

# The Complete Mitochondrial Genome Sequence and Characterization of Single-Nucleotide Polymorphisms in the Control Region of the Asian Seabass (*Lates calcarifer*)

G. Lin,<sup>1</sup> L.C. Lo,<sup>1</sup> Z.Y. Zhu,<sup>1</sup> F. Feng,<sup>1</sup> R. Chou,<sup>2</sup> G.H. Yue<sup>1</sup>

<sup>1</sup>Molecular Population Genetics Group, Temasek Life Sciences Lab, 1 Research Link, National University of Singapore, 117604, Singapore

<sup>2</sup>Agri-Food and Veterinary Authority of Singapore, Tower Block MND Complex, 069110, Singapore

Received: 21 April 2005 / Accepted: 25 June 2005 / Published online: 16 October 2005

## Abstract

We determined the complete mtDNA nucleotide sequence of *Lates calcarifer* using the shotgun sequencing method. The mitochondrial DNA (mtDNA) was 16,535 base pairs (bp) in length, and contained 13 protein coding genes, 22 transfer RNAs, 2 ribosomal RNAs, and one major noncoding control region (CR). The CR was unusually short at only 768 bp. A striking feature of the mitochondrial genome was the high G+C content (46.1%), which is among the highest in fish. The gene order was identical to that of a typical vertebrate. Phylogenetic analyses using concatenated amino acid sequences of 12 protein-coding genes of 30 fish species representing 14 suborders clearly showed *Lates calcarifer* was located in the cluster of fish species from the order Perciformes, supporting the traditional systematic classification. We characterized single-nucleotide polymorphisms (SNPs) in the CR by sequencing the complete CR of 25 individuals obtained from Australia and Singapore. A total of 68 SNPs were detected. Eighteen SNPs were fixed with alternative nucleotides in Australian and Singapore seabass, and these SNPs could be used for differentiating fish from the two countries.

**Keywords:** Barramundi — mtDNA — shotgun sequencing — single-nucleotide polymorphisms and phylogeny

## Introduction

The animal mitochondrial genome is a single, small, double-stranded circular DNA (Attardi, 1985). In

general, the size of the animal mitochondrial genome is approximately 16.5 kb and it usually contains 13 protein coding genes, 2 ribosomal RNAs (rRNAs; small 12 S and large 16 S), 22 transfer RNAs (tRNAs), and a major noncoding region that allows for the initiation of mitochondrial replication and transcription (Boore, 1999). Because of its compact size, multiple copy status in a cell, rapid evolutionary rate, and lack of recombination, mitochondrial DNA (mtDNA) has been extensively used as a marker for evolutionary and population genetic studies (Curole and Kocher, 1999). Because of its maternal inheritance, mtDNA showed some advantages over nuclear DNA markers in studies of population subdivision, gene flow, hybridization, and introgression (Jiggins et al., 1997; Chubb et al., 1998; McLean and Taylor, 2001; Stepien et al., 2001). Hence, the usefulness of the mitochondrial genome has received much attention in fisheries science. Many earlier studies of mtDNA concentrated on the polymorphism in the control region (CR) or protein coding genes by using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) (e.g., Hiendleder, 1996). Today the trend in more and more studies is to move to direct sequencing of genes (Katsares et al., 2003), CR (Lee et al., 1995), and complete mtDNA (Inoue et al., 2001; Machida et al., 2004).

Asian seabass *Lates calcarifer*, also called barramundi in Australia, is one of the nine *Lates* species of the family Centropomidae and is widely distributed in the coastal and freshwaters of the tropical Indowest Pacific, from the Persian Gulf to India to Northern Australia (Nelson, 1994). Cultured seabass is a commercially important fish species and considered to have a high potential in the tropical Asia Pacific region (Chou and Lee, 1997). When sold either

Correspondence to: Gen Hua Yue; E-mail: genhua@tll.org.sg

live or freshly frozen, its large size and delicate-flavored flesh command a premium price in the market. The fry is produced commercially in the hatchery and can be obtained throughout the year at reasonable prices because of its high stocking density and survival rate as well as the ease of managing the stock. Morphometric analysis indicated a slight difference between seabass from Australia compared with Asian seabass in Southeast Asia. Although some populations of seabass have been studied using microsatellites (Yue et al., 2002), protein polymorphism and partial mtDNA sequences (Chenoweth et al., 1998), the complete nucleotide sequence of the mitochondrial genome has not yet been uncovered.

In this study, we determined the complete mitochondrial genome (mtDNA) sequence of *Lates calcarifer* using shotgun sequencing of overlapping PCR products, and characterized single nucleotide polymorphisms (SNPs) in the CR by comparing sequences among 15 Asian seabass individuals from Australia and 10 from Singapore. All information reported in this article may facilitate in studies on population structure, genetic diversity, broodstock management and reconstruction of phylogenetic trees of life, as well as supply useful new markers for identification of the origins of fish and for the authentication of processed Asian seabass products.

## Materials and Methods

**Fish Sample and DNA Isolation.** A 25-day-old Asian seabass fingerling was obtained from Marine Aquaculture Centre, Agri-Food and Veterinary Authority, Singapore and the whole fingerling was immediately preserved in 100% ethanol. Total genomic DNA was extracted using a traditional phenol-chloroform method (Sambrook and Russell, 2001) and kept at  $-20^{\circ}\text{C}$ .

**PCR Amplification, Shotgun Sequencing, and Assembly of mtDNA.** The initial determination of the partial sequences of mtDNA was obtained using four universal primer pairs: L14841+H15149, L1091+H1478, L7450+H8055, and L9225+H9407 (Meyer, 1993). PCR was conducted on a PTC-100 PCR thermal cycler (MJ Research, Waltham, MA, USA) using the following program:  $94^{\circ}\text{C}$  for 2 min, followed by 35 cycles of  $94^{\circ}\text{C}$  for 30 s,  $55^{\circ}\text{C}$  for 30 s, and  $72^{\circ}\text{C}$  for 1 min, and a final extension at  $72^{\circ}\text{C}$  for 5 min. Each 25- $\mu\text{l}$  PCR reaction volume

contained 30 ng of DNA,  $1 \times$  PCR buffer (Finnzymes, Espoo, Finland), 200 nM of each primer, 50  $\mu\text{M}$  of each dNTP, and 1 U of DNA polymerase (Finnzymes). The primer pairs L14841+H15149, L1091+H1478, and L9225+H9407 amplified PCR products. The PCR products were cloned into a pGEM-T vector (Promega, San Luis Obispo, CA, USA) and sequenced using BigDye v3.0 chemicals (Applied Biosystems, Foster City, CA, USA) as described by Yue et al. (2004). The sequences obtained were aligned using Sequencher (GeneCodes, Ann Arbor, MI, USA), and three specific primer pairs: LcamtA1+B1, LcamtA2+B2, and LcamtA3+B3 (for sequences, see Appendix 1) were designed using PrimerSelect (DNASTAR, Madison, WI, USA). The expected length of the three PCR products was approximately 2.5 kb, 7.5 kb, and 7.5 kb, respectively. Long-distance PCR was carried out using Expand Long Template PCR system (Roche, Basel, Switzerland) on a PTC-100 thermal cycler (MJ Research). Each 50- $\mu\text{l}$  reaction volume contained 200  $\mu\text{M}$  of dNTPs,  $1 \times$  buffer 1 with 1.75 mM  $\text{MgCl}_2$ , 200 nM of each primer, 3.75 U of Taq Polymerase Enzyme Mix (Roche), and 50 ng of total DNA. The thermal cycling profile was pre-denaturation at  $94^{\circ}\text{C}$  for 2 min, 8 cycles of  $94^{\circ}\text{C}$  for 10 s,  $60^{\circ}\text{C}$  for 30 s, and  $68^{\circ}\text{C}$  for 6 min followed by 19 cycles of  $94^{\circ}\text{C}$  for 10 s,  $60^{\circ}\text{C}$  for 30 s, and  $68^{\circ}\text{C}$  for 6 min with a time increase of 20 s/cycle, then a final extension at  $68^{\circ}\text{C}$  for 10 min. Long-distance PCR products were electrophoresed on a 1% agarose gel (Bio-Rad, Hercules, CA, USA) and later stained with ethidium bromide for visualization via ultraviolet transillumination.

Long-distance PCR products of 2.5 kb, 7.5 kb, and 7.5 kb were sonicated using the Branson Digital Sonifier 450 (Labequip, Ontario, Canada) under 20% amplitude for 4 s, 6 s and 6 s, respectively. The sonicated DNA was separated on a 2% agarose gel, and fragments between 0.5 and 1.5 kb were excised, concentrated, and purified using GFX PCR DNA and Gel Band Purification Kit (Amersham, Piscataway, NJ, USA) according to the manufacturer's recommendations. The sonicated long-distance PCR products were treated with DNATerminator End Repair Kit (Lucigen, Middleton, WI, USA) according to the manufacturer's protocol, in which treatment of the products was carried out at room temperature for 30 min with a 50- $\mu\text{l}$  reaction volume containing 18  $\mu\text{l}$  of sterile distilled water, 10  $\mu\text{l}$  of  $5 \times$  end repair buffer, 2  $\mu\text{l}$  of end repair enzymes (Lucigen), and 20  $\mu\text{l}$  (approximately 500 ng) of sonicated DNA template. The treatment of the sonicated long PCR products was terminated by incubation at  $70^{\circ}\text{C}$  for 10 min. The treated PCR products were purified and

concentrated with GFX PCR DNA and Gel Band Purification Kit (Amersham) and ligated into pBluescript II KS (-) (Stratagene, La Jolla, CA, USA) with T4 DNA Ligase (Stratagene). The ligation was performed at room temperature for approximately 2 to 3 hours with an 11.1  $\mu$ l reaction volume containing 1.1  $\mu$ l of 10  $\times$  ligation buffer, 1  $\mu$ l of rATP (10 mM), 0.5  $\mu$ l of pBluescript II KS(-) (Stratagene) containing approximately 25 ng of plasmid, 1  $\mu$ l of T4 DNA ligase (4 U/ $\mu$ l), and 7.5  $\mu$ l (approximately 50 ng) of treated DNA template. The ligated DNA was cloned into XL-1 Blue competent cells (Stratagene) and cultured on Luria Bertani (LB) plates containing 100  $\mu$ g/L of ampicillin, 200  $\mu$ g/ $\mu$ l of X-galactoside, and 200  $\mu$ l of 0.5 M isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG). White colonies were picked and cultured in LB medium containing 125  $\mu$ g/ $\mu$ l of ampicillin overnight at 37°C, with shaking at 250 rpm. Colony PCR was performed on PTC-100 thermal cycler (MJ Research) as described by Yue et al. (2000). Sequencing of the PCR products was carried out using M13/M13 reverse primer, Bigdye v3.0 (Applied Biosystems), and the DNA sequencer ABI3730 xl (Applied Biosystems) according to the manufacturer's recommendation.

The DNA sequences were analyzed using EditSeq (DNASTAR). Flanking vector sequences were removed manually or automatically using Sequencher (GeneCodes) with a manual correction. Insert sequences were assembled by using the same software. Following the assembly, tRNA genes were identified by using the method described by Lowe and Eddy (1997). The locations of 13 protein-coding genes were identified by comparison of DNA or amino acid sequences of mitochondrial genomes of other bony fish, whereas the two tRNAs were determined by sequence homology and secondary structure as described by Johansen and Bakke (1996). The 5' ends of protein genes were inferred to be located at the first legitimate in-frame start codon (ATN, GTG, TTG, and GTT) that did not overlap with the preceding gene, except that with an upstream tRNA gene was limited to the most 3' nucleotide of the tRNA. Protein gene termini were inferred to be located at the first in-frame stop codon unless that codon was located within the sequence of a downstream gene. Otherwise, a truncated stop codon (T or TA) adjacent to the beginning of the downstream gene was designated as the termination codon and was assumed to be completed by polyadenylation after transcript cleavage. Base composition and codon usage were analyzed by using EditSeq and GeneQuest (DNASTAR). The complete mtDNA sequence of *Lates calcarifer* was deposited in GenBank under Accession No. DQ010541.

### ***SNP Detection and Characterization in the CR.***

The complete CR of 15 Australian seabass and 10 Singapore seabass individuals was amplified by PCR using the primer pair Dloop-A1 and Dloop-B1 (for sequences, see Appendix 1) located in the tRNA<sup>Thr</sup> and 12 S RNA regions flanking the whole CR. PCR was conducted in a 25- $\mu$ l reaction volume containing 40 ng of total DNA, 200 nM of each primer, 200  $\mu$ M dNTPs, 1 U of *Taq* DNA polymerase (Finzymes), and 1  $\times$  buffer with 1.5 mM MgCl<sub>2</sub> on a PTC-100 (MJ Research). The PCR program was 94°C for 2 min, then 34 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 48 s, followed by a final extension of 72°C for 5 min. The PCR products were separated on 1% agarose gels, and the strong band from each sample was excised and cleaned using glassmilk as described by Yue and Orban (2001). The cleaned PCR products were directly sequenced using two primers, Dloop-A1/Dloop-B1, and Bigdye v3.0 chemicals (Applied Biosystems) as described earlier. Forward and reverse sequences were assembled with Sequencher (GeneCodes) and manually checked by two persons independently. Sequences of the 25 individuals were aligned using Clustal\_X 1.85 (Thompson et al., 1997) and then analyzed with MEGA (Kumar et al., 2001). The CR nucleotide sequences of the 25 fish have been deposited in GenBank under the Accession Nos. DQ012409–DQ012433. Phylogenetic trees (NJ, ME, MP, and UPGMA) were constructed on the basis of the complete CR nucleotide sequences using MEGA (Kumar et al., 2001). Conserved sequences were detected by comparing seabass CR sequences with those of other fish species (Lee et al., 1995).

### ***Construction of Phylogenetic Trees.***

To determine the evolutionary position of the Asian seabass, the mtDNA sequences of 30 fish species belonging to 14 suborders (for details, see the legend to Figure 3) were selected and used. The analyses were based on concatenated sequences of the 12 heavy-strand encoded protein-coding genes. The sequence of ND6 encoded by the light strand was excluded in the analyses because of the deviating nucleotide and amino acid composition of this gene as compared to those encoded by the heavy strand. Amino acid and nucleotide sequences of the 12 protein coding genes were aligned with Clustal\_X 1.85 (Thompson et al., 1997). The data set was analyzed with maximum parsimony (MP), neighbor-joining (NJ), unweighted pair group method with arithmetic mean (UPGMA) and molecular evolution (ME) as implemented in the program MEGA (Kumar et al., 2001).

## Results and Discussion

**Long-Distance PCR and Shotgun Sequencing.** The long-distance PCR method described here was used to obtain the template for sequencing *Lates calcarifer* mtDNA. The whole mtDNA was amplified by long-distance PCR with 3 pairs of specific primers. In comparison with the traditional method (i.e., extraction of mtDNA from total DNA: e.g., Arnason and Johnsson, 1992) for sequencing mtDNA, long-distance PCR is quicker and simpler. Before specific primer pairs are designed, however, specific sequences of some conserved regions of the mtDNA should be

determined. By sequencing 96 clones with inserts between 0.50 and 1.5 kb in both directions, the whole mtDNA was covered about 6 times without any gaps. Using the ABI3730 xl sequencer, the 96 clones could be sequenced within 2 hours. Therefore long-distance PCR and shotgun sequencing are the simplest and quickest methods to obtain the complete sequence of mtDNAs.

**Genome Content and Base Composition.** The genome content of *Lates calcarifer* was in accordance to the length of other vertebrates [see review

**Table 1. Location of Features in the Mitochondrial Genome of *Lates calcarifer***

Gene/elements	Position		Codon		Size (bp)	5' intergenic space <sup>a</sup>
	From	To	Start	Stop		
Control region	0	768			768	0
<i>tRNA<sup>Phe</sup></i>	769	836			68	0
12 S <i>rRNA</i>	837	1,801			965	0
<i>tRNA<sup>Val</sup></i>	1,802	1,874			73	0
16 S <i>rRNA</i>	1,875	3,628			1,754	2
<i>tRNA<sup>Leu</sup></i>	3,631	3,703			73	0
ND1	3,704	4,678	ATG	TAA	975	3
<i>tRNA<sup>Ile</sup></i>	4,682	4,752			71	0
<i>tRNA<sup>Gln</sup></i>	4,753	4,824			72	0
<i>tRNA<sup>Met</sup></i>	4,824	4,893			70	0
ND2	4,894	5,940	ATG	TAA	1,047	0
<i>tRNA<sup>Trp</sup></i>	5,941	6,012			72	0
<i>tRNA<sup>Ala</sup></i>	6,013	6,082			70	1
<i>tRNA<sup>Asn</sup></i>	6,084	6,157			74	0
O <sub>L</sub>	6,158	6,191			34	1
<i>tRNA<sup>Cys</sup></i>	6,193	6,257			66	0
<i>tRNA<sup>Tyr</sup></i>	6,258	6,327			70	1
COX1	6,329	7,879	GTG	TAA	1,551	0
<i>tRNA<sup>Ser</sup></i>	7,880	7,950			71	4
<i>tRNA<sup>Asp</sup></i>	7,955	8,021			68	8
COX2	8,030	8,725	ATG	Taa	696	-1
<i>tRNA<sup>Lys</sup></i>	8,725	8,799			75	3
ATP8	8,803	8,969	ATG	TAA	168	-10
ATP6	8,960	9,643	ATG	TAA	684	-1
COX3	9,643	10,428	ATG	Taa	786	-1
<i>tRNA<sup>Gly</sup></i>	10,428	10,499			72	0
ND3	10,500	10,850	ATG	Taa	351	-2
<i>tRNA<sup>Arg</sup></i>	10,849	10,917			69	1
ND4L	10,919	11,215	GTG	TAA	297	-7
ND4	11,209	12,597	ATG	Taa	1,389	-8
<i>tRNA<sup>His</sup></i>	12,590	12,658			69	0
<i>tRNA<sup>Ser</sup></i>	12,659	12,726			71	5
<i>tRNA<sup>Leu</sup></i>	12,731	12,803			73	0
ND5	12,804	14,651	ATG	TAA	1,848	-4
ND6	14,648	15,178	ATG	TAG	531	0
<i>tRNA<sup>Glu</sup></i>	15,179	15,247			69	4
CYTB	15,252	16,392	ATG	Taa	1,143	3
<i>tRNA<sup>Thr</sup></i>	16,396	16,464			69	4
<i>tRNA<sup>Pro</sup></i>	16,469	16,535			71	

<sup>a</sup>Numbers correspond to the nucleotides separating adjacent genes. Negative numbers indicate overlapping nucleotides. Taa and Taa: incomplete stop codons.



```

ATATAATTAT ATATACATGT ATTCACACCA TGTATTATG CTAACCAATA AATGAGTATA 60
TGGTGTAAT GCATATATAT ATGTATTAAAT CAACATTTAC CTGTAATCAA AGGACATATG 120
      AT-repeats
TGCATTCAAT GGTACTCGTA AATGTAATGT ACAGTAACCT ATAATTAATA TACTTTAAAC 180
AATAATATTA CATATTAATC ATCAGCGATA ATAGAGCGTA GTGAGAGATC ACCAATTAGT 240
      TAS
AGGTATTGAG AGTGTGACG GTTCTTGATA GTCAAGGACA AACACAGCGT GGGGGTTACA 300
CAAATTGAAC TATTACTGGC ATTTGGTTCC TACCTCAGGT GCATGGATAA GAAAGAATTC 360
      CSB-D
CACTGACGTT CCTTCATAAA TGGCATAAGT TAATGGTGGG TAACATGAAT CCTCGTTACC 420
      CSB-1
CACCAAGCCG AGCATTCTTT CCAGCGGGTA AGGGGTTTAT AATTTTTTTC TTATCATCTC 480
ATTTGACATC TCAGAGTGCA CGCACAGAGC ATATAATATG GAAGTAGAAC ATAATTTTTT 540
TATGGAAGAG AATAGTCAGT ATGAATGGGG AAAAGGCTTG ACCTGAAGAA TTACATAATA 600
GGATATCACG AGCATAACGG GCTAACCTTC TACTCCACA TATCTATTAA TCCCCCCGG 660
TTTTTTTACG CACCCCCCCC CCCCCCCC AAAAACAGAG GAAGCCCCAC AATAACTTGA 720
      CSB-2
TTTTTAATCTA AAAATGCACT ATTTACACTA TTAAATATAT TTTAAGAA

```

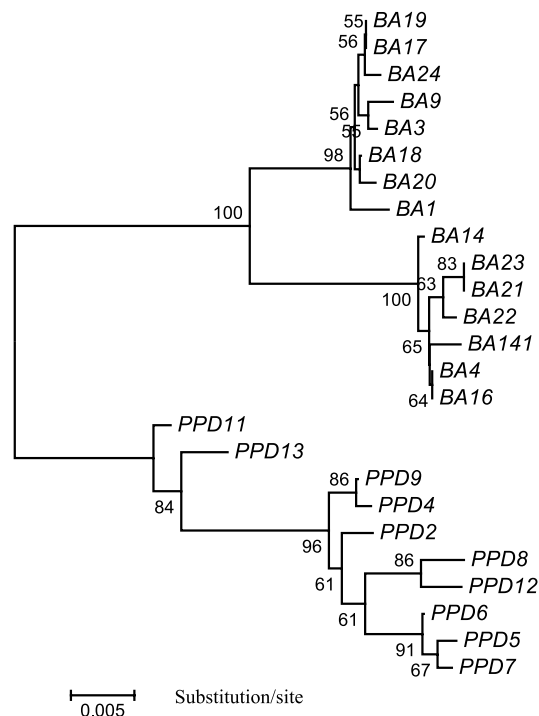
**Fig. 1.** The complete nucleotide sequences of mtDNA control region of *Lates calcarifer*. CSB-D: conserved sequence block D; CSB-1,-2,-3: conserved sequence blocks 1, 2, and 3. TAS: termination associated sequence.

see Boore (1999)], including 2 rRNA, 22 tRNA, 13 protein-coding genes, and a noncoding CR. The total length of the Asian seabass mitochondrial genome was 16,535 bp (Table 1). The G+C content (46.1%) of mtDNA of *Lates calcarifer* was higher than that of other fish species, such as *Takifugu rubripes* (44.3%) (Elmerot et al., 2002), zebrafish (39.9%) (Broughton et al., 2001), *Daphnia pulex* (37.7%) (Crease, 1999); it was also higher than invertebrate mtDNAs (the G+C content ranges in chordates from 44.4% to 37.8%; echinoderms, 41.1% to 38.7%; annelids, 38.4%; cnidarians, 37.5% to 35.5% (Noguchi et al., 2000), but similar to that of chicken (46.04%) (Nishibori et al., 2003) and Japanese quail (44.52%) (Nishibori et al., 2001). It seems that the high G+C content is associated with higher temperatures in tropical waters. In bacteria, previous studies showed a strong positive correlation between the G+C content of stem regions of rRNA and environmental growth temperature (Dalgaard and Garrett, 1993; Galtier and Lobry, 1997). A current study demonstrated that the relationship between the nucleotide content of structural RNAs and environmental growth temperature was not due entirely to phylogenetic history, but reflected a repeated selective response to the elevated environmental temperature (Wang and Hickey, 2002).

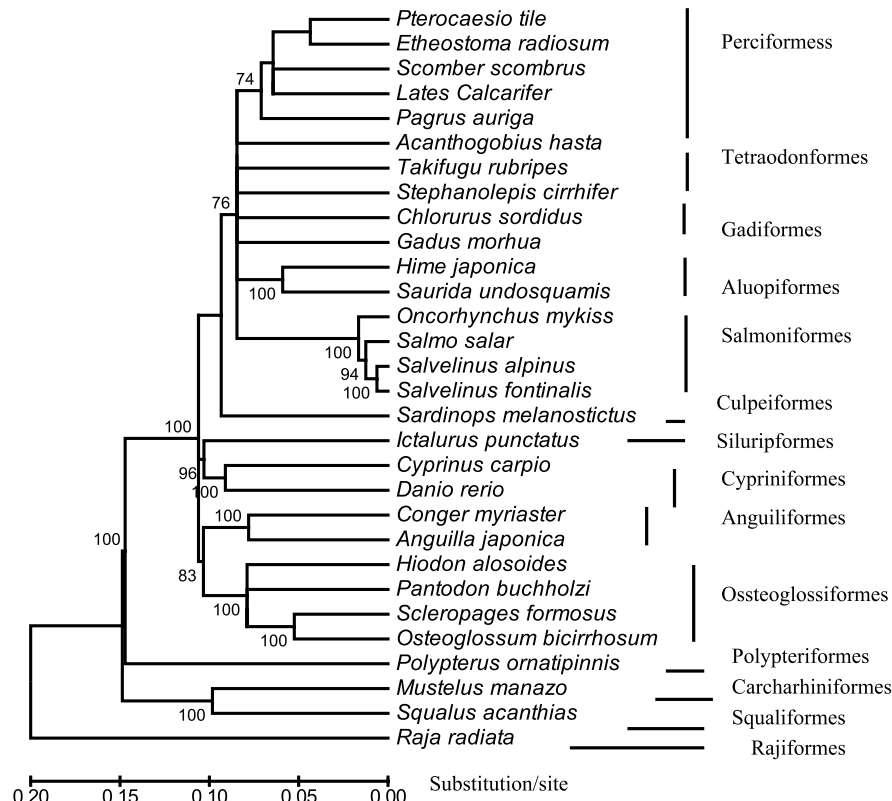
**Gene Order.** Although rearrangements of gene order in mtDNA have been seen in several taxa, such as *Balanoglossus carnosus* (Castresana et al., 1998) and *Branchiostoma lanceolatum* (Spruyt et al., 1998; Boore et al., 1999), the gene order of the Asian seabass was identical to that of the typical vertebrates. The 768-bp noncoding CR was located between the tRNA that encodes for proline (tRNA<sup>Pro</sup>) and phenylalanine (tRNA<sup>Phe</sup>) and the origin of light replication (O<sub>L</sub>: 34 bp) was located between a group of 5 tRNA genes (WANCY). Small

intergenic sequences varying from 1 to 8 nucleotides were present between some genes (Table 1).

**Protein-Coding Genes.** Typical of all bony fishes (Boore, 1999), all protein-coding genes has a methionine (ATG) start codon with the exception of COX1 and ND4L, which had an open reading



**Fig. 2.** A NJ tree generated with MEGA for 15 seabass individuals from Australia and 10 seabass individuals from Singapore, based on the complete nucleotide sequence of mtDNA control region. Scale bar is shown under the tree, while the bootstrap proportions created by 1,000 replicates are indicated on each branch. BA: individuals from Australia; PPD: individuals from Singapore. Gen-Bank Accession No. of the control region sequences: DQ012409–DQ012433.



**Fig. 3.** An NJ-tree reconstructed with MEGA, showing the evolutionary position of *Lates calcarifer* relative to that of 30 other fish species belonging to 14 suborders. The tree is based on concatenated amino acids sequences of 12 protein-coding genes. The scale bar is shown under the tree, whereas the bootstrap value (>50%) of 1,000 replicates is listed. The analyses included the following 30 species: *Pterocaesio tile* (NC-004408), *Pagrus auriga* (NC-005146), *Etheostoma radiosum* (NC-005254), *Scomber scombrus* (NC-006398), *Chlorurus sordidus* (NC-006355), *Acanthogobius hasta* (NC-006131), *Takifugu rubripes* (AJ421455), *Stephanolepis cirrifer* (NC-003177), *Gadus morhua* (X99772), *Hime japonica* (AB047821), *Saurida undosquamis* (NC-003162), *Oncorhynchus mykiss* (L29771), *Salmo salar* (U12143), *Salvelinus alpinus* (AF154851), *Salvelinus fontinalis* (AF154850), *Sardinops melanostictus* (AB032554), *Ictalurus punctatus* (AF482987), *Cyprinus carpio* (X61010), *Danio rerio* (AC024175), *Conger myriaster* (AB038381), *Anguilla japonica* (AB038556), *Hiodon alosoides* (AP004356), *Pantodon buchholzi* (AB043068), *Scleropages formosus* (DQ023143), *Osteoglossum bicirrhosum* (AB043025), *Mustelus manazo* (AB015962), *Squalus acanthias* (Y18134), *Raja radiata* (AF106038), *Polypterus ornatipinnis* (U62532) and *Lates calcarifer* (DQ010541).

frame of GTG (Table 1). The Asian seabass had and 1 TAG and 7 TAA stop codons and the remainder of the reading frames had an incomplete termination of either T or TA (Table 1). For those genes with an incomplete stop codon, the transcripts would be modified to form complete termination signal UAA via post-transcriptional polyadenylation. The overlapping reading frames on the same strand of ATP8-ATP6, ND4L-ND4, and ND5-ND6 shared 10, 7, and 4 nucleotides respectively, which are similar to Atlantic cod (Johansen and Bakke, 1996), floating goby (Kim et al., 2004), and walleye pollock (Yanagimoto et al., 2004). The G+C content (47%) in protein-coding genes is much higher than that in zebrafish (39%) (Broughton et al., 2001); Antarctic krill, *Euphausia superba* (32.2%) (Machida et al., 2004); and *Takifugu rubripes* (45.3%) (Elmerot et al.,

2002). The high G+C content in protein-coding genes might reflect a selective response to high environmental temperature.

**tRNA Genes and rRNA Genes.** The *Lates calcarifer* mitochondrial genome contained 22 tRNA genes interspersed between the rRNA and protein-coding genes, which is typical of mtDNAs of other vertebrates (Boore, 1999). The size of tRNA genes ranged from 66 to 75 nucleotides (Table 1), most of which were large enough for the encoded tRNAs to fold into cloverleaf secondary structures. The 12 S and 16 S rRNA genes were 965 and 1,754 nucleotides long, respectively. Their locations in the genome were identical to those of other teleosts (Elmerot et al., 2002), separated by the tRNA<sup>Val</sup> gene

and positioned between the genes of tRNA<sup>Phe</sup> and tRNA<sup>Leu</sup>.

**Noncoding Sequence.** Typically the light strand replication origin ( $O_L$ ) was positioned in a cluster of 5 tRNA genes known as the WANCY region between the tRNA<sup>Asn</sup> and tRNA<sup>Cys</sup> genes. In Asian seabass, the  $O_L$  comprised of 34 nucleotides, and interestingly the characteristic sequence motif 5'-GCCGG-3' at the base of the stem within tRNA<sup>Cys</sup> (Elmerot et al., 2002) was mutated to GGCGG sequence. The major noncoding CR in the Asian seabass mitochondrial genome was located between tRNA<sup>Pro</sup> and tRNA<sup>Phe</sup> genes. This region was 768 nucleotides long, shorter than that of other bony fishes (Lee et al., 1995), but similar to *Emmelichthys struhsakeri* (838 bp) (Miya et al., 2003) belonging to the order Perciformes. The CR had an average A+T nucleotide content of 64.32%. It had several conserved blocks: TAS, CSB-1, CSB-2, CSB-3, and CSB-D (Figure 1). An AT-repeat sequence was located at the 5' end of the CR, similar to that of the common carp (Chang et al., 1994). Tandemly repeated sequences present in the CR of some teleosts (Lee et al., 1995) were not observed in the Asian seabass.

**SNP Characterization of the CR and Evolutionary Relationship of Australian and Asian Seabass.** The length of the CR of the 25 individual fish ranged from 767 to 772 bp. The length difference among individuals was caused mainly by the length polymorphism of AT repeat sequences located at the 5' end of the CR. A total of 68 SNPs were detected in CR among the 25 seabass individuals from Australia and Singapore. Most SNPs were found in the first part of 5' end (1–300 bp: 39 SNPs) and in the region between 500 and 600 bp (16 SNPs), whereas only 7 and 9 SNPs were found in the region between 301 and 499 bp and between 601 and 768 bp respectively (for details, see Appendix 2), suggesting that the 5' end of the CR of Asian seabass is most polymorphic. Interspecies comparison of CR sequences had detected the same trend in other fish species (Lee and Kocher, 1995). At 18 positions, the SNPs were fixed with alternative nucleotides in both populations; therefore they can be used to differentiate the origin of the individuals, as well as for authentication of processed Asian seabass products. Among the 25 seabass individuals, a total of 22 haplotypes were detected. The seabass from Singapore were more diverse; each fish represented one haplotype, whereas the 15 seabass from Australia

represented only 12 haplotypes. The AT repeat sequence at the 5' end of the CR showed polymorphism among the 25 fish; the number of AT repeats ranged from 4 to 6.

Among the 68 variable sites, 65 are parsimony-informative. Based on the nucleotide sequences of the CR, 4 phylogenetic trees (NJ, ME, MP, and UPGMA) were constructed. All 4 trees showed the same genetic relationship among the 25 fish individuals. An NJ tree is shown in Figure 2. The seabass individuals from Australia and Singapore were clustered into two distinct branches with strong support of high bootstrap value (100%), suggesting distinct genetic differences between seabass from the two countries, which also correspond to their phenotypic differences.

**Evolutionary Position of *Lates calcarifer*.** Phylogenetic analyses were used to examine the evolutionary position of *Lates calcarifer*. The trees included 30 fish species representing a total of 14 suborders. Trees constructed using 4 different methods showed a similar topology. Figure 3 shows a neighbor-joining (NJ) tree based on phylogenetic analysis of the concatenated amino acids sequences of 12 mitochondrial protein-coding genes. The tree was rooted by 4 ancient fish species: *Raja raduata*, *Polyperus ornatipinnis*, *Squalus acanthia*, and *Mustelus manazo*. In agreement with the traditional systematic classification (Nelson, 1994), most fish orders and higher level systematic relationship were reconstructed as monophyletic. The tree constructed in this study is quite similar to the one reported by Elmerot et al. (2002). *Lates calcarifer* was located in the cluster of fish species from the order Perciformes, supporting the traditional taxonomic classification. However, the branch length in the cluster of Perciformes was much longer than that in cluster of Salmoniformes, suggesting that the divergence time among the species in Perciformes is longer than that among species in Salmoniformes. Because, according to Nelson (1994), the order Perciformes comprises more than 2,815 fish species, it is necessary to sequence more mtDNAs of fish species to learn the fine evolutionary relationships of those species.

### Acknowledgments

The authors thank AVA for supporting the project "Selective Breeding of Marine Food Fish" and Drs. Laszlo Orban and Wang Chungming for comments on the manuscript before submission.

## Appendix 1

### Specific Primers Used for Amplifying the Complete mtDNA and the Control Region of *Lates calcarifer*

Name	Sequences (5'-3')	Location	Annealing Temperature (°C)	Expected product length
LcamtA1	TGTCGCGACGTCAACTATGGATG	CYTB	63	2.5 kb
LcamtB1	GTAAGGGCCGAGAGGGGTGAGG	12 S RNA		
LcamtA2	CCGCCCCGGAATTACAAGCACTA	12 S RNA	63	7.5 kb
LcamtB2	GCAGGGGAACTGCTTCGATCACAATGGG	COX2		
LcamtA3	ACTCCTACATGGTCCCAACACAAGAT	COX2	63	7.5 kb
LcamtB3	CAGGGGAGAACATAGCCAACGAA	CYTB		
Dloop-A1	TAAATCGGCGGCCGGAGGTTT	tRNA <sup>thr</sup>	60	1.2 kb
Dloop-B1	CGTGGGCTCAGTGGGTTAGGTGTG	12 S RNA		

## Appendix 2

Single-nucleotide polymorphisms in the control region of *Lates calcarifer* mtDNA. In the upper part of the figure, the vertical numbers indicate the positions of nucleotides relative to sequences of the longest control region of the individual (PPD09). Below are shown the nucleotide changes. Dots represent identity with sequence of the individual BA23. Fish labeled BA are from Australia, while fish designated PPD are from Singapore.

		111	1111111122	222222223	3334445555	5555555555	55666667
	1111223	3556899000	1255677801	2448999995	5661471222	3445666678	99007883
	3681236166	7571218079	3723178735	7785135681	6067134089	6562023696	68570023
BA23	CCGGCCCGTT	GGAGATCAGT	ACACGTGGTA	CCAGGTTGTA	GGGGGCGTGG	GCCAAACTGA	AAAGTGAC
BA21	.....	.....	.....	.....	.....	.....	.....
BA22	.....	.....	.....	.....	.....C.....	.....G.....	.....
BA4	.....	.....	.....	.....	.....	.....G.....	.....G.....
BA16	.....	.....	.....	.....	.....	.....G.....	.....G.....
BA141	.....	.....	.....	.....	.....	.....TGC.....	.....G.....
BA14	.....	.....	.....	.....	.....T.....	.....G.....	.....G.....
BA19	.....C.....	.....GAGCT.....	.....T.T.....	.....C.T.A.CC.....	.....A.....	.....A.....G.....	.....G.....G.....
BA17	.....C.....	.....GAGCT.....	.....T.T.....	.....C.T.A.CC.....	.....A.....	.....A.....G.....	.....G.....G.....
BA24	.....C.....	.....GAGCT.....	.....T.T.....	.....C.T.A.CC.....	.....A.....	.....A.....TG.....	.....G.....G.....
BA9	.....C.....	.....GAGCT.....	.....T.T.....	.....C.T.A.CC.....	.....A.....	.....A.....GC.....	.....G.....G.....
BA3	.....C.....	.....GAGCT.....	.....T.T.....	.....C.T.A.CC.....	.....A.....	.....A.....GC.....	.....G.....G.....
BA18	.....C.....	.....GAGCT.....	.....T.T.....	.....C.T.A.CC.....	.....A.....	.....A.....G.....	.....G.....G.....
BA20	.....C.....	.....GAGCT.....	.....T.T.....	.....C.T.A.CC.....	.....A.....	.....A.....G.....	.....G.....G.....
BA1	.....C.....	.....GAGCT.....	.....T.T.....	.....C.T.A.CC.....	.....A.C.C.....	.....A.....G.....	.....C.....G.....
PPD9	TTAATAT..A	CAGA...AC	.TGTA..A..	.T.AACCAC.	.AA.ATACAT	ATT.G.T...G	.GCAGC.T
PPD4	TTAATAT..A	CAGA...AC	.TGTA..A..	.T.AACCAC.	.AA.ATACAT	ATT.G.T.T..G	.GCAGC.T
PPD5	TTAATAT...A	AAGA...CAC	GTGTA..AA.G	.T.AACCACG	.AA.ATACAT	ATT.G.T.T..G	.GCAGC.T
PPD7	TTAATATA..A	AAGA...CAC	GTGTA..AA.G	.T.AACCACG	.AA.ATACAT	ATT.G.T.T..G	.GCAGC.T
PPD6	TTAATAT...A	AAGA...CAC	GTGTA..AA.G	.T.AACCACG	.AA.ATACAT	ATT.G.T...G	.GCAGC.T
PPD8	TTAATATA..A	AAGA...CAC	GTGTAC..A..	.TGAACCA.G	AAAAATACAT	ATT.G.T...G	.GCAGC.T
PPD12	TTAATATA..A	AAGA...A..	.TGTA..A..	.TGAACCA.G	AAAAATACAT	ATT.G.T.T..G	.GCAGCCT
PPD2	TTAATATA..A	AAGA...AC	.TGTA..A..	.T.AACCAC.	.AA.ATACAT	ATT.G.TC..G	.GCAGC.T
PPD13	AATATACACT	AAGA...AC	.TGTA..AA.G	.T.AACCACG	.AA.ATACAT	ATT.G.T..G	.G.AG..T
PPD11	.....	AAGA...AC	.TGTA..A..	.T.AACCAC.	.AA.AT.CAT	ATTTG.T...G	.G.AG..T

## References

- Arnason U, Johnsson E (1992) The complete mitochondrial-DNA sequence of the Harbor seal, *Phoca-Vitulina*. *J Mol Evol* 34, 493–505
- Attardi G (1985) Animal mitochondrial DNA, an extreme example of genetic economy. *Int Rev Cytol* 93, 93–145
- Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Res* 27, 1767–1780
- Boore JL, Daehler LL, Brown WM (1999) Complete sequence, gene arrangement, and genetic code of mitochondrial DNA of the *Cephalochordate Branchiostoma floridae* (*Amphioxus*). *Mol Biol Evol* 16, 1010
- Broughton RE, Milam JE, Roe BA (2001) The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Res* 11, 1958–1967
- Castresana J, Feldmaier-Fuchs G, Yokobori S, Satoh N, Paabo S (1998) The mitochondrial genome of the *hemichordate Balanoglossus carnosus* and the evolution of *deuterostome mitochondria*. *Genetics* 150, 1115–1123
- Chang YS, Huang FL, Lo TB (1994) The complete nucleotide-sequence and gene organization of carp (*Cyprinus-Carpio*) mitochondrial genome. *J Mol Evol* 38, 138–155
- Chenoweth SF, Hughes JM, Keenan CP, Lavery S (1998) Concordance between dispersal and mitochondrial gene flow, isolation by distance in a tropical teleost, *Lates calcarifer* (Australian barramundi). *Heredity* 80, 187–197
- Chou R, Lee HB (1997) Commercial marine fish farming in Singapore. *Aquac Res* 28, 767–776



- Chubb AL, Zink RM, Fitzsimons JM (1998) Patterns of mtDNA variation in Hawaiian freshwater fishes. The phylogeographic consequences of amphidromy. *J Heredity* 89, 8–16
- Crease TJ (1999) The complete sequence of the mitochondrial genome of (*Cladocera, Crustacea*). *Gene* 233, 89–99
- Curole AP, Kocher TD (1999) Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends Ecol Evol* 14, 394–398
- Dalgaard JZ, Garrett RA (1993) "Archaeal hyperthermophile genes". In: Kates M, Kushner DJ, Matheson AT, eds. (Amsterdam: Elsevier) pp 535–563
- Elmerot C, Arnason U, Gojobori T, Janke A (2002) The mitochondrial genome of the pufferfish, *Fugu rubripes*, and ordinal teleostean relationships. *Gene* 295, 163–172
- Galtier N, Lobry JR (1997) Relationships between genomic G+C content, RNA secondary structures and optimal growth temperature in prokaryotes. *J Mol Evol* 44, 632–663
- Hiendleder S (1996) Molecular characterization of the sheep mitochondrial genome. *J Anim Breed Genet* 113, 293–302
- Inoue JG, Miya M, Aoyama J, Ishikawa S, Tsukamoto K, Nishida M (2001) Complete mitochondrial DNA sequence of the Japanese eel *Anguilla japonica*. *Fish Sci* 67, 118–125
- Jiggins CD, McMillan WO, King P, Mallet J (1997) The maintenance of species differences across a *Heliconius* hybrid zone. *Heredity* 79, 495–505
- Johansen S, Bakke I (1996) The complete mitochondrial DNA sequence of Atlantic cod (*Gadus morhua*), relevance to taxonomic studies among codfishes. *Mol Mar Biol Biotechnol* 5, 203–214
- Katsares V, Apostolidis A, Triantafyllidis A, Kouvatzi A, Triantaphyllidis C (2003) Development of mitochondrial DNA primers for use with homarid lobsters. *Mar Biotechnol* 5, 469–479
- Kim YJ, Kweon HS, Kim IC, Lee YM, Kim JM, Lee JS (2004) The complete mitochondrial genome of the floating goby, *Gymnogobius petschiliensis* (Perciformes, Gobiidae). *Mol Cells* 17, 446–453
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2, molecular evolutionary genetics analysis software. *Bioinformatics* 17, 1244–1245
- Lee WJ, Kocher TD (1995) Complete sequence of a sea lamprey (*Petromyzon-Marinus*) mitochondrial genome—early establishment of the vertebrate genome organization. *Genetics* 139, 873–887
- Lee WJ, Conroy J, Howell WH, Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. *J Mol Evol* 41, 54–66
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25, 955–964
- Machida RJ, Miya MU, Nishida M, Nishida S (2004) Large-scale gene rearrangements in the mitochondrial genomes of two calanoid copepods *Eucalanus bungii* and *Neocalanus cristatus* (*Crustacea*), with notes on new versatile primers for the srRNA and COI genes. *Gene* 332, 71–78
- McLean JE, Taylor EB (2001) Resolution of population structure in a species with high gene flow, microsatellite variation in the eulachon (*Osmeridae, Thaleichthys pacificus*). *Mar Biol* 139, 411–420
- Meyer A (1993) "Evolution of mitochondrial DNA in fishes". In: Hochachka PW, Mommsen TP, eds. (Amsterdam: Elsevier) pp 1–38
- Miya M, Takeshima H, Endo H, Ishiguro NB, Inoue JG, Mukai T, Satoh TP, Yamaguchi M, Kawaguchi A, Mabuchi K, Shirai SM, Nishida M (2003) Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Mol Phylogenet Evol* 26, 121–138
- Nelson J (1994) *Fishes of the World*. (New York: John Wiley & Sons)
- Nishibori M, Hayashi T, Tsudzuki M, Yamamoto Y, Yasue H (2001) Complete sequence of the Japanese quail (*Coturnix japonica*) mitochondrial genome and its genetic relationship with related species. *Anim Genet* 32, 380–385
- Nishibori M, Hanazono M, Yamamoto Y, Tsudzuki M, Yasue H (2003) Complete nucleotide sequence of mitochondrial DNA in chickens, White Leghorn and White Plymouth Rock. *Anim Sci* 47, 437–439
- Noguchi Y, Endo K, Tajima F, Ueshima R (2000) The mitochondrial genome of the brachiopod *Laqueus rubellus*. *Genetics* 155, 245–259
- Sambrook J, Russell DW (2001) *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press)
- Spruyt N, Delarbre C, Gachelin G, Laudet V (1998) Complete sequence of the amphioxus (*Branchiostoma lanceolatum*) mitochondrial genome, relations to vertebrates. *Nucleic Acids Res* 26, 3279–3285
- Stepien CA, Rosenblatt RH, Bargmeyer BA (2001) Phylogeography of the spotted sand bass, *Paralabrax maculatofasciatus*: Divergence of Gulf of California and Pacific Coast populations. *Evolution* 55, 1852–1862
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25, 4876–4882
- Wang HC, Hickey DA (2002) Evidence for strong selective constraint acting on the nucleotide composition of 16S ribosomal RNA genes. *Nucleic Acids Res* 30, 2501–2507
- Yanagimoto T, Kitamura T, Kobayashi T (2004) Complete nucleotide sequence and variation of mitochondrial DNA from 10 individuals of walleye pollock, *Theragra chalcogramma*. *Fish Sci* 70, 885–895
- Yue GH, Orban L (2001) Rapid isolation of DNA from fresh and preserved fish scales for polymerase chain reaction. *Mar Biotechnol* 3, 199–204
- Yue GH, Chen F, Orban L (2000) Rapid isolation and characterization of microsatellites from the genome of Asian arowana (*Scleropages formosus*, *Osteoglossidae*, *Pisces*). *Mol Ecol* 9, 1007–1009
- Yue GH, Li Y, Chao TM, Chou R, Orban L (2002) Novel microsatellites from Asian sea bass (*Lates calcarifer*) and their application to broodstock analysis. *Mar Biotechnol* 4, 503–511
- Yue GH, Ho MY, Orban L, Komen J (2004) Microsatellites within genes and ESTs of common carp and their applicability in silver crucian carp. *Aquaculture* 234, 85–98