Acta Oceanol. Sin., 2017, Vol. 36, No. 6, P. 52–60
DOI: 10.1007/s13131-017-0983-5
http://www.hyxb.org.cn
E-mail: hyxbe@263.net

Identification and molecular characterization of *Cathepsin L* gene and its expression analysis during early ontogenetic development of kuruma shrimp *Marsupenaeus japonicus*

QIAO Ying[1], WANG Jun[1], MAO Yong[1], LIU Min[1], SONG Xiaohong[1], SU Yongquan[1]*, WANG Chunzhong[2], ZHENG Zhipeng[2]

[1] State Key Laboratory of Marine Environmental Sciences, Xiamen University, Xiamen 361102, China

[2] Putian Tian-ran-xing Agricultural Development Co. Ltd., Fujian 351100, China

Received 16 February 2016; accepted 29 March 2016

Abstract

Cathepsin L gene is a member of the cysteine proteinase gene group. In this study *Cathepsin L* gene was isolated from Kuruma shrimp *Marsupenaeus japonicus* (*Mj-Cathepsin L*) and the full-length DNA sequence was 1 963 bp. *Mj-Cathepsin L* protein showed high homologies with other *Cathepsin L* proteins documented in vertebrates, mollusks and other crustaceans. Expression analysis of *Mj-Cathepsin L* gene in different tissues revealed that it was predominant in hepatopancreas. During early ontogenetic development stages *Mj-Cathepsin L* showed a development-regulated expression, and the *Mj-Cathepsin L* showed a molting stage-regulated expression during the five molting stages, inferring its role in the ontogenic development of *M. japonicus*. Two kinds of forms of *Mj-Cathepsin L* protein: pro-*Cathepsin L* and *Cathepsin L* were measured in hepatopancreas, stomach and intestine by Western Blotting.

Key words: *Cathepsin L* gene, larval development, molting cycle, tissue distribution, *Marsupenaeus japonicus*

Citation: Qiao Ying, Wang Jun, Mao Yong, Liu Min, Song Xiaohong, Su Yongquan, Wang Chunzhong, Zheng Zhipeng. 2017. Identification and molecular characterization of *Cathepsin L* gene and its expression analysis during early ontogenetic development of kuruma shrimp *Marsupenaeus japonicus*. Acta Oceanologica Sinica, 36(6): 52–60, doi: 10.1007/s13131-017-0983-5

1 Introduction

Cathepsin L, a lysosomal cysteine proteinase belongs to the papain superfamily, is an abundant endopeptidase in eukaryotic organisms (Roth et al., 2000). Studies revealed that *Cathepsin L* plays important roles in crucial biological processes such as carcinogenesis process, antigen presentation, exopeptidase maturation and embryogenesis (Nakagawa et al., 1998; Dahl et al., 2001; Miyamoto et al., 2011; Zhang et al., 2014). In crustaceans, *Cathepsin L* was first purified from gastrointestinal juice of the American lobster *Homarus americanus* (Laycock, 1989). To date, *Cathepsin L* has been reported in shrimps, such as *Litopenaeus* (*Penaeus*) *vannamei*, *Macrobrachium rosenbergii*, *Metapenaeus ensis*, *Nephrops norvegicus*, *Pandalus borealis* and *Penaeus monodon* (Le Boulay et al., 1995, 1996, 1998; Aoki et al., 2004; Hu and Leung, 2004; Glenn et al., 2005; Arockiaraj et al., 2013).

Studies on *Cathepsin L* in shrimps focused on its functions during the food digestion and the molting process. In *M. ensis*, *Cathepsin L* was found in the B cells of the hepatopancreas, the most important organ responsible for the immune, enzyme secretion and digestion (Hu and Leung, 2004). During the digestion-related cell differentiation in *M. ensis*, however, *Cathepsin L* mRNA was found in the F-cell rather than in the mature B-cell, indicated that the *Cathepsin L* digestive model during the rapid digestion-related cell differentiation in hepatopancreas (Hu and Leung, 2007). In *L. vannamei*, *Cathepsin L* transcript levels were correlated with the molting stages (Le Boulay et al., 1996). Though the *Cathepsin L* was found nothing significantly associated with body weight in *L. vannamei* populations (Glenn et al., 2005), but Qian et al. (2013) suggest *Cathepsin L* may be one of candidates which significantly affect the growth of *L. vannamei* and the distribution of *Cathepsin L* mRNA may be related to its specific functions in different tissues and its positive roles in regulation of shrimp muscle growth.

Kuruma shrimp *Marsupenaeus* (*Penaeus*) *japonicus* is a commercial important fishery and mariculture species in China. The objectives of this study were to analyze the molecular characters of *Cathepsin L* and to evaluate the role of *Cathepsin L* throughout the early ontogenetic developmental stages of *M. japonicus* for the first time, giving its significant changes on morphology, physiology and behavior.

2 Materials and methods

2.1 *Sample collection of Marsupenaeus japonicus*

To clone and analyze the tissue distribution of *Mj-Cathepsin L* gene, samples of *M. japonicus* (n=30, BW±SD, (21.0±2.54) g) were obtained from Dongshan (Fujian Province, China), and kept a concrete water tank with aeration for at least 7 d. Seven tissues (muscle, gill, heart, intestine, stomach, hepatopancreas and eye stalk) were collected, preserved immediately in Sample Protect-

Foundation item: The National High-tech R&D Program of China (863 Program) under contract No. 2012AA10A409-03; the Project of China Agriculture Research System under contract No. CARS-47; the Project of Xiamen Southern Ocean Research Center under contract No. 14CZY033HJ07; China Spark Program under contract No. 2015GA720002.
*Corresponding author, E-mail: yqsu@xmu.edu.cn

or for RNA (TaKaRa, Japan) and stored at –20°C for further RNA extraction.

Fertilized eggs of *M. japonicus* were collected from one broodstock and subsequently reared in a 5×5 m concrete tank following larval production procedures. The larvae were fed every four hours with prawn slices before post-larva 10 (P10) then switched to pellet feed. 0.2 g of fertilized egg, nauplius larva, zoeal larva (Z_1 to Z_3), mysis larva (M_1 to M_3) and post-larva (P_1–P_5, P_8, P_{11}, P_{14}, P_{17}, P_{21}, P_{27}) were collected and stored (as described above) for further analyses.

Determination of molt stage was based on the degree of setae development according to the method described by Oliveira Cesar (de Oliveira Cesar et al., 2006). Hepatopancreas were collected and preserved as mentioned above.

2.2 *RNA isolation and cDNA synthesis*

Total RNA was extracted from each 100 mg of hepatopancreas, intestine, muscle, heart, eye stalk, stomach and gill using the RNAiso Plus Reagent (TaKaRa, Code No. 9108, Japan) according to the manufacturer's protocol. The concentration and quality of the total RNA were measured by NanoDrop 1000 Spectrophotometer V3.7 (Thermo Scientific) and 1% agarose-gel electrophoresis, respectively. Only RNAs with absorbance ratios (A260:A280) greater than 1.8 were used for further analyses.

Total RNA (1 μg) was reverse-transcribed in a 25 μL reverse transcription reaction using the TaKaRa RNA PCR Kit (AMV) Ver.3.0 (TaKaRa, Code No. RR019A, Japan) according to the manufacturer's protocol for cDNA cloning.

2.3 *Mj-Cathepsin L DNA full-length amplification and sequencing*

The intermediate fragment of *Mj-Cathepsin L* was obtained via screening the hepatopancreas transcriptome of the *Marsupenaeus japonicus* which constructed by our laboratory. In order to confirm the intermediate fragment of *Mj-Cathepsin L* gene, a pair of primers *Cathepsin L*-F (5′-AGCAATGGCACAACTTCAAGGCTG-3′) and *Cathepsin L*-R (5′-CTACGGAAAAGATACAGCAAAGGCA-3′) was designed using PrimerPrimer 5.0 software. PCR products were visualized on 1.5% agarose-gel using Gene Finder™ (Zeeshan, China) staining. Amplicon was cloned directly into the pMD®19-T Vector Cloning Vector (TaKaRa, Code No. D102A, Japan) according to the manufacturer's instructions and subsequently transformed to DH5α bacterial competent cells. The plasmid DNA was isolated using E.Z.N.A.® Plasmid Mini Kit I (OMEGA Bio-Tek, Cat. No. D6942-01, USA) then used for the PCR assay to screen the positive colony before sending for commercial sequencing (Sangon Biotech Shanghai Co., China) to confirm its identity.

For the full-length amplification of *Mj-Cathepsin L* gene, rapid amplification of cDNA ends assays (3′-RACE and 5′-RACE) were employed. PCR amplification of the 3′ end was conducted using the specific primer CF3 (5′-GACGCCTCTCAACCTAGCCTCCAGT-3′), and the 3′ adaptor primer offered by the 3′-Full RACE Core Set (TaKaRa, Code No. 6106). In the PCR reaction, 1 μL CF3 and 3′ adaptor primers, and 0.6 μL cDNA were incubated with the Taq polymerase for 5 min at 95°C, followed by 32 cycles of 30 s at 95°C, 45 s at 68°C, 1 min at 72°C, and the final extension for 7 min at 72°C. The 5′ RACE Kit (2nd Generation, Roche) was employed for the 5′ cDNA preparation while the CR5 primer (5′-CCCGTCGTAGAGAAAGCCCAGCAGG-3′) was designed to amplify the complete 5′ sequence following to the manufacturer's guidelines. The screening and sequencing of PCR products followed the same methods above.

The genomic DNA was extracted from muscle using phenol-chloroform method. The specific primers genomeF (5′-GAAGTTCCTGTCAGTGTTGGCT-3′) and genomeR (5′-CACAGAACTCTAGACGAGCGGG-3′) were designed to amplify the *Mj-Cathepsin L* genomic DNA sequence based on the full-length *Mj-Cathepsin L* cDNA. The screening and sequencing of PCR products followed the same methods above.

2.4 *Sequence analysis*

Contig Express was employed to overlap all the confirmed sequences to get the full-length cDNA of *Mj-Cathepsin L*. The Open Reading Frame (ORF) was predicted using DNAMAN software (DNAMAN LynnonBiosoft, Santa Clara, CA, USA) and the signal peptide was predicted by the SignalP4.1 Server (http://www.cbs.dtu.dk/services/SignalP-3.0/). The propeptide and mature enzyme, catalytic C-H-N triad and the six Cys-disulfide-bridges were predicted via sequence alignment with homology sequences. The potential serine, threonine and tyrosine phosphorylation sites of *Mj-Cathepsin L* protein were predicted by NetPhos 2.0 Server (http://www.cbs.dtu.dk/services/NetPhos/).

Multiple *Cathepsin L* sequences from *M. japonicus*, *L. vannamei*, *M. ensis*, *P. camtschaticus*, *E. sinensis*, *H. americanus* CYSP2_HOMAM, *N. norvegicus* and *H. americanus* CYSP3_HOMAM, were aligned, and neighbor-joining (NJ) phylogenetic tree were performed using MEGA 6.0 program. The phylogenetic tree was tested by Bootstrap analysis of 1 000 replicates according to the literature method. Identity and similarity sequences were determined using the BLASTP algorithm with the MEGA version 6.0 stained with red colors by the ESPript 3.0 (http://espript. ibcp.fr/ESPript/ ESPript/index.php).

The genomic DNA sequence obtained was aligned with the cDNA to identify the exons and introns conforming to the splicing consensus GT-donor/AG-acceptor rule, and the genomic DNA sequences from other species were also consulted.

2.5 *Quantitative real-time PCR (qRT-PCR)*

Total RNA (1 μg) was reverse-transcribed using the PrimeScript™ RT Master Mix (Perfect Real Time) (TaKaRa, Japan) and stored at –20°C for the subsequent analysis. Two *Mj-Cathepsin L* gene-specific primers F (5′-TCTCAACCTAGCCTCCAGTTCTACC-3′) and R (5′-ATGCCGCAGTTGTTCTTCTTGTTGC-3′) were designed to amplify a product of 209 bp. In shrimp *glyceroldehyde-3 phosphate dehydrogenase* (GAPDH) and *Elongation factor 1-α* (*EF1-α*) were demonstrated to be the more stable genes than *β-actin* and *18S rRNA* (Dhar et al., 2009). Here the *Elongation factor 1-α* (*EF1-α*) was chosen to design a set of *EF1-α* primers Mj-*EF1-α*F (5′-GGAACTGGAGGCAGG ACC-3′) and Mj-*EF1-α*R (5′-AGCCACCGTTTG CTTCAT-3′) that approximately amplified a product of 158 bp, for the more similar amplification efficiency with the *Mj-Cathepsin L* gene.

The qRT-PCR amplifications were performed in triplicate using ABI 7500 Fast RT-PCR System (Applied Biosystems). Each reaction mixture had a total volume of 20 μL, which contained 10 μL of SYBR®PremixDimerEraserTM (2×), 0.4 μL ROX Reference Dye II (50×), 0.6 μL of each of the forward and reverse primers (10 mmol/L), 6.4 μL of PCR-grade water and 1 μL of the diluted cDNA (1:10). The 3-step PCR Standard Protocol was used which had an initial denaturation of 95°C for 30 s, followed by 40 cycles of 95°C for 5 s, 60°C for 30 s and 72°C for 30 s. Amplification efficiencies for *Mj-Cathepsin L* and *EF1-α* primers were determined via separate standard curves, and the specificities of the PCR amplification product were verified from the melting curves. The relative expression ratios of the target gene (*Mj-Cathepsin L*) versus the internal control gene in different tissues were calcu-

lated by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

2.6 *Prokaryotic expression of Mj-Cathepsin L recombinant protein*

The coding region of *Mj-Cathepsin L* was cloned into the prokaryotic expression ET-28a+ vector (Novegan, Germany) using the primers CTSLePF (5′-ATGGGTCGCGGATCCGAATTC ATGAAGTTCCTGTCAGTGTTGGCTT-3′) and CTSLePR (5′-CTCGAGTGCGGCCGCAAGCTTCACAGAACTCTAGACGAGCGGGTAC-3′) following the NovoRec PCR step directional cloning kit (Novoprotein, China). The recombinant *Mj-Cathepsin L*-pET-28a+ plasmid was transformed into *Escherichia coli* DH5α cells and screened the positive clones on Luria Broth (LB) culture medium supplied with 100 μg/mL of kanamycin. The positive clones were confirmed using sequencing (Sangon, China) and the confirmed plasmid was transformed into *E. coli* (BL21 DE3) pLysS cells for the expression of *Mj-Cathepsin L* recombinant protein. The log-phase bacterial cultures (5 mL) were added into the self-induced culture medium (500 mL) and incubated at 37°C with shaking at 150 r/min for 20 h. Before the induction, an aliquot of the cells was removed from the culture and cultured in the LB culture medium under the same condition for the none-induction control. The induction result was detected using 12% SDS-Polyacrylamide gel electrophoresis (SDS-PAGE).

2.7 *Purification of Mj-Cathepsin L recombinant protein*

The cultured cells were centrifuged at 5 000 r/min for 10 min at 4°C, and the supernatant was discarded. The precipitate was resuspended using the binding buffer with 0.2 mg/mL lysozyme, 20 μg/mL DNase, 1 mmol/L $MgCl_2$ and 1 mmol/L PMSF. The mixture was incubated at 22°C for 30 min and was ultrasonicated on ice in 5 s bursts, following 5 s rests between the bursts for approximately 90 cycles until the mixture was no longer viscous. After that, the mixture was centrifuged at 13 000 r/min for 20 min at 4°C and filtered using 0.45 μm micropore filter to discard the impurities for the subsequent operations.

The HisTrap™ HP column (1 mL) (GE Healthcare Life Sciences, Sweden) that loaded with Ni Sepharose High Performance Linking to the AKTA purifier 100 workstation (GE Healthcare) as an immobilized metal affinity chromatography system was used for the recombinant protein purification. At least 5 mL binding buffer (20 mmol/L sodium phosphate, 0.5 mol/L sodium chloride, 40 mmol/L imidazole, 6 mol/L urea, pH 7.4) was loaded onto the HisTrap™ HP column at the flow-rate of 1 mL/min for the equilibration of the column, then the sample was loaded at the flow-rate of 0.8 mL/min. The 5–10 mL binding buffer was used to wash the column at the flow-rate of 1 mL/min until the UV absorption baseline was reached. The recombinant protein was eluted using the elution buffer (20 mmol/L sodium phosphate, 0.5 mol/L sodium chloride, 500 mmol/L imidazole, 6 mol/L urea, pH7.4). The elution fractions were dialyzed in a linear gradient refolding buffer from 6 mol/L to 0 mol/L urea, finishing with the Milli-Q water. The purified *Mj-Cathepsin L* recombinant protein was frozen at –80°C for further studies.

For the obtain of Antiserum, 100 μg purified *Mj-Cathepsin L* recombinant protein was mixed with Freund's incomplete adjuvant (Sigma, St. Louis, MO) and injected three of eight weeks Kunming mice intra-peritoneally. After three additional boosters in the following three weeks (once a week) with 100 μg protein in Freund's complete adjuvant by the same route, the anti-sera was collected and preserved in –20°C until use.

2.8 *Western blot analysis*

Crude extracts from stomach, hepatopancreas and intestine of *M. japonicus* were prepared by homogenizing the tissues in PBS buffer, then centrifugated at 13 000 r/min to remove the residual debris. Crude protein quantification was performed using an Easy Protein Quantitative Kit (TransGen Biotech). Different tissue lysates were boiled in 2× loading buffer, run on the mini-protein II-system and were transferred to Immobilon-P polyvinylidene fluoride membranes (Millipore, Bedford, MA). The PVDF membrane was blocked in TBSTM (5% non-fat dried milk with 25 mmol/L Tris, pH 8.0, 125 mmol/L NaCl and 0.05% Tween 20 (v/v)) for 1 h at RT. Anti-*Mj-Cathepsin L* polyclonal mouse anti-sera at 1:1 000 and goat anti-mouse antibody conjugated to horseradish peroxidase (1:5 000) were used as antibody to incubate the PVDF membrane at 37°C for 1 h. Chemiluminescent detection was performed using the EasySee® Western Blot Kit (TransGen Biotech).

3 Results

3.1 *Mj-Cathepsin L full-length cDNA and DNA*

The full-length cDNA sequence of *Mj-Cathepsin L* (No. KJ871613) was obtained, with 1191 bp in length which encompassed a 5′-untranslated region (5′-UTR) of 15 bp, an ORF of 984 bp and a 3′-UTR region of 192 bp with a consensus downstream poly-(A) tail. The 984 bp of ORF region started with the initiation codon ATG on 16–18 base, ended with the terminal codon TGA on 997–999 base, and encoded a polypeptide of 327 amino acids (aa) with an estimated molecular mass of 35.9 kDa and the theoretical isoelectric point of the polypeptide is 5.34 (Fig. 1).

1 963 bp of *Mj-Cathepsin L* genomic DNA was obtained and the coding region consists of six exons separated by five introns. The interspecies comparison of the *Cathepsin L* orthologs in *Danio rerio*, *Homo sapiens*, *Mus musculus*, *Limpenaeus vannamei* and *Marsupenaeus japonicus* revealed that the sizes of exons and introns between the vertebrates ortholog and that of the arthropods are different (Fig. 2). The intron/exon distribution, exon sizes are highly conserved among *Penaeidae* family, except *Metapenaeus ensis*.

The confirmed *Mj-Cathepsin L* preproprotein contained a typical signal peptide sequence of 16 aa (Met_1–Ala_{16}), a propeptide of 93 aa (Ser_{17}–Thr_{109}) and mature enzyme domain of 217 aa (Leu_{110}–Val_{327}). Cys_{134}, His_{273} and Asn_{294} formed the predicted highly conserved catalytic triad and the six Cysteine forming conserved disulfide-bridges were located at Cys_{131}, Cys_{165}, Cys_{174}, Cys_{207}, Cys_{266} and Cys_{316} (Fig. 1). ERFNIN (Position 40–59), GNFD (Position 60–78) and GCNGG (Position 173–177) motifs for typical *Cathepsin L* proteases were marked above the alignment sequences with bold black capital letters (Fig. 3). The phosphorylation sites of *Cathepsin L* protein were at $Ser_{19, 37, 46, 81, 85, 142, 156, 158, 282}$, $Thr_{80, 123, 219}$ and Tyr43, 198, 289, 305 (Data not shown).

3.2 *Phylogenetic tree of Cathepsin L genes*

Multiple alignment of *Mj-Cathepsin L* with *Cathepsin L* amino acid sequences of other shrimps revealed the strong amino acid conservation in *Cathepsin L* proteins. Comparative analysis of the homology sequences revealed that the deduced amino acids sequence of *Mj-Cathepsin L* shared the highest identity with *L. vannamei* (90.5%), and 62.0%–79.8% identity to other shrimp *Cathepsin L* amino acids sequences (Table 1).

There were two large groups in the NJ tree of *Cathepsin L* proteins. One group consisted of two sub-groups, Crustacea and Arachnoidea, both belong to the Arthropoda (Fig. 4). *Marsupen-*

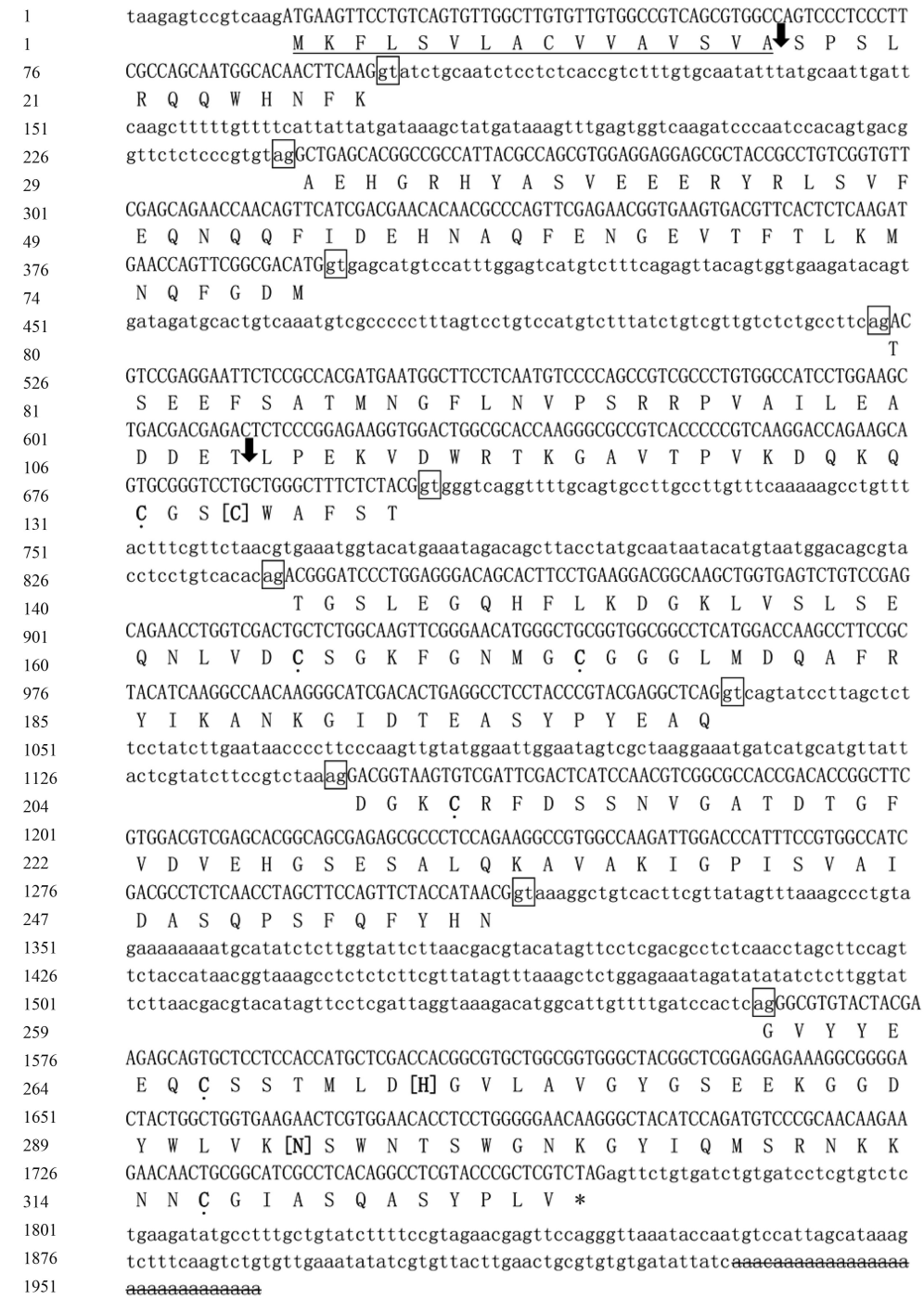

Fig. 1. Compiled genomic organization of the *Marsupenaeus japonicus Cathepsin L* gene. The open reading frame of *Mj-Cathepsin L* is marked with capital letters, the 15 bp 3′-UTR and 192 bp of the 5′-UTR are marked with lowercase letters. The putative signal peptide is underlined. Black down arrows indicate the cleavage site between signal peptide (SP) and propeptide, and between propeptide and mature enzyme. Black bold letters in the [ ] denote the catalytic C-H-N triad. The six Cys forming possible disulfide-bridges are marked with bold black and underlined with emphasis character. Stop condon is marked with *. Poly (A) sequence is marked with strikeout. The introns are in lowercase and intron splice sites GT-donor/AG-acceptor are boxed. Numbering of the nucleotide and the amino acid are given on the right-hand.

aeus japonicus was clustered with *L.vannamei* and formed the Crustacea sub-group with *N. norvegicus*, *H. americanus*, *M. ensis*, *P. camtschaticus* and *E. sinensis*. The two different parasite ticks *Rhipicephalus haemaphysaloides haemaphysaloides* and *Hyalomma anatolicum anatolicum* were clustered together forming the Arachnoidea sub-group. The other group is the Vertebrate sub-group, including *Salmo salar*, *Lates calcarifer* and *Homo sapiens*.

3.3 *Tissue distribution expression of Mj-Cathepsin L*

Mj-Cathepsin L mRNA was expressed significantly higher in hepatopancreas, followed by a moderate expression level in

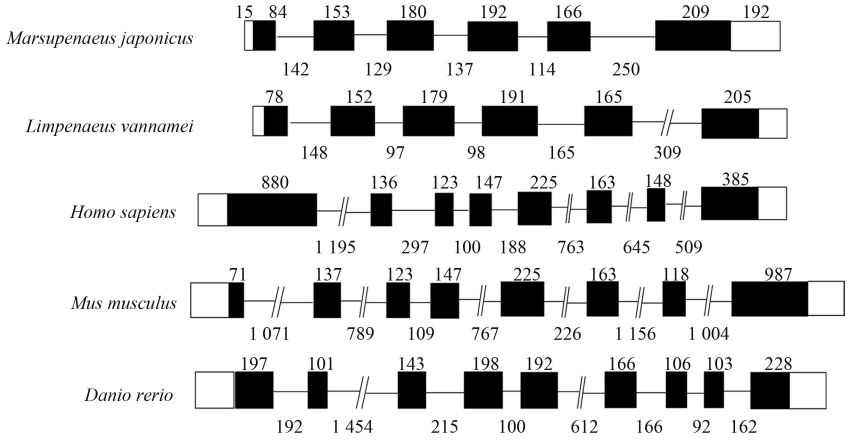

Fig. 2. Comparison of the gene structures of *Mj-Cathepsin L* genes with other species. The black boxes depict the coding regions, and the horizontal line represents the introns. The 5′ and 3′ UTR are shown as white boxes. The lengths of the elements are shown in nucleotides (bp). The genomic DNA sequences were obtained from the NCBI database (http://www.ncbi.nlm.nih.gov/) and ENSEMBL genomic database (http://www.ensembl.org).

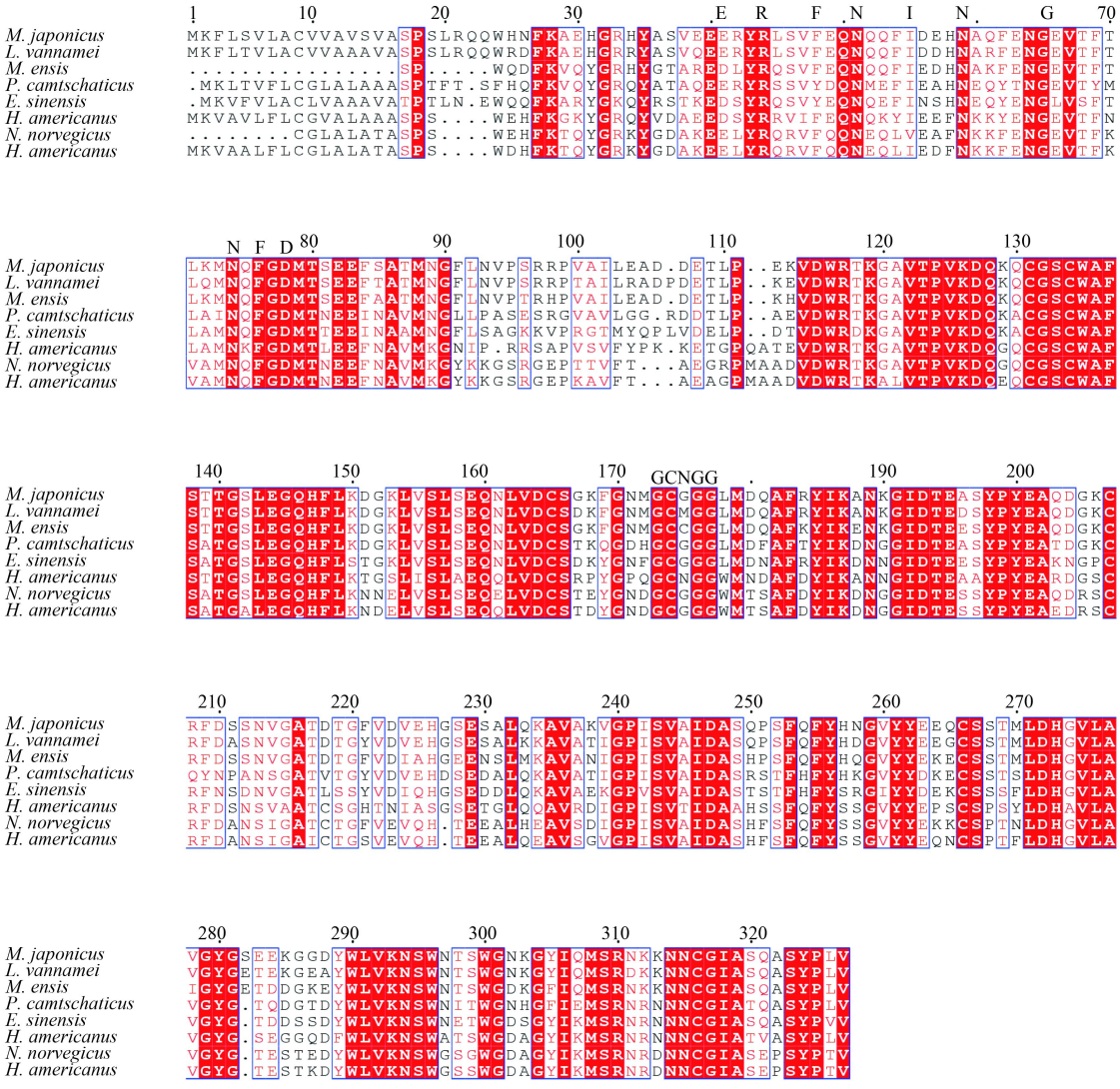

Fig. 3. Multiple alignment of the predicted amino acid sequence of *Mj-Cathepsin L* with other shrimp *Cathepsin L* amino acid sequences. Shading indicates residues that are identical (red paper-white lettered) and blue rim-red lettered indicates conserved residues. The homology sequences used were as the following order: *Litopenaeus vannamei* (CAA68066.1), *Metapenaeus ensis* (AAM96001.1), *Paralithodes camtschaticus* (ADQ73946.1), *Eriocheir sinensis* (ADO65978.1), *Homarus americanus* CYSP2_HOMAM (P25782.1), *Nephrops norvegicus* (CAA56915.1), and *Homarus americanus* CYSP3_HOMAM (P25784.1).

Table 1. Identity and genetic distance between *Mj-Cathepsin L* and other shrimp *Cathepsin L* amino acid sequences

	M. japonicus	*L. vannamei*	*M. ensis*	*P. camtschaticus*	*E. sinensis*	*H. americanus CYSP2*	*N. norvegicus*	*H. americanus CYSP3*
M. japonicus	–	0.174 0	0.286 8	0.638 0	0.737 4	0.835 1	0.787 9	0.833 5
L. vannamei	0.905	–	0.298 7	0.654 8	0.753 8	0.876 1	0.766 9	0.832 9
M. ensis	0.798	0.789	–	0.596 6	0.705 3	0.794 2	0.719 2	0.751 3
P. camtschaticus	0.685	0.685	0.662	–	0.681 6	0.902 7	0.784 4	0.852 0
E. sinensis	0.664	0.661	0.638	0.686	–	0.934 9	0.903 1	0.973 9
H. americanus CYSP2	0.623	0.606	0.609	0.597	0.588	–	0.680 9	0.691 9
N. norvegicus	0.617	0.621	0.643	0.622	0.581	0.669	–	0.150 9
H. americanus CYSP3	0.620	0.615	0.618	0.614	0.573	0.682	0.897	–

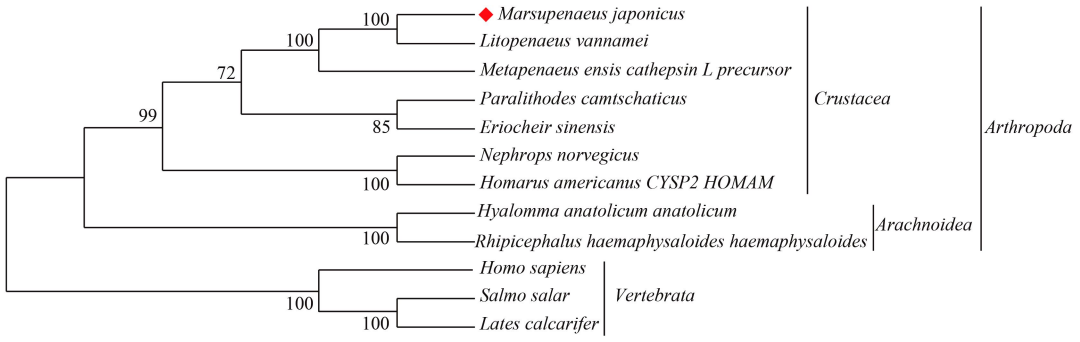

Fig. 4. Neighbor-joining phylogenetic tree of *Mj-Cathepsin L* and similar *Cathepsin L* proteins. Amino acid distances were calculated using the Poisson correction model in the MEGA program Version 6.0., and the phylogram topology was representative for 1000 bootstrap replicates. Sequences used in the alignment were: *Marsupenaeus japonicus* (No. KJ871613), *Litopenaeus vannamei* (No. CAA68066.1), *Metapenaeus ensis* (No. AAM96001.1), *Paralithodes camtschaticus* (No. ADQ73946.1), *Eriocheir sinensis* (No. ADO65978.1), *Homarus americanus* CYSP2_HOMAM (No. P25782.1), *Nephrops norvegicus* (No. CAA56915.1), *Rhipicephalus haemaphysaloides haemaphysaloides* (No. AAQ16117.1), *Hyalomma anatolicum anatolicum* (No. AFQ98384.1), *Salmo salar* (No. NP_001140018.1), *Lates calcarifer* (No. ABV59078.1), and *Homo sapiens* (No. NP_001903.1).

stomach, and relatively faint expression levels in intestine, muscle, heart, eye stalk and gill (P<0.05) (Fig. 5).

3.4 *Mj-Cathepsin L mRNA expression levels during early developmental stages and the molt cycle*

The relative expression levels of *Mj-Cathepsin L* mRNA were detected during the early developmental stages from fertilized egg to post-larva 27 (P_{27}) (Fig. 6). The expression levels of *Mj-*

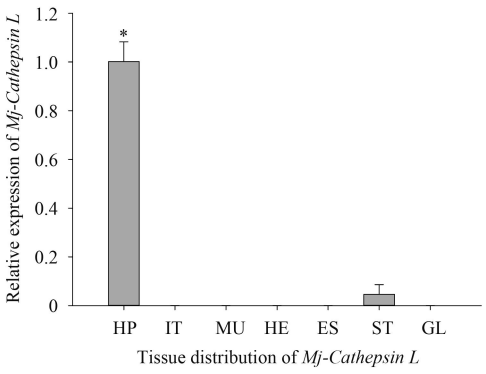

Fig. 5. Relative expression levels (n=3, mean±SD) of *Mj-Cathepsin L* gene in various adult tissues of *Marsupenaeus japonicus*. Relative expression levels of the target gene were normalized by EF1-α gene and calculated based on the $2^{-\Delta\Delta Ct}$ method using hepatopancreas as the calibrator. * above the bar indicates the significant difference at the level of 0.05. HP represents hepatopancreas, IT intestine, MU muscle, HE heart, ES eye stalk, ST stomach, and GL gill.

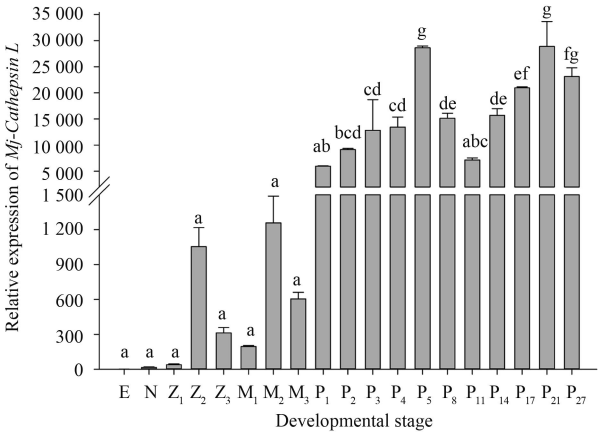

Fig. 6. Quantitative expressions (n=3, mean±SD) of *Mj-Cathepsin L* gene at early developmental stages of *Marsupenaeus japonicus*. Relative expression levels of the target gene were normalized against EF1-α gene and the fertilized eggs were chosen as the calibrator. Different letters above the bars indicate the significant differences at the level of 0.05. E represents fertilized eggs, N nauplius, Z_1–Z_3 zoea 1–3, M_1–M_3 mysis 1–3, and P_1–P_{27} post-larvae 1–27.

Cathepsin L divided into two parts; there were faintly expression levels before P_1 stage and maintained significantly high levels after P_1 stage (P<0.05). The expression levels increased steadily from P_1 to P_5 and sharply decreased at P_{8-11}; the expression level increased gradually to the highest value at P_{21} and fell back to a low level at P_{27} (P<0.05).

The transcription analysis revealed that *Mj-Cathepsin L* predominantly expressed during preecdysis (premolt) stages of the molt cycle, and maintained a relatively low level in other four stages (Fig. 7).

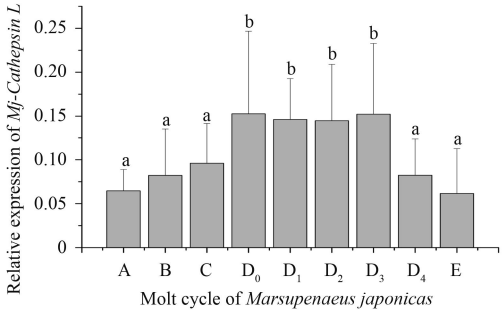

Fig. 7. Quantitative expressions of *Mj-Cathepsin L* gene during molt cycle. Stages A and B represent postecdysis/postmolt stages, Stage C intermolting stage, Stages D_0–D_4 preecdysis/ premolt stages, and Stage E ecdysis stage.

3.5 *Mj-Cathepsin L recombinant protein*

SDS-PAGE analysis showed that almost all the recombinant *Mj-Cathepsin L* was expressed as inclusion bodies contained in the cell lysate. The purified recombinant *Mj-Cathepsin L* product was detected using SDS-PAGE and showed a clear band with the predicted molecular mass of about 40.0 kDa (Fig. 8). The recombinant *Mj-Cathepsin L* was used to raise antibodies in mice for western blotting.

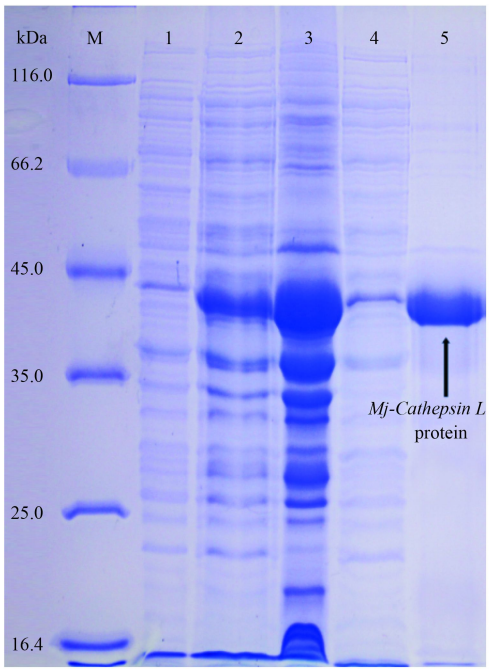

Fig. 8. Expression and purification analyses of recombinant *Mj-Cathepsin L*. SDS-PAGE analysis of expression and purification of recombinant *Mj-Cathepsin L*. M represents protein marker, 1 PET28a-*Mj-Cathepsin L* (BL21) (uninduced), 2 PET28a-*Mj-Cathepsin L* (BL21) (induced), 3 precipitation (the recombinant protein existed in the precipitations), 4 supernatant, and 5 the purified recombinant protein *Mj-Cathepsin L*. Left panel shows the position of the molecular mass markers (kDa).

3.6 *Western blotting analysis of Mj-Cathepsin L in stomach, hepatopancreas and intestine of M. japonicus*

Crude extracts of stomach, hepatopancreas and intestine were analyzed by western blot. The protein level of *Mj-Cathepsin L* in stomach, hepatopancreas and intestine of *M. japonicus* were detected using antibodies against recombinant *Mj-Cathepsin L*. The results showed that the anti-*Mj-Cathepsin L* serum positively identified two bands in ST and HP: the upper *Mj-Cathepsin L* protein (CathL) bands about 40 kDa and the lower mature-*Mj-Cathepsin L* protein (mat-CathL) bands about 22 kDa (Fig. 9).

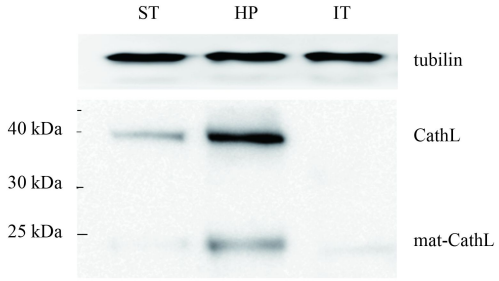

Fig. 9. The expression of *Mj-Cathepsin L* in different tissues of *Marsupenaeus japonicus* detected with the antibody against the recombinant *Mj-Cathepsin L* protein. ST represents stomach, HP hepatopancreas, and IT intestine. The upper lanes were probed with anti-tubulin as the protein-loading control, and the lower lanes represent the *Mj-Cathepsin L* protein (CathL) and mature-*Mj-Cathepsin L* protein (mat-CathL) levels.

4 Discussion

In this study, we for the first time identified *Cathepsin L* gene from *M. japonicus* and put insight into the *Mj-Cathepsin L* mRNA changes during ontogenetic development, which is the most crucial physiological and morphological processes in the life history of shrimps. The cDNA encoding *Mj-Cathepsin L* gene was 1 191 bp in full-length, and contained a typical signal peptide sequence of 16 aa (Met_1-Ala_{16}). The signal peptide is participated in the protein secretary process in many eukaryotic organisms including crustaceans (Le Boulay et al., 1998), indicating that the secretary mechanism of *Mj-Cathepsin L*. Interestingly, in some organisms the signal peptide is either absent or shorter in length (McIntyre et al., 1994; Chauhan et al., 1998), suggested that the secretary mechanism of *Cathepsin L* may be absent.

To our knowledges, *Cathepsin L* genes have diverse gene structures with the number of introns, ranging from zero to nine. The *Mj-Cathepsin L* with five introns was found to possess a relatively classic exon/intron structure among most of the shrimp. But interestingly some *Cathepsin L* genes have asymmetric distribution of introns and paralog of *Mj-Cathepsin L* in *Metapenaeus ensis* in the same family *Penaeidae* is intronless (Hu and Leung, 2006). The mechanism of the intron loss in *Cathepsin L* gene and the implicit information about deep evolutionary history need further research.

High similarity and identity of *Cathepsin L* genes in organisms including *M. japonicus* were revealed with the propeptide and mature enzyme, catalytic C-H-N triad and the six Cys-disulfide-bridges were predicted (Ma et al., 2010). ERF/WNIN-like motif is believed as the principal characteristic feature of *Cathepsin L* (Karrer et al., 1993). The deduced *Mj-Cathepsin L* protein presented in ERFNIN (Position 40–59), GNFD (Position 60–78) and GCNGG (Position 173–177) motifs which could be used to identify it as a member of the typical *Cathepsin L* family. The

phylogenetic analysis of the *Cathepsin L* genes showed that *Mj-Cathepsin L* clustered with crustaceans, suggesting the functional conservation of *Cathepsin L* in the same species group.

Cathepsin L was presumed to play an extracellular digestive role in guts of invertebrates including shrimps (Johnson and Rabosky, 2000) (Lima et al., 2001). *Mj-Cathepsin L* mRNA was predominantly expressed in hepatopancreas; same pattern was documented in *P. borealis* and *M. ensis* (Aoki et al., 2004; Hu and Leung, 2004). Hepatopancreas is considered as the most crucial organ of the crustaceans that participates in the secretion and excretion, storage of nutrients and even in the immune responses.

Morphological and physiological processes during ontogenetic development in shrimps consumed a quantity of energy for swim and predation, and young larvae took advantage of carbohydrates and lipids preferentially, conserved the proteins for the succeeding metamorphosis and development (Johnston, 2003; Darias et al., 2006). In this study, larvae were fed with the same prawn slices before P10 to reduce the diet effect to the expression of the *Mj-Cathepsin L* based on the assumption that the expression of *Cathepsin L* was regulated by the internal mechanism. *Mj-Cathepsin L* mRNA expression was demonstrated to be ontogenetic developmental-regulated in different larval stages.

First, the Nauplius stage, known as the none-feeding stage and started its preliminary differentiation of hepatopancreas, had a low expression of *Mj-Cathepsin L*. The nauplius larvae can secrete *Cathepsin L* to utilize the vitellogenin, promote the development and to prepare for the raptorial-feeding life after metamorphosis (Pan et al., 2006). Second, in Zoea and Mysis stages, the expressions of *Mj-Cathepsin L* presented an "M" shape trend with sharp peaks in Zoea 2 and Mysis 2 stages (Fig. 4). The digestive enzyme activities approached a high level from Z_3 to M_1 development in *M. japonicus* (Rodriguez et al., 1994). In this study the expressions of *Mj-Cathepsin L* mRNA in Z_3 to M_1 were relatively low, inferring that the *Mj-Cathepsin L* contribute to other functions than digestion during early development. Third, in the post-larval stages the carapace length showed a positive relationship with body length, responding with the increase of food intake (data not shown). The expression of *Mj-Cathepsin L* approached a relatively high level in the post-larval stages, with an obvious decrease from P_6 to P_{11} (Fig. 6), suggested that it may be associated with the behavioral change of *M. japonicas* from pelagic to benthic life around P_8 stage, and with the physiological changes of feeding habit, digestive system change and energy distribution (Lemos and Rodríguez, 1998). Interestingly, under the same food intake conditions, we noticed that mRNA expressions of *Mj-Cathepsin L* in different larvae stages are opposite to that of *α-amylase* in our previous studies, which is known as a major glucosidase involved in the digestive function of shrimps (unpublished data). This demonstrates the different nutrient intake in different larvae stages and suggests the development-regulated expression of the *Mj-Cathepsin L*.

Molting, a phenomenon of shedding the old cuticle and regenerating the new one, was considered as the most important process during larvae stages or even the whole life cycle. Le Boulay et al. (1996) demonstrated that variations in cysteine protease activity were correlated with the variations in the levels of specific mRNA during intermolt cycles of *Penaeus vannamei*. Here we found *Mj-Cathepsin L* predominantly expressed during preecdysis/premolt (D_0–D_3) stages of the molt cycle, and it maintained a relatively low level in other four stages. There was a gradually rising trend from Stages A to C, following the well-known Stage E (ecdysis). The expression of *Mj-Cathepsin L* during molt cycle suggested that *Mj-Cathepsin L* might be required for the molting process, especially in preecdysis/premolt (D) stages.

Western blotting confirmed the results of tissue distribution, we also demonstrated the nature *Mj-Cathepsin L* protein presents two forms in the shrimp tissues, the pro-*Mj-Cathepsin L* about 40 kDa and the mature-*Mj-Cathepsin L* (mat-*CathL*) about 22 kDa. The hepatopancreas showed two very strong bands as compared with stomach and intestine, and the upper pro-*Mj-Cathepsin L* was relatively stronger than lower mature-*Mj-Cathepsin L*. The results revealed the *Mj-Cathepsin L* protein mainly existed in the form of pro-*Mj-Cathepsin L*, but not the activity mature-*Mj-Cathepsin L*.

In summary, *Mj-Cathepsin L* gene was isolated from Kuruma shrimp *Marsupenaeus japonicus* and was demonstrated expressing in the hepatopancreas and stomach. The mRNA levels of *Mj-Cathepsin L* were studied during early ontogenetic development stages from eggs to post-larvae 28 and during the five molt cycles. The results suggest the *Mj-Cathepsin L* seems to be regulated by digest and molting processes during the early ontogenetic development stages of *Marsupenaeus japonicus*. Western blot also revealed two forms of *Mj-Cathepsin L* protein existing in three tissues. Further studies will focus on the localization of the *Mj-Cathepsin L*.

Acknowledgements

The authors thank Dong-shan Maoxin Aquaculture Company (Dongshan County, Fujian Province) for providing the experimental shrimp and equipment.

References

Aoki H, Ahsan M N, Watabe S. 2004. Molecular and enzymatic properties of a *cathepsin L*-like proteinase with distinct substrate specificity from northern shrimp (*Pandalus borealis*). J Comp Physiol B, 174(1): 59–69

Arockiaraj J, Gnanam A J, Muthukrishnan D, et al. 2013. Macrobrachium rosenbergii *cathepsin L*: molecular characterization and gene expression in response to viral and bacterial infections. Microbiol Res, 168(9): 569–579

Chauhan S S, Ray D, Kane S E, et al. 1998. Involvement of carboxy-terminal amino acids in secretion of human lysosomal protease *cathepsin L*. Biochemistry, 37(23): 8584–8594

Dahl S W, Halkier T, Lauritzen C, et al. 2001. Human recombinant pro-dipeptidyl peptidase I (cathepsin C) can be activated by cathepsins L and S but not by autocatalytic processing. Biochemistry, 40(6): 1671–1678

Darias M J, Murray H M, Gallant J W, et al. 2006. Characterization of a partial α-amylase clone from red porgy (*Pagrus pagrus*): expression during larval development. Comp Biochem Physiol B Biochem Mol Bio, 143(2): 209–218

de Oliveira Cesar J R, Zhao Baoping, Malecha S, et al. 2006. Morphological and biochemical changes in the muscle of the marine shrimp *Litopenaeus vannamei* during the molt cycle. Aquaculture, 261(2): 688–694

Dhar A K, Bowers R M, Licon K S, et al. 2009. Validation of reference genes for quantitative measurement of immune gene expression in shrimp. Mol Immunol, 46(8–9): 1688–1695

Glenn K L, Grapes L, Suwanasopee T, et al. 2005. SNP analysis of *AMY2* and *CTSL* genes in *Litopenaeus vannamei* and *Penaeus monodon* shrimp. Anim Genet, 36(3): 235–236

Hu Kejin, Leung P C. 2004. Shrimp *cathepsin L* encoded by an intronless gene has predominant expression in hepatopancreas, and occurs in the nucleus of oocyte. Comp Biochem Physiol B Biochem Mol Biol, 137(1): 21–33

Hu Kejin, Leung P C. 2007. Food digestion by *cathepsin L* and digestion-related rapid cell differentiation in shrimp hepatopancreas. Comp Biochem Physiol B Biochem Mol Biol, 146(1): 69–80

Hu K J, Leung P C. 2006. Complete, precise, and innocuous loss of multiple introns in the currently intronless, active *cathepsin L*-like genes, and inference from this event. Mol Phylogenet Evol, 38(3): 685–696

Johnson K S, Rabosky D. 2000. Phylogenetic distribution of cysteine proteinases in beetles: evidence for an evolutionary shift to an alkaline digestive strategy in Cerambycidae. Comp Biochem Physiol B Biochem Mol Biol, 126(4): 609–619

Johnston D J. 2003. Ontogenetic changes in digestive enzyme activity of the spiny lobster, *Jasus edwardsii* (Decapoda; Palinuridae). Marine Biology, 143(6): 1071–1082

Karrer K M, Peiffer S L, Ditomas M E. 1993. Two distinct gene subfamilies within the family of cysteine protease genes. Proc Natl Acad Sci U S A, 90(7): 3063–3067

Laycock M V, Hirama T, Hasnain S, et al. 1989. Purification and characterization of a digestive cysteine proteinase from the American lobster (*Homarus americanus*). Biochem J, 263(2): 439–444

Le Boulay C, Van Wormhoudt A, Sellos D. 1995. Molecular cloning and sequencing of two cDNAs encoding *cathepsin L*-related cysteine proteinases in the nervous system and in the stomach of the Norway lobster (*Nephrops norvegicus*). Comp Biochem Physiol B Biochem Mol Bio, 111(3): 353–359

Le Boulay C, Van Wormhoudt A, Sellos D. 1996. Cloning and expression of *cathepsin L*-like proteinases in the hepatopancreas of the shrimp *Penaeus vannamei* during the intermolt cycle. J Comp Physiol B, 166(5): 310–318

Le Boulay C, Sellos D, Van Wormhoudt A. 1998. *Cathepsin L* gene organization in crustaceans. Gene, 218(1–2): 77–84

Lemos D, Rodríguez A. 1998. Nutritional effects on body composition, energy content and trypsin activity of *Penaeus japonicus* during early postlarval development. Aquaculture, 160(1–2): 103–116

Lima A P C A, dos Reis F C G, Serveau C, et al. 2001. Cysteine protease isoforms from *Trypanosoma cruzi*, cruzipain 2 and cruzain, present different substrate preference and susceptibility to inhibitors. Mol Biochem Parasitol, 114(1): 41–52

Livak K J, Schmittgen T D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods, 25: 402–408

Ma Jianjun, Zhang Dianchang, Jiang Jingjing, et al. 2010. Molecular characterization and expression analysis of *cathepsin L*1 cysteine protease from pearl oyster *Pinctada fucata*. Fish Shellfish Immunol, 29(3): 501–507

McIntyre G F, Godbold G D, Erickson A H. 1994. The pH-dependent membrane association of pro*cathepsin L* is mediated by a 9-residue sequence within the propeptide. J Biol Chem, 269(1): 567–572

Miyamoto K, Iwadate M, Yanagisawa Y, et al. 2011. *Cathepsin L* is highly expressed in gastrointestinal stromal tumors. Int J Oncol, 39(5): 1109–1115

Nakagawa T, Roth W, Wong P, et al. 1998. Cathepsin L: critical role in Ii degradation and CD4 T cell selection in the thymus. Science, 280(5362): 450–453

Pan Luqing, Liu Hong yu, Xiao Guoqiang. 2006. A review on digestive enzyme of crustacean larvae. J Fishery Sci China, 13(3): 492–501

Qian Zhaoying, Li Xilian, Xin Jingjing, et al. 2013. PCR-SSCP Polymorphism of CTSL gene and its correlation with growth traits of *Litopenaeus vannamei* and the different mRNA expressions of CTSL. Haiyang Xuebao, 35(6): 121–127

Rodriguez A, Le Vay L, Mourente G, et al. 1994. Biochemical composition and digestive enzyme activity in larvae and postlarvae of *Penaeus japonicus* during herbivorous and carnivorous feeding. Marine Biology, 118(1): 45–51

Roth W, Deussing J, Botchkarev V A, et al. 2000. Cathepsin L deficiency as molecular defect of furless: hyperproliferation of keratinocytes and pertubation of hair follicle cycling. FASEB J, 14(13): 2075–2086

Zhang W, Wang S, Wang Q, et al. 2014. Overexpression of cysteine *cathepsin L* is a marker of invasion and metastasis in ovarian cancer. Oncol Rep, 31(3): 1334–1342