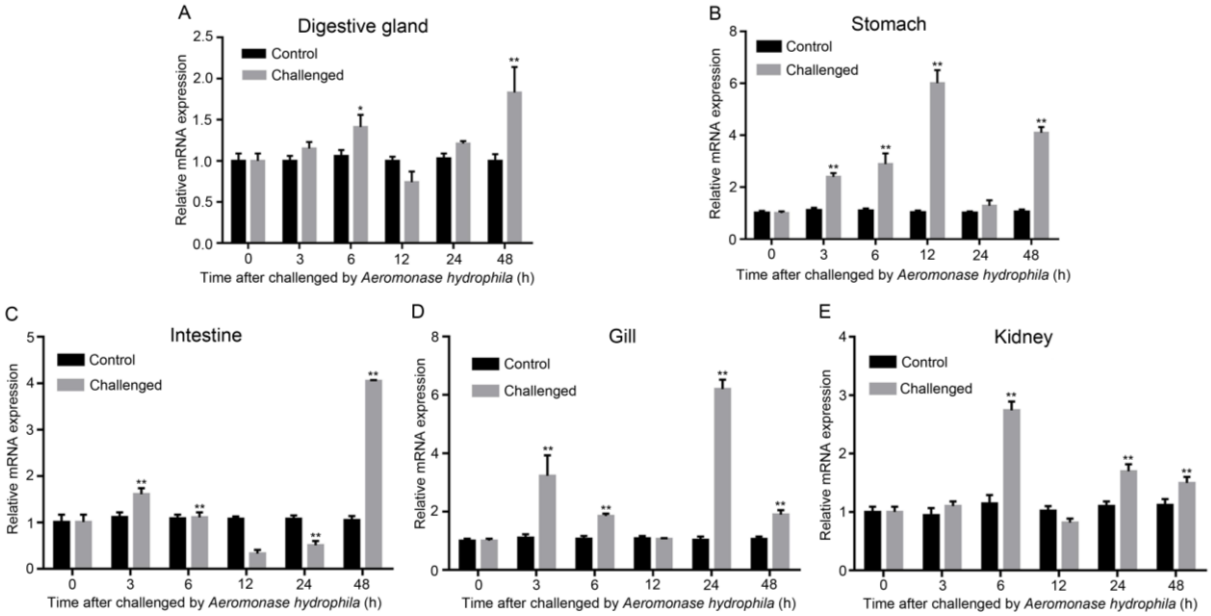

Fig 5 The HC-SPH gene expression profiles in tissues of mussels challenged by *A. hydrophila* infection. The transcripts were quantified using RT-qPCR in time-course manner and indicated by hours post challenge (hpc). The amplicons detected by qPCR in digestive gland (A), stomach (B), intestine (C), gill (D) and kidney (E) were normalized by control value (0 h) and adjusted to the scale for summarizing the result. The relative abundancy of the transcripts in a tissue at a time of hpc was compared with the control (0 h) to rule out the difference statistically. The up- and down-regulated expression levels of significant difference ($p < 0.05$, N=3) were indicated by single asterisk (*), and of highly significance ($p < 0.01$, N=3) by double asterisks (**)

number of distinct peptidase activities (serine endopeptidases) (Rawlings and Barrett, 1993) fell into three main activity types: trypsin-like, chymotrypsin-like and elastase-like SPs (Cooley *et al.*, 2001; Ovaere *et al.*, 2009; Madala *et al.*, 2010). The peptidases of the S1 family proteins are secretory precursors led by an N - terminal signal peptide that is cleaved to form the active enzyme. The cleavage initiates a structural rearrangement to produce a “new” N - terminus for binding. The phylogenetic tree of the S1 peptidase (Fig. 3) clearly showed that the HC-SPH was clustered to a group of trypsin-like type SPs roughly flanked by a group of elastase-like SPs at the bottom part and a large group of chymotrypsin-like proteases on top.

Gene expression in tissues regulated by bacterial infection suggested temporal and spatial active roles of HC-SPH *in immune - response*

Immune responses to infections are a series of cellular and molecular actions in invertebrate species (Peteiro *et al.*, 2007). Understanding of the reaction process is particularly important in molluscs because their susceptibility to microbial infection and transmitting diseases (Barcia and Ramos-Martinez, 2008) due to the lack of an adaptive immune system (Wang *et al.*, 2013). The mollusk defenses through a chemicophysical barriers preventing host invasion and the circulating hemocytes and secreted bio - factors to initiate immune responses (Hine, 1999; Garcia-Garcia *et al.*, 2008), including the activation of phenoloxidase system (Little *et al.*, 2005) and the complement system capable of recognizing and eliminating invading pathogens (Kardos *et al.*, 2008). The SPs are the central components of the complement system, specifically bind to and cleave the substrates in forms of super - molecular complexes with other SPs and non - enzymatic proteins (Harmat *et al.*, 2004). The expression patterns of SP in tissues would be altered to activate immune response stimulated by infection. The gene expression in physiological conditions examined by RT-PCR in this study showed the presence of HC-SPH gene transcripts in all detected tissues (Fig. 4B). In the time course RT-qPCR analysis of gene expression in five selected tissues of *A. hydrophila* challenged mussels showed that the trends of the HC-SPH expression were similar to the CfCUBSP in *Chlamys farreri* which indicated CfCUBSP might be involved in immune response (Yang *et al.*, 2017). These results implied that a temporal and spatial reinforce model for HC-SPH to response to foreign invaders that yet to be proved experimentally. The HC-SPH gene expressed at high levels in all secreting tissues at 48 hpc onwards when an immune acting system recruited and built up. The relatively low level of expression after the challenging and mechanisms of down - regulation before 24 hpc in the intestine remain to be further investigated.

In summary, the full - length cDNA coding for the SPH of the S1 family protein was cloned from *H. cumingii*. HC-SPH possessed a typical functional domain of a Tryp_SPc with characteristics of signal

peptide cleavage, active and substrate binding sites of the protease, which was further classified as a member of trypsin-like SP of S1 superfamily. Under the physiological conditions, the gene expression remained high in the digestive gland, but the expression showed a temporal complementary pattern of high levels in the digestive gland, kidney, stomach and low levels in the gills upon the regulation of bacterial infection. All tissues reached high level of the gene expression at 48 hpc to accomplish the recruiting of functional immune response. The molecular pathways in which the gene expression is regulated and the enzyme interact with its receptors to activate the immune-reactive network are to be further studied.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (No.31402289 and No.31040083) and the Science Foundation of Hunan Agricultural University(13YJ04). The authors are grateful to Dr. Shuliang Cui, Australian Phenomics Facility, John Curtin School of Medical Research, Australian National University, for reading and revising the first drafted manuscript.

References

Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schaffer AA, *et al*. Protein database searches using compositionally adjusted substitution matrices. FEBS J. 272: 5101-5109, 2005.

Apweiler R, Jesus Martin M, O'onovan C, Magrane M, Alam-Farugue Y. Reorganizing the protein space at the Universal Protein Resource (UniProt). Nucleic Acids Res. 40: D71-D75, 2012.

Barcia R, Ramos-Martinez JI. Effects of interleukin-2 on nitric oxide production in molluscan innate immunity. Invert. Surviv. J. 5: 43-49, 2008.

Blobel G. Intracellular protein topogenesis. Proc. Natl. Acad. Sci. U.S.A. 77: 1496-1500, 1980.

Blobel G. Protein targeting (Nobel lecture). Chembiochem. 1: 86-102, 2000.

Chang MX, Nie P, Wei LL. Short and long peptidoglycan recognition proteins (PGRPs) in zebrafish, with findings of multiple PGRP homologs in teleost fish. Mol. Immunol. 44: 3005-3023, 2007.

Charoensapsri W, Amparyup P, Hirono I, Aoki T, Tassanakajon A. Gene silencing of a prophenoloxidase activating enzyme in the shrimp, *Penaeus monodon*, increases susceptibility to *Vibrio harveyi* infection. Dev. Comp. Immunol. 33: 811-820, 2009.

Cooley J, Takayama TK, Shapiro SD, Schechter NM, Remold-O'Donnell E. The serpin MNEI inhibits elastase-like and chymotrypsin-like serine proteases through efficient reactions at two active sites. Biochemistry. 40: 15762-15770, 2001.

Cui Z, Liu Y, Wu D, Luan W, Wang S, Li Q, *et al*. Molecular cloning and characterization of a serine proteinase homolog prophenoloxidase-activating factor in the swimming crab *Portunus trituberculatus*. Fish Shellfish Immun. 29: 679-686, 2010.

Duffaud GD, Lehnhardt SK, E. MP, Inouye M. Structure and Function of the signal peptide. In: Knauf PA, Cook JS (eds). Membrane Protein Biosynthesis and Turnover. Academic Press, Florida, USA, pp. 65-104, 1985.

Garcia-Garcia E, Prado-Alvarez M, Novoa B, Figueras A, Rosales C. Immune responses of mussel hemocyte subpopulations are differentially regulated by enzymes of the PI 3-K, PKC, and ERK kinase families. Dev. Comp. Immunol. 32: 637-653, 2008.

Gerdol M, Gomez-Chiarri M, Castillo MG, Figueras A, Fiorito G, Moreira R, et al. Immunity in molluscs: recognition and effector mechanisms, with a focus on bivalvia. Advances in Comparative Immunology, Springer, 225-341, 2018.

Gerwick L, Reynolds WS, Bayne CJ. A precerebellin -like protein is part of the acute phase response in rainbow trout, *Oncorhynchus mykiss*. Dev. Comp. Immunol. 24: 597-607, 2000.

Gerwick L, Steinhauer R, Lapatra S, Sandell T, Ortuno J, Hajiseyedjavadi N, *et al*. The acute phase response of rainbow trout (*Oncorhynchus mykiss*) plasma proteins to viral, bacterial and fungal inflammatory agents. Fish Shellfish Immun. 12: 229-242, 2002.

Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen JA, *et al*. Evolution of sensory complexity recorded in a myxobacterial genome. Proc. Natl. Acad. Sci. U.S.A. 103: 15200-15205, 2006.

Gorbushin AM. Immune repertoire in the transcriptome of Littorina littorea reveals new trends in lophotrochozoan proto-complement evolution. Dev. Comp. Immunol. 84: 250-263, 2018.

Gorbushin AM. Derivatives of the lectin complement pathway in Lophotrochozoa. Dev. Comp. Immunol. 94: 35-58, 2019.

Harmat V, Gál P, Kardos J, Szilágyi K, Ambrus G, Végh B, *et al*. The structure of MBL-associated serine protease-2 reveals that identical substrate specificites of C1s and MASP-2 are realized through different sets of enzyme-substrate interactions. J. Mol. Biol. 342: 1533-1546, 2004.

Hine PM. The inter-relationships of bivalve haemocytes. Fish Shellfish Immun. 9: 367-385, 1999.

Huang TS, Wang H, Lee SY, Johansson MW, Soderhall K, Cerenius L. A cell adhesion protein from the crayfish *Pacifastacus leniusculus*, a serine proteinase homologue similar to *Drosophila masquerade*. J. Biol. Chem. 275: 9996-10001, 2000.

Jiang H, Liu S, Xiao TY, Xie M. Body color regulation of *Leptobotia taeniops* through tyrosinase gene expression. Appl. Ecol. Environ. Res. 17: 5525-5537, 2019.

Jitvaropas R, Amparyup P, Gross PS, Tassanakajon A. Functional characterization of a masquerade -like serine proteinase homologue from the black tiger shrimp *Penaeus monodon*. Comp. Biochem. Physiol. B. 153: 236-243, 2009.

Kardos J, Harmat V, Pallo A, Barabas O, Szilagyi K, Graf L, *et al*. Revisiting the mechanism of the autoactivation of the complement protease C1r

in the C1 complex: structure of the active catalytic region of C1r. Mol. Immun. 45: 1752-1760, 2008.
Kawabata S, Tokunaga F, Kugi Y, Motoyama S, Miura Y, Hirata M, *et al*. Limulus factor D, a 43-kDa protein isolated from horseshoe crab hemocytes, is a serine protease homologue with antimicrobial activity. FEBS Lett. 398: 146-150, 1996.
Klimovich AV, Gorbushin AM. Evolutionary radiation of cytotoxic effects in bivalve and gastropod haemolymph. J. Evol. Biochem. Physiol. 53: 470-479, 2017.
Kober L, Zehe C, Bode J. Optimized signal peptides for the development of high expressing CHO cell lines. Biotechnol. Bioeng. 110: 1164-1173, 2013.
Krem MM, Di Cera E. Molecular markers of serine protease evolution. EMBO J. 20: 3036-3045, 2001.
Krem MM, Di Cera E. Evolution of enzyme cascades from embryonic development to blood coagulation. Trends Biochem. Sci. 27: 67-74, 2002.
Kwon TH, Kim MS, Choi HW, Joo CH, Cho MY, Lee BL. A masquerade -like serine proteinase homologue is necessary for phenoloxidase activity in the coleopteran insect, *Holotrichia diomphalia* larvae. Eur. J. Biochem. 267: 6188-6196, 2000.
Lee KS, Kim BY, Choo YM, Jin BR. Dual role of the serine protease homolog BmSPH-1 in the development and immunity of the silkworm *Bombyx mori*. Dev. Comp. Immunol. 85: 170-176, 2018.
Li Q, Cui Z, Liu Y, Wang S, Song C. Identification and characterization of two novel types of non-clip domain serine proteases (PtSP and PtSPH1) from cDNA haemocytes library of swimming crab *Portunus trituberculatus*. Fish Shellfish Immun. 32: 683-692, 2012a.
Li Q, Cui Z, Liu Y, Wang S, Song C. Three clip domain serine proteases (cSPs) and one clip domain serine protease homologue (cSPH) identified from haemocytes and eyestalk cDNA libraries of swimming crab *Portunus trituberculatus*. Fish Shellfish Immun. 32: 565-571, 2012b.
Lin CY, Hu KY, Ho SH, Song YL. Cloning and characterization of a shrimp clip domain serine protease homolog (c-SPH) as a cell adhesion molecule. Dev. Comp. Immunol. 30: 1132-1144, 2006.
Little TJ, Hultmark D, Read AF. Invertebrate immunity and the limits of mechanistic immunology. Nat. Immunol. 6: 651-654. 2005.
Liu H, Liu Y, Song C, Ning J, Cui Z. Functional characterization of two clip-domain serine proteases in the swimming crab *Portunus trituberculatus*. Fish Shellfish Immun. 89: 98-107. 2019.
Liu HP, Chen RY, Zhang M, Wang KJ. Isolation, gene cloning and expression profile of a pathogen recognition protein: a serine proteinase homolog (Sp-SPH) involved in the antibacterial response in the crab *Scylla paramamosain*. Dev. Comp. Immunol. 34: 741-748, 2010.
Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods. 25: 402-408, 2001.
Lund V, Olafsen JA. Changes in serum concentration of a serum amyloid P -like pentraxin in Atlantic salmon, *Salmo salar* L., during infection and inflammation. Dev. Comp. Immunol. 23: 61-70, 1999.
Madala PK, Tyndall JD, Nall T, Fairlie DP. Update 1 of: Proteases universally recognize beta strands in their active sites. Chem. Rev. 110: PR1-31, 2010.
Niu D, Xiong Y, Peng M, Meng X, Lan T, Li J. Hemolytic reactions in the hemolymph of bivalve *Sinonovacula constricta* show complement -like activity. Fish Shellfish Immun. 79: 11-17, 2018.
Ovaere P, Lippens S, Vandenabeels P, Declercq W. The emerging roles of serine protease cascades in the epidermis. Trends Biochem. Sci. 34: 453-463, 2009.
Perona JJ, Craik CS. Structural basis of substrate specificity in the serine proteases. Protein Sci. 4: 337-360, 1995.
Peteiro LG, Labarta U, Fernandez-Reiriz MJ. Variability in biochemical components of the mussel (*Mytilus galloprovincialis*) cultured after Prestige oil spill. Comp. Biochem. Physiol. Toxicol. Pharmacol. 145: 588-594, 2007.
Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Meth. 8: 785-786, 2011.
Rawlings ND, Barrett AJ. Evolutionary families of peptidases. Biochem. J. 290: 205-218, 1993.
Ren Q, Qi YL, Hui KM, Zhang Z, Zhang CY, Wang W. Four invertebrate-type lysozyme genes from triangle-shell pearl mussel (*Hyriopsis cumingii*). Fish Shellfish Immun. 33: 909-915, 2012.
Ren Q, Xu ZL, Wang XW, Zhao XF, Wang JX. Clip domain serine protease and its homolog respond to Vibrio challenge in Chinese white shrimp, *Fenneropenaeus chinensis*. Fish Shellfish Immun. 26: 787-798, 2009.
Ren Q, Zhao X-F, Wang J-X. Identification of three different types of serine proteases (one SP and two SPHs) in Chinese white shrimp. Fish Shellfish Immun. 30:456-466, 2011.
Ross J, Jiang H, Kanost MR, Wang Y. Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. Gene. 304:117-131, 2003.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876-4882, 1997.
von Heijne G. Signal sequences. The limits of variation. J. Mol. Biol. 184: 99-105, 1985.
Wang H, Liang J, Zhao Y, Liu Q, Li Y, Yi Z, *et al*. Molecular cloning, characterization and expression analysis of trypsin-like serine protease from triangle-shell pearl mussel (*Hyriopsis cumingii*). Fish Shellfish Immun. 40: 603-608, 2014.

Wang L, Qiu L, Zhou Z, Song L. Research progress on the mollusc immunity in China. Dev. Comp. Immunol. 39: 2-10, 2013.

Wang L, Zhang H, Wang L, Zhang D, Lv Z, Liu Z, et al. The RNA-seq analysis suggests a potential multicomponent complement system in oyster *Crassostrea gigas*. Dev. Comp. Immun. 76: 209-219, 2017.

Wang Y, Jiang H. Purification and characterization of Manduca sexta serpin-6: a serine proteinase inhibitor that selectively inhibits prophenoloxidase-activating proteinase-3. Insect Biochem. Mol. Biol. 34: 387-395, 2004.

Wei Z, Wei Z, Sun W, Tran NT, Gong Y, Ma H, *et al*. Two novel serine proteases from Scylla paramamosain involved in the synthesis of anti-lipopolysaccharide factors and activation of prophenoloxidase system. Fish Shellfish Immun. 84: 322-332, 2019.

Xiao TY, Ge XK, Xu BH, Su JM, Zhang HY. Construction and analysis of subtracted cDNA library by suppression subtractive hybridization from *Hyriopsis cumingii* digestive gland. J. Fish. China. 33: 856-864, 2009.

Yang C, Wang L, Zhang H, Yi Q, Wang L, Wang H, *et al*. The first CUB-domain containing serine protease from *Chlamys farreri* which might be involved in larval development and immune response. Dev. Comp. Immunol. 76: 163-168, 2017.