

Sains Malaysiana 43(8)(2014): 1139–1148

Molecular Characterisation and Expression Analysis of Cathepsin D from the Asian Seabass *Lates calcarifer*

(Pencirian Molekul dan Analisis Pengekspresan Katepsin D daripada Ikan Siakap *Lates calcarifer*)

SHARIZA AZIZAN, KIEW-LIAN WAN* & ADURA MOHD-ADNAN

ABSTRACT

The lysosomal aspartic proteinase cathepsin D is an acute phase protein involved in various physiological processes, including vitellogenesis, yolk processing and immune responses. In this study, we characterised the cathepsin D from the Asian seabass Lates calcarifer and examined its expression profile during infection. The complete coding sequence of L. calcarifer cathepsin D consists of 1191 nucleotides, encoding a 396 amino acid protein molecule that is made up of a putative signal peptide, a leader peptide and a mature peptide. Phylogenetic analyses showed that two types of cathepsin D are present in the teleost lineage i.e. cathepsin D1 and D2, whereas higher vertebrates possess only one type of cathepsin D. L. calcarifer cathepsin D was clustered together with cathepsin D1 from other teleosts. Compared to mammalian sequences, L. calcarifer cathepsin D lacks the β -hairpin loop that forms the double chain and is present as a single chain peptide with conserved aspartic active sites like other fish. Both multiple sequence alignment and phylogenetic analysis indicated that the L. calcarifer cathepsin D sequence codes for cathepsin D1 and suggested that it shares the same functions with cathepsin D from other fish. Expression profiling analysis of cathepsin D in L. calcarifer infected with Aeromonas hydrophila showed that it is up-regulated in immune-related tissues such as gills, spleen and liver, suggesting that cathepsin D plays an important role in the innate immune response of L. calcarifer against pathogens.

Keywords: Acute phase protein; aspartic proteinase; expression profile; innate immune response

ABSTRAK

Katepsin D, sejenis proteinase aspartik lisosom merupakan protein fasa akut yang terlibat dalam pelbagai proses fisiologi, termasuk vitelogenesis, pemprosesan yolka dan gerak balas keimunan. Dalam kajian ini, kami telah mencirikan katepsin D daripada ikan siakap Lates calcarifer dan mengkaji profil pengekspresannya semasa infeksi. Jujukan pengekodan lengkap katepsin D L. calcarifer terdiri daripada 1191 nukleotida yang mengkod molekul protein bersaiz 396 asid amino yang merangkumi satu peptida isyarat putatif, satu peptida mendahului dan satu peptida matang. Analisis filogenetik menunjukkan bahawa terdapat dua jenis katepsin D hadir dalam susur galur teleost iaitu katepsin D1 dan D2, sementara vertebrata peringkat tinggi hanya mempunyai satu jenis katepsin D. Katepsin D L. calcarifer dikelompokkan bersama dengan katepsin D1 ikan teleost lain. Berbanding dengan jujukan mamalia, katepsin D L. calcarifer didapati tidak mempunyai jujukan gelung pin rambut β yang membentuk rantai ganda dua dan wujud sebagai rantai tunggal peptida dengan kehadiran tapak aktif aspartik terpulihara seperti ikan lain. Penjajaran jujukan berbilang dan analisis filogenetik menunjukkan bahawa jujukan katepsin D L. calcarifer mengkodkan katepsin D1 dan mencadangkan ia mempunyai fungsi yang sama dengan katepsin D ikan lain. Analisis profil pengekspresan katepsin D dalam L. calcarifer terinfeksi Aeromonas hydrophila mendedahkan bahawa pengekspresannya meningkat dalam tisu berkait-keimunan seperti insang, limpa dan hepar, mencadangkan bahawa katepsin D memainkan peranan yang penting dalam gerak balas keimunan semula jadi L. calcarifer terhadap patogen.

Kata kunci: Gerak balas keimunan semula jadi; profil pengekspresan; proteinase aspartik; protein fasa akut

INTRODUCTION

Cathepsin D is a lysosomal aspartic proteinase that is present in various tissues such as the spleen, liver, kidney, muscle and gills (Barret 1977; Metcalf & Fusek 1993). The cathepsin D gene is also known as a housekeeping gene because it is expressed constitutively (Riggio et al. 2000). Mammalian cathepsin D exists as a double chain; it has both a heavy and a light chain, with molecular mass of approximately 30 and 15 kDa, respectively (Metcalf & Fusek 1993). In fish, cathepsin D lacks a β -hairpin loop

and exists as a single chain (Momsen 2004; Nielsen & Nielsen 2001). Cathepsin D consists of cathepsin D, D1 and D2 (Feng et al. 2011; Rojo et al. 2010). Cathepsin D is found in higher vertebrates whereas cathepsin D1 and D2 have been found only in fish and invertebrates (Feng et al. 2011).

The major function of cathepsin D is protein degradation and the molecule has been reported to be involved in vitellogenesis, yolk processing and ovarian follicle growth and maturation (Brooks et al. 1997;

Carnevali et al. 1999). It is also believed to be involved in processing, secretion and activation of enzymes and hormones (Baldochi et al. 1993; Krieger & Hook 1992) and in physiological pathways including both extracellular proteolysis and intracellular catabolic proteolysis (Baricos et al. 1987; Gilberg 1988). Other studies have reported that cathepsin D plays a role in spawning in salmon *Salmo salar* (Mommmsen 2004) and muscle proteolysis in herring *Clupea harengus* (Nielsen & Nielsen 2001). Many studies have also suggested that cathepsin D is involved in the fish immune system (Jia & Zhang 2009; Liu et al. 2012). For example, the gene was shown to be highly expressed in immune-related organs of channel catfish following bacterial challenge (Feng et al. 2011). Cathepsin D has also been reported to be involved in the production of the antimicrobial peptide parasin I from histone H2A in amur catfish *Parasilurus asotus* (Cho et al. 2002).

The Asian seabass *Lates calcarifer* is a commercially important fish with a high potential market value in the tropical Asia Pacific region including Malaysia (Chou & Lee 1997; Nelson 1994). However, diseases outbreak is still a major concern in the *L. calcarifer* aquaculture industry (Hatha et al. 2005). Amongst them is haemorrhagic septicemia cause by the marine bacterium *Aeromonas hydrophila*, which brings about fatality in fish via the production of enterotoxin, dermonecrosis factor and hemolysin (Daskalov 2006).

In recent years, expressed sequence tags (ESTs), genome sequence survey, and microarray analysis have been used to study this fish species (Chong et al. 2011; Khoo et al. 2009; Mohd-Yusof et al. 2009; Tan et al. 2008). A number of genes have also been cloned and their activities were successfully studied (Lee et al. 2012; Mohamed-Jawad et al. 2012; Mohd-Padil et al. 2010). Although these studies shed insights into the immune system of *L. calcarifer*, our understanding on the molecular basis of host defense against pathogenic marine bacteria remains incomplete.

In this study, we have cloned and characterised the complete coding sequence of *L. calcarifer* cathepsin D. We further determined its expression profile using quantitative real-time PCR (qRT-PCR) in immune-related tissues, namely gills, spleen, liver and kidney, during *A. hydrophila* infection. The results provided clues on the roles of cathepsin D and indicated its functions in the *L. calcarifer* immune system.

MATERIALS AND METHODS

TISSUE SAMPLING

A pair of adult (male and female) *L. calcarifer* was obtained from the Marine Finfish Production and Research Center, Terengganu, Malaysia. For the generation of the complete coding sequence, the liver tissue was extracted from anaesthetised fish and kept in liquid nitrogen. For expression analysis, juvenile fish (average weight of 29

g) were used after acclimatisation to hatchery conditions for one week. Water salinity and oxygen density (DO) were maintained at 28 to 30 ppm and ~5.0 mg/L, respectively. A total of 144 fish were divided into three 300 L tanks with 48 fish per tank for the bacteria challenge. Subsequently, 100 mL of 1×10^7 colony forming units (CFUs) mL^{-1} of *A. hydrophila* (which was confirmed by serial dilution) suspended in LB broth were injected intraperitoneally into each fish in tanks 1 and 2. At the same time, the control fish in tank 3 were injected with 100 mL of sterile LB broth. Three fish were sacrificed from each tank every 6 h after injection up to 36 h. Immune related tissues i.e. gills, kidney, liver and spleen were harvested and stored in RNAlater (Qiagen, USA) and kept at -80°C until further use.

PRIMER DESIGN

The primers used for cloning and qRT-PCR were designed using Primer Premier 5.0 (Premier Biosoft, USA). Based on the *L. calcarifer* EST dataset (Mohd-Yusof et al. 2009), the primers for the cathepsin D sequence were designed to amplify the complete open reading frame (ORF). A pair of cathepsin D primers was then designed based on the ORF sequence to amplify a 150 bp fragment for expression analysis in liver, kidney, spleen and gills using qRT-PCR. Primers for housekeeping genes used as references in this experiment were designed based on Kumar et al. (2000) and Lee et al. (2012) (Table 1).

RNA EXTRACTION AND CDNA SYNTHESIS

Total RNA was extracted from fish tissues using TRIreagent (Molecular Research Centre Inc., USA) according to the manufacturer's protocol. The quality of the total RNA obtained was analysed using agarose gel electrophoresis and the quantity was determined using Nanodrop® ND 1000 (Thermo Scientific, USA). Approximately 1 μg of total RNA was used to synthesise the cDNA using RETRO Script® (Ambion, USA) based on the manufacturer's protocol. For expression analysis, total RNA was extracted using the same method, with additional treatment with DNase I (Qiagen, USA) to remove any possible DNA contamination. The quality and integrity of RNA (RNA integrity number; RIN) were determined using Bioanalyzer (Agilent Technologies, USA). RNA samples with RIN > 6.3 were selected for cDNA synthesis using the iScript™ cDNA Synthesis Kit (BioRad, USA) according to the manufacturer's instruction.

AMPLIFICATION AND CLONING OF CATHEPSIN D

Amplification of the complete ORF of *L. calcarifer* cathepsin D was carried out using a thermal cycler (Eppendorf, Germany). The PCR product of desired size was purified using the QIAquick® PCR Purification Kit (Qiagen, USA) and cloned into pJET1.2 using the CloneJET™ PCR Cloning Kit (Fermentas, USA). The recombinant plasmid was propagated in *Escherichia coli* JM107 and sequenced using an ABI 3100 sequencer (Applied Biosystems Inc., USA).

TABLE 1. Primer sequences used in this study

Name	Sequence
CatDF	5' CCTGGTCGTTTTTGCGGCTTTAG 3'
CatDR	5' CCGAGACAACAACAGAGTGGGCTTTGCTA 3'
CatDRtF	5' GTTCCTCACTGAGGTAGCCTGACAAACT 3'
CatDRtR	5' CCTCAGACCTTCAGTGTGGTGTTTG 3'
ElongaseF	5' TGGCTTATGCTGATTAACCC 3'
ElongaseR	5' TCTTTCCAATTAATTTGTGGCA 3'
18SF	5' TGGTTAATTCCGATAACGAACGA 3'
18SR	5' CGCCACTTGTCCCTCTAAGAA 3'
rpL8F	5' GGCAAACCCTCAACAATCAG 3'
rpL8R	5' CAGATTTAGTCTGGGTCATCTTCT 3'
β -actinF	5' CATCACACCTTCTACAACGA 3'
β -actinR	5' ACATGATCTGGGTCATCTTCT 3'

SEQUENCE ANALYSIS

The DNA sequence quality was screened using Phred (Ewing & Green 1998) with a threshold value of 30, the vector sequence was trimmed using Cross Match (<http://www.phrap.org>), coding region was determined using ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) and signal peptide was predicted using SignalP (Bendsten et al. 2004). Subsequently, the sequence was deposited in GenBank with the accession number JX082258. The sequence was then subjected to similarity search using the Basic Local Alignment Search Tool (BLAST) program (Altschul et al. 1997). The primary structure of the sequence was analysed using ProtParam (<http://web.expasy.org/protparam/>). All cathepsin D sequences from other species were retrieved from the BLAST similarity search results (Table 2). Multiple sequence alignment of the amino acid sequences was performed using MAFFT on the EMBL-EBI Server (<http://www.ebi.ac.uk/Tools/msa/mafft/>).

Phylogenetic trees based on amino acid sequences were generated using maximum likelihood (ML) and Bayesian inference implemented by PhyML 3.0 (Guindon et al. 2010) and MrBayes 3.12 (Ronquist & Huelsenbeck 2003), respectively. To examine the relationship between cathepsin D, D1 and D2, a rooted phylogenetic tree was performed using known vertebrates and invertebrates as an outgroup. Construction of a second phylogenetic tree was performed to specifically investigate the basis of *L. calcarifer* cathepsin D clustering within the fish clade using amphibian as an outgroup.

In ML analyses, the Le and Gascuel model (Le & Gascuel 2008) with the gamma distribution parameter was used for the tree construction. The robustness of the tree was evaluated using nonparametric bootstrapping with a value of 1000. MEGA 5.05 (Tamura et al. 2011) was used to view the ML trees and the computed consensus tree with a 50% majority rule. For Bayesian analyses, the trees were built using the *prset aamodelpr* model type

with *mixed* state frequencies. In these analyses, each run was conducted with 100000 generations and trees were sampled every 100 generations. The burn-in procedure discarded 25% of the first, which is 250 trees in these analyses and majority rule was used to construct the consensus tree.

QUANTITATIVE REAL-TIME PCR

Quantitative real-time PCR (qRT-PCR) was performed using Mastercycler® *ep realplex*⁴ (Eppendorf, Germany) to examine the expression profile of cathepsin D in tissues collected from *L. calcarifer* that had been infected with *A. hydrophila*. The reaction mixture consisted of a final volume of 20 μ L containing 10 μ L of SsoFast™ EvaGreen® Supermix (BioRad, USA), 1 μ L (~20 ng) of cDNA template, 2 μ L of the forward and reverse primer mix and 7 μ L of DNase free water (Qiagen, USA). The primers used are as listed in Table 1. The reaction was set according to the manufacturer's protocol as follows: 10 s of enzyme activation at 95°C followed by 40 cycles of 15 s of denaturation at 95°C, 1 min of annealing at 63.5°C and a melting curve analysis of 0.5°C increment starting from 65 up to 95°C. The five reference genes used for normalisation were elongase factor, 18S, rpL-8, β -actin, and GAPDH. The raw data were analysed using the relative expression software tool (REST) - multiple condition solver software (Pfaffl et al. 2002). The reaction efficiency was in the range of 90-110% and the correlation coefficient was ≥ 0.98 . The results were subjected to statistical analysis (randomisation test of REST) to compare infected and non-infected fish in three tanks. The results with *P* values < 0.05 were considered to be statistically significant. Gene expression was calculated using the equation fold change = $2^{-\Delta\Delta Ct}$ with the accepted cycle threshold, Ct value of 15-25. This Ct value range shows that the mixtures of reaction samples are cumulative enough to perform a reaction (Pfaffl et al. 2002).

TABLE 2. Sequence data used in multiple sequence alignment and phylogenetic analysis

Organism	Common name	GenBank accession no.
Mammal		
<i>Bos taurus</i>	Cow	DAA13526
<i>Sus scrofa</i>	Pig	NP_001032810
<i>Canis familiaris</i>	Dog	NP_00102792
<i>Papio anubis</i>	Baboon	NP_001162374
<i>Homo sapiens</i>	Human	NP_001900
<i>Mus musculus</i>	House mouse	NP_034113
Avian		
<i>Gallus gallus</i>	Chicken	NP_990508
Reptilian		
<i>Podarcis sicula</i>	Lizard	CAA08878
Amphibian		
<i>Xenopus tropicalis</i>	Western frog	NP_988964
<i>Xenopus laevis</i>	African frog	NP_001085308
Fish		
<i>Clupea harengus</i>	Atlantic herring	AAG27733
<i>Danio rerio</i>	Zebra fish	AAI54316
<i>Silurus asotus</i>	Cat fish	AAM62283
<i>Oncorhynchus mykiss</i>	Trout	NP_001118183
<i>Scophthalmus maximus</i>	Turbot	ABW70789
<i>Paralichthys olivaceus</i>	Japanese flounder	ACN43675
<i>Salmo salar</i>	Salmon	ACH70630
<i>Miichthys miiuy</i>	Miiuy croacker	ADP89523
<i>Takifugu rubripes D1</i>	Puffer fish	NP_001072052
<i>Ictalurus punctatus D1</i>	Channel catfish	AEC33270
<i>Hippoglossus hippoglossus D2</i>	Atlantic halibut	ABI85390
<i>Sparus auratus D2</i>	Red seabream	AAB88862
<i>Takifugu rubripes D2</i>	Puffer fish	BAD69802
<i>Ictalurus punctatus D2</i>	Channel catfish	ADO28602
<i>Danio rerio D2</i>	Zebra fish	AAH56836
<i>Lates calcarifer UKM</i>	Asian seabass	JX082258
Invertebrate		
<i>Drosophila melanogaster</i>	Fruit fly	NP_652013
<i>Homarus americanus</i>	American lobster	ACV53024
<i>Bombyx mori</i>	Silkworm	AAP50847
<i>Todarodes pacificus</i>	Japanese flying squid	BAD15111

RESULTS

ANALYSIS OF *L. CALCARIFER* CATHEPSIN D CODING SEQUENCE

The cDNA sequence of the putative *L. calcarifer* cathepsin D contains a 1191 bp ORF encoding a 396 amino acid peptide. Analysis of the ORF sequence indicated that the putative *L. calcarifer* cathepsin D consists of a signal peptide of 18 amino acids, a leader peptide extending for 43 amino acids and a mature peptide of 335 amino acids. Further analysis conducted using ProtParam showed that the mature peptide of the putative *L. calcarifer* cathepsin D has a molecular weight of 36.2 kDa and a PI value of 5.02, which are consistent with those of other fish cathepsin D (Jia & Zhang 2009). The total number of negatively charged residues (Asp and Glu) was 35, whereas there were 41 positively charged residues (Arg and Lys). The analysis indicated that it is a stable protein, with an instability index

value of a 37.27. SignalP analysis showed that the predicted cleavage position for the signal sequence is between Ala¹⁸ and Leu¹⁹. The putative *L. calcarifer* cathepsin D has two N-glycosylation sites and two aspartic acid active sites (DTG).

From the BLAST search against the GenBank database, a total 52 sequences of cathepsin D, D1 and D2 were retrieved. However, 23 of them were partial or truncated sequences and were discarded. Multiple sequence alignment showed a high level of amino acid sequence conservation between *L. calcarifer* and other vertebrates, especially fish (Figure 1). The two aspartic acid active sites (Asp¹¹⁵ and Asp³⁰⁴) were conserved in all vertebrate sequences. N-glycosylation sites are present in different numbers (1, 2 or 3) in different species. Most of the cathepsin D and D1 sequences had two N-glycosylation (Asn¹⁴³ and Asn²⁷²) sites, except for *Clupea harengus*, *Danio rerio* and *Oncorhynchus*

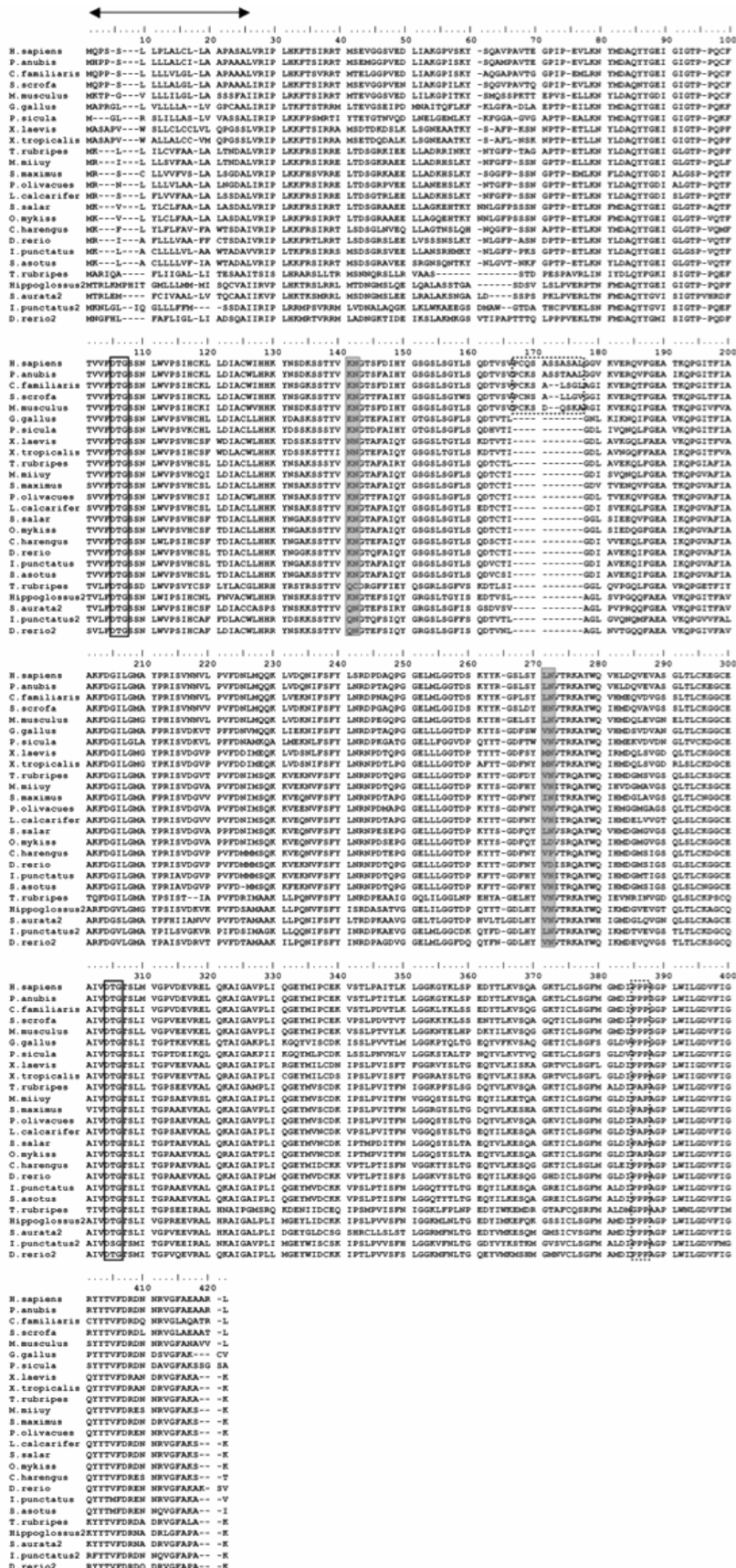


FIGURE 1. Alignment of organisms (mammal, avian, reptile, amphibian and fish). Amino acids of active sites (DTG) are enclosed in solid boxes, putative N-glycosylation sites are shaded grey, the β -hairpin loop specific for mammalian cathepsin D is shown in a dashed box, and polyproline loop is enclosed in dotted box. The signal peptide region is indicated by arrow

mykiss. The second N-glycosylation site was replaced by proline (Pro²⁷²) in *C. harengus* and aspartate (Asp²⁷²) in *D. rerio* and *O. mykiss*. Additionally, cathepsin D2 sequences have a third N-glycosylation site (Asn³⁵⁷) as previously reported (Mommensen 2004). However, in *T. rubripes* cathepsin D2, the third N-glycosylation site was replaced by Pro³⁵⁷. The N-terminal sequences of cathepsin D also showed high similarity. It has been reported that while the N-terminal sequence of cathepsin D is highly similar among species, it is different with the N-terminal sequence from other aspartic proteinase (cathepsin E, pepsin and renin), suggesting that it may be possible to classify aspartic proteinases based on their N-terminal sequences (Jia & Zhang 2009). The mammal cathepsin D sequences were longer than other species and showed the presence of a β -hairpin loop (167–177), the structure used to generate the double chain form of cathepsin D that is absent in other non-mammalian organisms. The *L. calcarifer* cathepsin D, like that of other fishes, lacked the β -hairpin loop and is present as single chain. The signal peptide region (1–25) was conserved according to the species involved (Figure 1). This region was dominated by hydrophobic amino acids such as methionine (M), leucine (L), alanine (A), glycine (G), isoleucine (I), phenylalanine (F), proline (P) and valine (V). This hydrophobic property is an important characteristic for signal recognition during the transport of protein (Dobberstein 1987). The sequences also contain endoplasmic reticulum membrane retention signals in the N-terminus (from position 2 to 5) and C-terminus (from position 416 to 423). These regions were variable and highly divergent among different vertebrate species.

The putative *L. calcarifer* cathepsin D sequence displayed high similarity with *Paralichthys olivaceus* and *Scophthalmus maximus*. Overall, the *L. calcarifer* cathepsin D sequence showed high conservation with the cathepsin D1 sequences of the other fish.

PHYLOGENETIC ANALYSIS

Two phylogenetic trees were constructed from all vertebrates and fish only sequences respectively and both showed similar topology for ML and Bayesian analyses (Figures 2 and 3). In Figure 2, the topology of the trees showed consistent taxonomy of organism hierarchy from lowest vertebrate to the highest. The phylogenetic tree construction involved six mammals, one avian, one reptile, two amphibians, 16 fish including *L. calcarifer* and an outgroup of four invertebrate sequences (Figure 2). It was divided into two vertebrate clades and the sequences within the vertebrate taxa were grouped according to the type of cathepsin D. Clade I consisted of cathepsin D of known vertebrates (mammals, avian, reptile and amphibians) and cathepsin D1 of the fish while clade II comprised only of cathepsin D2 from fish. The *L. calcarifer* cathepsin D sequence appeared to be grouped with cathepsin D1 from other fish. Further analysis showed that the fish cathepsin D sequences were separated based on fish habitat i.e. marine,

brackish or euryhaline and fresh water dwelling species (Figure 3). In clade I, cathepsin D1 from *T. rubripes*, *M. miiuy*, *P. olivaceus*, *S. salar*, *O. mykiss* and *C. harengus*, and cathepsin D from *L. calcarifer* were grouped as marine, brackish or euryhaline species while cathepsin D1 from three other species (*D. rerio*, *S. asotus* and *I. punctatus*) were grouped as freshwater species with a high supported value.

EXPRESSION ANALYSIS OF *L. CALCARIFER* CATHEPSIN D

Expression profiles of cathepsin D in immune-related tissues (gills, spleen, liver and kidney) were determined using qRT-PCR and the data were normalised using two different housekeeping genes for each tissue type. In gill and spleen tissues, the data were normalised with 18S and β -actin. In the liver tissue, 18S and elongase were used while in the kidney tissue, 18S and rpl8 were used to normalise the data. Different housekeeping genes were used because of their different expression stability in a given tissue type.

The analysis of cathepsin D expression in response to *A. hydrophila* showed an up-regulation of the gene in all immune-related tissues (gills, liver and spleen) with the exception of the kidney (Figure 4). In the kidney tissue (Figure 4(a)), expression of the cathepsin D gene was down-regulated at all time-points. In the gills tissue (Figure 4(b)), cathepsin D showed a fluctuated pattern of expression level. It has a low expression level at 6 h, increased to the background level at 12 and 18 h, decreased at 24 h, increased again to the highest level at 30 h and then decreased at 36 h. Expression of the cathepsin D gene in the liver tissue (Figure 4(c)) was down-regulated from 6 to 18 h but was then up-regulated starting from 24 h and reached the highest level at 36 h. In the spleen tissue (Figure 4(d)), cathepsin D was up-regulated with the lowest and highest level at 6 and 36 h, respectively.

DISCUSSION

The complete ORF sequence of the putative *L. calcarifer* cathepsin D was obtained by PCR amplification. The sequence is a single chain with a molecular weight of 36.2 kDa, which is consistent with the data for other fish species (Jia & Zhang 2009; Liu et al. 2012). *Lates calcarifer* cathepsin D lacks the required sequence (the β -hairpin loop) to generate the double chain form that exists in mammals (Figure 1). The alignment results showed high similarity among the cathepsin D in fish, which indicates they share similar features and have similar function(s).

Phylogenetic analyses showed that there are two types of cathepsin D present in the fish lineage, which are cathepsin D1 and D2. Analyses of the cathepsin D sequences grouped cathepsin D from *L. calcarifer* with cathepsin D1 from other fish, while cathepsin D2 from fish formed a separate clade with high confidence values. These findings are consistent with previous studies, which reported that more than one isoform of cathepsin D exist

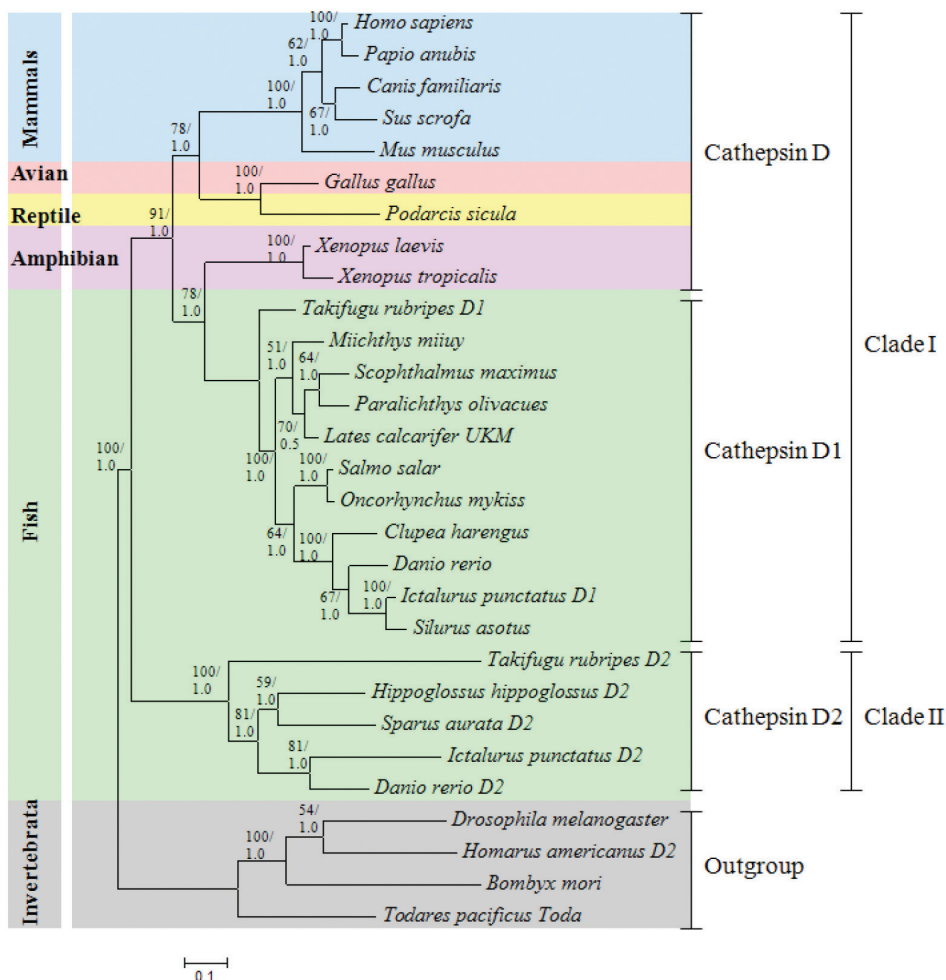


FIGURE 2. Bayesian phylogenetic tree based on cathepsin D amino acids. Values at the nodes represent the maximum likelihood percentages of 1000 replicates and Bayesian posterior probabilities. The classes are separated by colours: blue; mammal, pink; avian, yellow; reptile, purple; amphibian; green; fish and grey; invertebrate

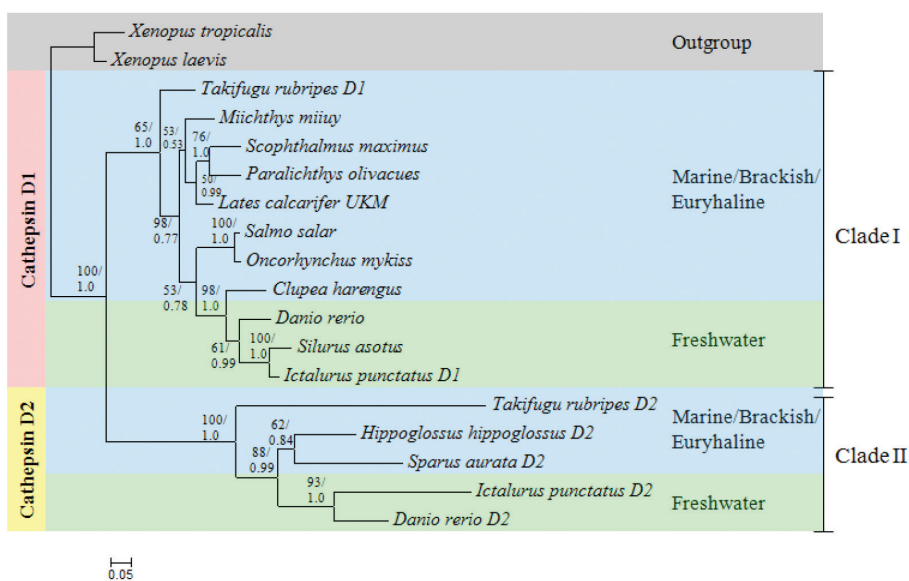


FIGURE 3. Bayesian phylogenetic tree based on cathepsin D1 and D2 of fish amino acids. Values at the nodes represent the maximum likelihood percentages of 1000 replicates and Bayesian posterior probabilities. The habitats are separated by colours: blue; marine/brakish/euryhaline and green; freshwater

(cathepsin D1 and D2) and formed a cluster in phylogenetic analyses (Carnevali et al. 2005; Kurokawa et al. 2005).

In pufferfish *Takifugu rubripes* and zebrafish *Danio rerio*, cathepsin D1 was shown to be expressed in all tissues and it is believed to be a typical housekeeping gene in the lysosome (Kurokawa et al. 2005; Riggio et al. 2000). However, cathepsin D2 was reported to be expressed specifically in the skin and gills of *T. rubripes* (Riggio et al. 2000) and the ovary of the mature red sea bream *Sparus aurata* (Carnevali et al. 1999). The presence of two types of paralogous cathepsin D was also reported in *D. rerio* and *S. maximus* and possibly due to entire genome duplication in teleosts (Mommssen 2004). Cathepsin D2 exists in teleosts and invertebrates and cathepsin D1 of teleost has a higher similarity to cathepsin D from higher vertebrates than to cathepsin D of teleost (Feng et al. 2011). This suggests that a duplication event occurred in the common ancestor of fish and tetrapods. Cathepsin D2 is subsequently lost in birds and mammals, thus explaining the tree topology (Feng et al. 2011). However, taken together, the basis of naming is still unclear.

Further phylogenetic analysis of cathepsin D from fish showed that they were grouped according to their habitat. Both cathepsin D1 and D2 clustered the fish based on marine/brackish or euryhaline and freshwater dwelling species. To date, this is the first study to report that cathepsin D from fish are divided into the type of habitat and the result were supported by high confidence values (Figure 3). This grouping pattern is similar to previous

studies of *L. calcarifer* immune related genes, including β -2-microglobulin and hepcidin and suggests that immune related genes evolved in response to the type of pathogens found in their environment (Lee et al. 2012; Mohd-Padil et al. 2010).

Overall, the topology showed that the *L. calcarifer* cathepsin D obtained in this study is derived from the same ancestral sequence, suggesting that the sequence is orthologous with other fish cathepsin D1 and could imply its similar functions. Reports of only cathepsin D1 is present in known fish species including *L. calcarifer* may be the result of insufficient study and it is possible that unidentified cathepsin D2 exists in other fish.

The up-regulation of cathepsin D expression in tested tissues (with the exception of the kidney) showed that it plays a certain function during *L. calcarifer* immune response. The result supported previous studies that reported the role of cathepsin D in fish immune response (Feng et al. 2011; Jia & Zhang 2009; Liu et al. 2012; Park et al. 1998). It has been suggested to be involved in various immune response activities, such as the activation of proteins destined for secretion and processing of antigens for presentation to the immune system (Mommssen 2004). Previous studies on the expression analysis of cathepsin D in the kidney have been reported in *I. punctatus* (Feng et al. 2011), *M. miiuy* (Liu et al. 2012) and *S. maximus* (Jia & Zhang 2009). The up-regulation of cathepsin D in those studies does not support the expression pattern of cathepsin D expressed in the kidney of *L. calcarifer*. The

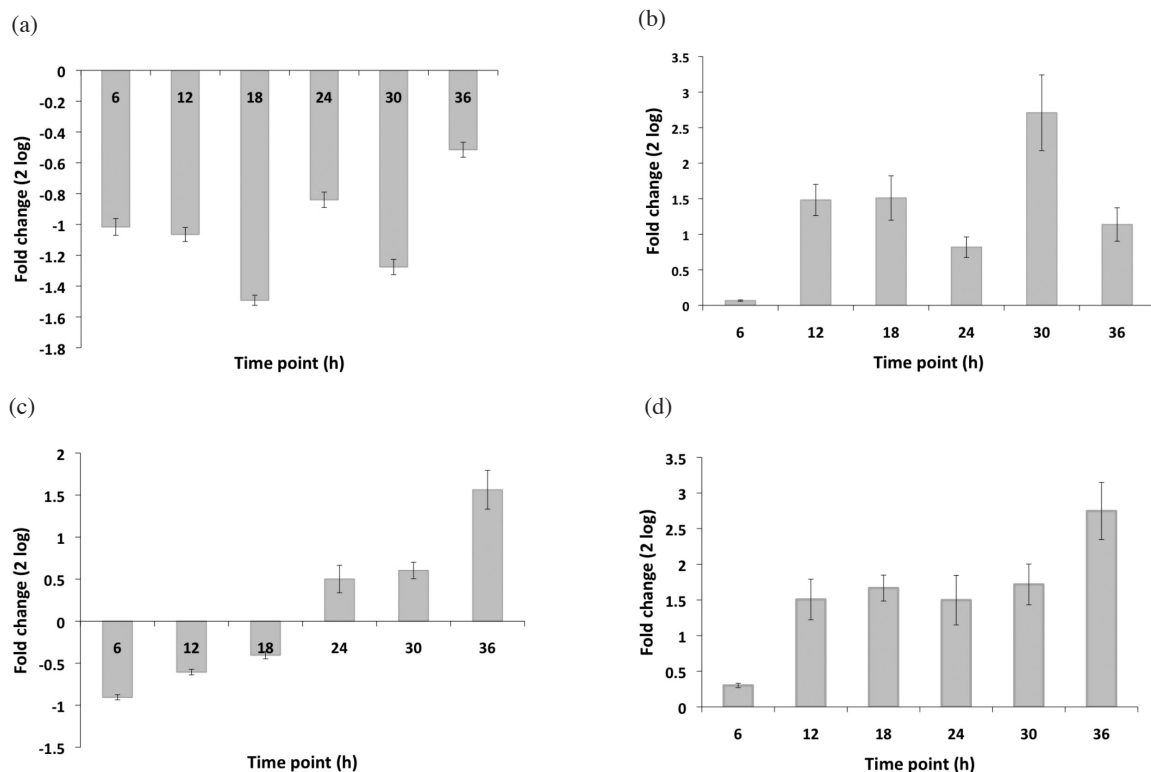


FIGURE 4. Expression profiles of the cathepsin D gene in different tissues of *L. calcarifer* (a) kidney, (b) gills, (c) liver and (d) spleen

down-regulation of cathepsin D in *L. calcarifer* is probably due to the abundance of iron regulator proteins in which the kidney functions in the osmoregulation of the body (Henderson et al. 1985).

In the gills, a mucosa-associated lymphoid tissue, the up-regulation observed was similar as reported in channel catfish *I. punctatus* (Feng et al. 2011). The expression level of cathepsin D fluctuated after the infection and reached the highest level during 30 h. In amur catfish *S. asotus*, it has been found that cathepsin D was highly expressed in the skin mucosa of wounded fish (Cho et al. 2002). Cathepsin D was reported to be involved in the production of parasin I, which contributes to the innate host defense of the fish against invading microorganisms (Mommensen 2004).

The up-regulation of cathepsin D in the liver was also reported in *I. punctatus* (Feng et al. 2011), *M. miiuy* (Liu et al. 2012) and *S. maximus* (Jia & Zhang 2009). In *L. calcarifer*, the expression level of cathepsin D was observed to increase from 24 h. The highest expression level of cathepsin D in *Silurus asotus* occurred in the liver tissue within 24 h after being infected (Feng et al. 2011). Cathepsin D was expressed in the liver 24 h post-infection suggesting its function in the degradation of various endocytosed proteins (e.g. albumin and myoglobin) that are correlated with antigen presentation. It was also suggested that cathepsin D is involved in the degradation of fibrinogen and in the release of phosphatase acid from its membrane-bound form into the lysosomal matrix (Hurley et al. 2000). In addition, cathepsin D has been implicated in the processing of prohormones and antigens in immune cells (Benes et al. 2008). The down-regulation of cathepsin D at the beginning of the challenge could be due to the abundant expression of the other stress proteins (Jia & Zhang 2009).

The up-regulation of cathepsin D observed in the spleen was consistent with its expression in *M. miiuy* (Liu et al. 2012) and *S. maximus* (Jia & Zhang 2009). The profile showed that cathepsin D had the highest expression in the spleen with approximately 2.8 fold change. The time-point of highest expression level supported previous study that reported the involvement of cathepsin D in the immune system (Jia & Zhang 2009). In the fish immune response, the increase of cathepsin D expression in the spleen suggests its function in the degradation of hemoglobin in the lysosomes and thus permitting reutilisation of the heme group (Riggio et al. 2000).

CONCLUSION

In conclusion, the results of this study showed that the cathepsin D of *L. calcarifer* is orthologous to the cathepsin D1 of other fish and has high similarity with the cathepsin D of other higher vertebrates compared to cathepsin D2 from fish. The cathepsin D sequences were phylogenetically grouped based on fish habitat indicating that its evolution may have been dependent on the type of pathogens present. The expression of cathepsin D in *L. calcarifer* was induced after challenged with *A. hydrophila*, suggesting that it plays a role in the immune response to microbial pathogens.

ACKNOWLEDGEMENTS

The authors thank the Marine Finfish Production and Research Center, Terengganu, Malaysia for technical assistance in fish sampling. This research was funded by the Ministry of Science, Technology and Innovation (MOSTI), Malaysia under Grant No. 07-05-MGI-GMB009.

REFERENCES

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25(17): 3389-3402.
- Baldocchi, R.A., Tan, L., King, D.S. & Nicoll, C.S. 1993. Mass spectrometric analysis of the fragments produced by cleavage and reduction of rat prolactin: Evidence that the cleaving enzyme is cathepsin D. *Endocrinology* 133: 935-938.
- Baricos, W.H., Zhou, Y.W., Fuerst, R.S., Barrett, A.J. & Shah, S.V. 1987. The role of aspartic and cysteine proteinase in albumin degradation by rat-kidney cortical lysosomes. *Archives of Biochemistry and Biophysics* 256(2): 687-691.
- Barret, A.J. 1977. Cathepsin D and other carboxyl proteinases. In *Proteinases in Mammalian Cells and Tissues*. New York: North Holland Publishing Company.
- Bendsten, J.D., Nielsen, H., von-Heijne, G. & Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* 340: 783-795.
- Benes, P., Vetvicka, V. & Fusek, M. 2008. Cathepsin D - Many functions of one aspartic protease. *Critical Reviews in Oncology/Hematology* 68: 12-28.
- Brooks, S., Tyler, C.R., Carnevali, O., Coward, K. & Sumpter, J.P. 1997. Molecular characterization of ovarian cathepsin D in the rainbow trout, *Onchorhynchus mykiss*. *Gene* 201(1-2): 45-54.
- Carnevali, O., Centonze, F., Brooks, S., Marota, I. & Sumpter, J.P. 1999. Molecular cloning and expression of ovarian cathepsin D in seabream, *Sparus aurata*. *Biology of Reproduction* 61: 785-791.
- Carnevali, O., Ciona, C., Tosti, L., Lubzens, E. & Maradonna, F. 2005. Role of cathepsin D in ovarian follicle growth and maturation. *General and Comparative Endocrinology* 146(3): 195-203.
- Cho, J.H., Park, I.Y., Kim, H.S., Lee, W.T., Kim, M.S. & Kim, S.C. 2002. Cathepsin D produces antimicrobial peptide parasin I from histone H2A in the skin mucosa of fish. *FASEB Journal* 16: 429-431.
- Chong, P.P., Mohd-Adnan, A. & Wan, K.L. 2011. Characterization of simple sequence repeats in the Asian Seabass, *Lates calcarifer* by random sequencing. *Sains Malaysiana* 40(5): 497-502.
- Chou, R. & Lee, H.B. 1997. Commercial marine fish farming in Singapore. *Aquaculture Research* 28: 767-776.
- Daskalov, H. 2006. The importance of *Aeromonas hydrophila* in food safety. *Food Control* 17: 474-483.
- Dobberstein, B. 1987. Structure and function of the signal recognition particle (SRP). *Molecular Biology Reports* 2(3): 213-217.
- Ewing, B. & Green, P. 1998. Base-calling of automated sequencer traces using Phred. II. error probabilities. *Genome Research* 8: 186-194.
- Feng, T., Zhang, H., Liu, H., Zhou, Z., Niu, D., Wong, L., Kucuktas, H., Liu, X., Peatman, E. & Liu, Z. 2011. Molecular

- characterization and expression analysis of the channel catfish cathepsin D genes. *Fish and Shellfish Immunology* 31(1): 164-169.
- Gilberg, A. 1988. Aspartic proteinases in fishes and aquatic invertebrates. *Comparative Biochemistry and Physiology Part B Biochemistry and Molecular Biology* 91: 425-435.
- Guindon, S., Dufayard, J., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* 59: 307-321.
- Hatha, M., Vivekanandhan, A.A., Joice, G.J. & Christol. 2005. Antibiotic resistance pattern of motile aeromonad from farm raised fresh water fish. *International Journal of Food Microbiology* 98(2): 131-134.
- Henderson, I.W., Hazon, N. & Hughes, K. 1985. Hormones, ionic regulation and kidney function in fishes. *Symposia of the Society Experimental Biology* 39: 245-265.
- Hurley, M.J., Larsen, L.B., Kelly, A.L. & McSweeney, P.L.H. 2000. The milk acid proteinase cathepsin D: A review. *International Dairy Journal* 10: 673-681.
- Jia, A. & Zhang, X.H. 2009. Molecular cloning, characterization and expression analysis of cathepsin D gene from turbot *Scophthalmus maximus*. *Fish and Shellfish Immunology* 26: 606-613.
- Khoo, C.K., Mohd-Adnan, A., Kua, B.C. & Abdul-Murad, A.M. 2009. Fabrication of *Lates calcarifer* cDNA microarray slide. *Sains Malaysiana* 38: 609-617.
- Krieger, T. & Hook, V.Y.H. 1992. Purification and characterization of a cathepsin D protease from bovine chromaffin granules. *Biochemistry* 31: 4223-4231.
- Kumar, R.S., Ijiri, S. & Trant, J.M. 2000. Changes in the expression of genes encoding steroidogenic enzymes in the channel catfish (*Ictalurus punctatus*) ovary throughout a reproductive cycle. *Biology of Reproduction* 63: 1676-1682.
- Kurokawa, T., Uji, S. & Suzuki, T. 2005. Identification of pepsinogen gene in the genome of stomachless fish, *Takifugu rubripes*. *Comparative Biochemistry and Physiology Part B Biochemistry and Molecular Biology* 140: 133-140.
- Le, S.Q. & Gascuel, O. 2008. An improved general amino acid replacement matrix. *Molecular Biology and Evolution* 25(7): 1307-1320.
- Lee, J.H., Wan, K.L. & Mohd-Adnan, A. 2012. Molecular characterization of hepcidin in the Asian seabass (*Lates calcarifer*) provides insights into its innate immune response. *Aquaculture* 330-333: 8-14.
- Liu, X., Shi, G., Cui, D., Wang, R. & Xu, T. 2012. Molecular cloning and comprehensive characterization of cathepsin D in the Miiuy croaker *Miichthys miiuy*. *Fish and Shellfish Immunology* 32: 464-468.
- Metcalf, P. & Fusek, M. 1993. Two crystal structures for cathepsin D: The lysosomal targeting signal and active site. *The EMBO Journal* 12(4): 1293-1302.
- Mohamed-Jawad, L.A.H., Rabu, A., Mohamed, R. & Mohd-Adnan, A. 2012. Phylogenetic characterization and the expression of recombinant C-reactive protein from the Asian seabass (*Lates calcarifer*). *Aquaculture* 338-341: 13-22.
- Mohd-Padil, H., Tajul-Arifin, K. & Mohd-Adnan, A. 2010. Characterization of the functional domain of β 2-microglobulin from the Asian seabass, *Lates calcarifer*. *PLoS One* 5(10): e13159.
- Mommsen, T.P. 2004. Salmon spawning migration and muscle protein metabolism: The August Krogh Principle at work. *Comparative Biochemistry and Physiology Part B Biochemistry and Molecular Biology* 139(3): 383-400.
- Mohd-Yusof, N.Y., Hoh, C.C., Mohd-Adnan, A. & Wan, K.L. 2009. Identification of immune-related genes by analysis of spleen expressed sequences tags from the Asian seabass, *Lates calcarifer*. *Sains Malaysiana* 38(6): 939-945.
- Nelson, J. 1994. *Fishes of the World*. New Jersey: John Wiley & Son.
- Nielsen, B.L. & Nielsen, H.H. 2001. Purification and characterization of cathepsin D from herring muscle (*Clupea harengus*). *Comparative Biochemistry and Physiology Part B Biochemistry and Molecular Biology* 128(2): 351-363.
- Park, I.Y., Park, C.B., Kim, M.S. & Kim, S.C. 1998. Parasin I, an antimicrobial peptide derived from histon H2A in the catfish, *Parasilurus asotus*. *FEBS Letters* 437: 258-262.
- Pfaffl, M.W., Horgan, G.W. & Dempfle, L. 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research* 30: 1-10.
- Riggio, M., Sscudiero, R., Filosa, S. & Parisi, E. 2000. Sex- and tissue-specific expression of aspartic proteinases in *Danio rerio* (zebrafish). *International Journal of Genes and Genomes Evolution* 260: 67-75.
- Rojo, L., Sotelo-Mundo, R., Garcia-Carreno, F. & Graf, L. 2010. Isolation, biochemical characterization, and molecular modeling of American lobster digestive cathepsin D1. *Comparative Biochemistry and Physiology Part B Biochemistry and Molecular Biology* 157(4): 394-400.
- Ronquist, F. & Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731-2739.
- Tan, S.L., Mohd-Adnan, A., Mohd-Yusof, N.Y., Forstner, M.R.J. & Wan, K.L. 2008. Identification and analysis of a preprochicken gonadotropin releasing hormone II (preprocGnRH-II) precursor in the Asian seabass, *Lates calcarifer*, based on an EST-based assessment of its brain transcriptome. *Gene* 411: 77-86.
- Shariza Azizan, Kiew-Lian Wan* & Adura Mohd-Adnan
School of Biosciences and Biotechnology
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor D.E.
Malaysia
- Kiew-Lian Wan*
Malaysia Genome Institute, Jalan Bangi
43000 Kajang, Selangor, D.E.
Malaysia
- *Corresponding author; email: klwan@ukm.edu.my

Received: 18 May 2013

Accepted: 26 November 2013