Gene 519 (2013) 18–25

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Genome-wide identification and divergent transcriptional expression of StAR-related lipid transfer (START) genes in teleosts

Huajing Teng [a,c,1], Wanshi Cai [a,b,1], Kun Zeng [c], Fengbiao Mao [a,b], Mingcong You [c], Tao Wang [c], Fangqing Zhao [a,*], Zhongsheng Sun [a,c,**]

[a] Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China
[b] Graduate University of the Chinese Academy of Sciences, Beijing 100049, China
[c] Institute of Genomic Medicine, Wenzhou Medical College, Wenzhou 325035, China

ARTICLE INFO

Article history:
Accepted 30 January 2013
Available online 13 February 2013

Keywords:
START genes
Gene loss
Fish-specific genome duplication
Transcriptional expression pattern

ABSTRACT

The lipid transfer reactions and the steroidogenic acute regulatory protein (StAR)-related lipid transfer (START) genes have a major role in lipid metabolism. However, START genes and their physiological functions in teleost fishes are relatively unknown. Through genome-wide screening, we identified and annotated 91 START genes in 5 teleost species. Although START domain-containing proteins are augmented in teleost genomes relative to tetrapod genomes, a similar number of genes are shared between them. Asymmetry of paralogous gene loss within the teleost START family and an extra copy of some START genes in teleosts resulting from fish-specific genome duplication have been demonstrated. A distinct transcriptional expression pattern within members of some START groups under different developmental stages suggests divergent functions within the same group in the developmental process. In addition, an asymmetric molecular evolution rate deviating from the neutral expectation has been observed in 7 of 14 teleost fish extra-duplicated pairs. The present study provides valuable information for increasing our understanding of the evolution and gene expression divergence under developmental stages of the START gene family in teleost fishes.

1. Introduction

Gene duplication, which has long been regarded as one of the major forces of evolution, can facilitate the acquisition of new functions of duplicated genes through neo-functionalization (Ohno, 1970) or partitioning of the ancestral gene functions between descendant duplicated genes by sub-functionalization (Conant and Wolfe, 2008; Force et al., 1999; He and Zhang, 2005; Lynch and Force, 2000). Tandem duplication, segmental duplication and whole-genome duplication (WGD) are major gene duplication mechanisms in eukaryotes. WGD is particularly intriguing because it has been regarded as a parsimonious evolutionary innovation of gene duplication (Haldane, 1932; Ohno, 1970; Taylor and Raes, 2004). It is well accepted that fish-specific genome duplication (FSGD) occurred prior to the teleost radiation (Amores et al., 1998; Christoffels et al., 2004; Jaillon et al., 2004; Teng et al., 2010; Vandepoele et al., 2004; Woods et al., 2005). However, the total number of genes in teleost species is not twice that present in tetrapod species, which prompted us to investigate the gene loss event after FSGD and to investigate whether fish-specific duplicated paralogs evolve at similar rates after duplication. As the results of various gene duplication and loss events, gene families provide a unique source for studying the evolutionary relationships between genes both within and between organisms. Changes in family size due to lineage-specific gene duplication or loss might provide insights into the evolutionary forces that have shaped eukaryotic genomes (Demuth et al., 2006). Thus, inferring an evolutionary scenario for a gene family is essential to understanding the phenotypic diversification of eukaryotic organisms (Hanada et al., 2009; Sato et al., 2009).

The steroidogenic acute regulatory protein (StAR)-related lipid transfer (START) domain, named after the mammalian 30 kDa StAR protein, is a protein module of around 200 amino acids implicated in the control of several aspects of lipid biology, including lipid trafficking, lipid metabolism and cell signaling (Alpy and Tomasetto, 2005; Soccio and Breslow, 2003). Mutation or misexpression of some START proteins was also reported to link to some pathological processes, including genetic disorders, autoimmune disease and cancer (Alpy and Tomasetto, 2005). Members of the START domain family have been shown to bind different ligands, such as sterols (e.g., StAR or STARD1) and lipids (e.g., PCTP or

Abbreviations: START, steroidogenic acute regulatory protein-related lipid transfer; WGD, whole-genome duplication; FSGD, fish-specific genome duplication; PH, pleckstrin homology; SAM, sterile alpha motif; RhoGAP, Rho-type GTPase-activating protein; 4HBT, 4-hydroxybenzoate thioesterase; HMM, hidden Markov model; d_N, nonsynonymous substitutions per nonsynonymous site; d_S, synonymous substitutions per synonymous site; RTK, receptor tyrosine kinase; Dr, *Danio rerio*; Tr, *Takifugu rubripes*; Tn, *Tetraodon nigroviridis*; Ol, *Oryzias latipes*; Ga, *Gasterosteus aculeatus*; Lc, *Latimeria chalumnae*; Xt, *Xenopus tropicalis*; Ac, *Anolis carolinensis*; Gg, *Gallus gallus*; Mm, *Mus musculus*; Hs, *Homo sapiens*; Ci, *Ciona intestinalis*.

* Corresponding author. Tel.: +86 10 64869325; fax: +86 10 64880586.
** Correspondence to: Z. Sun, Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China. Tel.: +86 10 64864959; fax: +86 10 64880586.
E-mail addresses: biofqzhao@gmail.com (F. Zhao), zsunusa@yahoo.com (Z. Sun).
[1] These authors contributed equally to this work.

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.01.058

STARD2), and exhibit enzymatic activity. Some other functional domains that were found associated with START in animals include pleckstrin homology (PH), sterile alpha motif (SAM), Rho-type GTPase-activating protein (RhoGAP), and 4-hydroxybenzoate thioesterase (4HBT) (Schrick et al., 2004; Soccio and Breslow, 2003). Ligand binding by START domain can regulate the activity of other domains within multi-domain proteins, such as the RhoGAP domain, the homeodomain and the thioesterase domain (Iyer et al., 2001; Ponting and Aravind, 1999). START domain is evolutionarily conserved in plants and animals. Fifteen START domain-containing proteins (STARD1–STARD15) have been identified in humans (Soccio and Breslow, 2003), and hundreds have been determined in invertebrates, bacteria and plants. However, only a very few START homologs have been reported in teleost fishes, which are the largest and most diverse group of vertebrates. The availability of sequenced and assembled genomes of zebrafish (*Danio rerio*), fugu (*Takifugu rubripes*) (Aparicio et al., 2002), Green Spotted Puffer (*Tetraodon nigroviridis*) (Jaillon et al., 2004), medaka (*Oryzias latipes*) (Kasahara et al., 2007) and three-spined stickleback (*Gasterosteus aculeatus*) has provided an opportunity for the genome-wide screening of START homologs and comparative analysis in teleost fish species.

In this study, we identified and annotated the START gene family members in teleosts through genome-wide screening and we investigated their transcript expression profile under experimental conditions, domain composition and phylogenetic relationships. In addition, a relative rate test was used to examine whether one of the duplicates has evolved at an accelerated rate following the duplication. With such an in-depth investigation, we expected to provide a detailed case in studying how genes evolve after gene duplication, and provide some data for future physiological function research of START genes in fishes.

2. Materials and methods

2.1. Data sets and phylogenetic analysis

Fifteen human START proteins were used as the query sequences in BLASTP and TBLASTN searches ($E<1e^{-5}$) against the NCBI (Maglott et al., 2010) or Ensembl databases (March 2011) (Flicek et al., 2011) of human, mouse, chicken, anole lizard, Western clawed frog, coelacanth, zebrafish, medaka, Green Spotted Puffer, three-spined stickleback, fugu and Sea squirts. Each matching sequence was used iteratively to search the databases until no new sequence was found. Additionally, a hidden Markov model (HMM) search (Johnson et al., 2010) was done in the proteome databases of the species listed above using the START domain (PFAM, PF01852). All protein sequences derived from the collected candidate START genes were further examined using the PFAM program (Mistry and Finn, 2007) with the default cut-off parameters. The amino acid sequence alignment of START domains was generated using MUSCLE (Edgar, 2004) with the default setting. A bootstrap consensus phylogenetic tree was constructed using the maximum-likelihood method in MEGA5 (Tamura et al., 2011) under the JTT + G model.

2.2. Analysis of synteny

All predicted genes within 20 Mb of each human or mouse START paralog were obtained using the BioMart mode in Ensembl (March 2011). Genes exhibiting orthologous relationship in both species (human/mouse) and supported by phylogenetic analysis were selected for syntenic analysis. Neighboring genes flanking the chicken, anole lizard, Western clawed frog, zebrafish, medaka, fugu, Green Spotted Puffer or three-spined stickleback START paralogs were obtained using the BioMart mode in Ensembl from dataset of the chicken (WASHUC2), anole lizard (AnoCar2.0), Western clawed frog (JGI_4.2), zebrafish (Zv9), medaka (MEDAKA1) fugu (FUGU4), Green Spotted Puffer (TETRAODON8) or three-spined stickleback (BROADS1) genome, respectively. Blocks of synteny were constructed on the basis of the orthologous relationship of genes among different species.

2.3. Transcriptional expression analysis of teleost START genes

The genome-wide microarray data of zebrafish (Domazet-Loso and Tautz, 2010), medaka (Iwahashi et al., 2009), Green Spotted Puffer (Chan et al., 2009) and three-spined stickleback were obtained from the NCBI Gene Expression Omnibus (GEO) with accession numbers GSE24616, GSE15380, GSE12976 and GSE34783, respectively. Extraction and filtration of each microarray data were processed as previously described (Chan et al., 2009; Domazet-Loso and Tautz, 2010; Iwahashi et al., 2009). Probe sets corresponding to the putative zebrafish STARTs were identified from Agilent Zebrafish (V2) Gene Expression Microarrays, NimbleGen Oryzias latipes_TIGR_rel5 (GFC023) 27 k array, Agilent custom 44 K Tetraodon array or Agilent-016492 three-spined stickleback 44 K 60 nt oligo array version 1.0. If more biological replicates were used in a specific experiment, such as 2–4 replicates in zebrafish, 5 replicates in three-spined stickleback and 3 replicates in medaka, mean of the expression values among the replicates were used. For genes with more than one set of probes, the mean of expression values was considered. Finally, the $\log_2$ transformed transcript intensity data were hierarchically clustered on the basis of the Euclidean distance with complete linkage in the Cluster program (de Hoon et al., 2004), and the relative transcript accumulation was represented in a color code with green or red showing the lower or higher levels of transcriptional expression, respectively.

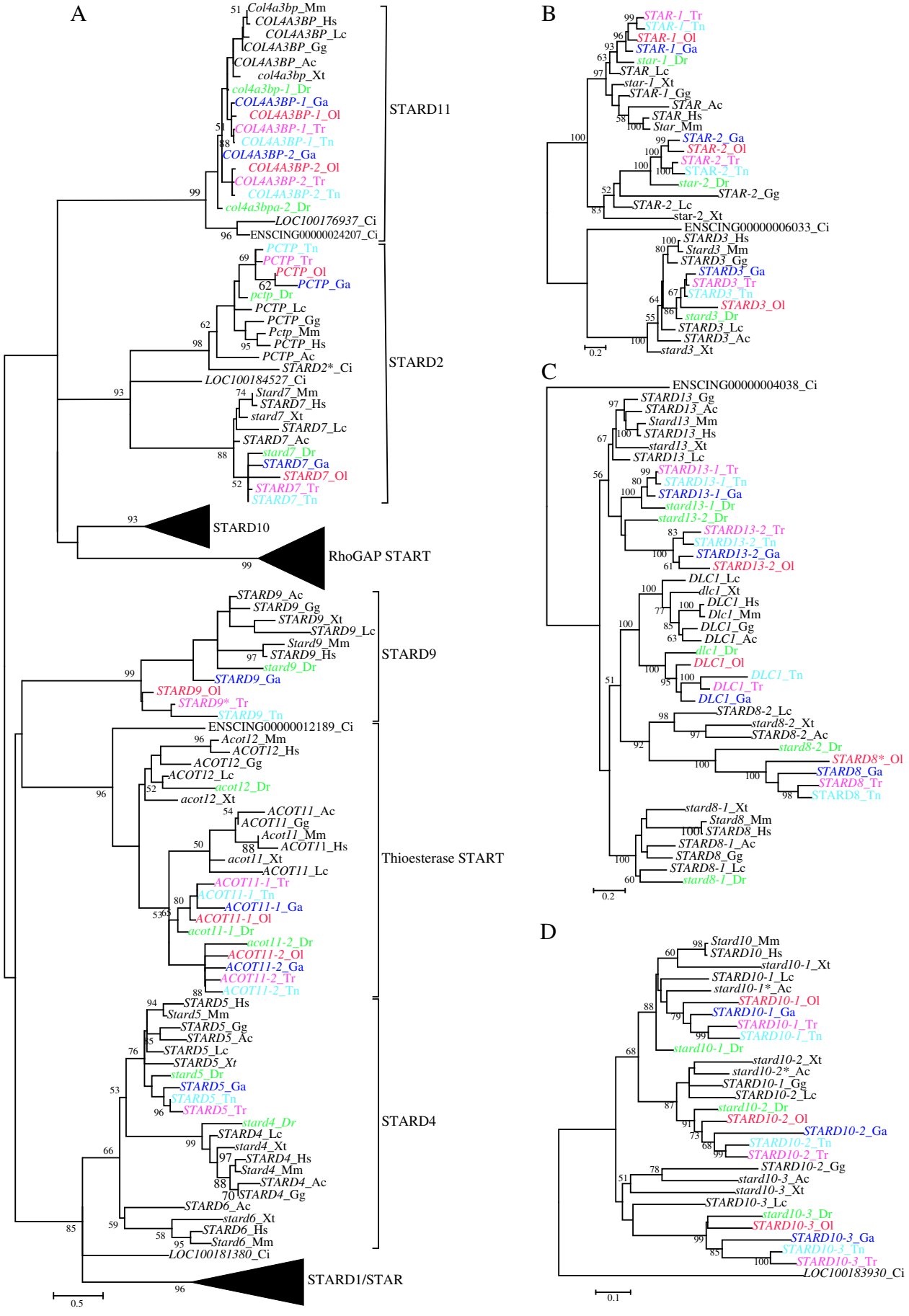
2.4. Divergence and relative rate test of duplicated teleost START pairs

The coding sequences of the duplicated teleost START pairs were aligned following the amino acid alignment by CodonAlign 2.0 (http://homepage.mac.com/barryghall/CodonAlign.html). Pairwise calculation of d_N/d_S between these teleost START pairs is estimated with the yn00 program of PAML4 (Yang, 2007). Further, nonparametric relative rate tests were done with amino acid sequences to investigate whether one of these teleost START pairs has evolved at an accelerated rate following the duplication using MEGA (Tamura et al., 2011). To test whether some sites were under positive selection, several site-specific models (M0, M1, M2, M3, M7 and M8) and branch site test 2 were used to detect positive selection using the codeml program implemented in PAML4 (Yang, 2007).

3. Results

3.1. Identification and phylogenetic analysis of START genes

Through extensive similarity-based searches, we identified 91 teleost START genes: 18 in Green Spotted Puffer, three-spined stickleback or fugu, 21 in zebrafish, and 16 in medaka (Supplemental Table 1). The length of STARTs in teleosts ranged from 198 to 1928 amino acid residues, and the number of exons ranged from 5 to 19 (Supplemental Table 1 and Fig. 1). In teleosts, about half of the START domain-containing proteins (48/91) are multi-domain proteins. Functional domains associated with START in teleosts include pleckstrin homology (PH) in STARD11s(COL4A3BPs), sterile alpha motif (SAM), Rho-type GTPase-activating protein (RhoGAP) in STARD8s, STARD12s(DLC1s) and STARD13s, and 4-hydroxybenzoate thioesterase (4HBT) in STARD14s(ACOT11s), consistent with earlier reports for non-teleosts (Alpy and Tomasetto, 2005; Schrick et al., 2004; Soccio and Breslow, 2003) (Supplemental Fig. 1). Although START domain-containing proteins are augmented in teleost genomes relative to mammalian genomes (Supplemental Table 1), similar gene numbers are found for each. In order to investigate the evolutionary relationship among these teleost START genes, 91 teleost START genes, 18 coelacanth and 80 tetrapod START orthologs, and 9 ascidian START genes were used in phylogenetic analysis by the maximum-likelihood method in MEGA5. The analysis unambiguously defined 8 START groups with high bootstrap values (Fig. 1); namely, the STARD1/STAR, STARD4, RhoGAP START, STARD2, STARD10, STARD11, thioesterase START and STARD9 group. The gene number of


A
Col4a3bp_Mm
COL4A3BP_Hs
COL4A3BP_Lc
COL4A3BP_Gg
COL4A3BP_Ac
col4a3bp_Xt
col4a3bp-1_Dr
COL4A3BP-1_Ga
COL4A3BP-1_Ol
COL4A3BP-1_Tr
COL4A3BP-1_Tn
COL4A3BP-2_Ga
COL4A3BP-2_Ol
COL4A3BP-2_Tr
COL4A3BP-2_Tn
col4a3bpa-2_Dr
LOC100176937_Ci
ENSCING00000024207_Ci
STARD11
PCTP_Tn
PCTP_Tr
PCTP_Ol
PCTP_Ga
pctp_Dr
PCTP_Lc
PCTP_Gg
Pctp_Mm
PCTP_Hs
PCTP_Ac
STARD2*_Ci
LOC100184527_Ci
Stard7_Mm
STARD7_Hs
stard7_Xt
STARD7_Lc
STARD7_Ac
stard7_Dr
STARD7_Ga
STARD7_Ol
STARD7_Tr
STARD7_Tn
STARD2
STARD10
RhoGAP START
STARD9_Ac
STARD9_Gg
STARD9_Xt
STARD9_Lc
Stard9_Mm
STARD9_Hs
stard9_Dr
STARD9_Ga
STARD9_Ol
STARD9*_Tr
STARD9_Tn
STARD9
ENSCING00000012189_Ci
Acot12_Mm
ACOT12_Hs
ACOT12_Gg
ACOT12_Lc
acot12_Dr
acot12_Xt
ACOT11_Ac
ACOT11_Gg
Acot11_Mm
ACOT11_Hs
acot11_Xt
ACOT11_Lc
ACOT11-1_Tr
ACOT11-1_Tn
ACOT11-1_Ga
ACOT11-1_Ol
acot11-1_Dr
acot11-2_Dr
ACOT11-2_Ol
ACOT11-2_Ga
ACOT11-2_Tr
ACOT11-2_Tn
Thioesterase START
STARD5_Hs
Stard5_Mm
STARD5_Gg
STARD5_Ac
STARD5_Lc
STARD5_Xt
stard5_Dr
STARD5_Ga
STARD5_Tn
STARD5_Tr
stard4_Dr
STARD4_Lc
stard4_Xt
STARD4_Hs
Stard4_Mm
STARD4_Ac
STARD4_Gg
STARD6_Ac
stard6_Xt
STARD6_Hs
Stard6_Mm
LOC100181380_Ci
STARD1/STAR
0.5
B
STAR-1_Tr
STAR-1_Tn
STAR-1_Ol
STAR-1_Ga
star-1_Dr
STAR_Lc
star-1_Xt
STAR-1_Gg
STAR_Ac
STAR_Hs
Star_Mm
STAR-2_Ga
STAR-2_Ol
STAR-2_Tr
STAR-2_Tn
star-2_Dr
STAR-2_Gg
STAR-2_Lc
star-2_Xt
ENSCING00000006033_Ci
STARD3_Hs
Stard3_Mm
STARD3_Gg
STARD3_Ga
STARD3_Tr
STARD3_Tn
STARD3_Ol
stard3_Dr
STARD3_Lc
STARD3_Ac
stard3_Xt
0.2
C
ENSCING00000004038_Ci
STARD13_Gg
STARD13_Ac
Stard13_Mm
STARD13_Hs
stard13_Xt
STARD13_Lc
STARD13-1_Tr
STARD13-1_Tn
STARD13-1_Ga
stard13-1_Dr
stard13-2_Dr
STARD13-2_Tr
STARD13-2_Tn
STARD13-2_Ga
STARD13-2_Ol
DLC1_Lc
dlc1_Xt
DLC1_Hs
Dlc1_Mm
DLC1_Gg
DLC1_Ac
dlc1_Dr
DLC1_Ol
DLC1_Tn
DLC1_Tr
DLC1_Ga
STARD8-2_Lc
stard8-2_Xt
STARD8-2_Ac
stard8-2_Dr
STARD8*_Ol
STARD8_Ga
STARD8_Tr
STARD8_Tn
stard8-1_Xt
Stard8_Mm
STARD8_Hs
STARD8-1_Ac
STARD8_Gg
STARD8-1_Lc
stard8-1_Dr
0.2
D
Stard10_Mm
STARD10_Hs
stard10-1_Xt
STARD10-1_Lc
stard10-1*_Ac
STARD10-1_Ol
STARD10-1_Ga
STARD10-1_Tr
STARD10-1_Tn
stard10-1_Dr
stard10-2_Xt
stard10-2*_Ac
STARD10-1_Gg
STARD10-2_Lc
stard10-2_Dr
STARD10-2_Ol
STARD10-2_Ga
STARD10-2_Tn
STARD10-2_Tr
STARD10-2_Gg
stard10-3_Ac
stard10-3_Xt
STARD10-3_Lc
stard10-3_Dr
STARD10-3_Ol
STARD10-3_Ga
STARD10-3_Tn
STARD10-3_Tr
LOC100183930_Ci
0.1

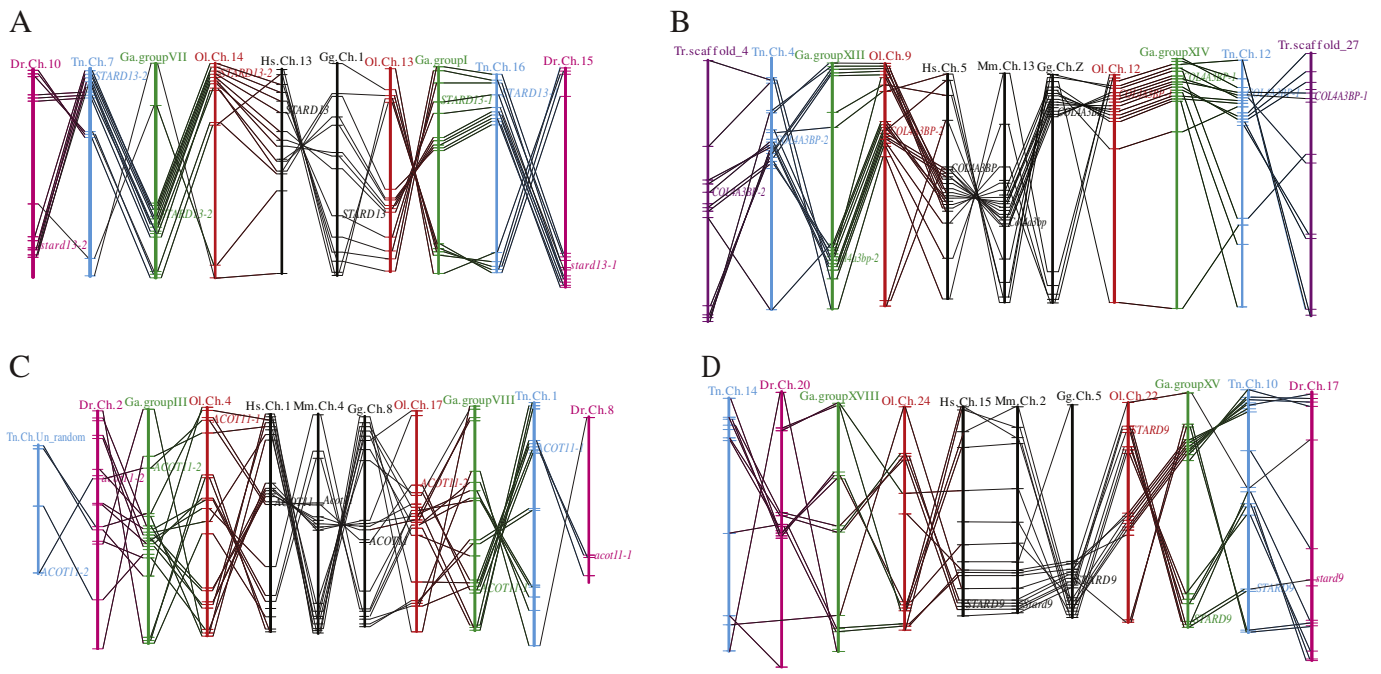


Fig. 2. Extra paralogons surrounding (A) *STARD13* (B) *COL4A3BP*, (C) *ACOT11* and (D) *STARD9* genes in teleost genomes. (A) The tetrapod START13 paralogon has two fish paralogons, the START13-1 paralogon and the START13-2 paralogon. Each of the *STOML3*, *DCLK1*, *SLC7A1*, *TRPC4*, *FREM2*, *SPG20*, *NBEA*, *B3GALTL*, *FRY* and *HMGB1* on human chromosome 13 near *START13* has two co-orthologs in medaka chromosomes 13 and 14, three-spined stickleback groups VII and I, Green Spotted Puffer chromosomes 16 and 7, zebrafish chromosomes 15 and 10. (B) The COL4A3BP-1 and COL4A3BP-2 fish paralogons are accompanied by one tetrapod START11 paralogon. *SV2C*, which is near *COL4A3BP* on human chromosome 5, has two co-orthologs located in medaka chromosomes 12 and 9, three-spined stickleback groups XIII and XIV, Green Spotted Puffer chromosomes 12 and 4, zebrafish chromosomes 5 and 21, fugu scaffolds 4 and 27. (C) Each ACOT11 paralogon defines a synteny with a high degree of conservation among the studied tetrapod and fish species. (D) Although no extra copy of START9 genes was found in some fish genomes, two conserved fish paralogons flanking each gene were observed, suggesting extra duplication of this STARD9 gene that occurred in the teleost ancestor, but was lost before the divergence of zebrafish. The positions of genes on chromosomes are not drawn to scale.

each group varies dramatically, ranging from 11 to 42. Some orthologs of tetrapod START genes, such as *STARD6*, are absent from all teleost fishes and some are absent from a certain fish lineage, such as *STARD4*_Tn and *STARD5*_Ol. Several START genes (*STARD1*, *STARD8*, *STARD10*, *COL4A3BD*, *STARD13* and *ACOT11*) have multiple copies in certain fish genomes, unlike mammalian genomes, indicating that extra rounds of duplication have occurred in the teleost lineage (Fig. 1).

3.2. Syntenic analysis

Orthologous genes flanking each START gene define a syntenic conservation among tetrapods and teleosts (Fig. 2; Supplemental Table 2, Supplemental Fig. 2 and Supplemental Fig. 3). Although some orthologs of tetrapod START genes, such as *STARD6*, are absent from all teleost fish genomes, syntenic regions of *STARD6* genes can be observed in teleosts (Supplemental Fig. 3). The same applies to some fish lineage-specific lost genes, e.g. *STARD4*_Tn, *ACOT12*_Tn, and *STARD*_Ol (Supplemental Fig. 3), suggesting that these genes were present in ancestral teleosts but were lost after the divergence of teleost fishes. Extra copies of some START genes, such as *STARD1*, *STARD8*, *STARD10*, *COL4A3BP*, *STARD13* and *ACOT11*, were observed in some fish genomes compared to mammalian genomes, indicating that extra duplication in the teleost lineage or mammalian-specific gene losses might have occurred in the evolutionary history of these genes (Postlethwait, 2007). Therefore, more tetrapod lineages (chicken, lizard and frog) were used in our synteny analysis to investigate their evolutionary history. Although frequent gene-linkage disruptions, micro-inversions or rearrangements occurred in teleosts, we could find some traces of extra paralogons around each of the STARD1, -8 or -10 genes in teleost and non-mammalian tetrapod genomes (Supplemental Fig. 2). For example, 2 chicken or fish paralogons STAR-1 and STAR-2 are accompanied by 1 mammalian STAR paralogon. It appears that extra paralogs of these 3 genes were generated in the earlier vertebrates before the divergence of the Teleostomi followed by mammalian-specific gene losses. Thus, the difference in gene number of STARD1, -8 or -10 genes between teleosts and mammals is most likely because of mammalian-specific gene losses (Postlethwait, 2007). However, extra paralogons around each of the *COL4A3BP*, *STARD13* and *ACOT11* genes were found in teleost genomes but not in tetrapod genomes. For example, the tetrapod START13 paralogon has 2 fish paralogons, START13-1 and START13-2 (Fig. 2). Each of *STOML3*, *DCLK1*, *SLC7A1*, *TRPC4*, *FREM2*, *SPG20*, *NBEA*, *B3GALTL*, *FRY* and *HMGB1* on human chromosome 13 near START13 has two co-orthologs on medaka chromosomes 13 and 14, three-spined stickleback groups VII and I, Green Spotted Puffer chromosomes 16 and 7 and zebrafish chromosomes 15 and 10, which suggests that extra copies of the COL4A3BP, STARD13 and ACOT11 genes in teleosts might be the results of FSGD (Kasahara et al., 2007). Although no extra copy of the STARD9 gene was found in some fish genomes, 2 conserved fish paralogons flanking each gene were also

Fig. 1. (A) Phylogenetic analysis of the START gene family and detailed phylogeny of (B) STARD1/STAR, (C) RhoGAP START and (D) STARD10. The bootstrap consensus phylogenetic tree was constructed using the maximum-likelihood method in MEGA5, and the numbers indicate the percentage bootstrap support. The symbol '◂' represents the compressed START group of STARD1/STAR, RhoGAP START or STARD10, which are expanded in (B), (C) or (D), respectively. *Gene predicted using FGENESH + software (http://linux1.softberry.com/berry.phtml). Dr, *Danio rerio*; Tr, *Takifugu rubripes*; Tn, *Tetraodon nigroviridis*; Ol, *Oryzias latipes*; Ga, *Gasterosteus aculeatus*; Lc, *Latimeria chalumnae*; Xt, *Xenopus tropicalis*; Ac, *Anolis carolinensis*; Gg, *Gallus gallus*; Mm, *Mus musculus*; Hs, *Homo sapiens*; Ci, *Ciona intestinalis*.

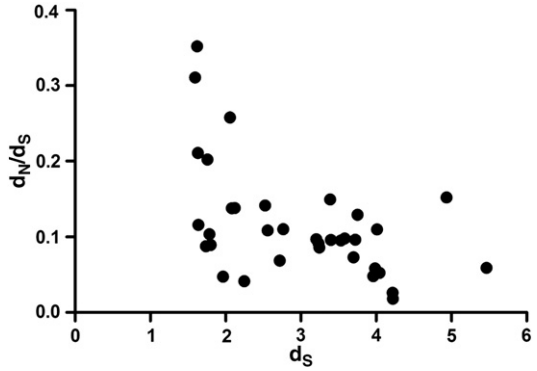


Fig. 3. Individual d_S and d_N/d_S of paralogous START gene pairs in the teleost.

observed. It is possible that extra duplication of this STARD9 gene occurred in the teleost ancestor but was lost before divergence of the zebrafish.

3.3. Divergence and relative rate test of extra-duplicated teleost START pairs

Modes of selection can be estimated by the ratio of the number of nonsynonymous substitutions per nonsynonymous site (d_N) to the number of synonymous substitutions per synonymous site (d_S), i.e. $d_N/d_S>1$ indicates positive selection; $d_N/d_S<1$ indicates purifying selection; and $d_N/d_S=1$ indicates neural evolution (Yang, 2007). The combination of phylogenetic and syntenic analyses revealed extra copies of *COL4A3BP*, *STARD13* and *ACOT11* genes in teleosts. These teleost genes were selected for further evolutionary analysis. Although positive selection has been pervasive during vertebrate evolution (Studer et al., 2008), none of the three site-specific positive selection models or branch site test 2 of PAML predicted any site under positive selection for any teleost START gene listed above with probabilities >95% (data not shown), and pairwise comparison of d_N/d_S between these duplicate pairs was markedly <1 (Fig. 3 Supplemental Table 3), suggesting that these ancient duplicates likely have been subject to purifying selection. An asymmetric molecular evolution rate deviating from the neutral expectation occurs in 7 of 14 fish-specific duplicated teleost START pairs (Table 1), suggesting that one paralog might evolve faster than another after duplication.

3.4. Differential transcript profiling of teleost START genes under experimental conditions

The level at which a gene is expressed under some conditions can provide useful clues to gene function. To examine the transcript abundance patterns of the START genes, we used a comprehensive expression analysis with the publicly available microarray data for zebrafish (Domazet-Loso and Tautz, 2010), medaka (Iwahashi et al., 2009), Green Spotted Puffer (Chan et al., 2009) and three-spined stickleback. All of the 18 three-spined stickleback START genes and 8 of the Green Spotted Puffer START genes were expressed in all detected tissues, but the mRNA level of different genes peaked in different tissues. The difference of steady-state levels of START transcripts between tissues was greater than that between environmental populations. In zebrafish, 14 START genes were detected to be expressed during the ontogenetic progression phase, and expression profiles of them defined clearly different developmental stages (Fig. 4). Differences in transcript abundance levels of these START genes, such as the 400-fold difference of mean expression level between *stard10-3* and *stard10-2* during these stages, were observed. It indicated that the contributions of different STARTs to growth and development might be associated with their expression levels. According to the transcript profiling, 3 zebrafish START gene clusters were observed and the 8 START groups defined in our earlier phylogenetic analysis were all included in these clusters except the data-deficient STARD9 groups, but these clusters were not highly related with gene phylogeny. Contrary expression patterns within members of the STARD1/STAR, STARD4 and STARD10 groups in zebrafish, STARD1/STAR in Green Spotted Puffer and STARD2 in medaka were observed, suggesting divergent functions within the same group during the developmental process.

Transcriptional expression analysis of fish-specific duplicated START paralogs revealed that zebrafish COL4A3BP paralogs and three-spined stickleback ACOT11 paralogs have divergent expression patterns. Of the 3 duplicated zebrafish START10 paralogs identified by our phylogenetic study, *stard10-1* and *stard10-2*, but not *stard10-3*, have similar transcriptional expression patterns (Fig. 4). It appears that the transcriptional expression patterns of the paralogs have diverged during long-term evolution, suggesting functional diversification of duplicated genes.

4. Discussion

As the major organic constituents of fish, lipids function as major sources of metabolic energy for growth, reproduction and migration

Table 1
Tajima relative rate tests of teleost START duplicate genes[a].

Test group	*Mt*[b]	*M1*[c]	*M2*[d]	χ^2	*P*[e]
*col4a3bp-2*_Dr/*col4a3bp-1*_Dr with *COL4A3BP*_Hs	389	68	26	18.77	0.00001
*COL4A3BP-1*_Dr/*COL4A3BP-2*_Tn with *COL4A3BP*_Hs	490	43	48	0.27	0.60018
*COL4A3BP-1*_Ol/*COL4A3BP-2*_Ol with *COL4A3BP*_Hs	495	40	45	0.29	0.58759
*COL4A3BP-1*_Ga/*COL4A3BP-2*_Ga with *COL4A3BP*_Hs	488	31	35	0.24	0.62246
*COL4A3BP-1*_Tr/*COL4A3BP-2*_Tr with *COL4A3BP*_Hs	496	40	44	0.19	0.66252
*stard13-1*_Dr/*stard13-2*_Dr with *STARD13*_Hs	705	55	74	2.80	0.09436
*STARD13-1*_Tn/*STARD13-2*_Tn with *STARD13*_Hs	606	73	149	26.02	0.00000
*STARD13-1*_Ga/*STARD13-2*_Ga with *STARD13*_Hs	616	62	161	43.95	0.00000
*STARD13-1*_Tr/*STARD13-2*_Tr with *STARD13*_Hs	619	71	155	31.22	0.00000
*acot11-1*_Dr/*acot11-2*_Dr with *ACOT11*_Hs	255	25	18	1.14	0.28575
*ACOT11-1*_Tn/*ACOT11-2*_Tn with *ACOT11*_Hs	356	36	66	8.82	0.00297
*ACOT11-1*_Ol/*ACOT11-2*_Ol with *ACOT11*_Hs	362	47	56	0.79	0.37519
*ACOT11-1*_Ga/*ACOT11-2*_Ga with *ACOT11*_Hs	375	41	64	5.04	0.02480
*ACOT11-1*_Tr/*ACOT11-2*_Tr with *ACOT11*_Hs	368	42	70	7.00	0.00815

[a] The Tajima relative rate test was used to examine the equality of evolutionary rate between teleost START duplicate pairs.
[b] *Mt* is the sum of identical sites and divergent sites in all three sequences tested.
[c] *M1* is the number of unique differences in the first paralog.
[d] *M2* is the number of unique differences in the second paralog.
[e] If $P<0.05$ the test rejects the equal substitution rates between the two duplicates.

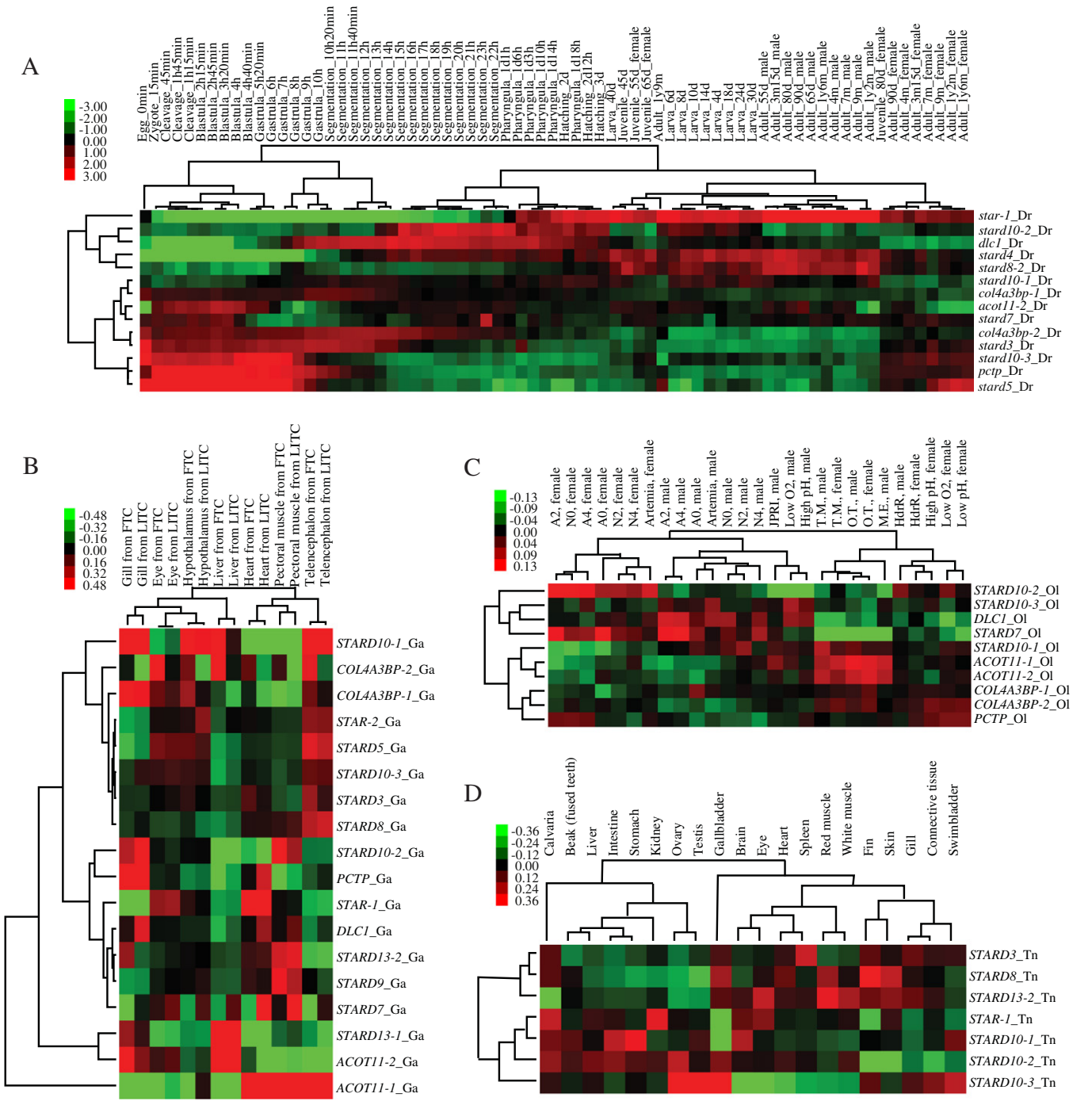


Fig. 4. Relative transcript abundance profiles of the teleost START genes under different conditions. (A) Transcript profiling of the zebrafish START genes (GEO GSE24616) at different developmental stages. (B) Transcriptional expression pattern of three-spined stickleback (GEO GSE12976) START genes in different tissues in marine (LITC) and freshwater (FTC) populations. (C) Transcript abundance pattern of START genes (GEO GSE15380) in different medaka strains (HdrR; JPRI) under different test conditions (A0, aeration and a static water supply; A2, aeration and two times semistatic; A4, aeration and four times semistatic; N0, nonaeration and static; N2, nonaeration and two times semistatic; N4, nonaeration and four times semistatic) or feeding types (Artemia, *Artemia nauplii*; T.M., tetramine; O.T., otohime; M.E., medakanoesa). (D) Transcriptional expression pattern of Green Spotted Puffer (GEO GSE34783) STARTs in different tissues. The transcript abundance levels for the teleost START genes were clustered using hierarchical clustering based on Euclidean distance with complete linkage in the Cluster program. Each row corresponds to the normalized expression profile of a particular gene and their names are shown. The relative transcript accumulation is represented in a color code with green or red showing the lower or higher levels of transcriptional expression, respectively.

(Tocher, 2003). In addition, the fatty acids of fish lipids are rich in ω3 long chain, highly unsaturated fatty acids that have particularly important roles in animal nutrition. However, the metabolic processes regulating deposition and mobilization of fat in fish species are poorly understood (Mommsen et al., 1999). Lipid transfer reactions and START genes have a major role in lipid metabolism, and its disorder is potentially linked to some cardiovascular diseases in human (Tall et al., 1986). However, START genes and their physiological functions in fish are relatively unknown. Here, we present a comparative genomic study of START paralogs in the teleost lineage, and asymmetric evolution and

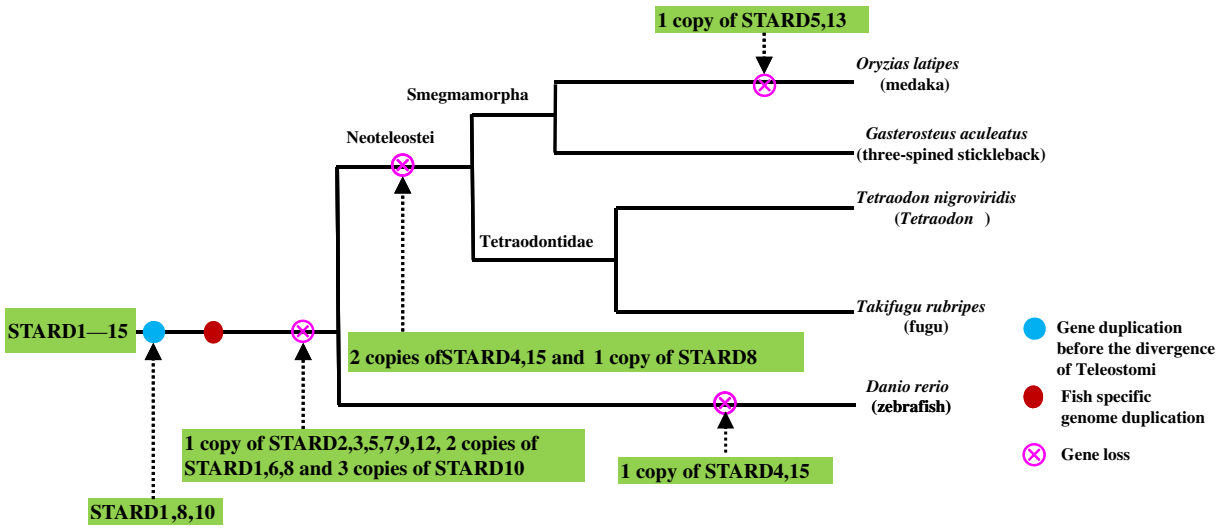


Fig. 5. Hypothetical scenarios of teleost START evolution. The inferred evolutionary events (gene duplication and gene loss) are indicated on the respective branches.

distinct transcriptional expression of these genes in teleosts have been observed. According to our phylogenetic and syntenic analyses, we provide a hypothetical scenario of teleost START evolution (Fig. 5). It is likely that before the divergence of tetrapods and teleosts, 3 (STARD1/8/10) of the 15 ancestral STARTs gave rise to extra duplication followed by mammalian-specific gene losses. After that, extra paralogs of STARTs were generated in the teleost lineage during the FSGD event, followed by the loss of many copies of START genes. After the divergence of zebrafish and Neoteleostei, STARD4/15 and one copy of STARD8 were lost from the Neoteleostei. During successive divergence of the Tetraodontidae and Smegmamorpha and speciation, one copy of STARD5/13 was lost in medaka. This led to the preservation of different ancient STARTs in different teleost fish species.

Eight not 6 START groups were found in this study, because more START genes were used in this study compared to earlier reports (Alpy and Tomasetto, 2005; Soccio and Breslow, 2003). Asymmetry of paralogous gene retention was found between or within each teleost START group. For example, in the STARD1/STAR group, all STARD1 but not all STARD3 were retained in the teleost genome; all teleost STARD6 but partially STARD4 lost in the STARD4 group. Members within the same group might have similar functions. This was confirmed by the observation that (1) the STARD1/STAR group (STARD1 and STARD3) has similar biophysical and functional properties (Tuckey et al., 2004) and mice lacking the STARD3 appear normal and show no defect in steroidogenesis (Kishida et al., 2004); and (2) expression of STARD4 or STARD5 stimulates steroidogenesis by P450scc and liver X receptor reporter gene activity, indicating that both of them function in cholesterol metabolism (Soccio et al., 2005). It is possible that the lost genes were functionally unimportant or redundant to the teleost, or compensation within the same group was available. Intriguingly, the clusters of gene/transcript expression profiles define clearly different developmental stages or environmental conditions, whereas they are not highly related to gene phylogeny. The transcriptional expression differences within members of the STARD1/STAR group were observed in different tissues of Green Spotted Puffer and three-spined stickleback. Earlier studies demonstrated that STARD1 and STARD3 are differentially localized in cells (Alpy et al., 2001) and STARD3 can function in steroidogenesis in organs that do not express STARD1, such as the placenta (Watari et al., 1997). It is possible that tissue-specific expression and subcellular localizations within the same group lead to the observed expression difference during development, because whole fertilized eggs, embryos or larvae were used in the expression study (Domazet-Loso and Tautz, 2010).

It is worthy to mention that 3 teleost STARD10 paralogs were found in our phylogenetic study. Two STARD10 genes that reside in the same chromosome were found in 3 of the 5 teleost fishes, with the exception of incompletely assembled fugu scaffolds or medaka ultracontig. We speculate that tandem duplication occurs in this gene. Because of frequent gene-linkage disruptions, micro-inversions or rearrangements in teleosts, we cannot find further evidence to support the hypothesis of the occurrence of tandem duplication during the evolutionary history of this gene. The expression pattern of duplicated genes can provide useful clues to gene function, and will be of benefit to understanding the driving force and the functional consequence of paralogs (Prince and Pickett, 2002). Zebrafish START10 paralogs have inconsistent transcriptional expression patterns during developmental phases. The transcript abundance of *STARD10-2* and *STARD10-3* peaked in the ovary and testis of Green Spotted Puffer. Diversified STARD10 paralogs might have a role in energy metabolism by mobilizing phosphatidylcholine during development in the testis (Yamanaka et al., 2000). We suspect that the expression difference increases the adaptability of duplicated genes to environmental changes, thus conferring a possible evolutionary advantage.

Asymmetric evolution might be an indicator of neo-functionalization. Some duplicated genes exhibiting asymmetric protein sequence evolution have been reported (Brunet et al., 2006; Conant and Wagner, 2003; Jordan et al., 2004; Lynch and Force, 2000; Nembaware et al., 2002; Van de Peer et al., 2001). This asymmetry has been regarded as a contribution to Ohno's model (Kellis et al., 2004), which proposes that the slow copy maintains an ancestral role and rate of change; while the fast copy evolves to optimize novel functions (Ohno, 1970). Our study revealed fish-specific duplicated extra copies of *col4a3bp*, *stard13* and *acot11* genes in teleosts. When these genes were selected for further evolutionary analysis, we found that an asymmetric molecular evolution rate deviating from the neutral expectation occurs in 7 of 14 duplicated pairs (Table 1). Similar to our analysis, the duplicated teleost HoxA clusters or type III receptor tyrosine kinase (RTK) genes were characterized as evolvement in an asymmetric manner (Braasch et al., 2006; Wagner et al., 2005). These results indicate that asymmetric divergence of fish-specific paralogs might be a common feature, and this feature might contribute to some fish-specific behavior or the diversity of teleost fishes.

In conclusion, asymmetric evolution and divergent transcriptional expression of START genes have occurred in teleost genomes. This detailed analysis of the START gene family in teleost fishes has provided a case in studying how genes evolve after gene duplication, and might

provide some insights into the physiological function divergence of START genes in fishes.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2013.01.058.

Acknowledgments

This work was supported by the grants from the National Natural Science Foundation of China (31200888), the International S&T Cooperation Program of China (2011DFA30670) and the Wenzhou Science and Technology Program (S20100054).

References

Alpy, F., Tomasetto, C., 2005. Give lipids a START: the StAR-related lipid transfer (START) domain in mammals. J. Cell Sci. 118, 2791–2801.
Alpy, F., et al., 2001. The steroidogenic acute regulatory protein homolog MLN64, a late endosomal cholesterol-binding protein. J. Biol. Chem. 276, 4261–4269.
Amores, A., et al., 1998. Zebrafish hox clusters and vertebrate genome evolution. Science 282, 1711–1714.
Aparicio, S., et al., 2002. Whole-genome shotgun assembly and analysis of the genome of fugu rubripes. Science 297, 1301–1310.
Braasch, I., Salzburger, W., Meyer, A., 2006. Asymmetric evolution in two fish-specifically duplicated receptor tyrosine kinase paralogons involved in teleost coloration. Mol. Biol. Evol. 23, 1192–1202.
Brunet, F.G., et al., 2006. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. Mol. Biol. Evol. 23, 1808–1816.
Chan, E.T., et al., 2009. Conservation of core gene expression in vertebrate tissues. J. Biol. 8, 33.
Christoffels, A., Koh, E.G., Chia, J.M., Brenner, S., Aparicio, S., Venkatesh, B., 2004. Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. Mol. Biol. Evol. 21, 1146–1151.
Conant, G.C., Wagner, A., 2003. Asymmetric sequence divergence of duplicate genes. Genome Res. 13, 2052–2058.
Conant, G.C., Wolfe, K.H., 2008. Turning a hobby into a job: how duplicated genes find new functions. Nat. Rev. Genet. 9, 938–950.
de Hoon, M.J., Imoto, S., Nolan, J., Miyano, S., 2004. Open source clustering software. Bioinformatics 20, 1453–1454.
Demuth, J.P., De Bie, T., Stajich, J.E., Cristianini, N., Hahn, M.W., 2006. The evolution of mammalian gene families. PLoS One 1, e85.
Domazet-Loso, T., Tautz, D., 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. Nature 468, 815–818.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Flicek, P., et al., 2011. Ensembl 2011. Nucleic Acids Res. 39, D800–D806.
Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531–1545.
Haldane, J.B.S., 1932. The Causes of Evolution. Harper & Brothers, New York, London.
Hanada, K., Kuromori, T., Myouga, F., Toyoda, T., Shinozaki, K., 2009. Increased expression and protein divergence in duplicate genes is associated with morphological diversification. PLoS Genet. 5, e1000781.
He, X., Zhang, J., 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169, 1157–1164.
Iwahashi, H., Kishi, K., Kitagawa, E., Suzuki, K., Hayashi, Y., 2009. Evaluation of the physiology of Medaka as a model animal for standardized toxicity tests of chemicals by using mRNA expression profiling. Environ. Sci. Technol. 43, 3913–3918.
Iyer, L.M., Koonin, E.V., Aravind, L., 2001. Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. Proteins 43, 134–144.
Jaillon, O., et al., 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 431, 946–957.
Johnson, L.S., Eddy, S.R., Portugaly, E., 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinforma. 11, 431.
Jordan, I.K., Wolf, Y.I., Koonin, E.V., 2004. Duplicated genes evolve slower than singletons despite the initial rate increase. BMC Evol. Biol. 4, 22.
Kasahara, M., et al., 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature 447, 714–719.
Kellis, M., Birren, B.W., Lander, E.S., 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature 428, 617–624.
Kishida, T., et al., 2004. Targeted mutation of the MLN64 START domain causes only modest alterations in cellular sterol metabolism. J. Biol. Chem. 279, 19276–19285.
Lynch, M., Force, A., 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics 154, 459–473.
Maglott, D., Ostell, J., Pruitt, K.D., Tatusova, T., 2010. Entrez gene: gene-centered information at NCBI. Nucleic Acids Res. 39, D52–D57.
Mistry, J., Finn, R., 2007. Pfam: a domain-centric method for analyzing proteins and proteomes. Methods Mol. Biol. 396, 43–58.
Mommsen, T.P., Vijayan, M.M., Moon, T.W., 1999. Cortisol in teleosts: dynamics, mechanisms of action, and metabolic regulation. Rev. Fish Biol. Fish. 9, 211–268.
Nembaware, V., Crum, K., Kelso, J., Seoighe, C., 2002. Impact of the presence of paralogs on sequence divergence in a set of mouse-human orthologs. Genome Res. 12, 1370–1376.
Ohno, S., 1970. Evolution by Gene Duplication. Springer-Verlag, New York.
Ponting, C.P., Aravind, L., 1999. START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. Trends Biochem. Sci. 24, 130–132.
Postlethwait, J.H., 2007. The zebrafish genome in context: ohnologs gone missing. J. Exp. Zool. B Mol. Dev. Evol. 308, 563–577.
Prince, V.E., Pickett, F.B., 2002. Splitting pairs: the diverging fates of duplicated genes. Nat. Rev. Genet. 3, 827–837.
Sato, Y., Hashiguchi, Y., Nishida, M., 2009. Temporal pattern of loss/persistence of duplicate genes involved in signal transduction and metabolic pathways after teleost-specific genome duplication. BMC Evol. Biol. 9, 127.
Schrick, K., Nguyen, D., Karlowski, W.M., Mayer, K.F., 2004. START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. Genome Biol. 5, R41.
Soccio, R.E., Breslow, J.L., 2003. StAR-related lipid transfer (START) proteins: mediators of intracellular lipid metabolism. J. Biol. Chem. 278, 22183–22186.
Soccio, R.E., Adams, R.M., Maxwell, K.N., Breslow, J.L., 2005. Differential gene regulation of StarD4 and StarD5 cholesterol transfer proteins. Activation of StarD4 by sterol regulatory element-binding protein-2 and StarD5 by endoplasmic reticulum stress. J. Biol. Chem. 280, 19410–19418.
Studer, R.A., Penel, S., Duret, L., Robinson-Rechavi, M., 2008. Pervasive positive selection on duplicated and nonduplicated vertebrate protein coding genes. Genome Res. 18, 1393–1402.
Tall, A., Sammett, D., Granot, E., 1986. Mechanisms of enhanced cholesteryl ester transfer from high density lipoproteins to apolipoprotein B-containing lipoproteins during alimentary lipemia. J. Clin. Invest. 77, 1163–1172.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Taylor, J.S., Raes, J., 2004. Duplication and divergence: the evolution of new genes and old ideas. Annu. Rev. Genet. 38, 615–643.
Teng, H., et al., 2010. Evolutionary mode and functional divergence of vertebrate NMDA receptor subunit 2 genes. PLoS One 5, e13342.
Tocher, D.R., 2003. Metabolism and functions of lipids and fatty acids in teleost fish. Rev. Fish. Sci. 11, 107–184.
Tuckey, R.C., Bose, H.S., Czerwionka, I., Miller, W.L., 2004. Molten globule structure and steroidogenic activity of N-218 MLN64 in human placental mitochondria. Endocrinology 145, 1700–1707.
Van de Peer, Y., Taylor, J.S., Braasch, I., Meyer, A., 2001. The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. J. Mol. Evol. 53, 436–446.
Vandepoele, K., De Vos, W., Taylor, J.S., Meyer, A., Van de Peer, Y., 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc. Natl. Acad. Sci. U. S. A. 101, 1638–1643.
Wagner, G.P., et al., 2005. Molecular evolution of duplicated ray finned fish HoxA clusters: increased synonymous substitution rate and asymmetrical co-divergence of coding and non-coding sequences. J. Mol. Evol. 60, 665–676.
Watari, H., et al., 1997. MLN64 contains a domain with homology to the steroidogenic acute regulatory protein (StAR) that stimulates steroidogenesis. Proc. Natl. Acad. Sci. U. S. A. 94, 8462–8467.
Woods, I.G., et al., 2005. The zebrafish gene map defines ancestral vertebrate chromosomes. Genome Res. 15, 1307–1314.
Yamanaka, M., et al., 2000. Molecular cloning and characterization of phosphatidylcholine transfer protein-like protein gene expressed in murine haploid germ cells. Biol. Reprod. 62, 1694–1701.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.