Acta Oceanol. Sin., 2013, Vol. 32, No. 8, P. 35-43
DOI: 10.1007/s13131-013-0339-8
http://www.hyxb.org.cn
E-mail: hyxbe@263.net

Characterization of the *cyp19a1a* gene from a BAC sequence in half-smooth tongue sole (*Cynoglossus semilaevis*) and analysis of its conservation among teleosts

SHAO Changwei[1,2], LIU Geng[3], LIU Shanshan[2], LIU Changlin[2], CHEN Songlin[2]*

[1] Ocean University of China, Qingdao 266071, China
[2] Yellow Sea Fisheries Research Institute, Chinese Academy of Fisheries Sciences, Qingdao 266071, China
[3] BGI-Shenzhen, Shenzhen 518000, China

Received 11 June 2012; accepted 22 December 2012

©The Chinese Society of Oceanography and Springer-Verlag Berlin Heidelberg 2013

Abstract
The *cyp19a1a* gene encodes an aromatase that plays a key role in sex differentiation of the gonad. The first bacterial artificial chromosome (BAC) sequence of half-smooth tongue sole (*Cynoglossus semilaevis*) containing the intact *cyp19a1a* gene was reported and the conservation and synteny of the *cyp19a1a* gene among teleosts were analyzed in the study. The BAC is 107 367 bp in size, with an overall guanine-cytosine (GC) content of 43.44%, and contains 4.38% transposable elements. Nine genes were predicted, including seven functional genes and two hypothetical genes. The *cyp19a1a* gene of all tested teleosts has nine exons and eight introns, and potential binding sites flanking the transcriptional start site are conserved. The expression pattern among teleosts is also similar during ovarian differentiation. Synteny analysis revealed a conserved gene cluster *PKH4B-SL9A5-FHOD3-CEBPG-CEBPA* among teleosts. These findings suggest that, among teleosts, *cyp19a1a* genes not only have similar genomic structures, but also have conserved functions. The genomic environment of *cyp19a1a* in tongue sole is not universal in teleosts, reflecting the particular evolution of tongue sole *cyp19a1a* after it diverged from the other teleosts.
Key words: BAC, *cyp19a1a*, conservation, gene synteny, half-smooth tongue sole

Citation: Shao Changwei, Liu Geng, Liu Shanshan, Liu Changlin, Chen Songlin. 2013. Characterization of the *cyp19a1a* gene from a BAC sequence in half-smooth tongue sole (*Cynoglossus semilaevis*) and analysis of its conservation among teleosts. Acta Oceanologica Sinica, 32(8): 35–43, doi: 10.1007/s13131-013-0339-8

1 Introduction

Cytochrome P450 aromatase (P450arom) is a terminal enzyme of the aromatase complex that catalyzes the synthesis of androgens to estrogens. Estrogens are generally regarded as exogenous hormones responsible for reproductive physiology in vertebrates, especially sex determination and differentiation in teleosts. For example, inhibiting aromatase activity can result in various degrees of masculinization, including complete phenotypic and functional males in some teleost fishes (Huang et al., 2009), such as Chinook salmon, *Oncorhynchus tshawytscha* (Piferrer et al., 1994); Nile tilapia, *Oreochromis niloticus*; rainbow trout, *Oncorhynchus mykiss* (Guiguen et al., 1999); Japanese flounder, *Paralichthys olivaceus* (Kitano et al., 2000); and zebrafish, *Danio rerio* (Uchida et al., 2004). Thus, P450 aromatase is considered a key steroid for ovarian differentiation and sex reversal during the critical developmental stage of some teleost species.

The *cyp19* gene, which encodes P450arom, is believed to be present as a single copy in the genome of most mammals and its tissue-specific expression is achieved by alternative splicing and/or different promoter usage (Simpson et al., 1993). In humans, for example, a single *cyp19* gene is located on chromosome 15, containing nine coding exons and one noncoding exon. The multiple tissue-specific first exons in human *cyp19* produce different transcripts in various tissues and are responsible for its tissue-specific expression (Mendelson et al., 1987; Simpson et al., 1993). By contrast, most teleost species have two separate and distinctive genes designated as *cyp19a1a* and *cyp19a1b*, resulting from the fish-specific genome duplication (FSGD) (Callard and Tchoudakova, 1997). The *cyp19a1a* gene, which encodes P450aromA, is predominantly present in the gonad, while the *cyp19a1b* gene, which encodes P450aromB, is mostly brain-specific. For most teleost species, the initiation codon of *cyp19a1a* is located in exon 1 (Chang et al., 2005; Galay-Burgos et al., 2006; Kanda et al., 2006; Zhang et al., 2008). By contrast, the start codon is in the exon 2 of *cyp19a1b*, which is consistent with the organization of the mammalian *cyp19* gene (Chang et al., 2005; Dalla Valle et al., 2005). The two *cyp19* genes have different tissue specificities, expression patterns, enzymatic activities, and regulation modes, as reported in zebrafish, *D. rerio* (Kishida and Callard, 2001; Trant et al., 2001); European sea bass, *Dicentrarchus labrax* (Blazquez and Piferrer, 2004); Nile tilapia, *O. niloticus* (Chang et al., 2005); Atlantic halibut, *Hippoglossus hippoglossus* (Matsuoka et al., 2006; van Nes et al., 2005); common carp, *Cyprinus carpio* (Barney et al., 2008); red-spotted grouper, *Epinephelus akaara* (Huang et al.,

Foundation item: The National Nature Science Foundation of China under contract Nos 41006107, 31072202 and 31130057; the National High-Technology R&D Program of China (863 Program) under contract No. 2012AA10A403-2; the Special Fund for Agro-scientific Research in the Public Interest of China under contract No. 200903046; Taishan Scholar Project Fund of Shandong of China.
*Corresponding author, E-mail: chensl@ysfri.ac.cn

2009); yellowtail clownfish, *Amphiprion clarkii* (Kobayashi et al., 2010); and catfish, *Clarias gariepinus* (Sridevi et al., 2012). The existence of two *cyp19* genes was first reported in goldfish, *Carassius auratus*. *Cyp19a1b* was discovered as a highly abundant 3 kb transcript in the brain by hybridization, while a 1.9 kb *cyp19a1a* cDNA was isolated through a stepwise PCR cloning strategy in goldfish. The sequence identity between these two genes is only 62%, which is less than that of goldfish *cyp19a1a* and *cyp19a1a*, compared to that of other teleosts (68%–72%); however, presumptive functional domains are highly conserved (Tchoudakova and Callard, 1998).

In half-smooth tongue sole (*Cynoglossus semilaevis*), as in other teleost species, the *cyp19* gene also has two isoforms encoding brain and gonadal aromatases, respectively (Deng et al., 2009a). The tongue sole has a distinct trait whereby the females grow two to four times faster than males. More importantly, the natural population exhibits a skewed sex ratio, with almost 70% of offspring being male because of spontaneous sex reversal (Ji et al., 2010). A high male ratio is undesirable for aquaculture, resulting in overall reduction of output and heavy economic loss. Although the tongue sole employs a ZW/ZZ system of genetic sex determination (GSD), the phenotypic sex is influenced by both genetic and environmental factors, including temperature, sex hormones, and other environmental factors. Treatment with high temperature (28°C) during the critical developmental stage leads to about 70% sex-reversed ZW male individuals (Deng et al., 2007; Tian et al., 2011). The same phenomenon was observed using 80 μg/L methyl-testosterone (MT), which is an inhibitor of aromatizable estrogens, suggesting that the aromatase was involved in the sex differentiation of tongue sole (Chen et al., 2008).

The importance of the aromatase in sex differentiation, particularly in teleosts, makes it necessary to understand the common features of the *cyp19* gene among independent teleost lineages, such as conserved regulation mechanisms and genomic environment. To date, however, most studies of the *cyp19* gene have focused on the expression pattern, promoter activity, and function in sexual differentiation in a single species. Only a few studies have addressed structural and functional conservation of the *cyp19* gene among teleosts (Castro et al., 2005; Guiguen et al., 2010).

In the present study, we report the sequencing and characterization of a BAC clone that contains the full-length *cyp19a1a* gene of tongue sole. Comparison of *cyp19a1a* genes, including their genomic structures, expression patterns, and gene synteny suggested that the *cyp19a1a* gene is relatively conserved among teleosts.

2 Materials and methods

2.1 *Tongue sole samples*

Half-smooth tongue sole at all stages of development were collected from Shandong Huanghai Aquaculture Co., Ltd. For the early stages, including the 4 days post hatching (dph), 8 dph, 25 dph, and 48 dph of tongue sole fries, the undifferentiated gonads situating at the anterior part of the whole intestine were removed and then cut into several parts using fine forceps. The suitability of isolated gonad parts was assessed using the sex-specific marker (CseF-SSR1) and gonad-specific genes (*dmrt1* and *cyp19a1a*) (Chen et al., 2012; Deng et al., 2008, 2009a). Gonads collected from five to ten fries and three adults of the same age and same gender were pooled and stored at –80°C for RNA isolation.

2.2 *BAC sequence and assembly*

A BAC clone, Hind025N15, containing the intact *cyp19a1a* gene was previously identified by Overgo hybridization, originating from the BAC libraries of tongue sole (Shao et al., 2010). BAC DNA purified by Qiagen Q-Tip100 was sheared by nebulization to an average size of 0.6 kb. After end-filling, DNA fragments were size-fractionated and cloned into the plasmid of pUC18. The resulting plasmids were electroporated into DH5α competent cells (TAKARA). Clones were picked into 96-well trays filled with LB culture medium, grown for 12 h and frozen until needed. Approximately 1 992 clones were sequenced with the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) on the ABI 3730 sequencing analyzer. Sequences were manually edited, aligned, and assembled using Sequencer software (PHRED and PHRAP) (Ewing et al., 1998) (http://www.phrap.org/).

2.3 *Analysis of the sequence data*

Repetitive elements were identified by RepeatMasker (Version 3.2.6) against the Repbase TE library (Version 2008-08-01). Then the BAC sequence was masked by RepeatMasker to perform gene predictions using two *de novo* prediction software programs, Genscan (Salamov and Solovyev, 2000) and Augustus (Stanke and Waack, 2003), trained by human genome annotation. The predicted genes were checked by the domain search software interpro (http://www.ebi.ac.uk/interpro/). A gene with significant homology [E less than (e-20)] to a known protein was classified according to the human protein name, while the other gene without significant homology to any protein was classified as a hypothetical protein. The homology searches were carried out using the Basic Blast program, at http://www.ncbi.nlm.nih.gov/blast/.

The synteny comparisons across different teleost species, including *Takifugu rubripes, Tetraodon nigroviridis, Oryzias latipes, Gasterosteus aculeatus*, and *D. rerio* were carried out *in silico* by means of reciprocal BLAST 'best-hit' searches using the ENSEMBL database (http://www.ensembl.org). Orthologous gene location data were retrieved from the genomes of teleosts, and were arranged by reference to the order of BAC genes.

For the *cyp19a1a* gene, the deduced amino acid sequences of tongue sole, together with *cyp19a1a* amino acid sequences reported for other teleost species, were aligned by ClustalW Version 1.6 using default settings (Thompson et al., 1994). Furthermore, the conserved domain was identified by GENEDOC program Version 2.6.02 (http://www.psc.edu/biomed/genedoc/). A phylogenetic tree was constructed with the neighbor joining method using Mega 5 (Tamura et al., 2011) and the *cyp19* gene from chicken and human served as an out group to root the tree. The exon-intron boundaries, as well as the intron sequences, were determined by comparing the genomic sequence with the deduced amino acid sequence using Genewise2-2-0 (Birney et al., 2004). The transcription start site and putative transcriptional factor binding sites were identified by MatInspector (http://www.genomatix.de).

2.4 *Expression of the cyp19a1a gene during the sex differentiation stages*

Total RNA of gonads at different developmental stages was isolated and reversely transcribed as described previously (Dong et al., 2011; Chen et al., 2001). A pair of primers (F:

GGTGAGGATGTGACCCAGTGT; R: ACGGGCTGAAATCGCAAG) for quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis was designed using the Primer3 program. The final PCR reactions contained 0.4 mmol/L of each primer; 10 μl SYBR Green (Invitrogen) and 80 ng template of cDNA. qRT-PCR was performed on ABI PRISM 7500 Real-Time PCR System using Hotstart *Taq* polymerase (Qiagen) in a final volume of 20 μl and β-actin gene was used as internal reference. All reactions were subjected to: 95°C for 35 s followed by 40 cycles at 95°C for 5 s, 60°C for 34 s, and then 95°C for 15 s, 60°C for 1 min, and 95°C for 15 s. Melting curve analysis was applied to all reactions to ensure homogeneity of the reaction product. The result was analyzed using 7500 System SDS Software.

3 Results

3.1 *Characterization of the BAC sequence*

The sequence of BAC Hind025N15 (GenBank accession No. JQ003881) was 107 367 bp and assembled from approximately 1 589 useful reads, with average length of 548 bp, covering about 7.69 times the length of the BAC. Two gaps in a region that could not be fully sequenced and/or assembled probably represented a rich-repeat sequence. The GC content of the whole BAC sequence was 43.44% and the sequence consisted of about 5.14% transposable elements (TEs). DNA transposons were predominant and made up 2.72% of the entire BAC clone, covering 6 390 bp. Retrotransposons accounted for 2.42% of the BAC length, including 1.21% long interspersed nuclear element (LINE, 2 850 bp) and 0.13% short interspersed nuclear element (SINE, 307 bp). A 2 553 bp length of long terminal repeat (LTR) alone accounted for 1.08% of all transposable elements (Table 1).

Genscan and Augustus software predicted nine genes from the BAC Hind025N15 sequence. Seven of them had homologous genes when analyzed by BLASTP, BLASTN, and BLASTX, while the other two genes (*G0001* and *G0002*) were defined as hypothetical genes. The gene order along the BAC was *CP19A, IDH3A, PKH4B, SL9A5, FHOD3, G0001, CEBPG, CEBPA*, and *G0002*. The average size of the nine genes in this BAC was 5 469 bp and the gene density was 1 gene/11 929 bp. The first gene was the target gene, *cyp19a1a*, which was 2 457 bp and located at 3 854 nt to 6 311 nt on Hind025N15 (Table 1). The next three functional genes were 2 448 bp (*IDH3A*), 13 532 bp (*PKH4B*) and 6 150 bp (*SL9A5*) long, respectively (Table 1). *IDH3A* encodes isocitrate dehydrogenase 3 (NAD+) alpha, an enzyme that catalyzes the oxidative decarboxylation of isocitrate to alpha-ketoglutarate (alpha-KG), requiring nicotinamide adenine dinucleotide (NAD) as a cofactor (Huh et al., 1996). *PKH4B* encodes pleckstrin homology domain-containing family G member 4B, which is involved in the process of rho protein signal transduction (Nagase et al., 2001). *SL9A5* encodes solute carrier family 9 member 5, which is involved in pH regulation to eliminate acids generated by active metabolism (Klanke et al., 1995). The fifth gene, *FHOD3*, is the largest gene (17 716 bp) and encodes formin homology 2 domain containing 3, which might play a role in the regulation of the actin cytoskeleton (Iskratsch et al., 2010). The others two small genes (*CEBPG* and *CEBPA*) encode CCAAT/enhancer binding proteins (C/EBPs), which are DNA-binding proteins that recognize two different motifs: the CCAAT homology motif common to many promoters and the enhanced core homology motif common to many enhancers (Hattori et al., 2003).

Table 1. Characterization of a BAC sequence in half-smooth tongue sole

BAC	Size/bp	Coverage (×)	Total reads No.	Useful reads No.	Average length/bp	Contig No.	GC/%
Hind025N15	107 367	7.69	1 992	1 589	548	3	43.44
Repeat type	DNA	LINE	SINE	LTR	Other	Unknown	Total
Length/bp	6 390	2 850	307	2 553	0	0	10 289
Genome/%	2.72	1.21	0.13	1.08	0	0	5.14
Protein gene	CP19A	IDH3A	PKH4B	SL95A	FHOD3	CEBPG	CEBPA
Gene length/bp	2 457	2 448	13 532	6 150	17 716	491	935

3.2 *Genomic organization of cyp19a1a*

The exon/intron organization of the tongue sole *cyp19a1a* gene was determined by comparing the BAC sequence with *cyp19a1a* cDNA obtained by homology cloning. In addition, the genomic structures of other teleost *cyp19a1a* genes were confirmed using the whole genome sequence and cDNA of *cyp19a1a*. All the teleost *cyp19a1a* genes have nine exons and eight introns, which are inserted at exactly the same positions as found in human (Tanaka et al., 1995) (Fig. 1b). All donor and acceptor sites of these introns were GT and AG, respectively, following the GT/AG rule. The total length of the introns of *cyp19a1a* in different teleosts varied from 702 bp to 13 968 bp, while the total length of exons varied from 1 402 to 1 578 bp, suggesting that the variation in the size of *cyp19a1a* among teleosts is accounted for by the varying lengths of their introns (Fig. 1b). *Takifugu* and *Tetraodon cyp19a1a* genes span only 2 224 and 2 256 bp, respectively, being much smaller than the zebrafish *cyp19a1a* gene (15 519 bp), because of their extremely small introns (758 bp and 702 bp, respectively).

The deduced amino acid sequence of the tongue sole *cyp19a1a* was aligned with other teleost P450aromA homologs (Fig. 2), and a phylogenetic tree was constructed using human and chicken P450aromA as out groups and including P450aromB sequences. In the resulting tree, the brain and ovary aromatase present in teleosts forms two main clades (Fig. 3). Sequence alignment demonstrated five conserved putative functional domains, namely the membrane spanning region, the substrate binding region, the distal helix I region, the steroid binding region, and the heme-binding helix region (Fig. 2). Comparison of the deduced amino acid sequence revealed that the tongue sole P450aromA shared a high sequence identity to P450aromA of medaka (77.7%), *Takifugu* (76.4%), *Tetraodon* (75.1%), stickleback (76.3%), Japanese flounder (77.0%), southern flounder (76.1%), and barfin flounder (75.0%). However, it had a relatively lower sequence identity to P450aromA of zebrafish (68.1%) and lower levels of overall sequence identity with humans (58.4%) and chickens (60.2%). The P450aromA and P450aromB sequences of tongue sole were only 64.7% identical.

To further explore the conservation of the teleost *cyp19a1a*, we performed a detailed comparison of 1 000 bp upstream of the translation start sites from zebrafish, tongue sole,

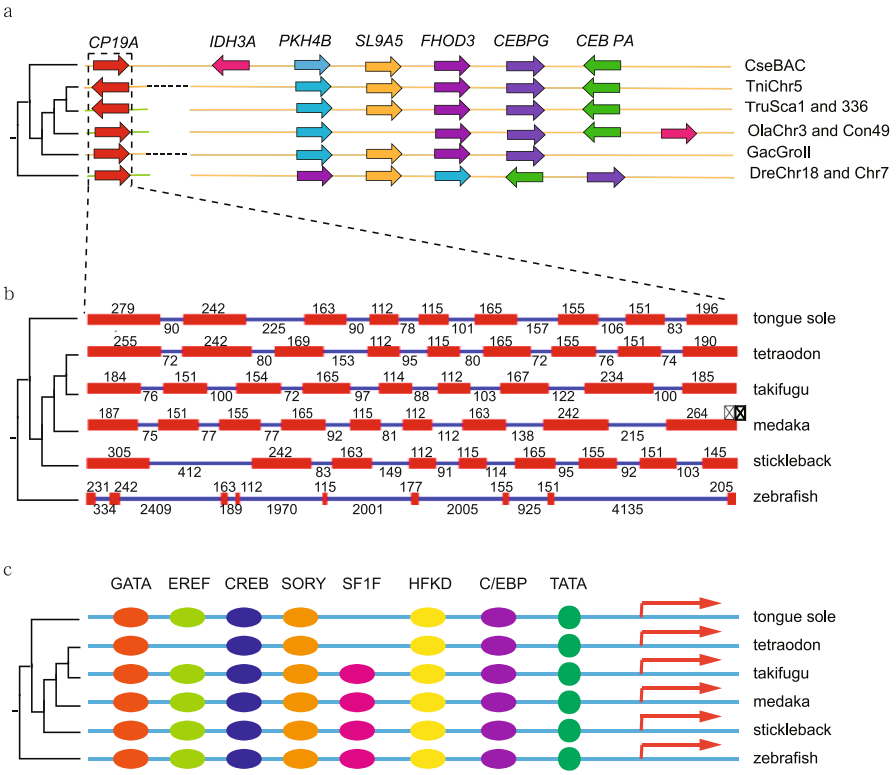

Fig.1. Comparison of the *cyp19a1a* gene among six different teleost species. a. Teleost gene synteny. Genomic organization and gene synteny comparisons across teleosts for genes annotated from tongue sole BAC clone Hind025N15. Genes are depicted by colored arrows; the arrowhead indicates the orientation for each gene. Gene names are indicated above the tongue sole genes, and orthologs across species are depicted in the same colors. The black dashed lines from the genomic region of *Takifugu* and stickleback indicate that the *cyp19a1a* and the other orthologs had a large distance. For the genomic regions of *Tetraodon*, zebrafish, and medaka, the orthologs distributed on the different chromosomes or scaffolds are indicated by two color lines. CseBAC: BAC sequence of tongue sole; TniChr5: Chromosome 5 of *Takifugu*; TruScaf1 and 336: scaffold 1 and scaffold 336 of *Tetraodon*; OlaChr3 and Con49: Chromosome 3 and ultracontig 49 of medaka; GacGroII: Group II of stickleback; DreChr18 and Chr7: Chromosome 18 and Chromosome 7 of zebrafish. b. Exon and intron structure of *cyp19a1a*. Gene exon and intron structures are depicted for six teleost species. Exons for *cyp19a1a* are represented by red boxes connected by introns, which are depicted as blue lines. Exon and intron lengths are labeled in base pairs; exon lengths are indicated directly below each exon, and intron lengths are indicated directly above each intron. c. Transcription factor binding sites in the *cyp19a1a* promoter region. The green circle represents the conserved TATA box and the other colored circles with the names above the tongue sole BAC sequence represent the potential binding sites predicted by MatInspector. The red arrows indicate the transcriptional start sites. Transcription factor binding sites are absolutely conserved in different teleost species but the order and spacing of base pairs are poorly conserved. Thus, the schematic map representing the conservation of promoter regions was depicted after artificial contraposition.

medaka, *Takifugu*, *Tetraodon*, and stickleback. We identified a highly conserved TATA box and several potential regulatory elements, including C/EBP (CCAAT binding factors), FKHD (fork head domain factors), SF1F (vertebrate steroidogenic factor), SORY (SOX/SRY-sex/testis determining and related HMG box factors), CREB (cAMP-responsive element binding proteins), EREF (estrogen response elements), and GATA (GATA binding factors) among the six different teleost species (Fig. 1c). However, a binding site for SF-1, which is an important transcription factor regulating *cyp19* genes in human ovaries, was not identified in the 5′ flanking region of tongue sole and *Tetraodon cyp19a1a*. In addition, the EREF (the binding site of the estrogen receptor) motif was not detected on the *Tetraodon cyp19a1a* sequence. MatInspector detected other potential binding sites that were not conserved in most teleost species, such as an aryl-hydrocarbon receptor (AhR) recognition site, a nerve growth factor inducible-B protein, and a doublesex and mab-3 related transcription factor 1 (DMRT).

3.3 *Expression of the cyp19a1a gene in tongue sole*

The expression of *cyp19a1a* in tongue sole during developmental stages from 4 dph to 2 years old was quantified by qRT-PCR. The expression of the *cyp19a1a* gene increased significantly in gonad samples of developing juveniles at 70 dph, with a significant dimorphic expression pattern between males and females. *Cyp19a1a* expression increased to a peak level of 160 dph in both males and females. The expression decreased thereafter in females and persisted in the mature ovary, while in males the expression was lower relative to females (Fig. 4).

3.4 *Gene synteny*

Seven protein-coding genes of BAC Hind025N15 in tongue sole were further mapped onto other teleost genomes by BLASTP and the locations of the ortholog in the respective genomes were retrieved. The results showed that the gene order and orientation in the *PKH4B-SL9A5-FHOD3-CEBPG-CEBPA* cluster was conserved among five teleosts (medaka, *Tetraodon*,

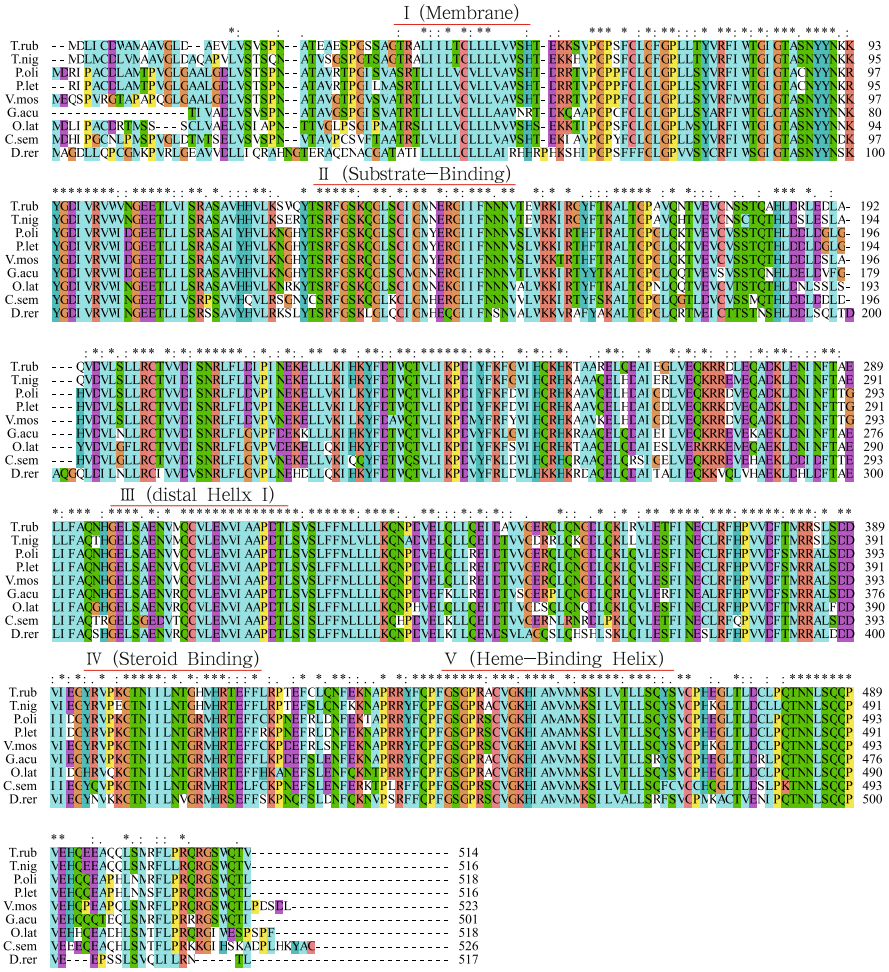

Fig.2. Amino acid sequence alignment of *cyp19a1a* proteins from different teleost species. *T. rub: Takifugu rubripes* (BAF93506); *T. nig: Tetraodon nigroviridis* (CAF99837); *P. oli: Paralichthys olivaceus* (BAA74777.1); *P. let: Paralichthys lethostigma* (AAX55671.1); *V. mos: Verasper moseri* (ACI04549); *G. acu: Gasterosteus aculeatus* (ACN70015); *O. lat: Oryzias latipes* (BAA11656); *C. sem: Cynoglossus semilaevis* (ABL74474); and *D. rer: Danio rerio* (AAK00643). The five conserved regions are covered by red lines and marked with Roman numerals I–V. The identity is indicated with stars, and gaps used to maximize the alignment are shown with dashes.

Takifugu, stickleback, and tongue sole), except that *SL95A* and *CEBPA* are missing in medaka and stickleback, respectively (Fig. 1a). Although zebrafish have the same genes as other teleost species in this region, there is no obvious conservation because of the intensive interchromosomal arrangement that caused gene disorder, including the pair of *FHOD3* and *PKH4B* and the pair of *CEBPA* and *CEBPG*. Besides, no significant similarity to *IDH3A* was identified in the *Tetraodon*, *Takifugu*, stickleback, and zebrafish genomes. However, in the medaka and tongue sole genomes, it was detected in the corresponding region where the synteny has not been maintained. The *cyp19a1a* (*CP19A*) gene was distributed on the same chromosome with conserved cluster in *Tetraodon* (Chromosome 5) and stickleback (Group II), but it is at least 4 Mb distant from the proximal gene *PKH4B* of the conserved cluster. In contrast, it was detected on the different chromosomes and/or scaffolds from the conserved cluster in *Takifugu* (Scaffold 336 versus Scaffold 1), medaka (Ultracontig 49 versus Chromosome 3), and zebrafish (Chromosome 7 versus Chromosome 18) (Fig. 1a).

4 Discussion

BAC clones are very useful for identifying specific gene organization across relatively long chromosomal distances and thus facilitate comparative analysis among different species, especially for non-model species with no available whole genome sequence (Yasukochi et al., 2011). Here, we reported a BAC clone containing the entire *cyp19a1a* gene from tongue sole, which exhibits a special function during embryonic development. We performed a comparative analysis among different lineages of teleost species, and revealed the conservation of this gene.

Cyp19a1a, which encodes P450aromatase, catalyzes the synthesis of estrogens and is a well-studied gene found in species ranging from teleosts to mammals (Diotel et al., 2010). In most teleosts, the structure and function of the *cyp19a1a* gene has been well conserved during evolution (Callard and Tchoudakova, 1997; Castro et al., 2005). The genomic structure of the tongue sole *cyp19a1a* gene, which contains nine exons and eight introns, is almost identical to those of other teleost species. Interestingly, although the gene size varies among different teleost species, there are only minor variations in the sizes of the CDS region (1 402–1 578 bp), while large dissimilarities were observed in the introns of different species, in particular between zebrafish (13 968 bp) and *Tetraodon* (702 bp). Thus, much of the difference in gene size among teleost species is accounted for by the differences in the sizes of their introns.

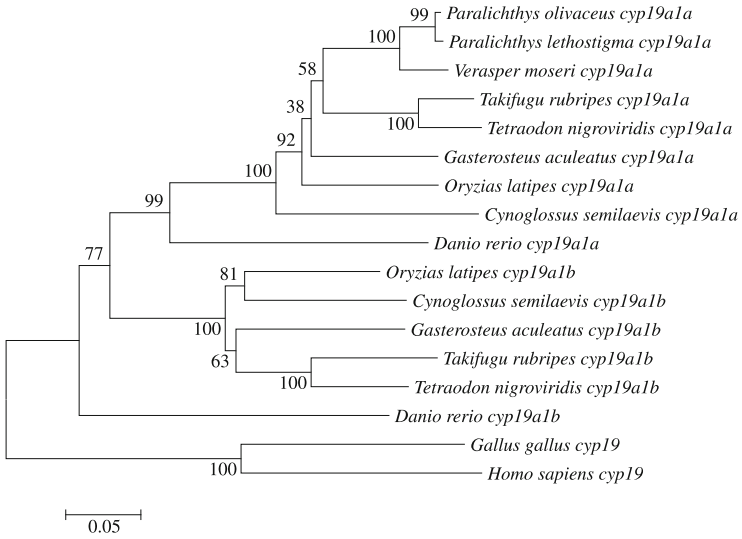

Fig.3. Phylogenetic tree of vertebrate aromatase genes. The scientific names of the organism, followed by the gene name, are shown. Distances are used to construct the phylogenetic tree and bootstrap values are based on 1 000 resampling replicates. The accession number of different organisms was followed by *Takifugu rubripes cyp19a1a* (BAF93506), *Takifugu rubripes cyp19a1b* (BAF93507), *Tetraodon nigroviridis cyp19a1a* (CAF99837), *Tetraodon nigroviridis cyp19a1b* (CAG05537), *Oryzias latipes cyp19a1a* (BAA11656), *Oryzias latipes cyp19a1b* (AAP83449), *Gasterosteus aculeatus cyp19a1a* (ACN70015), *Gasterosteus aculeatus cyp19a1b* (ENSGACP00000007910), *Cynoglossus semilaevis cyp19a1a* (ABL74- 474), *Cynoglossus semilaevis cyp19a1b* (ABM90641), *Paralichthys olivaceus* (BAA74777.1), *Paralichthys lethostigma* (AAX55671.1), *Verasper moseri* (ACI04549), *Danio rerio cyp19a1a* (AAK00643), *Danio rerio cyp19a1b* (AAK00642), *Homo sapiens* (AAR37047), and *Gallus gallus* (AAA48738).

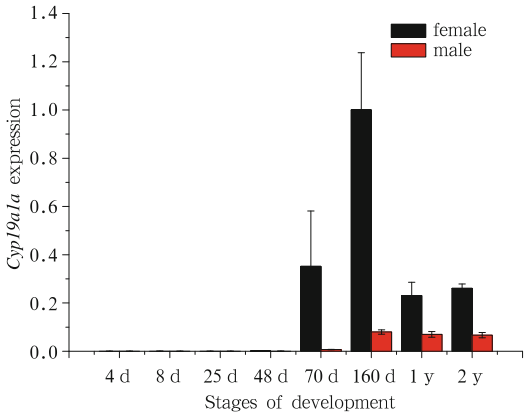

Fig.4. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis of *cyp19a1a* during developmental stages in female and male tongue sole. Vertical bars show mean ± standard error (SE) (n=3).

A cross-species comparison of P450aromA sequences among teleosts indicated five regions of greatest conservation, including the membrane spanning region (I), substrate binding region (II), distal helix I region (III), steroid binding region (IV), and heme-binding helix region (V). Conceivably, the steroid binding region, also known as the aromatase-specific domain, is functionally essential in all tested species, while I, II, III, and V have evolved conserved functions within the teleosts (Huang et al., 2009). Interestingly, similar to goldfish P450aromA (Callard and Tchoudakova, 1997), the tongue sole P450aromA sequence, despite having only about 64.7% identity with its ortholog P450aromB, possessed high identities with other teleost species (68.1%–77.7%), reflecting their conserved expression pattern in the ovary and brain, respectively.

DNA sequences denoting *cis*-acting elements that regulate the expression of the *cyp19a1a* gene by the binding of transcription factors are located within a region spanning ~1 000 bp flanking the transcription start site. The available genome sequences for teleost species allowed us to examine the mechanism by which the *cyp19a1a* promoter controls the level of transcription initiation, which largely determines the gene expression in different teleost species. The major findings were that the TATA box was consistently conserved with respect to position, and the most highly conserved regulatory elements (EREF, C/EBP, FKHD, SF1F, SORY, CREB, and GATA) clearly indicated a common mechanism of regulation in the tested teleost species. Imperfect consensus estrogen response elements (EREs) were identified in the *Tetraodon* sequence, as they were in the case of the gonadal aromatase promoter from the orange-spotted grouper sequence (Zhang et al., 2012). It would be interesting to investigate whether there is another regulatory model for the control of gonadal development in *Tetraodon* or whether this represents a mistake by the prediction software. As a binding site of transcriptional factor SF1 that mediated *cyp19a1a* expression during vitellogenesis in Nile tilapia (Yoshiura et al., 2003), SF1F was also not detected in the promoter of tongue sole and *Tetraodon cyp19a1a*. Further investigations are required to clarify this issue through additional experiments and as is also the case with the ERE in *Tetraodon*. FOXL2 is known as the earliest sexually dimorphic marker in ovarian development and is essential for granulosa cell differentiation and ovarian maintenance (Baron et al., 2004). Several other studies implicated FOXL2 as a candidate regulator of the gonadal aromatase gene as shown in rainbow trout (Kanda et al., 2006) and Nile tilapia (Wang et al., 2007). The other conserved potential sites, including C/EBP, SORY, CREB, and GATA have previously been descri-

bed in gonadal aromatase promoters from tilapia, European seabass, and medaka (Chang et al., 2005; Dalla Valle et al., 2002; Tanaka et al., 1995). Together, these studies contribute to our understanding of the *cyp19a1a* transcriptional regulation mechanism, but how the promoter controls transcriptional modulation and chromatin structure remodeling is still poorly understood and there is no complete analysis of the *cyp19a1a* promoter.

In general, the expression level of *cyp19a1a* is consistent with the level of estrogen, which is essential for ovarian development. A previous study revealed that the tongue sole *cyp19a1a* transcript was mainly expressed in the ovaries of females (ZW♀) and expressed at a low level in the gonads of pseudo-males (ZW♂), induced by both MT immersion and high temperature treatment (Deng et al., 2009a). This result indicates that the *cyp19a1a* gene is indispensable for ovarian differentiation in tongue sole. Thus, in the present study, we determined the expression pattern of *cyp19a1a* during the stages of sex determination and differentiation in tongue sole. The sexual dimorphism between females and males at 70 dph (the stage of gonadal differentiation) was accompanied by a spike in *cyp19a1a* expression, indicating the pivotal role of *cyp19a1a* in ovarian differentiation in tongue sole. Specific over-expression of *cyp19a1a* in vitellogenic follicles during oogenesis has been also observed in many fish species, including Nile tilapia (Sudhakumari et al., 2005), rainbow trout (Guiguen et al., 1999; Vizziano et al., 2007), European seabass (Blazquez et al., 2008), and medaka (Nakamoto et al., 2006; Patil and Gunasekera, 2008). The similar expression pattern of the *cyp19a1a* gene among teleost species also suggests its involvement in ovarian differentiation.

Comparative genomic analysis has been used extensively to identify similar genomic features and establish the extent of gene co-linearity among different species with close taxonomic relationships (Deng et al., 2009b). Our analysis of gene synteny between tongue sole BAC genes and their counterparts in five other teleost species revealed a conserved cluster *PKH4B-SL9A5-FHOD3-CEBPG-CEBPA*, suggesting a microcolinearity between teleost genomes. However, the *cyp19a1a* gene was not always in the conserved syntenic region, but was distributed far from the common conserved gene cluster in the *Tetraodon* and stickleback genomes, or even located on different chromosomes and/or scaffolds in the medaka, *Takifugu*, and zebrafish genomes. The scaffolds containing the counterparts of the tongue sole genes in medaka and *Takifugu* have not been mapped onto a chromosome; therefore, we lack sufficient information to conclude whether the *cyp19a1a* gene and the conserved cluster are distributed linearly on the same chromosome. To illustrate this issue, we analyzed regions upstream and downstream of the *cyp19a1a* gene in five other teleost species, based on the available genome sequences. Unexpectedly, two conserved gene clusters around the *cyp19a1a* gene were detected. One was the *DMXL2-GLDN-CP19A-TNFAIP8L3* cluster, which was detected in the syntenic region of medaka, stickleback, *Takifugu, Tetraodon*, and zebrafish genomes. The other cluster was *SCG3-LYSMD2-TMOD2-LEO1-MAPK6-USP50*, which was conserved in the medaka, *Takifugu, Tetraodon*, and zebrafish genomes, but not in the stickleback genome (data not shown). Furthermore, in human, all genes of the two conserved clusters exhibit a linear arrangement on human chromosome 15, indicating that the *cyp19a1a* gene has a common genomic environment (data not shown). Based on existing data, the different manner of *cyp19a1a* gene arrangement occurring in the syntenic region of tongue sole may have been caused by insertions, translocations, and/or duplication of the *cyp19a1a* gene after tongue sole diverged from the other teleost species. Taken together, despite the discrepancies existing in the syntenic region between tongue sole and other teleost species, all teleosts still share a highly conserved gene content.

In conclusion, BAC clone Hind025N15 represents the first large genomic sequence of tongue sole. Analysis of the TEs, GC content, and gene organization of the BAC sequence provides the first glimpse of the whole genome sequence profile of tongue sole. Comparative analysis revealed that the *cyp19a1a* gene has common conserved features, such as the exon/intron pattern, the potential binding sites upstream of the transcriptional start site, and its function in ovarian differentiation among teleosts. Although the *cyp19a1a* gene of tongue sole did not follow the common synteny, the detected gene cluster (*PKH4B-SL9A5-FHOD3-CEBPG-CEBPA*) was conserved in all tested fish species. These characteristics will make the BAC clone an excellent starting platform for future functional studies of the *cyp19a1a* gene. Additional whole genome sequencing efforts will provide further insight into the evolutionary history of the *cyp19a1a* gene in tongue sole.

References

Barney M L, Patil J G, Gunasekera R M, et al. 2008. Distinct cytochrome P450 aromatase isoforms in the common carp (*Cyprinus carpio*): sexual dimorphism and onset of ontogenic expression. Gen Comp Endocrinol, 156: 499–508

Baron D, Cocquet J, Xia Xuhua, et al. 2004. An evolutionary and functional analysis of FoxL2 in rainbow trout gonad differentiation. J Mol Endocrinol, 33: 705–715

Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. Genome Res, 14: 988–995

Blazquez M, Gonzalez A, Papadaki M, et al. 2008. Sex-related changes in estrogen receptors and aromatase gene expression and enzymatic activity during early development and sex differentiation in the European sea bass (*Dicentrarchus labrax*). Gen Comp Endocrinol, 158: 95–101

Blazquez M, Piferrer F. 2004. Cloning, sequence analysis, tissue distribution, and sex-specific expression of the neural form of P450 aromatase in juvenile sea bass (*Dicentrarchus labrax*). Mol Cell Endocrinol, 219: 83–94

Callard G V, Tchoudakova A. 1997. Evolutionary and functional significance of two CYP19 genes differentially expressed in brain and ovary of goldfish. J Steroid Biochem Mol Biol, 61: 387–392

Castro L F, Santos M M, Reis-Henriques M A. 2005. The genomic environment around the Aromatase gene: evolutionary insights. BMC Evol Biol, 5: 43

Chang X, Kobayashi T, Senthilkumaran B, et al. 2005. Two types of aromatase with different encoding genes, tissue distribution and developmental expression in Nile tilapia (*Oreochromis niloticus*). Gen Comp Endocrinol, 141: 101–115

Chen Songlin, Deng Siping, Ma Hongyu, et al. 2008. Molecular marker-assisted sex control in half-smooth tongue sole (*Cynoglossus semilaevis*). Aquaculture, 283: 7–12

Chen Songlin, Hong Yunhan, Scherer S J, et al. 2001. Lack of ultraviolet-light inducibility of the medaka fish (*Oryzias latipes*) tumor suppressor gene. Gene, 264: 197–203

Chen Songlin, Ji Xiangshan, Shao Changwei, et al. 2012. Induction of mito-gynogenetic diploids and identification of WW superfemale using sex-specific SSR markers in half-smooth tongue sole (*Cynoglossus semilaevis*). Mar Biotechnol, 14: 120–128

Dalla Valle L, Lunardi L, Colombo L, et al. 2002. European sea bass (*Dicentrarchus labrax L.*) cytochrome P450arom: cDNA cloning, expression and genomic organization. J Steroid Biochem Mol Biol, 80: 25–34

Dalla Valle L, Toffolo V, Vianello S, et al. 2005. Genomic organization of the CYP19b genes in the rainbow trout (*Oncorhynchus mykiss Walbaum*). J Steroid Biochem Mol Biol, 94: 49–55

Deng Siping, Chen Songlin. 2008. Molecular cloning, characterization and RT-PCR expression analysis of Dmrt1α from half-smooth tongue-sole, *Cynoglossus semilaevis*. Journal of Fishery Science of China, 15: 577–584

Deng Siping, Chen Songlin, Tian Yongsheng, et al. 2007. Gonadal differentiation and effects of temperature on sex determination in half-smooth tongue sole, *Cynoglossus semilaevis*. Journal of Fishery Science of China, 14: 714–719

Deng Siping, Chen Songlin, Xu Jianyong, et al. 2009a. Molecular cloning, characterization and expression analysis of gonadal P450 aromatase in the half-smooth tongue-sole, *Cynoglossus semilaevis*. Aquaculture, 287: 211–218

Deng Zhiyong, Li Yin, Xia Ran, et al. 2009b. Structural analysis of 83-kb genomic DNA from *Thellungiella halophila*: sequence features and microcolinearity between salt cress and *Arabidopsis thaliana*. Genomics, 94: 324–332

Diotel N, Le Page Y, Mouriec K, et al. 2010. Aromatase in the brain of teleost fish: expression, regulation and putative functions. Front Neuroendocrinol, 31: 172–192

Dong Xiaoli, Chen Songlin, Ji Xiangshan, et al. 2011. Molecular cloning, characterization and expression analysis of Sox9a and Foxl2 genes in half-smooth tongue sole (*Cynoglossus semilaevis*). Acta Oceanologica Sinica, 30: 68–77

Ewing B, Hillier L, Wendl M C, et al. 1998. Base-calling of automated sequencer traces using phred: I. Accuracy assessment. Genome Res, 8: 175–185

Galay-Burgos M, Gealy C, Navarro-Martin L, et al. 2006. Cloning of the promoter from the gonadal aromatase gene of the European sea bass and identification of single nucleotide polymorphisms. Comp Biochem Physiol A Mol Integr Physiol, 145: 47–53

Guiguen Y, Baroiller J F, Ricordel M J, et al. 1999. Involvement of estrogens in the process of sex differentiation in two fish species: the rainbow trout (*Oncorhynchus mykiss*) and a tilapia (*Oreochromis niloticus*). Mol Reprod Dev, 54: 154–162

Guiguen Y, Fostier A, Piferrer F, et al. 2010. Ovarian aromatase and estrogens: a pivotal role for gonadal sex differentiation and sex change in fish. Gen Comp Endocrinol, 165: 352–366

Hattori T, Ohoka N, Inoue Y, et al. 2003. C/EBP family transcription factors are degraded by the proteasome but stabilized by forming dimer. Oncogene, 22: 1273–1280

Huang Wei, Zhou Li, Li Zhi, et al. 2009. Expression pattern, cellular localization and promoter activity analysis of ovarian aromatase (*Cyp19a1a*) in protogynous hermaphrodite red-spotted grouper. Mol Cell Endocrinol, 307: 224–236

Huh T L, Kim Y O, Oh I U, et al. 1996. Assignment of the human mitochondrial NAD+ -specific isocitrate dehydrogenase alpha subunit (IDH3A) gene to 15q25.1–>q25.2 by in situ hybridization. Genomics, 32: 295–296

Iskratsch T, Lange S, Dwyer J, et al. 2010. Formin follows function: a muscle-specific isoform of FHOD3 is regulated by CK2 phosphorylation and promotes myofibril maintenance. J Cell Biol, 191: 1159–1172

Ji Xiangshan, Chen Songlin, Ma Hongyu, et al. 2010. Natural sex reversal of female *Cynoglossus semilaevis* in rearing populations. Journal of Fisheries of China, 34: 322–335

Kanda H, Okubo T, Omori N, et al. 2006. Transcriptional regulation of the rainbow trout CYP19a gene by FTZ-F1 homologue. J Steroid Biochem Mol Biol, 99: 85–92

Kishida M, Callard G V. 2001. Distinct cytochrome P450 aromatase isoforms in zebrafish (*Danio rerio*) brain and ovary are differentially programmed and estrogen regulated during early development. Endocrinology, 142: 740–750

Kitano T, Takamune K, Nagahama Y, et al. 2000. Aromatase inhibitor and 17alpha-methyltestosterone cause sex-reversal from genetical females to phenotypic males and suppression of P450 aromatase gene expression in Japanese flounder (*Paralichthys olivaceus*). Mol Reprod Dev, 56: 1–5

Klanke C A, Su Y R, Callen D F, et al. 1995. Molecular cloning and physical and genetic mapping of a novel human Na+/H+ exchanger (NHE5/SLC9A5) to chromosome 16q22.1. Genomics, 25: 615–622

Kobayashi Y, Horiguchi R, Miura S, et al. 2010. Sex- and tissue-specific expression of P450 aromatase (*cyp19a1a*) in the yellowtail clownfish, *Amphiprion clarkii*. Comp Biochem Physiol A Mol Integr Physiol, 155: 237–244

Matsuoka M P, van Nes S, Andersen O, et al. 2006. Real-time PCR analysis of ovary- and brain-type aromatase gene expression during Atlantic halibut (*Hippoglossus hippoglossus*) development. Comp Biochem Physiol B Biochem Mol Biol, 144: 128–135

Mendelson C R, Evans C T, Simpson E R. 1987. Regulation of aromatase in estrogen-producing cells. J Steroid Biochem, 27: 753–757

Nagase T, Kikuno R, Ohara O. 2001. Prediction of the coding sequences of unidentified human genes. XXI. The complete sequences of 60 new cDNA clones from brain which code for large proteins. DNA Res, 8: 179–187

Nakamoto M, Matsuda M, Wang Deshou, et al. 2006. Molecular cloning and analysis of gonadal expression of Foxl2 in the medaka, *Oryzias latipes*. Biochem Biophys Res Commun, 344: 353–361

Patil J G, Gunasekera R M. 2008. Tissue and sexually dimorphic expression of ovarian and brain aromatase mRNA in the Japanese medaka (*Oryzias latipes*): implications for their preferential roles in ovarian and neural differentiation and development. Gen Comp Endocrinol, 158: 131–137

Piferrer F, Zanuy S, Carrillo M, et al. 1994. Brief treatment with an aromatase inhibitor during sex differentiation causes chromosomally female salmon to develop as normal, functional males. J Exp Zool, 270: 255–262

Salamov A A, Solovyev V V. 2000. Ab initio gene finding in Drosophila genomic DNA. Genome Res, 10: 516–522

Shao Changwei, Chen Songlin, Scheuring C F, et al. 2010. Construction of two BAC libraries from half-smooth tongue sole *Cynoglossus semilaevis* and identification of clones containing candidate sex-determination genes. Mar Biotechnol, 12: 558–568

Simpson E R, Mahendroo M S, Means G D, et al. 1993. Tissue-specific promoters regulate aromatase cytochrome P450 expression. J Steroid Biochem Mol Biol, 44: 321–330

Sridevi P, Chaitanya R K, Dutta-Gupta A, et al. 2012. FTZ-F1 and FOXL2 up-regulate catfish brain aromatase gene transcription by specific binding to the promoter motifs. Biochim Biophys Acta, 1819: 57–66

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics, 19(Suppl 2): ii215–225

Sudhakumari C C, Senthilkumaran B, Kobayashi T, et al. 2005. Ontogenic expression patterns of several nuclear receptors and cytochrome P450 aromatases in brain and gonads of the Nile tilapia *Oreochromis niloticus* suggests their involvement in sex differentiation. Fish Physiol Biochem, 31: 129–135

Tamura K, Peterson D, Peterson N, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol, 28: 2731–2739

Tanaka M, Fukada S, Matsuyama M, et al. 1995. Structure and promoter analysis of the cytochrome P-450 aromatase gene of the teleost fish, medaka (*Oryzias latipes*). J Biochem, 117: 719–725

Tchoudakova A, Callard G V. 1998. Identification of multiple CYP19 genes encoding different cytochrome P450 aromatase isozymes in brain and ovary. Endocrinology, 139: 2179–2189

Thompson J D, Higgins D G, Gibson T J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res, 22: 4673–4680

Tian Yongsheng, Wang Di, Xu Ying, et al. 2011. Effects of rearing temperature on growth and sex determination in the half-smooth tongue sole (*Cynoglossus semilaevis*). Journal of Fisheries of China, 35: 176–182

Trant J M, Gavasso S, Ackers J, et al. 2001. Developmental expression of cytochrome P450 aromatase genes (CYP19a and CYP19b) in zebrafish fry (*Danio rerio*). J Exp Zool, 290: 475–483

Uchida D, Yamashita M, Kitano T, et al. 2004. An aromatase inhibitor or high water temperature induce oocyte apoptosis and depletion of P450 aromatase activity in the gonads of genetic female zebrafish during sex-reversal. Comp Biochem Physiol A Mol Integr Physiol, 137: 11–20

van Nes S, Moe M, Andersen O. 2005. Molecular characterization and expression of two cyp19 (P450 aromatase) genes in embryos, larvae, and adults of Atlantic halibut (*Hippoglossus hippoglossus*). Mol Reprod Dev, 72: 437–449

Vizziano D, Randuineau G, Baron D, et al. 2007. Characterization of early molecular sex differentiation in rainbow trout, *Oncorhynchus mykiss*. Dev Dyn, 236: 2198–2206

Wang Deshou, Kobayashi T, Zhou Linyan, et al. 2007. Foxl2 upregulates aromatase gene transcription in a female-specific manner by binding to the promoter as well as interacting with ad4 binding protein/steroidogenic factor 1. Mol Endocrinol, 21: 712–725

Yasukochi Y, Tanaka-Okuyama M, Kamimura M, et al. 2011. Isolation of BAC clones containing conserved genes from libraries of three distantly related moths: a useful resource for comparative genomics of Lepidoptera. J Biomed Biotechnol, 2011: 165894

Yoshiura Y, Senthilkumaran B, Watanabe M, et al. 2003. Synergistic expression of Ad4BP/SF-1 and cytochrome P-450 aromatase (ovarian type) in the ovary of Nile tilapia, *Oreochromis niloticus*, during vitellogenesis suggests transcriptional interaction. Biol Reprod, 68: 1545–1553

Zhang Weimin, Lu Huijie, Jiang Haiyan, et al. 2012. Isolation and characterization of *cyp19a1a* and *cyp19a1b* promoters in the protogynous hermaphrodite orange-spotted grouper (*Epinephelus coioides*). Gen Comp Endocrinol, 175: 473–487

Zhang Yang, Zhang Weimin, Yang Huiyi, et al. 2008. Two cytochrome P450 aromatase genes in the hermaphrodite ricefield eel *Monopterus albus*: mRNA expression during ovarian development and sex change. J Endocrinol, 199: 317–331