View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by AJOL - African Journals Online

academicJournals

Vol. 14(29), pp. 2282-2299, 22 July, 2015
DOI: 10.5897/AJB2014.14001
Article Number: E7E03D954258
ISSN 1684-5315
Copyright © 2015
Author(s) retain the copyright of this article
http://www.academicjournals.org/AJB

African Journal of Biotechnology

Full Length Research Paper

Diversity analysis of the immunoglobulin M heavy chain gene in Nile tilapia, *Oreochromis niloticus* (Linnaeus)

Chiraporn Phuyindee[1,4], Sasimanas Unajak[2] and Prapansak Srisapoome[3,4]*

[1]Interdisciplinary Graduate Program in Genetic Engineering, Kasetsart University, Bangkok, Thailand.
[2]Department of Biochemistry, Faculty of Sciences, Kasetsart University, Bangkok, Thailand.
[3]Department of Aquaculture, Faculty of Fisheries, Kasetsart University, Bangkok, Thailand.
[4]Center for Advanced Studies for Agriculture and Food, Kasetsart University Institute for Advanced Studies, Kasetsart University, Bangkok, Thailand 10900 (CASAF, NRU-KU, Thailand).

Received 25 June, 2014; Accepted 13 July, 2015

A full-length cDNA encoding the immunoglobulin (IgM) heavy chain gene of Nile tilapia was successfully cloned using the 5' and 3' RACE techniques. The complete cDNA of the Nile tilapia IgM heavy chain gene is 1,921 bp in length and has an open reading frame (ORF) of 1,740 bp, which corresponds to 580 amino acid residues. The deduced amino acid sequence of the Nile tilapia IgM heavy chain includes a typical secretory IgM heavy chain designated "*On*-sIgM" and a variable region that is connected to 4 constant regions to form the L_H-V_H-$C_\mu 1$-$C_\mu 2$-$C_\mu 3$-$C_\mu 4$ pattern. Comparisons of the nucleotide and amino acid sequences of *On*-sIgM with IgM heavy chains of other organisms showed the highest similarity scores of 62.6 and 55.4%, respectively, to the orange-spotted grouper (*Epinephelus coioides*). Structural analysis of 126 cDNAs encoding variable domains of the IgM heavy chain revealed that at least 9 V_H families, 6 D_H segments and 4 J_H families were utilized using several mechanisms to generate the repertoire of antigen-binding domains. Variation analysis of the variable domains indicated that the amino acid sequences of the framework regions (FRs) were less variable than those of the complementarity determining regions (CDRs), among which the most variable was CDR3. Tissue expression profile analysis using quantitative real-time RT-PCR of healthy Nile tilapia showed that the IgM heavy chain gene was ubiquitously expressed in all 13 tested tissues, but the highest expression level was observed in the head kidney, followed by the spleen, intestine and peripheral blood leukocytes (PBLs). Furthermore, Southern blot analysis of the constant region of the IgM heavy chain gene of 3 different fishes indicated that Nile tilapia genomes may contain 2 copies of the IgM gene.

Key words: Nile tilapia, IgM heavy chain, variable region, diversity, secreted form, southern blot.

INTRODUCTION

Nile tilapia (*Oreochromis niloticus*) is a freshwater fish that is cultured worldwide and is an important economic aquatic animal. The global production of tilapia was 3.6 million tons in 2011, and yearly increases are thought to continue. In 2010, tilapia production was focused in 2 leader countries (China and Egypt), with productions of 1,331,890 and 557,049 tons, respectively. In 2014, 174,872 tons were produced in Thailand as result of

increased aquaculture throughout the country, ranging from earthen ponds to intensive cage-culture systems, making Thailand the 5th largest world tilapia producer (FAO, 2014). This activity has also been promoted by related industries and supply chains, such as hatcheries, feed manufacturers and distributors, to support the huge domestic consumption. However, this intensive culture system has a high risk of disease outbreaks. The bulk of fish deaths occurs due to protozoan, fungal and bacterial infections, typically due to *Aeromonas hydrophila, Streptococcus agalactiae* and *Flavobacterium columnare*, and result in lost tilapia yields in culture systems (Mohamed and Refat, 2011; Pridgeon et al., 2011; Rodkhum et al., 2011). To circumvent these problems, fish immunity must improve to increase the efficiency of disease prophylactic and therapeutic methods. The immune systems of vertebrates govern homeostasis, prevention and surveillance and are generally divided into 2 parts: innate and adaptive immune systems. The adaptive or acquired immune system, which was discovered in cartilaginous fish, is distinguished from the innate immune system by antibody (immunoglobulin) production by plasma B cells and the functions of cytotoxic T cells associated with the degranulation process. Antibody production and degranulation are potent and effective methods used to specifically eliminate pathogenic infection. Immunoglobulins are important molecules in jawed vertebrates, ranging from gnathostomes to tetrapods, but are not found in invertebrates (Flajnik, 2002; Flajnik and Du Pasquier, 2004). The fundamental functions of immunoglobulins include toxic neutralization, the promotion of phagocytosis by opsonization and activation of the complement system (Walport, 2001; Holland and Lambris, 2002). An immunoglobulin molecule is composed of 2 heavy chains and 2 light chains that are joined by inter- and intra-disulfide bonds.

In bony fish, 3 major isotypes of immunoglobulins exist: IgM, IgD and IgT/IgZ (Hikima et al., 2011; Salinas et al., 2011). Immunoglobulin isotypes are determined by the constant region (C_H), which also dictates the effector function of the molecule in different types of immune responses. The variable region contains the antigen-binding site (Roitt et al., 2001) and is located at the N-terminus of the heavy and light chains. Heavy chains are composed of a variable segment (V_H), a diversity segment (D_H), which is not found in the light chain, and a joining region (J_H). The variable regions of the heavy and light chains consist of 4 framework regions (FR1-4) and 3 complementarity determining regions (CDR1-3), or hyper-variable regions. The CDRs are highly variable in Nucleotide sequence because of considerable contact with antigens (Pilstrom and Bengten, 1996).

The mechanisms used to generate the diverse immunoglobulins in higher vertebrates can be summarized into at least 7 events that consist of combinatorial diversity, junctional imprecision, junctional diversity, gene conversion, secondary $V_{H/L}$ gene recombination, somatic hypermutation and heavy/light chain pairing. During B cell development, the diversity of antigen-binding elements begins with rearrangement mediated by recombination-activating gene (*RAG*), which initiates the assembly of the antigenic binding domain of immunoglobulins, Artemis (DNA repair proteins) and terminal deoxynucleotidyl transferase (TdT), which are utilized for P and N nucleotide addition, respectively (Lieber, 1992; Kuo and Schlissel, 2009).

In Osteichthyes, secreted IgM (sIgM) is one of the major proteins in the serum and is generated during immune responses against pathogenic infection. IgM is classified as the primordial immunoglobulin of the adaptive immune response and is found in monomeric and tetrameric forms in circulating blood (Acton et al., 1971; Wilson and Warr, 1992). IgM can exist in 2 forms, sIgM and membrane-bound (mIgM), which are generated *via* alternative RNA splicing of the primary transcript of the μ gene (Ross et al., 1998). sIgM consists of the variable region and 4 constant domains in the heavy chain, whereas mIgM contains variable region, 3 constant domains and 2 additional transmembrane domains (T_M1 and T_M2) and acts as a B cell receptor for initial antigen binding (Dylke et al., 2007). To date, the cloning and characterization of the IgM gene has intensively been reported in holostean, cartilaginous and teleost fish (Rauta et al., 2012). However, information about the mechanisms important for generating diversity for antigen binding is reported in some teleost fish but still lacking in Nile tilapia.

The aim of this study was to increase the understanding of the teleost immune system, specifically IgM, which is the most vital humoral molecule for adaptive immune responses.

This study performed molecular characterization of the full-length cDNA of the IgM heavy chain gene of Nile tilapia and the diverse expression of its variable domain were intensively investigated. In addition, tissue distribution analysis was performed using quantitative real-time RT-PCR, and genomic structural analysis of the gene was performed using Southern blot analysis. Information from the current study may provide a better understanding of the adaptive immune system of Nile tilapia.

*Corresponding author. E-mail: ffispssp@ku.ac.th. Tel: +662 579 2924; Fax: +662 561 3984.

Author(s) agree that this article remain permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Table 1. Oligonucleotide primers used for PCR analysis.

Gene name	Oligonucleotide primer	Sequence 5' → 3'	Amplicon size	Experiment
IgM heavy chain	- IgMuF1	AGGAGACAGGACTGGAATGCACAA	-	3' RACE-PCR
IgM heavy chain	- IgMuR1	TTGTGCATTCCAGTCCTGTCTCCT	-	5' RACE-PCR, Variable domain analysis
IgM heavy chain	- IgMuF2	GGATGATACCTATACTGCCTCCTG	174	Real-time PCR
IgM heavy chain	- IgMuR22	AATCTAGTCTGATCATTCAGGTCA	174	Real-time PCR
ß-actin	- ß-actinF2	ACAGGATGCAGAAGGAGATCACAG	155	Real-time PCR
ß-actin	- ß-actinR2	GTACTCCTGCTTGCTGATCCACAT	155	Real-time PCR
-	- UPM-long	CTAATACGACTCACTATAGGGCAAGCATGG TATCAACGCAGAGT	-	RACE-PCR
-	- UPM-short	AAGCAGTGGTATCAACGCAGAGT	-	RACE-PCR
Constant region of IgM heavy chain	- SB F	GGATGATACCTATACTGCCTCCTG	533	Southern blot
Constant region of IgM heavy chain	- SB R	GGTGAACAACACAGAAGCGTGT	533	Southern blot

MATERIALS AND METHODS

Experimental animals

Healthy adult Nile tilapia weighing 500 to 600 g were obtained from the Department of Aquaculture, Faculty of Fisheries, Kasetsart University. The fish were maintained in aerated water tanks and fed with commercial feed twice a day for a week.

Cloning of the full-length cDNA of the IgM heavy chain gene of Nile tilapia

Total RNA from the head kidney and spleen of an adult Nile tilapia was extracted using TRIzol reagent (Gibco BRL, USA) according to the manufacturer's instructions. The mRNAs were consequently prepared using a QuickPrep Micro mRNA Purification Kit (Amersham Biosciences, USA). Five hundred micrograms of mRNA from each organ were pooled, 1 µg of mixed mRNA was used per reaction, and 5' and 3' first-strand cDNA were synthesized using the BD Smart RACE cDNA Amplification Kit (Clontech, USA). The cDNAs were then used as templates for 5' and 3' RACE PCRs, which were conducted using the specific primers IgMuF1 and IgMuR1, respectively (Table 1). These primers were designed from the EST clone HK0156 encoding the partially constant region of the Nile tilapia IgM (GenBank accession no. FF279636). The PCR conditions included pre-denaturation for 5 min at 95°C; 25 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 90 s; and a final elongation step at 72°C for 5 min. The 5' and 3' RACE PCR products were purified using the HiYield™ Gel/PCR Fragments Extraction Kit (RBC Bioscience, Taiwan), ligated into the pGEM T-Easy cloning vector (Promega, USA) and transformed into *Escherichia coli* strain JM 109, which was grown on Luria Bertani (LB) agar containing ampicillin (0.01 g/mL), IPTG (100 mM) and X-gal (50 mg/mL). Each plate of transformants was incubated at 37°C for 18 h. Positive clones, that is, white colonies, were selected, and plasmids were extracted using the Plasmid DNA Extraction Manual Kit (Bio Excellence, Thailand). Nucleotide sequencing of the selected clones in the 5' and 3' directions was performed by Macrogen, Inc. (Korea) using the M13F and M13R primers with the Thermo Sequence Fluorescent Labeled Primer Cycle Sequencing Kit (Amersham Pharmacia Biotech).

Characterization of the full-length cDNA of the IgM heavy chain gene of Nile tilapia

After sequencing, the nucleotide sequences were screened for vector contamination, and vector sequences were removed using VecScreen (http://www.ncbi.nlm.nih.gov/VecScreen/ VecScreen.html). The nucleotide sequences from the 3' and 5' fragments were multiply aligned to find overlapping regions and compared with nucleotide and amino acid sequences of other vertebrate IgMs in the GenBank database using the BLASTN and X programs (http://blast.ncbi.nlm.nih.gov/ Blast.cgi). The coding sequences, conserved residues and signature motifs that were important for immunoglobulin functions and structure were determined using the IMGT (International ImMunoGeneTics Information System) database (http://www.ebi.ac.uk/imgt/) and other publications. The full-length cDNA of the IgM heavy chain in Nile tilapia was examined to predict its signal peptide sequence using the SignalP 4.0 Server (http://www.cbs.dtu.dk/services/ SignalP/). The similarity and identity of the nucleotides and amino acids of the IgM heavy chain of Nile tilapia and other vertebrates were calculated using the MatGat 2.02 program (http://bitincka.com/ledion/matgat).

Phylogenetic analysis

The deduced amino acid sequences of the constant domain of the IgM heavy chain, consisting of the Cµ2-Cµ4 domains, in Nile tilapia and other vertebrates (gnathostomes to mammals) were multiply aligned using CLUSTALW. Cµ1 was excluded because it is known as the high conserved region resulted from the evolutionary duplicated to generate other C regions. The IgD heavy chains of Mandarin fish, *Siniperca chuatsi* (ACO88906), and grouper fish, *Epinephelus coioides* (AEN71108), were used as outgroups for the phylogenetic tree. Then, the evolution of the IgM heavy chain gene was determined using the UPMGA method by performing 1,000 bootstrap resampling replicates with the MEGA program, version 5.05 (http://www.megasoftware.net).

Construction of a cDNA library of the variable domain of the IgM heavy chain gene

The cDNA library was constructed using 5' RACE PCR with the

specific primer IgMuR1 (Table 1). Briefly, the previously prepared, ready-to-use, first-strand cDNA template for 5' RACE PCR was amplified, cloned and sequenced using the same protocols described above.

Diversity analysis of the variable domain

After sequencing, the entire nucleotide sequences of randomly selected clones were analyzed for homology with other sequences available in the GenBank database using the BLASTN and BLASTX programs, as previously described. A representative sequence from each redundant group was arbitrarily chosen for further family classification. The resulting 126 cDNA sequences were analyzed to find the leader sequence, FR and CDR according to the IMGT standardization numbering. Each of the V_H families, D_H segments and J_H families was classified using the CAP3 program (http://bioweb.pasteur.fr/seqana/ interface/cap3. html). The V_H family was grouped based on the percentages of nucleotide sequence identity in the same V_H family greater than 80% (Brodeur and Riblet, 1984). Then, the similarity and identity of the V_H, D_H and J_H amino acids were calculated using MatGat 2.02, and multiple alignments were performed using the CLUSTALW program. To examine the degree of sequence variability in the variable region of Nile tilapia IgM, the deduced amino acid sequences were multiple-aligned and calculated as the position variability using the Kabat and Wu method (Kabat and Wu, 1971) and Shannon analysis (Stewart et al., 1997).

Tissue distribution of IgM heavy chain gene by quantitative real-time PCR

Total RNA from the brain, gills, gonad, heart, head kidney, intestine, liver, muscle, skin, spleen, stomach, peripheral blood leukocytes and trunk kidney of a healthy Nile tilapia was extracted using TRIzol reagent (Gibco BRL, USA). The contaminating genomic DNA was digested with RNAse-free DNAse I (Fermentas, USA), and first-strand cDNA synthesis was performed using 1 µg of total RNA from each tissue with the RevertAid First Strand cDNA Synthesis Kit (Fermentas, USA). First-strand cDNA from the 13 tissues was quantitatively examined using the IgMuF2 and IgMuR22 primers (Table 1). The expression levels were normalized to the expression level of beta-actin mRNA using the β-actinF2 and β-actinR2 primers (Table 1). The quantitative real-time RT PCR was conducted using an Mx Pro™ 3005P QPCR (Stratagene, USA), and the mRNA expression of the IgM heavy chain and beta-actin genes was detected using Brilliant II SYBR Green qPCR Master Mix (Stratagene, USA). The cycling conditions consisted of 95°C for 10 min, followed by 40 cycles of 95°C for 30 s, 55°C for 1 min and 72°C for 1 min. For each sample, 3 replicates were performed for the IgM heavy chain and β-actin mRNAs. Standard curves were constructed to examine the efficiency and specificity of both specific primer sets. PCR efficiencies for IgM heavy chain and β-actin genes were 2.015 and 2.023, respectively. The relative expression ratio of the IgM heavy chain gene in Nile tilapia was calculated according to the $2^{-\Delta\Delta C_T}$ formula (Livak and Schmittgen, 2001). Statistical analysis was performed using the SPSS program, version 13.0. Differences in the expression levels of the Nile tilapia IgM heavy chain gene in the 13 tissues were analyzed using one-way analysis of variance (ANOVA), and the means were compared using Duncan's new multiple range test. The significance level was established at $P<0.05$.

Southern blot analysis

Genomic DNA was isolated from the whole blood of 3 different Nile tilapias and was subjected to phenol-chloroform treatment, as described by Taggart et al. (1992). Ten micrograms of DNA from each fish were completely digested with the *Eco* RI and *Pst* I restriction enzymes, and electrophoresis in a 1% agarose gel was performed to separate the DNA fragments. The DNA fragments were then transferred to a nitrocellulose membrane using the capillary blotting method with 20X SSC, and the membrane was dried and baked at 80°C for 2 h in a hot-air oven. Probes specific for the Cµ2-Cµ3 constant regions were prepared by PCR using the designed primers SBF and SBR (Table 1). PCR probes were labeled with Digoxigenin-11-dUTP using the DIG-High Prime DNA Labeling and Detection Starter Kit I (Roche, Germany) according to the instruction manual. The membranes were incubated in hybridization solution (DIG Easy Hyb) with denatured, DIG-labeled DNA probe at 68°C overnight. After hybridization, the membranes were stringently washed twice in ample 2X SSC, 0.1% SDS at 25°C for 5 min and in 0.5X SSC, 0.1% SDS at 68°C for 15 min under constant agitation. Then, immunological detection of the membrane was carried out following the procedure recommended by the manufacturer. Finally, color detection with NBT/BCIP was performed to investigate the intensity of the bands by photography.

RESULTS

Cloning and characterization of a full-length cDNA encoding the IgM heavy chain gene in Nile tilapia

A complete full-length cDNA of the IgM heavy chain gene in Nile tilapia was successfully cloned using 3' and 5' RACE PCR. The full-length cDNA was 1,921 nucleotides in length and composed of a 45-nucleotide 5' untranslated region (UTR) that was followed by the open reading frame (ORF) beginning with ATG, the first translated codon. The length of the ORF was 1,740 bp and encoded 580 amino acids, and the leader peptide was predicted to consist of 26 amino acids. Translation terminated at nucleotide position 1,786, which encoded TAG, the stop codon. The length of the 3' UTR was 90 nucleotides and included the polyadenylation signal (AATAAA) and poly A tail (Figure 1). The deduced amino acid sequence of the Nile tilapia IgM heavy chain gene included a typical heavy chain sequence for secretory IgM, which was termed "*On*-sIgM". Its organization began with 1 variable region and 4 constant regions that formed a L_H-V_H-$C_\mu1$-$C_\mu2$-$C_\mu3$-$C_\mu4$ pattern, which is different from the teleost fish IgM membrane-bound form that is generally rearranged as L_H-V_H-$C_\mu1$-$C_\mu2$-$C_\mu3$-T_M1-T_M2 (Saha et al., 2005; Tian et al., 2009). The potential N-linked glycosylation sites were found as NSS in the Cµ2, NKT in the Cµ3 and 2 NTTs in the Cµ4 domain (Figure 1). Comparisons of *On*-sIgM (GenBank accession number KC677037) with known IgM heavy chain cDNAs of other higher vertebrates showed that the nucleotide identity scores were between 38.0 to 47.3% and the amino acid identity and similarity scores ranged from 24.2 to 28.8% and 45.7 to 51.5%, respectively (Table 2). On the other hand, comparisons of *On*-sIgM with known IgM heavy chain cDNAs of other cartilaginous and teleost fishes indicated that the nucleotide identity scores were

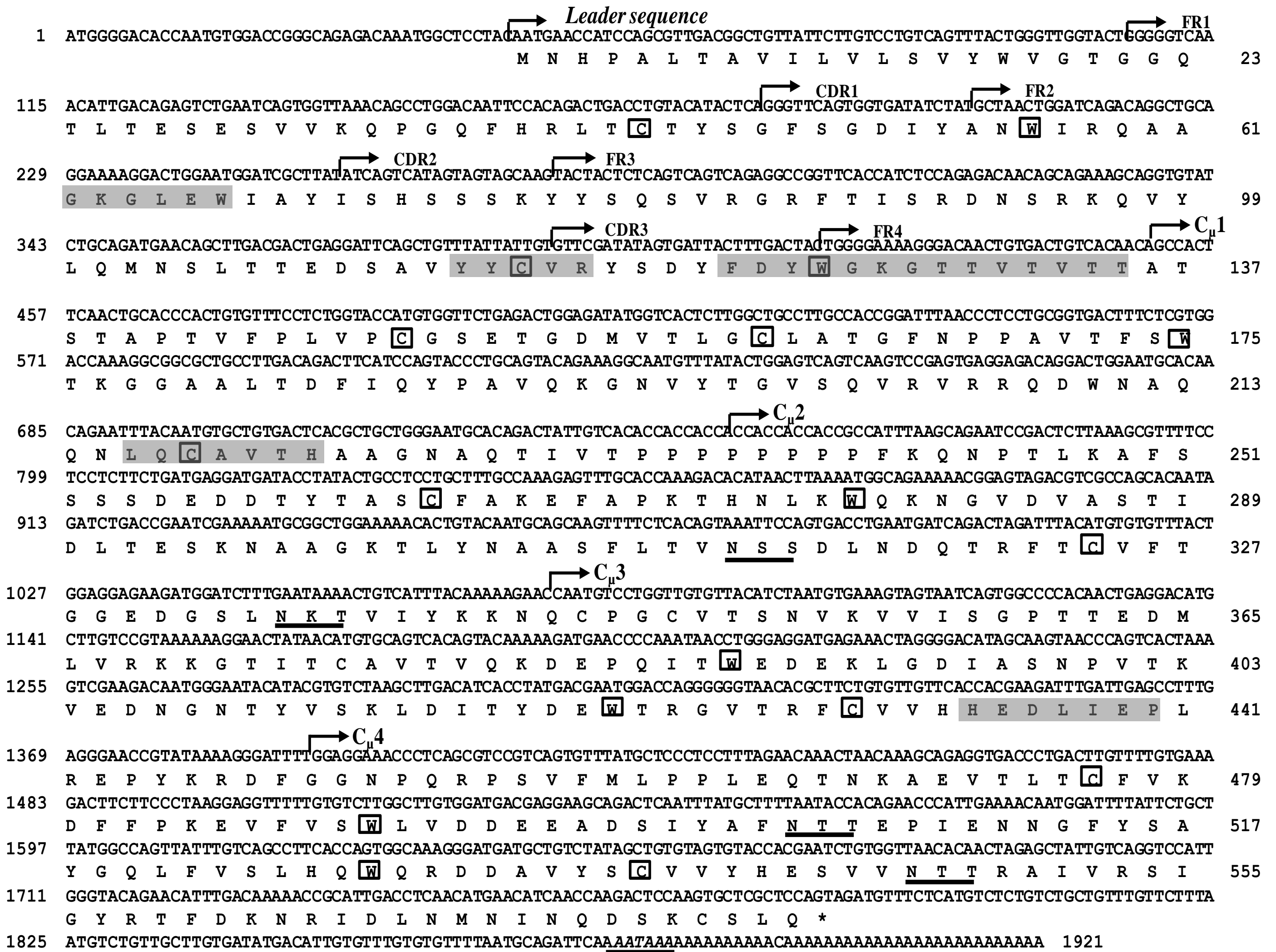

Figure 1. Nucleotide and deduced amino acid sequences of *On*-sIgM. The predicted amino acid sequence is marked under the nucleotide sequence. The conserved cysteine and tryptophan residues are boxed. The conserved blocks GKGLEW at FR2, YYCVR at FR3 and FDYWGKGTTVTVTT at FR4 and the immunoglobulin signature motif, LQCAVTH, are highlighted in gray. Four potential glycosylation sites are underlined. A typical polyadenylation signal, AATAAA, is italicized and underlined. The TAG stop codon is indicated with an asterisk.

between 38.0 to 67.0% and the amino acid identity and similarity scores ranged from 26.4 to 55.4% and 46.1 to 75.0%, respectively. Noticeably, the greatest amino acid similarity to *On*-sIgM (75.0%) was with the closely related orange-spotted grouper, *E. coioides* (Table 2).

Evolutionary relationship between the Nile tilapia IgM heavy chain gene and other vertebrates

The relationship between the Nile tilapia IgM heavy chain gene and other vertebrates was examined by phylogenetic analysis using the deduced amino acid of the IgM heavy chain constant region, Cμ2-Cμ4. In the evolutionary tree, all of the IgM heavy chain genes were clearly separated from the IgD heavy chain genes of the Mandarin and grouper fishes, which were used as the outgroups of the tree. The tree could be split into 2 major clusters that included superclasses Tetrapoda and Pisces. The first group (superclass Tetrapoda) was composed of human, dolphin, cow, mouse, rat, platypus, salamander, duck, chicken, turtle and newt. Interestingly, classes Chondrichthyes (cartilaginous fish) and Sarcopterygii (lobe-finned fish; lungfish) were also grouped into this branch. Only class Osteichthyes (bony fish) was grouped into the second group. *On*-sIgM was classified into the group of Osteicthyes in superclass Pisces and was closely related to the orange-spotted grouper (order Perciformes), which was also similar based on homology analysis (Figure 2).

Table 2. Comparisons of Nile tilapia IgM heavy chain sequences with those of other vertebrates.

Name	Accession number	Identity (%)		Similarity (%)
		Nucleotide	Amino acid	
Higher vertebrates				
Human	CAA47708	40.2	28.8	47.0
Dolphin	AAG40853	40.2	26.9	45.7
Cow	AAN60017	47.3	26.4	46.2
Mouse	CAC20701	41.2	25.8	46.2
Platypus	AA037747	38.0	28.2	46.9
Salamander	CAE02685	41.4	27.2	47.6
Duck	AAA68605	39.3	24.2	51.5
Chicken	CAA25762	38.8	25.9	47.9
Cartilaginous fish				
Antarctic skate	ACU11614	47.7	28.2	46.1
Nurse shark	AAT76789	38.1	28.9	49.0
Teleost fish				
African lungfish	AAO52809	38.0	26.4	47.2
Long nose gar	AAC59688	41.9	33.9	52.1
Bowfin	AAC59687	52.1	34.4	54.0
European eel	ABM87939	52.0	34.7	57.5
Zebrafish	AAT67447	43.5	34.2	55.1
Grass Carp	ABD76396	43.6	34.1	54.2
Haddock	CAH04753	46.1	34.8	55.5
Rainbow trout	AAB27359	50.6	41.2	64.9
Atlantic salmon	AAB24064	50.8	40.3	64.0
Japanese pufferfish	BAD26619	52.2	45.1	63.6
Snakehead	ACF49353	60.3	48.5	67.2
Atlantic halibut	AAF69488	56.0	49.2	65.4
Japanese flouder	BAB60868	56.6	49.2	66.1
Orange-spotted grouper	AAX78211	62.6	55.4	75.0
Mandarin fish	AAQ14845	60.2	52.2	67.4
Tristan klipfish	ACH87158	67.0	51.6	69.6
Black rockcod	AAL99934	58.2	49.1	65.9
Antarctic fish	ABW77756	56.5	48.6	68.2
Ploughfish	ABY54906	57.9	48.8	66.1
Antarctic fish	ABW77754	56.8	48.6	65.5
Blackfin icefish	AAL99930	64.4	49.9	67.2
Antarctic fish	ABW81218	57.6	48.7	66.5

Structural and diversity analyses of the variable domain of the IgM heavy chain gene in Nile tilapia

The putative V_H, D_H and J_H segments of the non-redundant 126 cDNA clones (GenBank accession number KC708098- KC708223) could be classified into 9 families, 6 segments and 4 families, respectively, based on the percent nucleotide identity. In the V_H domain classification, the range of nucleotide identity for each V_H family was between 51.5 to 66.5%. The nucleotide and amino acid identity between the clones within each family ranged from 80.1 to 99.7% and 80.2 to 99.1%, respectively. Families V_H II and V_H IV were more frequently employed than other families because they showed utilized frequencies of 30.2 and 26.9%, respectively. The V_H I, III, V, VI and VII families exhibited utilized frequencies of 10.3, 18.2, 7.9, 3.2 and 1.6%, respectively; however, the V_H VIII and V_H IX families possessed only

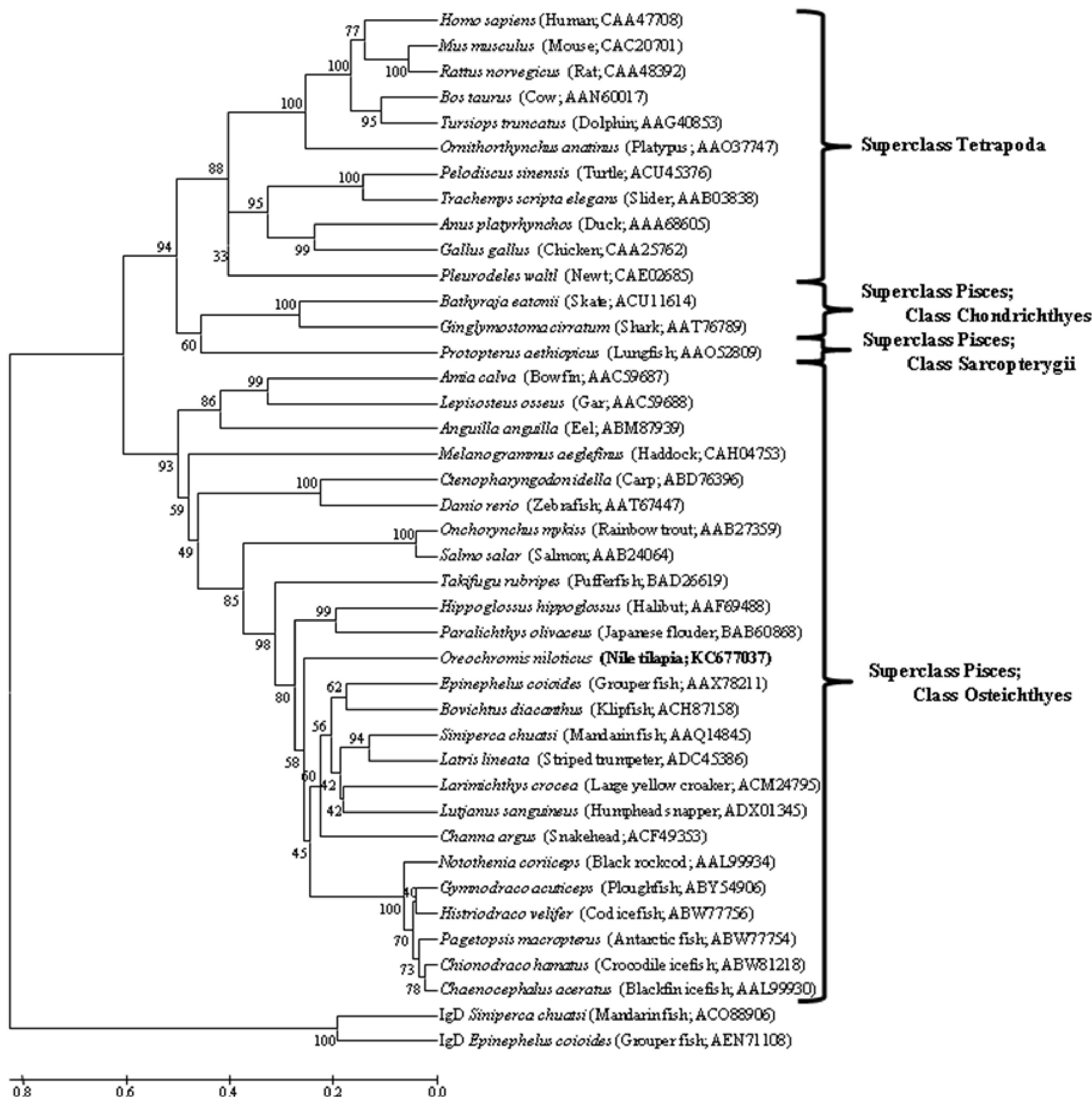

Figure 2. Phylogenetic tree showing the relationships between the Nile tilapia IgM heavy chain amino sequences and those of other known vertebrates (Cμ2-Cμ4). The numbers at the relevant branches refer to bootstrap values of 1,000. Values indicate the percentage along the branch. Common names and accession numbers for the sequences are indicated in parentheses behind their scientific names. The IgD heavy chain ($C\delta_x$-δ_n) of the Mandarin fish *Siniperca chuatsi* (Basilewsky, 1855) (ACO88906) and grouper fish *Epinephelus coioides* (Hamilton, 1822) (AEN71108) were used as outgroups.

1 clone (0.8%) that was used for V_H gene rearrangement of the IgM heavy chain (Figure 3). Arbitrary classification of the D_H segments placed them into 6 groups, with the core nucleotide sequences in each group as follows: GCGGCG, TGGGA, GGCTAC, GGTGCT, GACGAA and TACAA. Additionally, P and N nucleotide additions were investigated for the entire group. In particular, palindromic sequence additions were discovered in the following

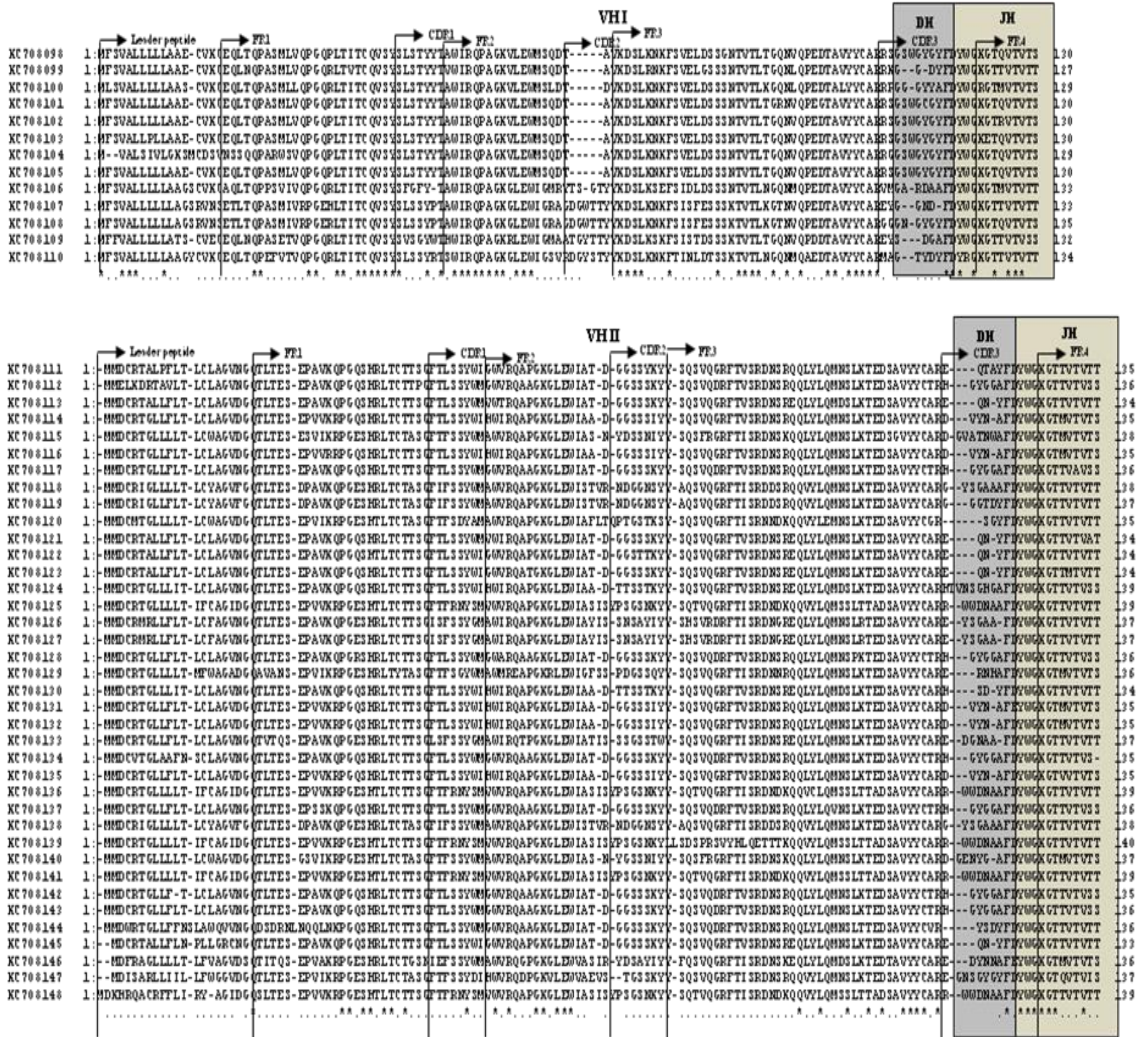

Figure 3. Classification of the variable region of the Nile tilapia IgM heavy chain. The FR and CDR domains were identified using the IMGT database. Dashes indicate gaps that were introduced for maximal alignment, and the amino acid identity is indicated by asterisks. Dots indicate residues conserved in most sequences. The accession number of each sequence is shown before the first point of the sequence.

forms: GCGC, CGCG, GCGGCGCCGC, GCTAGC, CGTACG, CCGG, CAGCTG, AGCT, ACGT, TTTAAA, GGATCC and AATT. The highest utilized frequency was observed for the D_H VI segment (28.6%), which coincided with the highest redundancy clones (21.4%), while the D_H IV segments showed the lowest utilized frequency of 3.2% (Figure 4).

The boundary of the J_H segment was determined as the first codons encoding the FDYWG motif; as a result, the J_H segment could be classified into 4 major groups. The J_HIII segment in the FDYWGKGTTVTVTT form was the most frequently (43.7%) used in rearrangements, while the least frequently used segment was J_HI (7.9%) (Figure 5). Moreover, the diversity and variance of the CDR3

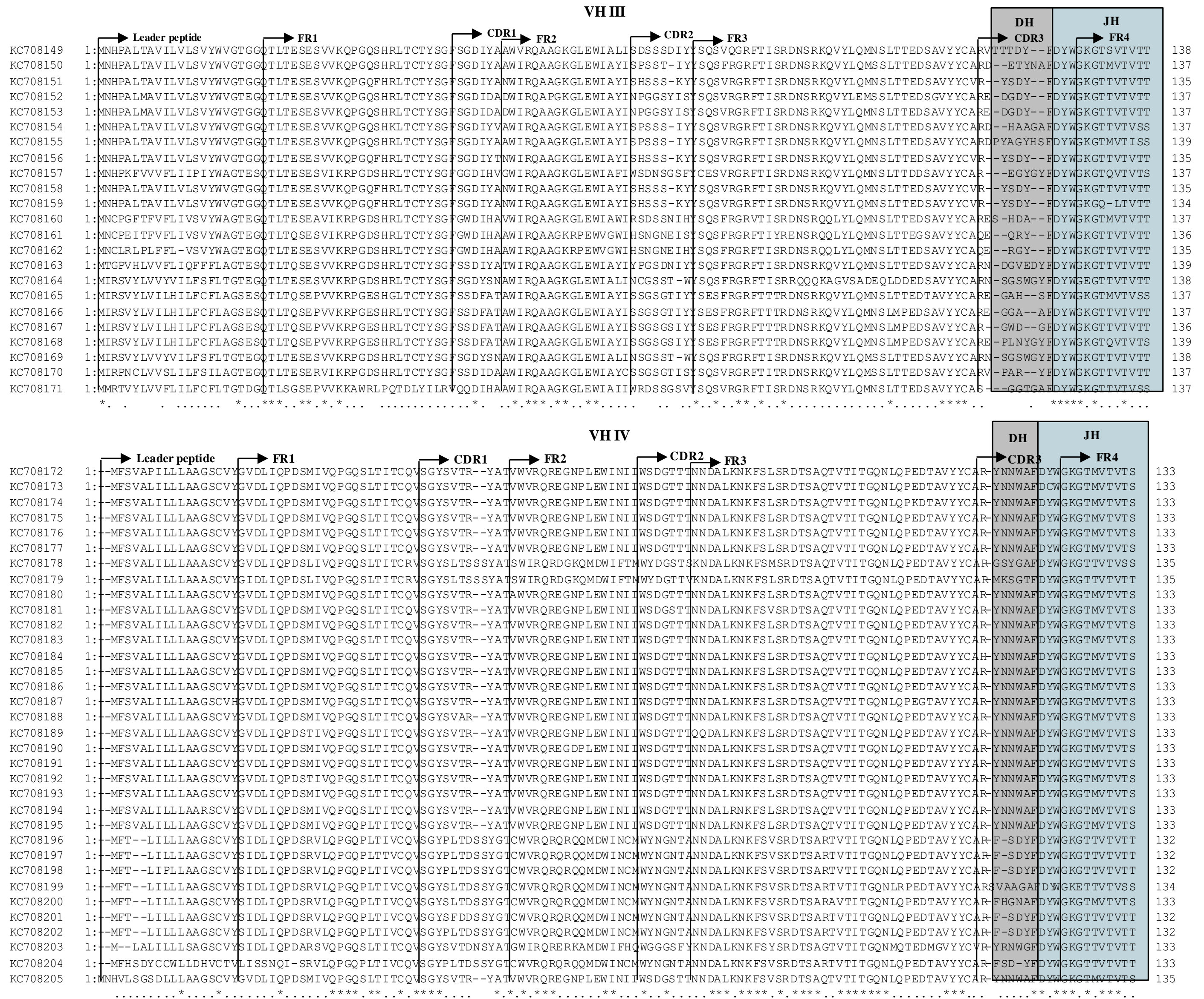

Figure 3. Contd.

region of 126 cDNA clones encoding the Nile tilapia IgM heavy chain gene were examined. The results showed that its nucleotide length ranged from 24 to 42 bp (8 to 14 amino acid residues), and 10 amino acid residues of CDR3 were the most frequently used to create the diverse repertoire of the variable domain. The average length and length variability of the CDR3 region were 10.97 and 1.99, respectively (Figure 6).

Additionally, diversity analysis of the variable domain of the Nile tilapia IgM heavy chain was relatively characterized using the Kabat and Wu method and Shannon analysis, which are general mathematical tools used to estimate variability. The result of these methods coincidently indicated that the amino acid sequences of the FRs were distinctively less variable than those of the CDRs; in particular, CDR3 showed the highest variability at position 106, followed by CDR2 and CDR1 (Figure 7).

Tissue distribution of the Nile tilapia IgM heavy chain gene

Quantitative real-time RT PCR analyses of the expression profile of the Nile tilapia IgM heavy chain gene indicated that the mRNA transcripts were expressed in 13 tissues. The highest expression level was observed in the head kidney (12.4-fold greater compared to the brain), which significantly differed from

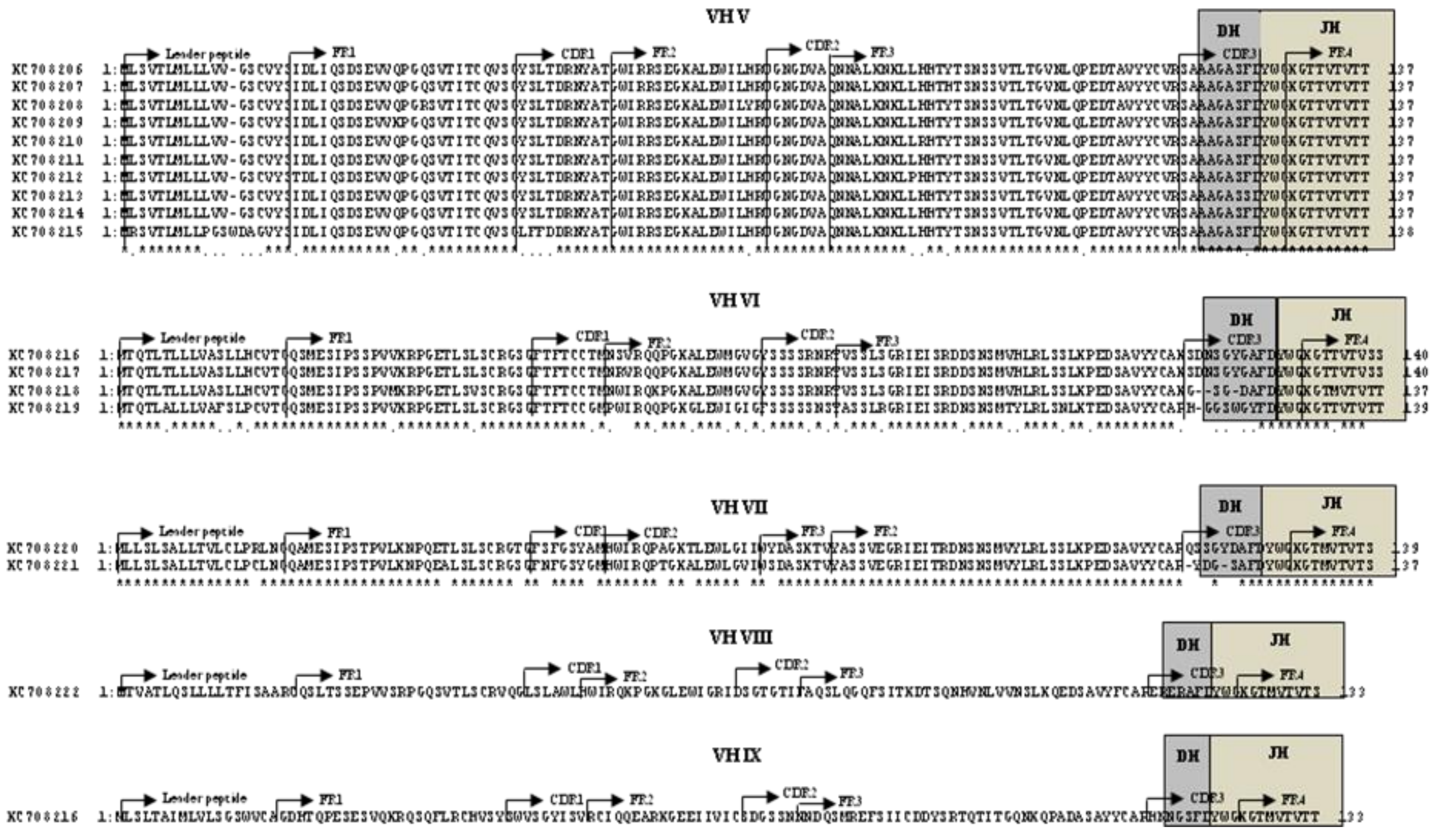

Figure 3. Contd.

that of other tissues, and was followed by the spleen, intestine and peripheral blood leukocytes (PBLs). Low expression was observed in the muscle and heart (Figure 8).

Southern blot analysis

Structural analysis of the constant region of the Nile tilapia IgM heavy chain gene was performed after digesting the genomes from 3 different fishes with the *Eco* RI and *Pst* I restriction enzymes. The hybridization of a specific probe ($C_\mu2$-$C_\mu3$ exon) illustrated that the bands appeared in the same pattern and with the same size and intensity in each fish genome. The sizes of these bands were approximately 20 and 7 kb for the *Eco* RI digestion and 22 and 10 kb for the *Pst* I digestion (Figure 9).

DISCUSSION

Molecular cloning and characterization of a cDNA encoding the IgM heavy chain gene in Nile tilapia revealed that 3' RACE PCR only amplified 1 distinctive band and this band was identical to only the secreted form of the IgM heavy chain. In this experiment, the specific primer that was used for the 3' RACE PCR was designed based on the Cµ1 region. Theoretically, the mIgM, IgD and IgZ/T transcripts, which normally contain the Cµ1 domain (Hikima et al., 2011), must be simultaneously amplified with sIgM, but we could not identify all of them in this experiment. It is possible that those mRNA levels were low in the spleen and head kidney, that most mRNAs were primary IgM-IgD transcripts or that the lengths of these mRNAs were too long to amplify. Additionally, these mRNAs may have short half-lives. Based on our these results, attempts to find other forms of IgM and other heavy chain isotypes failed with the currently used techniques, indicating that sIgM was the most abundant group of immunoglobulin heavy chain transcripts. Thus, further study is needed for a more complete understanding of the immunoglobulin heavy chains in this fish. The organization of *On*-sIgM found in this current study was rearranged to form a leader sequence and a variable region, which was followed by a constant region (Cµ1-Cµ4). This type of rearrangement is present in the secreted and soluble IgM forms found in the circulating blood. In contrast, the teleost IgM membrane-bound form, which possesses a transmembrane (TM) domain, contains a constant region that is rearranged as Cµ1-Cµ3-TM. To compare the

Accession	VH Tyr Cys (Y C)	DH	JH Phe Asp Tyr (F D Y)
		DH I	
KC708218	TAC TGT GCC AAA	GGAAGCGGCGATGCT	TTT GAC TAC
KC708166	TAT TGT GCT CGA	GAAGGCGGCGCTGCT	TTT GAT TAC
KC708126	TAT TGT GCC CGA	GAGTATAGCGGCGCTGCT	TTT GAT TAC
KC708118	TAT TGT GCT CGG	GGGTATAGCGGCGCCGCTGCT	TTT GAT TAC
KC708215	TAT TGT GTG CGT	AGCGCGGCAGCTGGGGCTAGC	TTT GAC TAC
KC708138	TAT TGT GCT CGG	GGGTATAGCGGCGCCGCT	TTT GAT TAC
KC708142	TAT TGT ACT CGA	CACGGATACGGCGGTGCT	TTT GAC TAT
KC708110	TAC TGT GCC AGA	ATGGCGGGAACGTACGACTAC	TTT GAC TAC
KC708165	TAT TGT GCT CGA	GAAGGGGCGCATTCT	TTT GAC TAC
KC708107	TAC TGT GCC AGA	GAGTATGGCGGGAACGAC	TTT GAC TAC
		DH II	
KC708106	TAC TGT GCC AGA	GTGATGGGAGCGCGGGACGCTGCT	TTT GAC TAC
KC708148	TAT TGT GCT CGA	CGCTGGTGGGATAACGCTGCT	TTT GAT TAC
KC708133	TAT TGT GCC CGA	GAGGATGGGAACGCTGCT	TTT GAT TAC
KC708221	TAC TGT GCC AGA	TACGATGGGAGCGCT	TTT GAC TAC
KC708167	TAT TGT GCT CGA	GGCTGGGACGGT	TTT GAC TAC
KC708163	TAT TGT GCC AGA	AACGATGGGGTGGAGGACTAC	TTT GAC TAC
KC708203	TAC TGT GTG AGA	TACCGGAACTGGGGC	TTT GAC TAC
		DH III	
KC708108	TAC TGT GCC AGA	GGAGGCGGGAACGGGTATGGCTAC	TTT GAC TAC
KC708155	TAT TGT GCT CGA	GACCCCTACGCGGGCTACCATTCT	TTT GAC TAC
KC708120	TAT TGT GGT CGA	AGCGGCTAC	TTT GAC TAC
KC708162	TAT TGT GCC CAA	GAGCGCGGCTAC	TTT GAC TAC
KC708100	TAC TGT GCC AGA	AGGAGaggggtGGCTACTACGCT	TTT GAC TAC
KC708169	TAC TGT GCT CGG	AATAGTGGCAGCTGGGGCTAC	TTT GAC TAC
KC708115	TAT TGT GCT CGA	GACGGAGTGGCTACGAACTGGGC	TTC GAC TAC
KC708220	TAC TGT GCC AGA	CAGAGCAGTGGCTACGACGCT	TTT GAC TAC
KC708216	TAC TGC GCC AAA	TCGGATAACAGTGGCTACGGTGCT	TTT GAC TAT
KC708147	TAT TGT GCT CGA	GagggtAACAGTGGCTATGGCTAC	TTT GAC TAC
KC708219	TAC TGT GCC AGA	CACGGTGGCAGCTGGGGCTAC	TTT GAC TAC
KC708157	TAT TGT GCT CGA	GAGGGCTATGGCTAC	TTT GAC TAC
KC708168	TAT TGT GCT CGA	GAACCTTTAAACTATGGCTAC	TTT GAC TAC
KC708104	TAC TGT GCC AGA	aggggtGGTAGCTGGGGCTATGGGTAC	TTT GAC TAC
KC708102	TAC TGT GCC AGA	aggagtGGTAGCTGGGGCTATGGGTAC	TTT GAC TAC
KC708111	TAT TGT GCT CGA	GAGCAAACGGCCTAC	TTT GAC TAC
KC708170	TAC TGT GCT CGA	GTTCCGGCGCGCTAC	TTT GAC TAC
KC708161	TAT TGT GCC CAA	GAGCAGCGCTAC	TTT GAC TAC
KC708119	TAT TGT GCT CGG	GGTGGTGGAACCGACTAC	TTT GAC TAC
KC708152	TAT TGT GCT CGA	GAGGACGGAGACTAC	TTT GAC TAC
KC708196	TAC TGT GCA CGT	TTCTCCGACTAC	TTT GAC TAC
KC708099	TAC TGT GCC AGA	AGGAAAGGTGGCGACTAC	TTT GAC TAC
KC708149	TAT TGT GCT CGA	GTCACTACTACGGACTAC	TTT GAC TAC
KC708130	TAT TGT GCT CGA	CACAGGGACTAC	TTT GAC TAC
KC708159	TAT TGT GTT CGA	TATAGTGATTAC	TTT GAC TAC
		DH IV	
KC708199	TAC TGT GCA CGT	TCAGTGGCAGCTGGTGCT	TTT GAC TAT
KC708124	TAT TGT GCT CGA	CACACTGTTAACAGTGGCCACGGTGCT	TTT GAC TAT
KC708109	TAT TGT GCT CGA	GAGTATAGCGACGGTGCT	TTT GAC TAT
KC708171	TAC TGT GCC AGT	GGTGGAACCGGTGCT	TTT GAC TAT
KC708154	TAT TGT GCT CGA	GACCACGCAGCGGGTGCT	TTT GAC TAT
KC708178	TAC TGT GCA CGT	GGATCCTACGGTGCT	TTT GAC TAT
KC708160	TAT TGT GCC AGA	GAGAGCCACGATGCT	TTT GAC TAC
		DH V	
KC708150	TAT TGT GCT CGA	GACGAAACGTACAACGCT	TTT GAC TAC
KC708121	TAT TGT GCT CGA	GAGCAAAATTAC	TTT GAC TAC
KC708140	TAT TGT GCT CGA	GACGGGGAAAATTACggggct	TTT GAC TAC
KC708222	TAC TGT GCA CGG	GAGCGGGAACGCGCT	TTT GAC TAC
KC708179	TAC TGT GCA CGT	ATGAAGAGCGGCACC	TTT GAC TAC
KC708129	TAT TGT GCT CGG	GAGCGGAACCACGCT	TTT GAC TAC
		DH VI	
KC708205	TAC TGT GCA CGT	TATAACAACTGGGCT	TTT GAC TAC
KC708223	TAC TGC GCT AGA	CATAACAACGGGTCT	TTT GAC TAC
KC708132	TAT TGT GCT CGA	GATGTATACAACGCT	TTT GAC TAC
KC708135	TAT TGT GCT CGA	GACGTATACAACGCT	TTT GAC TAC
KC708146	TAT TGT GCT CGA	GAGGATTACAACAACGCT	TTT GAC TAC

Figure 4. Classification of the VH/DH/JH junctions. Core DH nucleotides are shaded in gray. Palindromic sequences are underlined, and inverted (D-D joining) sequences are shown in small letters. The accession number of each sequence is shown before the first point of the sequence..

amino acid sequences of *On*-sIgM to those of other vertebrates, we selected the Cμ1-Cμ4 regions for multiple alignments, while the Cμ2-Cμ4 regions were used for phylogenetic analyses because the Cμ1 gene was

JH segments

```
JH I
KC708098  TTTGACTACTGGGGGAAAGGAACACAAGTCACAGTAACTTCT
KC708102  -------------------------G----------------
KC708147  -------------------------------------T----
KC708103  -------------------A----------------------

JH II
KC708205  TTTGACTACTGGGGGAAAGGTACAATGGTTACAGTCACATCA
KC708173  -------G----------------------------------
KC708177  -----------------------G------------------
KC708115  --C---------------------------------------
KC708131  -----G--------------C--------C--G--T------
KC708223  --------------------G------------------A--
KC708114  --------------------C--------C--G--T------
KC708135  --------------------C---G----C--G--T------
KC708150  --------------A-----C--------C--G--T------
KC708100  ----------------G------------------C------
KC708155  --------------------------------TA-C-GT---
KC708218  --------------------------------T------A--
KC708106  --------------------------------T--C---A--
KC708165  --------------------------------T--C-GT---

JH III
KC708111  TTTGACTACTGGGGAAAAGGGACAACTGTGACTGTCACAACA
KC708159  -----------------G--ACA-CTGACTGTCACA---T--
KC708163  --------------------------C---------------
KC708164  ---------------G--------------------------
KC708213  --C---------------------------------------
KC708120  --------------G---------------------------
KC708121  ------------------------------------G-----
KC708123  ---------------------------A--------------
KC708162  --------------------A---------------------
KC708110  ---------A--------------------------------
KC708149  ------------------------T-C---------------

JH IV
KC708112  TTTGACTATTGGGGAAAGGGGACAACAGTCACTGTTTCATCA
KC708118  -----T--C--------A--A--C--G--------CA--A--
KC708117  ------------------------------G-----------
KC708199  -------------------A----------------------
```

Figure 5. Nucleotide sequences of the four putative J_H segments. Nucleotide identities of the J_H III segment are indicated as dashes. Nucleotide deletion is indicated as highlight letter and nucleotide addition is underlined (for clone KC708159 compared with clone KC708111). The accession number of each sequence is shown before the first point of the sequence.

evolutionally duplicated to form other C genes (Bengten et al., 2002).

The results of the evolutionary relationship analysis showed that the IgM heavy chain of cartilaginous fish was grouped in the higher vertebrate cluster because these primitive fish do generate humoral immune responses (Coscia et al., 2012). Interestingly, the cartilaginous fish IgM heavy chain gene of the lung fish (subclass Dipnoi) was also clustered in the higher vertebrate branch because the lung fish is closely related to crossopterygians (coelacanths), which are ancestors of amphibians and higher vertebrates. Likewise, Ota et al.

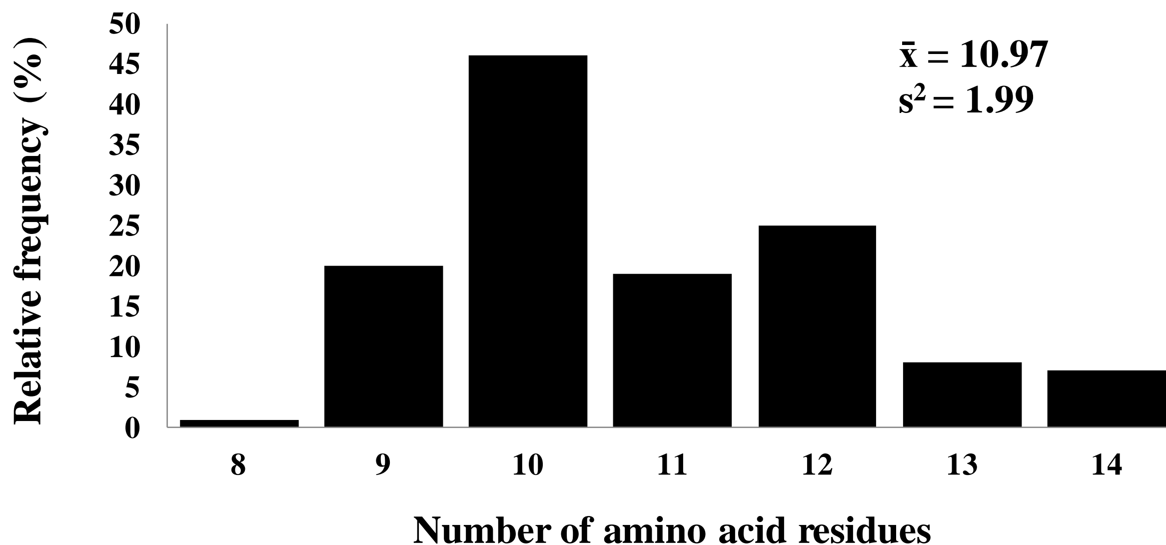

Figure 6. The length distributions of the CDR3 regions. The CDR3 regions were calculated from the 126 amino acid sequences of the variable domain of the IgM heavy chain. The average number and variance of the amino acid number are indicated by $\bar{x}$ and s^2, respectively.

Figure 7. Variability plots of the 126-amino-acid sequences of the variable domains of the IgM heavy chain of Nile tilapia. The calculation was performed using the variability formulas of the (a) Kabat and Wu method (Kabat and Wu, 1971) and (b) Shannon analysis (Stewart *et al*. 1997).

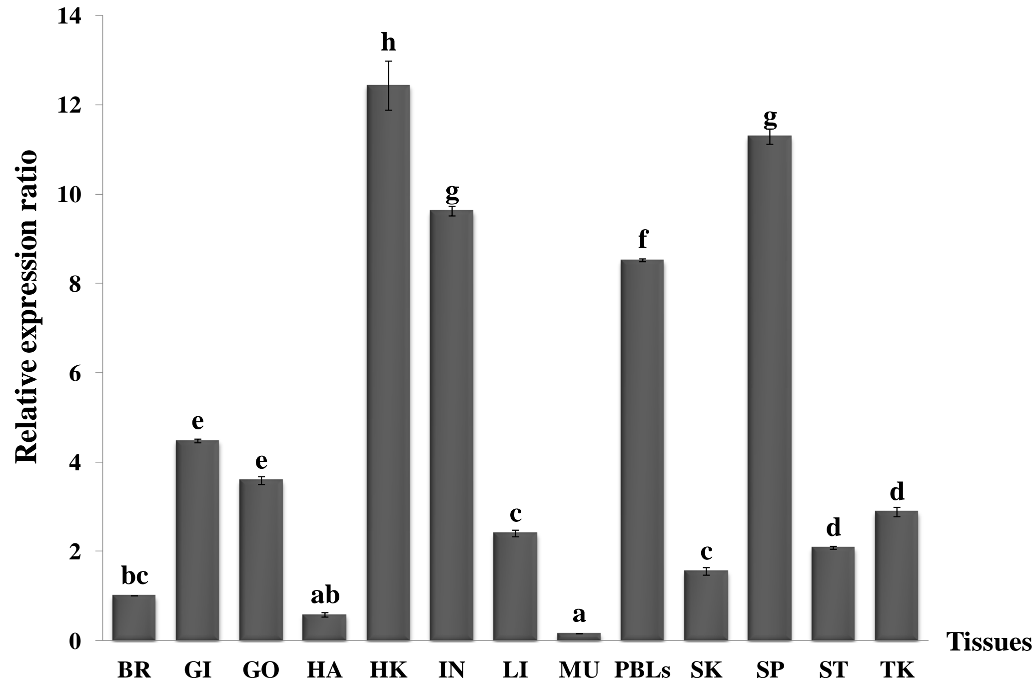

Figure 8. Quantitative real-time PCR analysis of the IgM heavy chain in 13 tissues of Nile tilapia. Significant differences are indicated with different letters on each bar (P < 0.05). BR; brain, GI; gills, GO; gonad, HA; heart, HK; head kidney, IN; intestine, LI; liver, MU; muscle, PBLs; peripheral blood leukocytes, SK; skin, SP; spleen, ST; stomach, TK; trunk kidney.

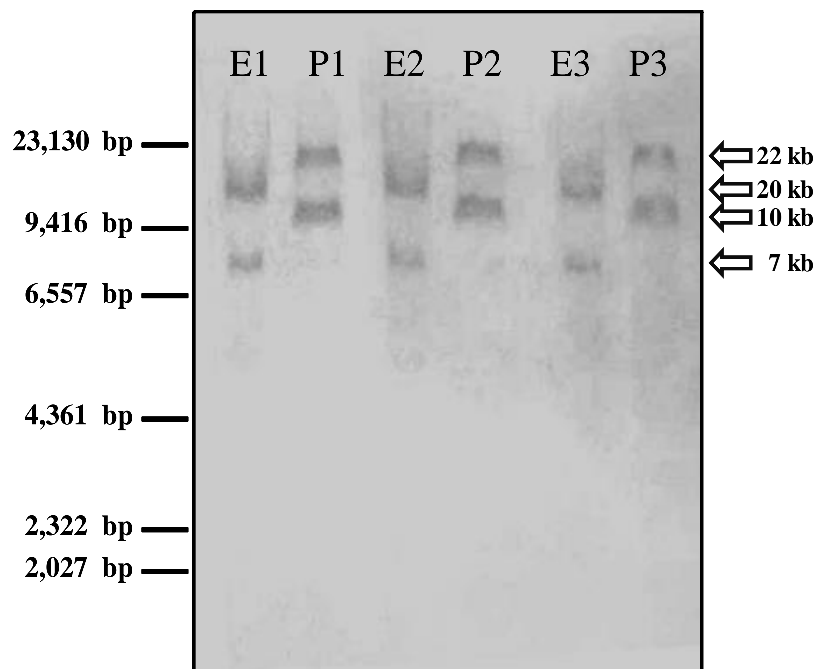

Figure 9. Southern blot hybridization of the Cμ2-Cμ3 constant region of the IgM heavy chain gene in Nile tilapia. Genomic DNA of 3 different fishes was isolated from whole blood and digested with the Eco RI (E) and Pst I (P) restriction enzymes. The band sizes were estimated by comparison with the lambda Hind III ladder shown on the left.

(2002) found that the IgM heavy chain genes of lung fishes (lobe-finned fishes) were closely related to those of tetrapods rather than neopterygians, which are primitive bony fish including bowfins and sturgeons.

When constructing the variable domain cDNA library of *On*-sIgM, the primer was designed in the Cμ2 region to eliminate contamination with the variable domains of IgD transcripts. The classification of putative V_H, D_H and J_H segments indicated that the V_H II family (30.2%) was an important group for the IgM heavy chain gene repertoire. The V_H II family demonstrated different initiation codons, AUGAUG, compared with other families, and it was suggested that these codons may encourage simpler and faster translation initiation, in accordance with the report of Kozak (1998) and Coscia and Oreste (2003). Recently, many studies on the diversity of the variable domain of the immunoglobulin heavy chain have analyzed the mRNA and genomic levels in higher and lower vertebrates. Interestingly, these compiled data have demonstrated that the numbers of V_H family members in higher vertebrates tend to be low, although some organisms, such as mouse and frog, demonstrate large numbers of V_H family members. However, the V_H family numbers in teleost fish were somewhat higher than those of higher vertebrates, possibly because the diversity generation mechanisms of the variable domain of the IgM heavy chain in higher vertebrates is more variable than that of teleost fish. Hence, large numbers of V_H family members were necessary to generate diversity of the variable domain in teleost fish, especially in Nile tilapia.

Recently, many studies on the diversity of the variable domain of the Ig heavy chain gene have demonstrated the mRNA and genomic levels in higher and lower vertebrates (Table 3). Most of the V_H families in teleost fish have been mainly studied at the transcriptional level. Nile tilapia demonstrated 9 V_H families, which was moderate amount compared to those of vertebrate V_H families (Table 3). These numbers were close to human, that is, 7 V_H families (Matsuda et al., 1998) and an Atlantic charr, that is, 8 V_H families (Andersson and Matsunaga, 1998). Atlantic salmon revealed the greatest number of V_H families, that is, 18 families (Yasuike et al., 2010) followed by zebrafish, channel catfish and rainbow trout, that is, 14, 13 and 13 families, respectively (Danilova et al., 2005; Yang et al., 2003; Brown et al., 2006). Furthermore, when 9 V_H families of Nile tilapia were analyzed a distinctive distribution of P and N nucleotide addition, inversion (D-D joining) and nucleotide deletion was observed in the generation of putative D_H segment diversity, similar to the report of Hsu et al. (1989) and Coscia and Oreste (2003). Additionally, the J_H segments were rarely diverse, with the nucleotide sequences differing by only 1 to 9 residues, and these differences may be the effect of allelic variants (Stenvik et al., 2000). Moreover, nucleotide addition and deletion of the J_H III segment by 2 essential enzymes, TdT and RAG, were also observed in clone number KC708159 (Figure 6). CDR3 region analysis demonstrated that the diversity of the CDR3 regions had a rather low exhibition level. Commonly, the length distribution and variance values of the CDR3 regions in higher vertebrates are higher than those of lower vertebrates, especially cold-blooded vertebrates. Short CDR3 regions in cold-blooded vertebrates may restrict diversity generation of the antigen-binding site of antibody molecules (Roman et al., 1995). Additionally, highly specific affinity binding with an antigen was discovered for CDR3 regions with high variability (Casali and Schettino, 1996; Kabat and Wu, 1991). Diversity analysis of the variable domain of *On*-sIgM indicated that the amino acid sequences of the FRs were less variable than those of the CDRs, with the greatest variability observed for CDR3.

Comparison of variable domain residues using Shannon's method and Kabat and Wu analysis showed that the calculated amino acid variability (according to the Kabat and Wu method) was more sharply shown in the variable pattern. Based on the current data, the diversity generation mechanisms of the variable domain repertoire of *On*-sIgM were likely obtained from combinatorial diversity, junctional imprecision and junctional diversity. At least 9 V_H families, 6 D_H segments and 4 J_H families of *On*-sIgM were used to generate diversity through random linkage resulting from RAG enzyme activity (Tonegawa, 1983). Moreover, Artemis and TdT may be used to promote junctional diversity through the deletion and addition of P and N nucleotides at the $V_H/D_H/J_H$ junction site (CDR3 region). However, it may be expected that other mechanisms, such as somatic hypermutation, secondary $V_{H/L}$ gene recombination and heavy/light chain pairing, occur to increase the antigen binding capability, antibody diversity and antigen recognition in Nile tilapia immune responses.

In this experiment, quantitative real-time RT-PCR was employed to study the expression profile of *On*-sIgM in various tissues of Nile tilapia. The highest expression level was found in the head kidney, followed by the spleen, intestine and PBLs, as these organs act as major lymphoid organs. Generally, the head kidney, spleen and intestine are acknowledged as hematopoietic tissues that play crucial roles in blood cell generation (Abbas et al., 2007). Moreover, blood-borne antigens are stored at the germinal center within the spleen, where multiple defense mechanisms emerge to recognize and neutralize antigens using specific antibodies (Grontvedt and Espelid, 2003; Saha et al., 2005). Therefore, these organs may provide larger population numbers of pro-B cells, pre-B cells, immature B cells and mature B cells than other organs. However, we found *On*-sIgM transcripts in other non-lymphoid organs, which suggested that large numbers of mature B cells may normally circulate and infiltrate into these organs (Mao et al., 2012).

Surprisingly, Southern blot analysis of the constant region of the IgM heavy chain gene in 3 different fishes indicated that the Nile tilapia diploid genome might

Table 3. The number of V_H gene families in teleost fish and other vertebrates.

Vertebrate species	V_H gene families	Study levels		References
		Transcriptional	Genomic	
Nile tilapia	9	/		This study
Pufferfish	2		/	Peixoto and Brenner, 2000
Zebrafish	14	/		Danilova et al., 2005
Emeral rockcod	2	/		Coscia and Oreste, 2003
Channel catfish	13	/	/	Yang et al., 2003
Atlantic cod	4	/		Stenvik et al., 2000
Atlantic salmon	18	/		Yasuike et al., 2010
Rainbow trout	13	/		Brown et al., 2006
Atlantic charr	8	/		Andersson and Matsunaga, 1998
Goldfish	3		/	Wilson et al., 1991
Sturgeon	3		/	Lundqvist et al., 1998
Nurse shark	5	/		Rumfelt et al., 2004
Frog	11	/		Haire et al., 1990
Chicken	1		/	Ota and Nei, 1995
Rabbit	1		/	Mage et al., 1984
Pig	1		/	Sun et al., 1994
Mouse	15		/	Mainville et al., 1996
Human	7		/	Matsuda et al., 1998
Nile tilapia	9	/		This study
Pufferfish	2		/	Peixoto and Brenner, 2000
Zebrafish	14	/		Danilova et al., 2005
Emeral rockcod	2	/		Coscia and Oreste, 2003
Channel catfish	13	/	/	Yang et al., 2003
Atlantic cod	4	/		Stenvik et al., 2000
Atlantic salmon	18	/		Yasuike et al., 2010
Rainbow trout	13	/		Brown et al., 2006
Atlantic charr	8	/		Andersson and Matsunaga, 1998
Goldfish	3		/	Wilson et al., 1991
Sturgeon	3		/	Lundqvist et al., 1998
Nurse shark	5	/		Rumfelt et al., 2004
Frog	11	/		Haire et al., 1990
Chicken	1		/	Ota and Nei, 1995
Rabbit	1		/	Mage et al., 1984
Pig	1		/	Sun et al., 1994
Mouse	15		/	Mainville et al., 1996
Human	7		/	Matsuda et al., 1998

contain 2 copies of this gene. This finding was confirmed by cloning, sequencing and a search for *Eco* RI or *Pst* I restriction sites within the intron linking the Cμ2 and Cμ3 exons, which were not found between these exons. Moreover, the nucleotide length between the Cμ2 and Cμ3 exons was determined to be approximately 104 bp (GenBank accession no. KJ558374). Generally, the length of the intron between the Cμ2 and Cμ3 exons of the IgM gene in teleost fish is not larger than 4 kb (Bengten et al., 2002; Srisapoome et al., 2004). Hence, these results indicate that Nile tilapia might possess a pseudo Cμ gene, similar to the channel catfish *Ictalurus punctatus*, or a true second Cμ gene cluster to increase the diversity of the immunoglobulin heavy chain gene.

Conclusion

The results of recent studies imply crucial functional roles for IgM, for which diversification at variable domains is generated through a number of variations that increase antigen recognition through the actions of the TdT and

RAG enzymes. Our work suggests that fish such as the Nile tilapia may possess additional Cμ loci in their genomes to create more diverse Ig heavy chains, which may be important for the generation of specific immune responses against pathogens.

Conflict of interests

The author(s) did not declare any conflict of interest.

ACKNOWLEDGEMENTS

This work was partially supported by the Center for Advanced Studies for Agriculture and Food, Institute for Advanced Studies, Kasetsart University under the Higher Education Research Promotion and the National Research University Project of Thailand, Office of the Higher Education Commission, Ministry of Education, Thailand.

REFERENCES

Abbas AK, Lichtman AH, Pillai S (2007). Cellular and molecular immunology. W. B. Saunder company, Philadelphia.

Acton RT, Weinheimer PF, Hall SJ, Niedermeier W, Shelton E, Bennett JC (1971). Tetrameric immune macroglobulins in three orders of bony fishes. Proc. Natl. Acad. Sci. U.S.A. 68(1):107-111.

Andersson E, Matsunaga T (1998). Evolutionary stability of the immunoglobulin heavy chain variable region gene families in teleost. Immunogenetics, 47(3):272-277.

Bengten E, Quiniou SMA, Stuge TB, Katagiri T, Miller NW, Clem LW, Warr GW, Wilson M (2002). The IgH locus of the channel catfish, *Ictalurus punctatus*, contains multiple constant region gene sequences: different genes encode heavy chains of membrane and secreted IgD. J. Immunol. 169(5):2488-2497.

Brodeur PH, Riblet R (1984). The immunoglobulin heavy-chain variable region (Igh-V) locus in mouse. I. One hundred IgH-V genes comprise seven families of homologous genes. Eur. J. Immunol. 14(10):922-930.

Brown GD, Kaattari IM, Kaattari SL (2006). Two new Ig V_H gene families in *Oncorhynchus mykiss*. Immunogenetics 58(11):933-936.

Casali P, Schettino EM (1996). Structure and function of natural antibodies. Curr. Top. Microbiol. Immunol. 210: 167-179.

Coscia MR, Oreste U (2003). Limited diversity of the immunoglobulin heavy chain variable domain of the emerald rockcod *Trematomus bernacchii*. Fish Shellfish Immunol. 14(1):71-92.

Coscia MR, Cocca E, Giacomelli S, Cuccaro F, Oreste U (2012). Immunoglobulin from Antarctic fish species of Rajidae family. Mar. Genomics. 5:35-41.

Danilova N, Bussmann J, Jekosch K, Steiner LA (2005). The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. Nat. Immunol. 6(3): 295-302.

Dylke J, Lopes J, Lawson MD, Machtaler S, Matsuuchi L (2007). Role of the extracellular and transmembrane domain of Igα/ß in assembly of the B cell antigen receptor (BCR). Immunol. Lett. 112(1): 47-57.

FAO (2014) Fisheries Statistic. Food and Agricultural Organization of the United Nation, Rome, Italy.

Flajnik MF (2002). Comparative analyses of immunoglobulin genes: surprises and portents. Nat. Rev. Immunol. 2(9): 688-698.

Flajnik MF, Du Pasquier LD (2004). Evolution of innate and adaptive immunity: can we draw a line? Trends Immunol. 25(12):640-644.

Grontvedt RN, Espeild S (2003). Immunoglobulin producing cells in the spotted wolfish, *Anarhichas minor* (Olafsson): localization in adults and during juvenile development. Dev. Comp. Immunol. 27(6-7): 1-10.

Haire RN, Amemiya CT, Suzuki D, Litman GW (1990). Eleven distinct V_H gene families and additional patterns of sequence variation suggest a high degree of immunoglobulin gene complexity in a lower vertebrate, *Xenopus laevis*. J. Exp. Med. 171(5): 1721-1737.

Hikima J, Jung TS, Aoki T (2011). Immunoglobulin genes and their transcriptional control in teleosts. Dev. Comp. Immunol. 35(9): 924-936.

Holland MCH, Lambris JD (2002). The complement system in teleosts. Fish Shellfish Immunol. 12: 399-420.

Hsu E, Schwager J, Alt FW (1989). Evolution of immunoglobulin genes: V_H families in the amphibian *Xenopus*. Proc. Natl. Acad. Sci. U.S.A. 86(20): 8010-8014.

Kabat EA, Wu TT (1971). Attempts to locate complementary-determining residues in the variable positions of light and heavy chains. Ann. N. Y. Acad. Sci. 190: 382-393.

Kabat EA, Wu TT (1991). Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of V_H and V_L genes, minigenes, and complementarity-determining regions to binding of antibody combining sites. J. Immunol. 147(5): 1709-1719.

Kuo TC, Schlissel MS (2009). Mechanisms controlling expression of the RAG locus during lymphocyte development. Curr. Opin. Immunol. 21(2): 173-178.

Kozak M (1998). Primer extension analysis of eukaryotic ribosome-mRNA complexes. Nucleic Acids Res. 26(21): 4853-4859.

Lieber MR (1992). The mechanism of V(D)J recombination: A balance of diversity, specificity and stability. Cell 70(6): 873-876.

Livak KJ, Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}{}_T$ method. Methods 25(4): 402-408.

Lundqvist ML, Strömberg S, Pilström L (1998). Ig heavy chain of the sturgeon *Acipenser baeri*: cDNA sequence and diversity. Immunogenetics 48(6): 372-82.

Mage RG, Bernstein KE, McCartney-Francis N, Alexander CB, Young-Cooper GO, Padlan EA, Cohen GH (1984). The structural and genetic basis for expression of normal and latent V_Ha allotypes of the rabbit. Mol. Immunol. 21(11): 1067-1081.

Mainville CA, Sheehan KM, Klaman LD, Giorgetti CA, Press JL Brodeur PH (1996). Deletional mapping of fifteen mouse V_H gene families reveals a common organization for three Igh haplotypes. J. Immunol. 156(3): 1038-1046.

Mao MG, Lei JL, Alex PM, Hong WS, Wang KJ (2012). Characterization of RAG1 and IgM (mu chain) marking development of the immune system in red-spotted grouper (*Epinephelus akaara*). Fish Shellfish Immunol. 33(4): 725-735.

Matsuda F, Ishii K, Bourvagnet P, Kuma K, Hayashida H, Miyata T, Honjo T (1998). The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. J. Exp. Med. 188(11): 2151-2162.

Mohamed HM, Refat NAGA (2011). Pathological evaluation of probiotic, *Bacillus subtilis*, against *Flavobacterium columnare* in *Tilapia nilotica* (*Oreochromis niloticus*). J. Am. Sci. 7: 244-256.

Ota T, Nei M (1995). Evolution of immunoglobulin V_H pseudogenes in chickens. Mol. Biol. Evol. 12(1):94-102.

Ota T, Rast JP, Litman GW, Amemiya CT (2002). Lineage-restricted retention of a primitive immunoglobulin heavy chain isotype within the Dipnoi reveals an evolutionary paradox. Proc. Natl. Acad. Sci. U.S.A 100(5): 2501-2506.

Peixoto BR, Brenner S (2000). Characterization of approximately 50 kb of the immunoglobulin *VH* locus of the Japanese pufferfish, *Fugu rubripes*. Immunogenetics, 51(6):443-451.

Pilstrom L, Bengten E (1996). Immunoglobulin in fish - genes, expression and structure. Fish Shellfish Immunol. 6: 243-262.

Pridgeon JW, Aksoy M, Klesius PH, Li Y, Mu X, Srivastava K, Reddy G (2011). Identification and expression profiles of multiple genes in Nile tilapia in response to bacterial infections. Vet. Immunol. Immunopathol. 144(1-2): 111-119.

Rauta PR, Nayak B, Das S (2012). Immune system and immune responses in fish and their role in comparative immunity study: A model for higher organisms. Immunol. Lett. 148(1): 23-33.

Rodkhum C, Kayansamruaj P, Pirarat N (2011). Effect of water temperature on susceptibility to *Streptococcus agalactiae* serotype Ia infection in Nile tilapia (*Oreochromis niloticus*). Thai J. Vet. Med. 41:309-314.

Roitt I, Brostoff J, Male D (2001). Immunology, 6th edition, Gower Medical Publishing: London.

Roman T, De Guerra A, Chrlemagne J (1995). Evolution of specific antigen recognition: size reduction and restricted length distribution of the CDRH3 regions in the rainbow trout. Eur. J. Immunol. 25(1):269-273.

Ross DA, Wilson MR, Miller NW, Clem LW, Warr GW (1998). Evolutionary variation of immunoglobulin μ heavy chain RNA processing pathways: origins, effects, and implications. Immunol. Rev. 166:143-151.

Rumfelt LL, Lohr RL, Dooley H, Flajnik MF (2004). Diversity and repertoire of IgW and IgM VH families in the newborn nurse shark. BMC Immunol. 5:8.

Saha NR, Suetake H, Suzuki Y (2005). Analysis and characterization of the expression of the secretory and membrane forms of IgM heavy chains in the pufferfish, *Takifugu rubripes*. Mol. Immunol. 42(1):113-124.

Salinas I, Zhang YA, Sunyer JO (2011). Mucosal immunoglobulins and B cells of teleost fish. Dev. Comp. Immunol. 35(12): 1346-1365.

Srisapoome P, Ohira T, Hirono I, Aoki T (2004). Genes of the constant regions of functional immunoglobulin heavy chain of Japanese flounder, *Paralichthys olivaceus*. Immunogenetics 56(4):292-300.

Stenvik J, Lundback AS, Jorgensen TO, Pilstrom L (2000). Variable region diversity of the Atlantic cod (*Gadus morhua* L.) immunoglobulin heavy chain. Immunogenetics 51(8- 9):670-680.

Stewart JJ, Lee CY, Ibrahim S, Watts P, Shlomchik M, Weigert M, Litwin S (1997). A Shannon entropy analysis of immunoglobulin and T cell receptor. Mol. Immunol. 34(15):1067-1082.

Sun J, Kacskovics I, Brown WR, Butler JE (1994). Expressed swine VH genes belong to a small VH gene family homologous to human VHIII. J. Immunol. 153(12):5618-5627.

Taggart JB, Hynes RA, Prodohl PA, Ferguson A (1992). A simplified protocol for routine total DNA isolation from salmonid fishes. J. Fish Biol. 40:963-965.

Tian C, Chen X, Ao J (2009). The up-regulation of large yellow croaker secretory IgM heavy chain at early phase of immune response. Fish Physiol. Biochem. 36(3):483-490.

Tonegawa S (1983). Somatic generation of antibody diversity. Nature 302:575-581.

Walport MJ (2001). Complement, first of two parts. N. Engl. J. Med. 344:1058-1066.

Wilson MR, Middleton D, Warr GW (1991). Immunoglobulin *VH* genes of the goldfish, *Carassius auratus*: a re-examination. Mol. Immunol. 28(4-5):449-457.

Wilson MR, Warr GW (1992). Fish immunoglobulins and the genes that encode them. Annu. Rev. Fish. Dis. 2:201-221.

Yang F, Ventura-Holman T, Waldbieser GC, Lobb C J (2003). Structure, genomic organization, and phylogenetic implications of six new VH families in the channel catfish. Mol. Immunol. 40(5): 247-260.

Yasuike M, de Boer J, von Schalburg KR, Cooper GA, McKinnel L, Messmer A, So S, Davidson WS, Koop BF (2010). Evolution of duplicated IgH loci in Atlantic salmon, *Salmo salar*. BMC Genomics 11:486.