Origin and evolution of GATA2a and GATA2b in teleosts: insights from tongue sole, Cynoglossus semilaevis

Jinxiang Liu, Jiajun Jiang, Zhongkai Wang, Yan He, Quanqi Zhang

Background: Following the two rounds of whole-genome duplication that occurred during deuterostome evolution, a third genome duplication occurred in the lineage of teleost fish and is considered to be responsible for much of the biological diversification within the lineage. GATA2, a member of GATA family of transcription factors, is an important regulator of gene expression in hematopoietic cell in mammals; yet the role of this gene or its putative paralogs in ray-finned fishes remains relatively unknown. Methods: In this study, we attempted to identify GATA2 sequences from the transcriptomes and genomes of multiple teleosts using the bioinformatic tools MrBayes, MEME, and PAML. Following identification, comparative analysis of genome structure, molecular evolution rate, and expression by real-time qPCR were used to predict functional divergence of GATA2 paralogs and their relative transcription in tissues of female and male tongue sole (Cynoglossus semilaevis). Results: Two teleost GATA2 genes were identified in the transcriptomes of tongue sole and Japanese flounder (Paralichthys olivaceus). Synteny and phylogenetic analysis confirmed that the two genes likely originated from the teleost-specific genome duplication. Additionally, selection pressure analysis predicted these gene duplicates to have undergone purifying selection with divergent new functions. This was further supported by differential expression pattern of GATA2a and GATA2b observed in tissues of female and male tongue sole. Discussion: Our results indicate that two GATA2 genes originating from the first teleost-specific genome duplication have remained transcriptionally active in some fish species and have likely undergone neofunctionalization. This knowledge provides novel insights into the evolution of the teleost GATA2 genes and constituted important groundwork for further research on the GATA gene family.

Origin and evolution of GATA2a and GATA2b in teleosts: insights from tongue sole, *Cynoglossus semilaevis*

Jinxiang Liu, Jiajun Jiang, Zhongka Wang, Yan He, Quanqi Zhang

Key Laboratory of Marine Genetics and Breeding, Ministry of Education, Ocean University of China, Qingdao, Shandong, China

Corresponding Author:

Quanqi Zhang

No 5 Yushan Road, Qingdao, Shandong, 266003, China

Email address: qzhang@ouc.edu.cn

ABSTRACT

Background: Following the two rounds of whole-genome duplication that occurred during deuterostome evolution, a third genome duplication occurred in the lineage of teleost fish and is considered to be responsible for much of the biological diversification within the lineage. GATA2, a member of GATA family of transcription factors, is an important regulator of gene expression in hematopoietic cell in mammals; yet the role of this gene or its putative paralogs in ray-finned fishes remains relatively unknown.

Methods: In this study, we attempted to identify GATA2 sequences from the transcriptomes and genomes of multiple teleosts using the bioinformatic tools MrBayes, MEME, and PAML. Following identification, comparative analysis of genome structure, molecular evolution rate, and expression by real-time qPCR were used to predict functional divergence of GATA2 paralogs and their relative transcription in tissues of female and male tongue sole (*Cynoglossus semilaevis*).

Results: Two teleost GATA2 genes were identified in the transcriptomes of tongue sole and Japanese flounder (*Paralichthys olivaceus*). Synteny and phylogenetic analysis confirmed that the two genes likely originated from the teleost-specific genome duplication. Additionally, selection pressure analysis predicted these gene duplicates to have undergone purifying selection and possible divergent new functions. This was supported by differential expression pattern of GATA2a and GATA2b observed in tissues of female and male tongue sole.

Discussion: Our results indicate that two GATA2 genes originating from the first teleost-specific

genome duplication have remained transcriptionally active in some fish species and have likely undergone neofunctionalization. This knowledge provides novel insights into the evolution of the teleost GATA2 genes and constituted important groundwork for further research on the GATA gene family.

INTRODUCTION

GATA transcription factors are evolutionarily conserved proteins that bind the consensus motif WGATAR in gene regulatory regions (Evans *et al*. 1988; Whitelaw *et al*. 1990). GATA proteins are characterized by the conserved N-terminal and C-terminal zinc finger motifs. The N-terminal zinc finger is required for DNA binding, whereas the C-terminal zinc finger stabilizes binding and physical interaction with other co-factors (Yang & Evans 1992). All GATA proteins are essential to animal developmental processes, including germ layer specification, hematopoiesis, and cardiogenesis (Sorrentino *et al*. 2005). All the GATA family members can induce reprogramming and substitute for *Oct4* (Shu *et al*. 2015).

GATA has been identified in vertebrates, invertebrates, fungi, and plants (Lowry & Atchley 2000), as well as protostomes and deuterostomes (Patient & McGhee 2002). The GATA gene family, including GATA123 and GATA456 subfamilies (Gillis *et al*. 2008), has undergone significant expansion after whole-genome duplication in vertebrate lineages. To date, two GATA genes have been identified in the sea urchin *Strongylocentrotus purpuratus*, and two in the hemichordate *Saccoglossus kowalevskii*, the urochordate *Ciona intestinalis*, and the cephalochordate *Branchiostoma floridae*. Meanwhile, six GATA transcription factors (GATA1 to GATA6) have been found in tetrapods and teleosts (Gillis *et al*. 2009).

Previous studies have verified multiple rounds of whole-genome duplication in vertebrate lineages, which may play a significant role in vertebrate evolution (Hoegg & Meyer 2005; Hoffmann *et al*. 2012; Hughes 1999). Interestingly, a third whole-genome duplication event (3R) occurred in teleosts (Amores *et al*. 1998; Postlethwait *et al*. 1998). Teleost-specific genome duplication (TGD) provided more gene copies, contributing to the evolutionary and phenotypic diversification of teleosts. TGD-derived gene duplicates supported the cause–effect relationship between gene copy number and species diversity (Siegel *et al*. 2007). The duplicated genes might possess great divergence from their ancestors, as demonstrated by the changes in evolutionary rates, expression patterns, and regulatory mechanisms observed across the teleost lineage (Braasch *et al*. 2006; Hoegg & Meyer 2007; Mulley *et al*. 2006). Duplicated genes have three main fates, that is, nonfunctionalization, subfunctionalization, and neofunctionalization (Force *et al*. 1999).

In teleosts, research investigating GATA2 has been minimal. To better understand the origination and functional divergence of GATA2 in teleost, this study aimed to investigate GATA2 gene(s) from the transcriptome of tongue sole, Japanese flounder, and other teleosts. Following identification of two GATA2 genes in the tongue sole, chromosomal synteny and phylogenetic analysis of these genes was performed to investigate the origin and evolution of GATA2 in teleosts. Then, analysis of genomic structure, molecular positive selection, and expression pattern of the two GATA2 genes in tongue sole were performed to identify potential changes in functionality for the duplicated GATA2 genes within the teleost lineage. This study provides evidence to support the GATA family expansion theory that the increase of GATA members follows the whole-genome duplication. It also lays the foundation for further evolutionary and functional studies of

the GATA gene family in teleosts.

MATERIALS AND METHODS

Ethics Statement

All research was conducted in accordance with the Institutional Animal Care and Use Committee of the Ocean University of China and with the China Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training (State science and technology commission of the People's Republic of China for No. 2, October 31, 1988. http://www.gov.cn/gongbao/content/2011/content_1860757.htm).

Fish

Healthy tongue sole (three females and males) of one-year-old were chosen from a larger cohort population. The fish were anesthetized (MS-222 at 30μg/mL) and then killed by breaking verterbra. Brain, heart, intestine, kidney, liver, spleen, and gonad tissues were collected in triplicate from each fish. All of the samples were immediately frozen using liquid nitrogen and stored at −80 °C for total RNA extraction.

Identification of GATA gene family sequences in the tongue sole

GATA gene family members were identified from Amazon molly (*Poecilia formosa*), fugu (*Takifugu rubripes*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*), tetraodon (*Tetraodon nigroviridis*), and tilapia (*Oreochromis niloticus*) whose genomes are completely sequenced and available from the Ensembl database. The retrieved sequences were used as query sequences in BLAST searches. The mRNA sequences of GATA genes were identified using tBLASTn analysis from the tongue sole transcriptome previously sequenced by our laboratory.

105 The transcriptome was generated from a total of 749,954 reads using a single 454 sequencing run
106 and assembled into 62,632 contigs, of which 26,589 sequences were successfully annotated (Wang
107 *et al*. 2014b). These fragments were used to search for the corresponding chromosomal regions
108 containing in the tongue sole genome from NCBI (GenBank accession: PRJNA73987).
109 *Cs*GATA2a was found on scaffold385_11, and *Cs*GATA2b was identified on scaffold57_8.

110 **Identification of GATA gene family sequences in the Japanese flounder**

111 The sequences retrieved from tongue sole and the other six teleosts listed above were used as query
112 sequences to search for *Po*GATA genes. The sequences were identified from the Japanese flounder
113 transcriptome through tBLASTn analysis (Wang *et al*. 2014a). An unpublished Japanese flounder
114 genome was used to search for the DNA sequences of *Po*GATA2a and *Po*GATA2b (Supplemental
115 seq file).

116 **GATA2 sequence alignment and phylogenetic analysis**

117 The sequence alignments of GATA2a and GATA2b were based on their predicted peptide
118 sequences using Clustal X with default parameters (Chenna *et al*. 2003). Phylogenetic trees were
119 constructed to confirm the ortholog and paralog relationships of both duplicates. The sequences
120 used to construct gene trees were retrieved from Ensembl and NCBI (species names, gene names,
121 and accession numbers are available in Table S1). The most appropriate substitution model of
122 molecular evolution was determined using JModelTest v2.1.4 (Darriba *et al*. 2012). To confirm
123 the tree topologies, a Bayesian tree and a maximum likelihood tree were respectively constructed
124 using MrBayes v3.2.2 (Huelsenbeck & Ronquist 2001; Ronquist *et al*. 2012) and phyML v3.1
125 (Guindon *et al*. 2010). MrBayes was run for 400,000 generations with two runs and four chains in

parallel and a burn-in of 25%. PhyML was run for 1000 replications. Other parameters were based on the result of JModelTest.

Tests for positive selection in GATA2a and GATA2b

A Bayesian tree was constructed using MrBayes based on GATA2a and GATA2b. The tree includes all species used for positive selection analyses (Table S1). The TIM3+I+G model with base frequencies and substitution rate matrix estimated from the parameters (as suggested by JModelTest) was used. The standard site model in CODEML of PAML v4.7 was used to calculate selection pressures (Yang 2007). The site model employed ML estimation of the ratio of nonsynonymous to synonymous substitutions ($d_N/d_{S=}\omega$) and nested likelihood ratio tests (LRTs) on a phylogeny tree.

Genomic structure, motif, and synteny analysis of teleost GATA2 paralogs

Diagrams of exon–intron structures were obtained using the online Gene Structure Display Server 2.0 (GSDS: http://gsds.cbi.pku.edu.cn) with CDS and genomic sequences (Hu *et al*. 2015). Motifs in the candidate GATA2 DNA sequences were identified using MEME (Bailey *et al*. 2009). The Synteny Database (Catchen *et al*. 2009) was used to generate dotplots of the human GATA2 gene region on chromosome Hsa3 and the genome of zebrafish to analyze the syntenic conservation between fish and human chromosomes.

RNA isolation, cDNA synthesis, and qRT-PCR

Total RNA was extracted from tissue samples with Trizol reagent (Invitrogen, Carlsbad, CA, USA) in accordance with the manufacturer's instructions. DNA was removed using DNase I (TaKaRa, Dalian, China) treated with 2h at 37°C, and the protein was digested using an RNAclean

147 RNA Kit (Biomed, Beijing, China). The quality and quantity of the extracted RNA were identified
148 via electrophoresis and Nanophotometer® Pearl (Implen GmbH, Munich, Germany). Frist-strand
149 cDNA was synthesized using the PrimeScript™ RT-PCR Kit (TaKaRa) in accordance with the
150 manufacturer's instructions.
151 Quantitative Real-time was conducted on a LightCycler 480 (Roche, Forrentrasse, Switzerland).
152 The respective primer pairs for GATA2a and GATA2b were Cs-GATA2a-RT and Cs-GATA2b-
153 RT (Table S2), which were designed by IDT (http://www.idtdna.com/Primerquest/Home/Index)
154 in the 3′ UTR of both genes. Standard curves were established from a serial dilution of plasmids
155 containing GATA2a, GATA2b, and reference gene RPL17 fragments. Efficiency values (91.58%,
156 88.25%, and 92.10%, respectively) were calculated by standard curves (Boyle *et al*, 2009). cDNAs
157 from three females and males were diluted as templates (10 ng/μL) for sample assessment. The
158 SYBR Green master mix (Roche, Switzerland) was used as the PCR detection system. Three
159 biological replicates were used for each tissue, including no-template controls and each replicate
160 consisted of a sample pool of three tissues. Thermocycling consisted of an initial polymerase
161 activation of 30 s at 94 °C, followed by 40 cycles at 94 °C for 15 s and 60 °C for 45 s. Product
162 specificity was ensured through melting curve analysis which consisted of 40 cycles. RPL17 of
163 tongue sole was used as the reference gene to normalize the expression which has been shown to
164 be stably expressed between male and female tongue sole in multiple tissue types (Liu *et al*. 2014).
165 The sizes of GATA2a, GATA2b, and RPL17 amplicons were 121bp, 122bp, and 114bp, and melt
166 curve starting temperatures were 60°C, 61°C, and 60°C, respectively. Data were analyzed through
167 the $2^{-\Delta\Delta Ct}$ method.

168 **Statistical analysis**

169 qRT-PCR data were statistically analyzed using one-way ANOVA on log10-transformed data
170 followed by LSD test using SPSS 20.0, and $P < 0.05$ was considered to indicate statistical
171 significance. All data were expressed as mean ± standard error of the mean (SEM).

172 **RESULTS**

173 **Identification of GATA genes**

174 We identified GATA1, GATA2a, GATA2b, GATA3, GATA4, GATA5, and GATA6 from the
175 transcriptomes of tongue sole and Japanese flounder via tBLAST to infer the origin and
176 evolutionary history of the GATA gene family in teleosts. Other GATA genes were searched from
177 Ensembl and NCBI. The GATA family can be divided into the GATA123 and GATA456
178 subfamilies. Protein analysis showed that the GATA family in teleosts comprised two conserved
179 zinc finger motifs at the N-terminal and the C-terminal domains. However, the different GATA
180 paralogs in teleosts had varied lengths. Seven GATA genes, including two GATA2 genes
181 (GATA2a and GATA2b), were detected in the teleost GATA family. Only six GATA genes were
182 detected in tetrapods.

183 **Phylogenetic relationships and evolution of the GATA gene family**

184 The identified DNA sequences were analyzed to investigate the evolutionary relationship of
185 GATA genes among various teleosts using multiple sequence alignment with Clustal X. A
186 phylogenetic tree of the GATA gene family was constructed using MrBayes and phyML based on
187 the alignment results. The two programs inferred similar topologies, which indicated that the

188 GATA gene family could be divided into seven well-conserved clades and two subfamilies in
189 teleosts (Figure 1).
190 Our results also indicated distinct ancestral relationship within each subfamily of the GATA gene
191 family. A close relationship was observed between GATA2a/b and GATA3 within the GATA123
192 subfamily, and between GATA5 and GATA6 within the GATA456 subfamily.
193 **Phylogenetic analysis of teleost-specific GATA2a and GATA2b**
194 Multiple amino acid alignment was conducted to explore the origin, generation, and differentiation
195 of GATA2a and GATA2b in teleosts. The sequence similarity between GATA2a and GATA2b
196 was 82.55%, with two highly conserved zinc finger motifs. Sequence alignment suggested the
197 occurrence of two GATA2b-specific mutations in the N-terminal and C-terminal zinc fingers.
198 Specifically, serine was dehydroxymethyled into glycine in the N-terminal zinc finger motif, and
199 alanine was demethylated into glycine in the C-terminal zinc finger motif (Figure 2 and Figure
200 S1). The N-terminal zinc finger motif can stabilize the binding and physically interact with other
201 co-factors, and the C-terminal zinc finger motif is required for DNA binding. Thus the
202 dehydroxymethylation and demethylation mutations might trigger protein structure alteration, and
203 further affect the molecular functions in biological processes.
204 The phylogenetic trees of GATA2a and GATA2b in teleosts were constructed using MrBayes and
205 phyML, with the human GATA2 sequence as an outgroup. The two trees were similar in topology
206 with minimal bootstrap differences. Results indicated that teleost GATA2 genes could be divided
207 into two well-conserved clusters: GATA2a and GATA2b (Figure 3), implying that GATA2a and
208 GATA2b in teleosts were probably generated from the same ancestor.

209 **Genomic structures of teleost GATA2**

210 Gene structure graphics were constructed by the online program Gene Structure Display Server to
211 analyze the evolutionary mechanism of GATA2. The graphics showed that both GATA2a and
212 GATA2b had five exons in CDS, except for *Pf*GATA2b, which had an extra intron dividing the
213 second exon into two segments. The lengths of each corresponding exon were highly conserved,
214 but intron lengths varied among species. The GATA2a and GATA2b in fugu and tetraodon had
215 the shortest intron lengths. In most teleosts, the GATA2b gene was longer than the GATA2a gene,
216 which might infer that the two subtypes of GATA2 had undergone gene differentiation, that is,
217 they originated from a common ancestor but diverged into two genes differing in protein structure
218 and functions (Figure S2A). This inference was further supported by motif prediction on teleost
219 GATA2 by MEME. Four main motifs (motifs 1, 2, 3, and 4) were predicted in both GATA2a and
220 GATA2b. An additional motif 4 was predicted at the end of GATA2b in most teleosts (Figure
221 S2B).

222 **Synteny analysis of teleost GATA2 paralogs**

223 Chromosomal synteny analysis was carried out between human and zebrafish to test whether that
224 GATA2 paralogs originated from whole-genome duplication. Conserved synteny dotplots showed
225 that the GATA2a and GATA2b regions in zebrafish shared conserved synteny. The zebrafish
226 GATA2a region on Dre11 shared conserved synteny neither with the zebrafish GATA2b region
227 Dre6 nor with the human GATA2 region Hsa3 (Figure 4). Previous studies confirmed that these
228 chromosomes originated from the common ancestral chromosome and duplicated during the
229 teleost-specific genome duplication (Kasahara *et al*. 2007; Nakatani *et al*. 2007).

Gene neighborhood analysis showed highly conserved synteny within GATA2a or GATA2b and between the two genes. In teleosts, the genes near GATA2a, except for some genes lost in tetraodon duplication era, were mostly conserved and shared the same direction (Figure 5A). Long fragments consisting of several genes were lost in the upstream and downstream regions of GATA2b in Amazon molly and fugu, but the other genes remained conserved. Comparison of the upstream genes of GATA2a and GATA2b revealed that a fragment including four genes was conserved, albeit in opposite directions (indicated by blank pentagons) (Figure 5B). A gene in the upstream region and a two-gene string in the downstream region (indicated by blank pentagons) were also highly conserved between GATA2a and GATA2b. These results implied that the genes neighboring teleost GATA2a or GATA2b were highly conserved, and more conserved among teleosts after duplication.

Molecular evolution of teleost GATA2a and GATA2b

In general, phenotypic differences can arise from mutations affecting protein functions or changes in gene regulation (Stainier *et al*. 1996). Therefore, we examined the coding sequence evolution in two GATA2 paralogs to test for positive selection and potential functional changes in teleosts. The site models in PAML were used to assess different selective pressures. The estimation of positive selection based on the phylogenetic trees is shown in Figure S3. Three model pairs (M0/M3, M1a/M2a, and M7/M8) were selected and compared with the site-specific codeml model to test whether variable ω ratios occurred at amino acid sites. The parameters and the LRT results are listed in Table 1.

In GATA2, M3 (discrete) was significantly better than M0 (one-ratio) ($P < 0.05$). Thus, M0 was

rejected, indicating the extreme variation in selection pressure among amino acid sites. Overall, the GATA2 sequences had undergone positive selection. Additional tests with M1a (neutral) and M2a (selection), M7 (beta)/M8 (beta & ω) and M8a were conducted using the chi2 program in PAML. The LRT significantly differed in the M7/M8 pair of GATA2b ($P < 0.05$). One candidate amino acid site for positive selection (356P, $P < 0.05$) was identified (356P*) through the Bayes Empirical Bayes (BEB) method of M8. No site under positive selection was identified in GATA2a. Then, the relationship between amino acid sites under positive selection and function divergence was analyzed. The site 356P with a posterior probability > 0.95 was located in the C-terminal zinc finger in GATA2b, indicating that GATA2b, especially its motif, had experienced a strong selective pressure, which might develop mechanism adapting to water environment.

Expression levels of GATA2a and GATA2b in tissue

Quantitative real-time PCR using RNA extracted from multiple tongue sole tissues was performed to test if transcription regulation of GATA2a and GATA2b had undergone divergence in teleosts. Both genes were expressed in all tissues tested but possessed distinct levels of expression. Heart and the brain showed higher relative expression of GATA2a/b than other somatic tissues in both sexes, and extraordinarily high GATA2b expression was found in the heart (Figure S4). A sexual dimorphic expression pattern was observed in the gonads. In the ovary, GATA2a expression was hardly observed and GATA2b expression was very low (Figure 6A), while in the testis, GATA2a expression was moderate and GATA2b expression was relatively high (Figure 6B).

DISCUSSION

Expansion of vertebrate GATA transcription factor genes during multiple whole-genome duplications

GATA transcription factors play crucial roles in regulating the development and differentiation processes including hematopoiesis, cardiogenesis, and germ layer specification (Holtzinger & Evans 2005; LaVoie 2003). In the present study, seven GATA genes were identified from both tongue sole and Japanese flounder transcriptomes.Indeed, all teleosts analyzed in this study possessed seven GATA genes, including six GATA genes shared with tetrapods and an additional teleost-specific GATA2 duplication. As teleosts have undergone a unique 3R genome duplication, some gene families became larger in teleosts than in tetrapods or chondrichthyes. Thus, the present results are consistent with previous reports that GATA gene family expansion occurred through genome duplication and that clade-specific conserved losses of duplicated paralogs occurred after duplication (Gillis *et al*. 2009).

Phylogenetic analysis suggested that the GATA gene family had undergone distinct expansion that separated the GATA123 and GATA456 subfamilies, both of which were subsequently expanded. Our results differed from findings on the evolution of the GATA gene family in protostomes but agreed with those on the evolution of vertebrates. In protostomes, only the GATA456 subfamily appeared to have undergone expansion (Gillis *et al*. 2008). By contrast, the GATA123 and GATA456 subfamilies both expanded in deuterostomes through the retention of duplicated GATA genes during multiple whole-genome duplications (Dehal & Boore 2005). Our molecular phylogenetic analysis, together with the conserved syntenic paralogs (Gillis *et al*. 2009), provided evidence to support the expansion through genome duplication.

Origin of GATA2 paralogs

Several molecular mechanisms, such as gene duplication, exon shuffling, gene fission and fusion, retrotranspositon, and mobile elements, have been proposed to understand the origin of new genes (Long *et al*. 2003). Gene duplication events, including single-gene duplication, segmental duplication, and genome duplication, are crucial to produce new genes (Bailey *et al*. 2002; Samonte & Eichler 2002). In the present study, the results of chromosomal synteny analysis and gene-neighborhood synteny analysis indicated that the two GATA2 paralogs were generated through genome duplication in teleosts.

The fate of newborn genes is diverse. Some scholars believed that a number of a duplicate gene pairs eventually become nonfunctional and that most duplicates eventually perish as pseudogenes (Bailey *et al*. 1978). Gene duplicates possibly acquire new functions (neofunctionalization) or undergo subfunctionalization and are preserved in a lineage (Force *et al*. 1999; Kimura & King 1979; Li 1980). During whole-genome duplication of yeast, arabidopsis, rice and tetraodon, all of the genes were duplicated, but only 10%–30% of new genes were preserved, and others were lost in evolution (Byrne & Wolfe 2005; Paterson *et al*. 2006). In the present study, ohnolog-gone-missing (ogm) was observed throughout the evolution of the GATA gene. Based on phylogenetic analysis, we conjectured that GATA1-ogm and GATA4-ogm occurred after 2R, which is consistent with a former study (Gillis *et al*. 2009). Most GATA paralogs, except GATA2, were lost after 3R in teleosts. GATA2 may have been preserved in the evolutionary process because of environmental pressure and further supports that the two GATA2 paralogs originated from 3R duplication.

Structures of the GATA2a and GATA2b genes

The structure of GATA genes is generally conserved, as shown in protostomes and deuterostomes relative to vertebrate transcriptomes (Gillis *et al*. 2008; Gillis *et al*. 2009). In the present study, we examined the conservation of the exon/intron structures of GATA2a and GATA2b in teleosts. The genomic structure of GATA2 was conserved; all GATA2 genes, except for *Pf*GATA2b, contained five exons in CDS. The lengths of the five exons were conserved, but the lengths of the introns varied. Introns are important indicators in eukaryotic evolution, where the gain and loss of introns reflect positive correlation or negative correlation with the coding-sequence evolution rate (Carmel *et al*. 2007; Slamovits & Keeling 2009). The intron lengths of GATA2a were generally shorter than those of GATA2b, suggesting that the two GATA2 genes had diverged. Meanwhile, motif prediction showed an additional motif 4 in GATA2b. This motif might separate GATA2b from GATA2a functionally, which was consistent with the phylogenetic results. These results implied that GATA2a and GATA2b in teleosts separated from each other and generated different structures and functions. We inferred that the sequence of GATA2a and GATA2b had been changed under selection pressure.

Potential for functional divergence of GATA2a and GATA2b

In general, new genes evolve with rapid changes in their sequence and structure (Wang *et al*. 2002; Zhang *et al*. 2002), and mutation is the initial condition in evolution. Positive Darwinian selection may be another important force driving the evolution of new genes (Ohta 1994; Walsh 1995). The evolutionary rates of gene pairs that originated from duplication are usually different, and the rapid evolution of one of the gene pairs is a general phenomenon (Johnson *et al*. 2001; Wang *et al*.

2002). In the present study, the teleost GATA2 phylogenetic tree provided evidence that the evolutionary rate of GATA2b was faster than that of GATA2a under current environmental pressure. Thus, GATA2b has likely diverged from an ancestral GATA2 more similar to present GATA2a paralog.

The amino acid sequences of GATA genes contain the well-conserved N-terminal and C-terminal zinc finger motifs, which significantly contribute to structure and function. In the present study, the two zinc finger motifs were highly conserved in GATA2a and GATA2b in teleosts. Two mutated amino acid sites were found located in the two zinc finger motifs in GATA2b relative to GATA2a. Measuring the rate of relaxation and determining the presence of amino acid residue under positive selection are crucial to determine whether positive selection has driven the evolution of the GATA2 paralogs and whether or not selection constraints affect GATA2 genes after duplication in teleosts. The results of selection pressure analysis provided evidence of purifying selection, and one site (356P) in GATA2b was predicted to have undergone a strong positive selection. Interestingly, this site was located in the C-terminal zinc finger motif, which has been inferred to play an important role during evolution. Therefore, this positively selected site might affect the binding activity of GATA2b or even affect the selection of binding sites, resulting in the functional divergence between GATA2a and GATA2b.

In the current study, transcriptional analysis was performed using qRT-PCR. The expression patterns of GATA2a and GATA2b were similar in most somatic tissues, but sexual dimorphic expression was apparent, especially in the spleen and the gonad. Previous studies have shown that new genes have evolved in conjunction with rapid changes in expression (Wang *et al*. 2002; Zhang

355 *et al*. 2002), and the differential expression of these genes was believed to be the first step in
356 functional divergence. The classical model for the evolution of duplicate genes identifies two
357 possibilities: one is that one of the duplicated genes u degenerates by accumulating deleterious
358 mutations; the other is that one duplicate acquires a new adaptive function (Ohno 1970). However,
359 the duplication–degeneration–complementation (DDC) model predicts that the duplicate gene
360 preservation involves the partitioning of ancestral functions rather than the evolution of new
361 functions (Force *et al*. 1999). Moreover, the expression levels of GATA2b in the brain, the
362 pituitary gland, and the gonad differed between females and males in tilapia (Zhang 2009). Based
363 on the results of our present study, we hypothesize that the differential transcription of the GATA2
364 paralogs in tongue sole follow the DDC model; that is, GATA2a and GATA2b partitioned the
365 ancestral functions of GATA2 in teleosts. GATA2a might have maintained the functions of
366 GATA2 in hemopoiesis and in the multiplication and differentiation of hematopoietic stem cells,
367 whereas GATA2b might have acquired some functions related to sexual differentiation and gonad
368 development or sexual maturation. These results provide preliminary evidence that the duplicated
369 GATA2 genes may have undergone neofunctionalization in teleosts.

370 **CONCLUSIONS**

371 In summary, we investigate the origin of teleost GATA2a/b genes and reports for the first time
372 that two GATA2 genes are present in teleosts as a result of TGD. In addition, our results indicate
373 possible neofunctionalization of the duplicated GATA2 genes, providing novel insight into the
374 teleost GATA gene family and future functional studies of GATA2 in fish.

375 **REFERENCES**

376 Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang
377 YL *et al*. . 1998. Zebrafish hox clusters and vertebrate genome evolution. *Science*
378 282:1711-1714.

379 Bailey GS, Poulter RT, and Stockwell PA. 1978. Gene duplication in tetraploid fish: model for
380 gene silencing at unlinked duplicated loci. *Proceedings of the National Academy of*
381 *Sciences of the United States of America* 75:5575-5579.

382 Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW,
383 and Eichler EE. 2002. Recent segmental duplications in the human genome. *Science*
384 297:1003-1007.

385 Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, and Noble WS.
386 2009. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research*
387 37:W202-W208.

388 Boyle B, Dallaire N, and MacKay J. 2009. Evaluation of the impact of single nucleotide
389 polymorphisms and primer mismatches on quantitative PCR. *BMC Biotechnology* 9:75-75.

390 Braasch I, Salzburger W, and Meyer A. 2006. Asymmetric evolution in two fish-specifically
391 duplicated receptor tyrosine kinase paralogons involved in teleost coloration. *Molecular*
392 *Biology and Evolution* 23:1192-1202.

393 Byrne KP, and Wolfe KH. 2005. The Yeast Gene Order Browser: Combining curated homology
394 and syntenic context reveals gene fate in polyploid species. *Genome Research* 15:1456-
395 1461.

396 Carmel L, Rogozin IB, Wolf YI, and Koonin EV. 2007. Evolutionarily conserved genes
397 preferentially accumulate introns. *Genome Research* 17:1045-1050.
398 Catchen JM, Conery JS, and Postlethwait JH. 2009. Automated identification of conserved synteny
399 after whole-genome duplication. *Genome Research* 19:1497-1505.
400 Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, and Thompson JD. 2003.
401 Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*
402 31:3497-3500.
403 Darriba D, Taboada GL, Doallo R, and Posada D. 2012. jModelTest 2: more models, new
404 heuristics and parallel computing. *Nat Meth* 9:772-772.
405 Dehal P, and Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate.
406 *PLoS Biology* 3:e314.
407 Evans T, Reitman M, and Felsenfeld G. 1988. An erythrocyte-specific DNA-binding factor
408 recognizes a regulatory sequence common to all chicken globin genes. *Proceedings of the*
409 *National Academy of Sciences of the United States of America* 85:5976-5980.
410 Force A, Lynch M, Pickett FB, Amores A, Yan YL, and Postlethwait J. 1999. Preservation of
411 duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531-1545.
412 Gillis WQ, Bowerman BF, and Schneider SQ. 2008. The evolution of protostome GATA factors:
413 molecular phylogenetics, synteny, and intron/exon structure reveal orthologous
414 relationships. *BMC Evolutionary Biology* 8:1-15.
415 Gillis WQ, St John J, Bowerman B, and Schneider SQ. 2009. Whole genome duplications and
416 expansion of the vertebrate GATA transcription factor gene family. *BMC Evolutionary*

Biology 9:207-207.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, and Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* 59:307-321.

Hoegg S, and Meyer A. 2005. Hox clusters as models for vertebrate genome evolution. *Trends in Genetics* 21:421-424.

Hoegg S, and Meyer A. 2007. Phylogenomic analyses of KCNA gene clusters in vertebrates: why do gene clusters stay intact? *BMC Evolutionary Biology* 7:139-139.

Hoffmann FG, Opazo JC, and Storz JF. 2012. Whole-genome duplications spurred the functional diversification of the globin gene superfamily in vertebrates. *Molecular Biology and Evolution* 29:303-312.

Holtzinger A, and Evans T. 2005. Gata4 regulates the formation of multiple organs. *Development* 132:4005-4014.

Hu B, Jin J, Guo AY, Zhang H, Luo J, and Gao G. 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31:1296-1297.

Huelsenbeck JP, and Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754-755.

Hughes AL. 1999. Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. *Journal of Molecular Evolution* 48:565-576.

Johnson ME, Viggiano L, Bailey JA, Abdul-Rauf M, Goodwin G, Rocchi M, and Eichler EE.

2001. Positive selection of a gene family during the emergence of humans and African apes. *Nature* 413:514-519.

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y *et al*. . 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447:714-719.

Kimura M, and King JL. 1979. Fixation of a deleterious allele at one of two "duplicate" loci by mutation pressure and random drift. *Proceedings of the National Academy of Sciences of the United States of America* 76:2858-2861.

LaVoie HA. 2003. The role of GATA in mammalian reproduction. *Experimental Biology and Medicine* 228:1282-1290.

Li WH. 1980. Rate of gene silencing at duplicate loci: A theoretical study and interpretation of data from tetraploid fishes. *Genetics* 95:237-258.

Liu C, Xin N, Zhai Y, Jiang L, Zhai J, Zhang Q, and Qi J. 2014. Reference gene selection for quantitative real-time RT-PCR normalization in the half-smooth tongue sole (*Cynoglossus semilaevis*) at different developmental stages, in various tissue types and on exposure to chemicals. *PLoS ONE* 9:e91715.

Long M, Betran E, Thornton K, and Wang W. 2003. The origin of new genes: glimpses from the young and old. *Nat Rev Genet* 4:865-875.

Lowry JA, and Atchley WR. 2000. Molecular evolution of the GATA family of transcription factors: Conservation within the DNA-Binding domain. *Journal of Molecular Evolution* 50:103-115.

459 Mulley JF, Chiu CH, and Holland PWH. 2006. Breakup of a homeobox cluster after genome
460 duplication in teleosts. *Proceedings of the National Academy of Sciences of the United*
461 *States of America* 103:10369-10372.
462 Nakatani Y, Takeda H, Kohara Y, and Morishita S. 2007. Reconstruction of the vertebrate
463 ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome*
464 *Research* 17:1254-1265.
465 Ohno, S., 1970. Evolution by gene duplication. Springer-Verlag, New York.
466 Ohta T. 1994. Further examples of evolution by gene duplication revealed through DNA sequence
467 comparisons. *Genetics* 138:1331-1337.
468 Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, and Estill JC. 2006. Many gene
469 and domain families have convergent fates following independent whole-genome
470 duplication events in *Arabidopsis, Oryza, Saccharomyces* and *Tetraodon*. *Trends in*
471 *Genetics* 22:597-602.
472 Patient RK, and McGhee JD. 2002. The GATA family (vertebrates and invertebrates). *Current*
473 *Opinion in Genetics & Development* 12:416-422.
474 Postlethwait JH, Yan YL, Gates MA, Horne S, Amores A, Brownlie A, Donovan A, Donovan A,
475 Egan ES, Force A, Gong Z, Goutel C *et al.* . 1998. Vertebrate genome evolution and the
476 zebrafish gene map. *Nature Genetics* 19:303-303.
477 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard
478 MA, and Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference
479 and model choice across a large model space. *Systematic Biology* 61:539-542.

480 Samonte RV, and Eichler EE. 2002. Segmental duplications and the evolution of the primate
481 genome. *Nat Rev Genet* 3:65-72.
482 Shu J, Zhang K, Zhang M, Yao A, Shao S, Du F, Yang C, Chen W, Wu C, Yang W *et al*. . 2015.
483 GATA family members as inducers for cellular reprogramming to pluripotency. *Cell Res*
484 25:169-180.
485 Siegel N, Hoegg S, Salzburger W, Braasch I, and Meyer A. 2007. Comparative genomics of
486 ParaHox clusters of teleost fishes: gene cluster breakup and the retention of gene sets
487 following whole genome duplications. *BMC Genomics* 8:312-312.
488 Slamovits CH, and Keeling PJ. 2009. Evolution of ultrasmall spliceosomal introns in highly
489 reduced nuclear genomes. *Molecular Biology and Evolution* 26:1699-1705.
490 Sorrentino RP, Gajewski KM, and Schulz RA. 2005. GATA factors in *Drosophila* heart and blood
491 cell development. *Seminars in Cell & Developmental Biology* 16:107-116.
492 Stainier DY, Fouquet B, Chen JN, Warren KS, Weinstein BM, Meiler SE, Mohideen MA,
493 Neuhauss SC, Solnica-Krezel L, Schier AF *et al*. . 1996. Mutations affecting the formation
494 and function of the cardiovascular system in the zebrafish embryo. *Development* 123:285-
495 292.
496 Walsh JB. 1995. How often do duplicated genes evolve new functions? *Genetics* 139:421-428.
497 Wang W, Brunet FG, Nevo E, and Long M. 2002. Origin of sphinx, a young chimeric RNA gene
498 in *Drosophila* melanogaster. *Proceedings of the National Academy of Sciences of the*
499 *United States of America* 99:4448-4453.
500 Wang W, Wang J, You F, Ma L, Yang X, Gao J, He Y, Qi J, Yu H, Wang Z *et al*. . 2014a. Detection

501 of alternative splice and gene duplication by RNA sequencing in Japanese flounder,
502 *Paralichthys olivaceus*. *G3: Genes|Genomes|Genetics* 4:2419-2424.
503 Wang W, Yi Q, Ma L, Zhou X, Zhao H, Wang X, Qi J, Yu H, Wang Z, and Zhang Q. 2014b.
504 Sequencing and characterization of the transcriptome of half-smooth tongue sole
505 (*Cynoglossus semilaevis*). *BMC Genomics* 15:470.
506 Whitelaw E, Tsai SF, Hogben P, and Orkin SH. 1990. Regulated expression of globin chains and
507 the erythroid transcription factor GATA-1 during erythropoiesis in the developing mouse.
508 *Molecular and Cellular Biology* 10:6596-6606.
509 Yang HY, and Evans T. 1992. Distinct roles for the two cGATA-1 finger domains. *Molecular and*
510 *Cellular Biology* 12:4562-4570.
511 Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and*
512 *Evolution* 24:1586-1591.
513 Zhang J, Zhang YP, and Rosenberg HF. 2002. Adaptive evolution of a duplicated pancreatic
514 ribonuclease gene in a leaf-eating monkey. *Nat Genet* 30:411-415.
515 Zhang W. 2009. Molecular cloning, gene expression of GATA-2, p450(11β) and their possible
516 roles in gonadal differentiation and gametogenesis in tilapia. M. Phil. Thesis, Southwest
517 University.

518

519 **FIGURE LEGENDS**

520 **Figure 1** Phylogenetic analyses of vertebrate GATA gene family. Phylogenetic tree constructed
521 using MrBayes with the TPM1uf+I+G model; MCMC = 400,000 generations. Values at the tree

522 nodes represent posterior probabilities. *On-Oreochromis niloticus, Cs-Cynoglossus semilaevis,*
523 *Tn-Tetraodon nigroviridis, Mm-Mus musculus, Gg-Gallus gallus; Hs-Homo sapiens; Nv-*
524 *Nematostella vectensis.*

525

526 **Figure 2** Partial multiple sequence alignment of the deduced GATA2a and GATA2b protein
527 sequences. In teleosts, two conserved zinc finger motifs were found in GATA2a and GATA2b
528 (underlined sequences). Two GATA2b-specific mutations were identified in zinc finger motifs
529 (star-shaped site). The proline site (arrowhead) in the N-terminal zinc finger motif had undergone
530 positive selection.

531

532 **Figure 3** Phylogenetic analysis of teleost GATA2. (A) Phylogenetic tree constructed based on
533 GATA2a and GATA2b in teleosts by using MrBayes with the TIM3+I+G model; MCMC =
534 400,000. (B) Maximum likelihood phylogenetic tree constructed using phyML with the
535 TIM3+I+G model. PhyML was run for 1000 replications. Numbers at the nodes are bootstrap
536 support values with a percentage based on 1000 replicates. *Tn-Tetraodon nigroviridis, On-*
537 *Oreochromis niloticus, Po-P. olivaceus, Tr-T. rubripes, Pf-P. formosa, Ga-G. aculeatus. Ol-*
538 *Oryzias latipes, Cs-Cynoglossus semilaevis.*

539

540 **Figure 4** Chromosome synteny analysis of teleost GATA2 paralogs. Dotplots of the human
541 GATA2 gene region on human chr3 show double conserved synteny to the two GATA2 paralogs
542 in zebrafish on chromosomes Dre6 (GATA2b) and Dre11 (GATA2a).

543

544 **Figure 5** Chromosomal segments showing the conserved syntenic blocks containing GATA2a and
545 GATA2b in teleosts. The genes are represented by colored pentagons, and the gene names are
546 indicated on top. Color pentagons indicate the same gene in different species and its respective
547 genomic position in relation to several other genes. The pentagon's direction indicates the gene
548 direction compared with the reference gene. The empty spaces indicate a region with other genes
549 or the absence of the gene in the genome. The blank pentagons indicate conserved genes between
550 GATA2a and GATA2b.

551

552 **Figure 6** Expression of GATA2a (A) and GATA2b (B) in tongue sole tissues relative of RPL17.
553 Data are shown as mean ± SEM (n = 3). Values with different asterisks indicate statistical
554 significance ($P < 0.05$). B: brain; H: heart; I: intestine; K: kidney; L: liver; S: spleen; G: gonad.

555 **SUPPLEMENTAL INFORMATION**

556 **Supplemental Figure Legends**

557 **Figure S1** Multiple sequence alignment of the deduced GATA2a and GATA2b protein sequences.

558

559 **Figure S2** Phylogenetic relationships, exon–intron structure, and motif structures of GATA2
560 genes. (A) ML phylogenetic tree and exon–intron structures of the GATA2 genes. Box: exon;
561 lines: introns. The lengths of boxes and lines are scaled based on gene length. (B) MEME motif
562 search results. Conserved motifs are indicated in numbered color boxes.

563

564 **Figure S3** Phylogenetic tree of teleost GATA2 genes used in PAML analysis. (A) Phylogenetic
565 tree constructed based on GATA2a sequences by using MrBayes with the TPM2uf+G model to
566 assess selection pressure; MCMC = 200,000. (B) Phylogeny for site model constructed based on
567 GATA2b sequences by using MrBayes with the TIM2+I model (MCMC = 200,000).

568

569 **Figure S4** Relative expression of GATA2a (A) and GATA2b (B) in female and male tissues. Data
570 are shown as mean ± SEM (n = 3). Values with different superscripts indicate statistical significance
571 ($P < 0.05$).

572 **Supplemental Tables**

573 **Table S1** Database ID of the sequences used in this study.

574 **Table S2** Primers used for qRT-PCR.

575 **Supplemental Seq file**

576 Supplemental seq file.txt

Figure 1(on next page)

Phylogenetic analyses of vertebrate GATA gene family.

Phylogenetic tree constructed using MrBayes with the TPM1uf+I+G model; MCMC = 400,000 generations. *On-Oreochromis niloticus, Cs-Cynoglossus semilaevis, Tn-Tetraodon nigroviridis, Mm-Mus musculus, Gg-Gallus gallus; Hs-Homo sapiens; Nv-Nematostella vectensis.*

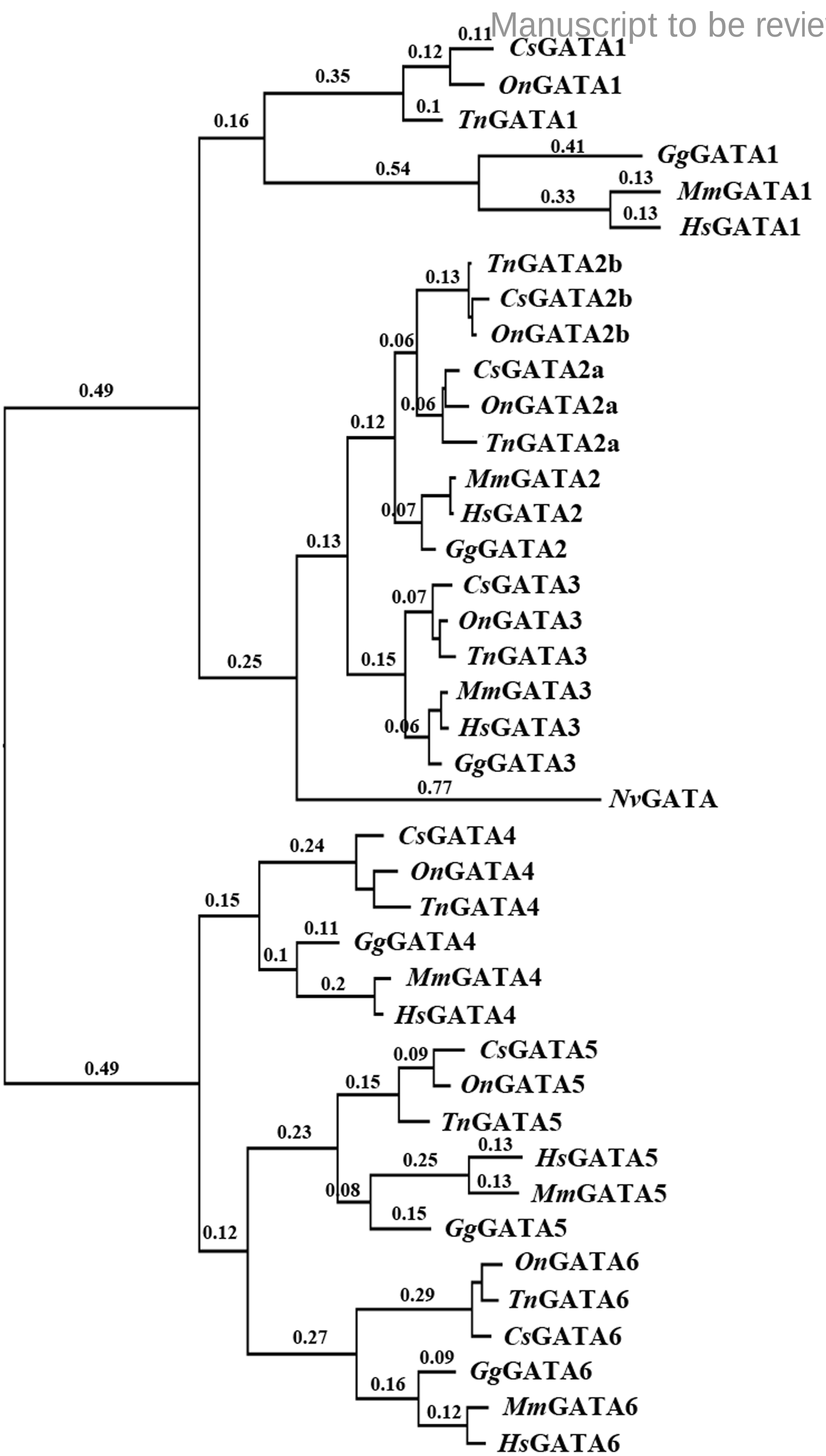


Figure 2(on next page)

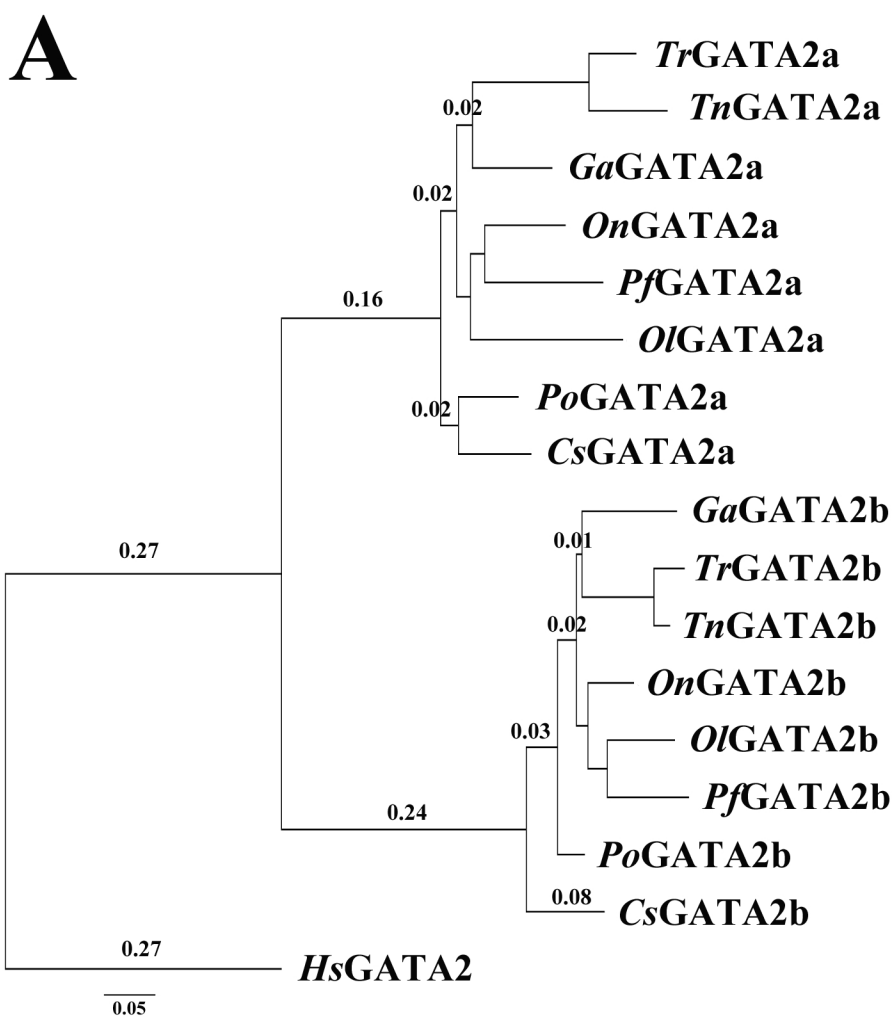
Partial multiple sequence alignment of the deduced GATA2a and GATA2b protein sequences.

In teleosts, two conserved zinc finger motifs were found in GATA2a and GATA2b (underlined sequences). Two GATA2b-specific mutations were identified in zinc finger motifs (star-shaped site). The proline site (arrowhead) in the N-terminal zinc finger motif had undergone positive selection.

Figure 3(on next page)

Phylogenetic analysis of teleost GATA2 .

(A) Phylogenetic tree constructed based on GATA2a and GATA2b in teleosts by using MrBayes with the TIM3+I+G model; MCMC = 400,000. (B) Maximum likelihood phylogenetic tree constructed using phyML with the TIM3+I+G model. PhyML was run for 200 replications. *Po-P.olivaceus, Tr-T. rubripes, Pf-P. formosa, Ga-G. aculeatus.*


A
TrGATA2a
TnGATA2a
GaGATA2a
OnGATA2a
PfGATA2a
OlGATA2a
PoGATA2a
CsGATA2a
GaGATA2b
TrGATA2b
TnGATA2b
OnGATA2b
OlGATA2b
PfGATA2b
PoGATA2b
CsGATA2b
HsGATA2
0.02
0.02
0.16
0.02
0.27
0.01
0.02
0.03
0.24
0.08
0.27
0.05

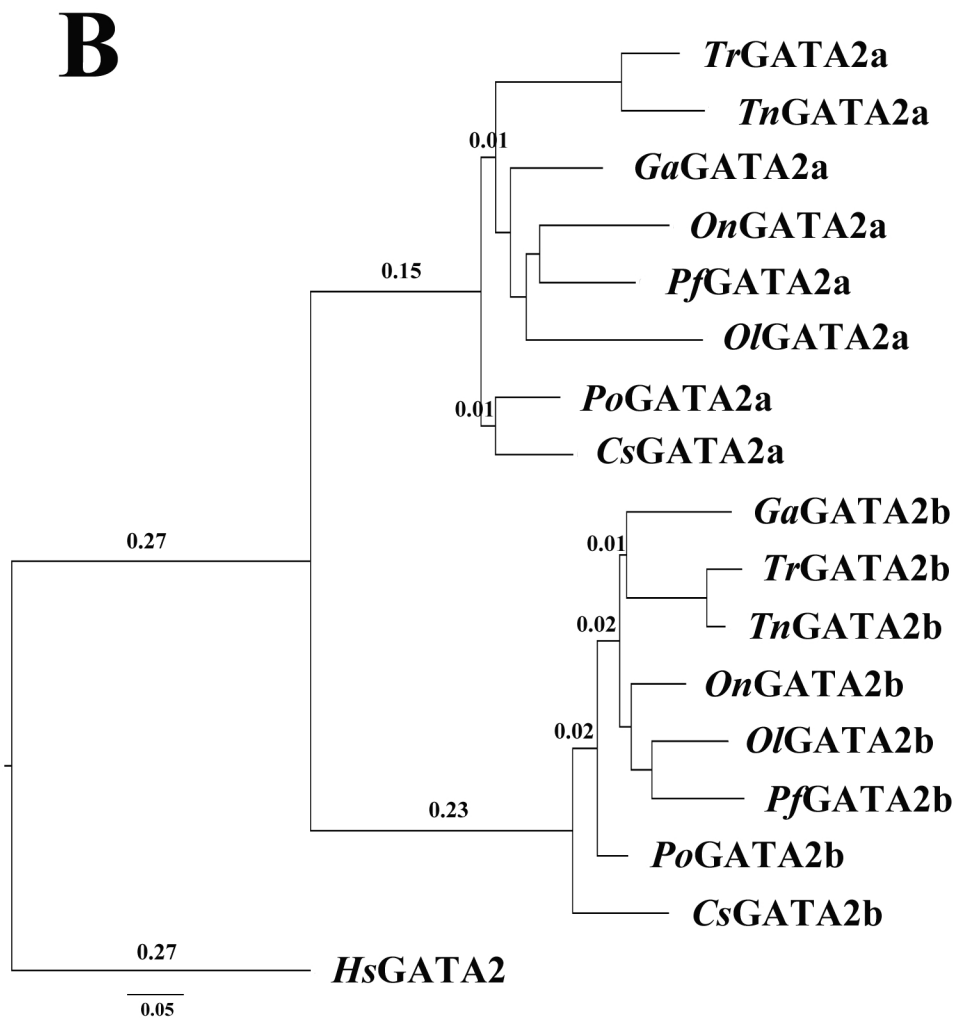

B
TrGATA2a
TnGATA2a
GaGATA2a
OnGATA2a
PfGATA2a
OlGATA2a
PoGATA2a
CsGATA2a
GaGATA2b
TrGATA2b
TnGATA2b
OnGATA2b
OlGATA2b
PfGATA2b
PoGATA2b
CsGATA2b
HsGATA2
0.01
0.15
0.01
0.27
0.01
0.02
0.02
0.23
0.27
0.05

Figure 4(on next page)

Chromosome synteny analysis of teleost GATA2 paralogs.

Dotplots of the human GATA2 gene region on human chr3 show double conserved synteny to the two GATA2 paralogs in zebrafish on chromosomes Dre6 (GATA2b) and Dre11 (GATA2a).

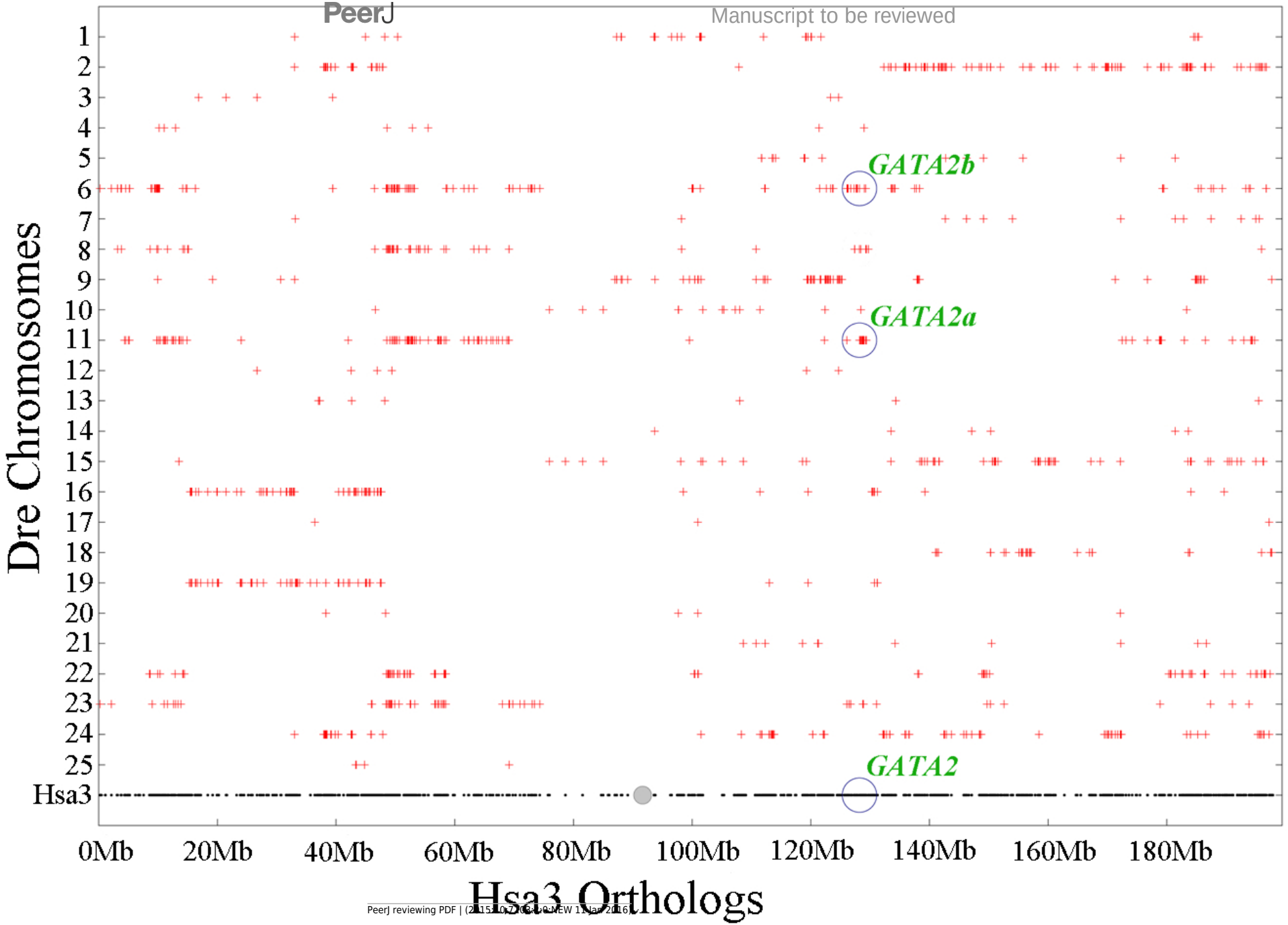

GATA2b
GATA2a
GATA2
Dre Chromosomes
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
Hsa3
0Mb
20Mb
40Mb
60Mb
80Mb
100Mb
120Mb
140Mb
160Mb
180Mb
Hsa3 Orthologs

Figure 5(on next page)

Chromosomal segments showing the conserved syntenic blocks containing GATA2a and GATA2b in teleosts.

The genes are represented by colored pentagons, and the gene names are indicated on top. Color pentagons indicate the same gene in different species and its respective genomic position in relation to several other genes. The pentagon's direction indicates the gene direction compared with the reference gene. The empty spaces indicate a region with other genes or the absence of the gene in the genome. The blank pentagons indicate conserved genes between GATA2a and GATA2b.

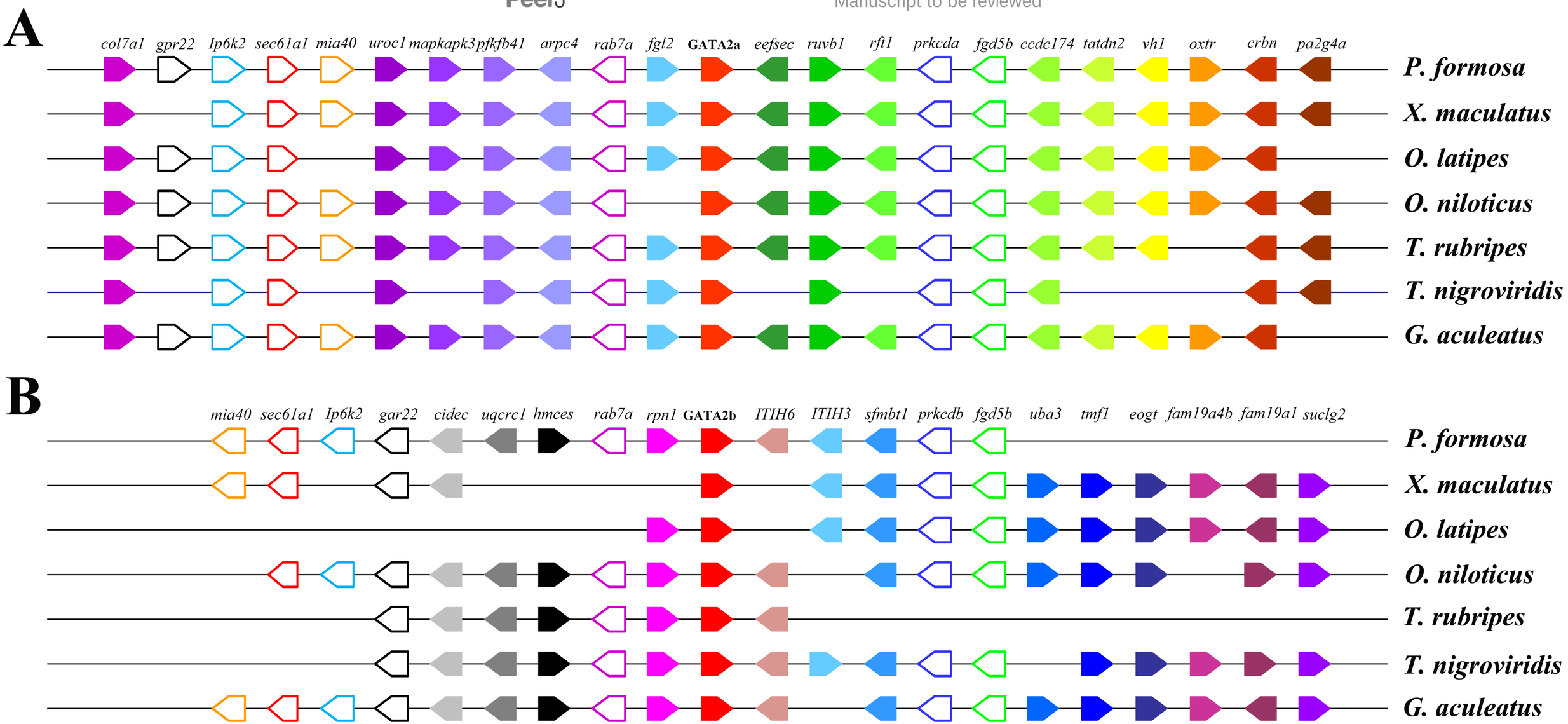
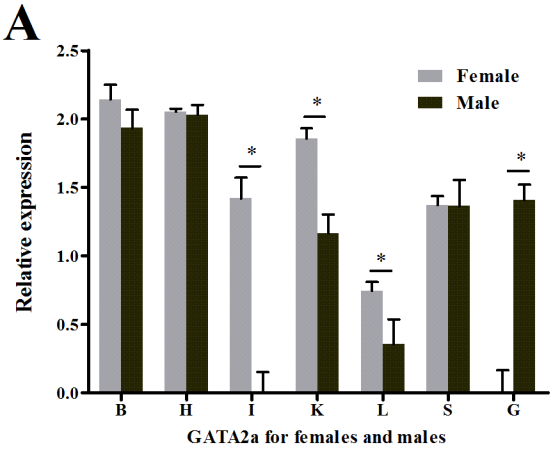

A
col7a1 gpr22 Ip6k2 sec61a1 mia40 uroc1 mapkapk3 pfkfb41 arpc4 rab7a fgl2 GATA2a eefsec ruvb1 rft1 prkcda fgd5b ccdc174 tatdn2 vh1 oxtr crbn pa2g4a
P. formosa
X. maculatus
O. latipes
O. niloticus
T. rubripes
T. nigroviridis
G. aculeatus
B
mia40 sec61a1 Ip6k2 gar22 cidec uqcrc1 hmces rab7a rpn1 GATA2b ITIH6 ITIH3 sfmbt1 prkcdb fgd5b uba3 tmf1 eogt fam19a4b fam19a1 suclg2
P. formosa
X. maculatus
O. latipes
O. niloticus
T. rubripes
T. nigroviridis
G. aculeatus

Figure 6(on next page)

Relative expression levels of GATA2a and GATA2b in tongue sole tissues.

(A) Relative expression level of GATA2a in males and females. (B) Relative expression level of GATA2b in males and females. B: brain; H: heart; I: intestine; K: kidney; L: liver; O: ovary; T: testis; S: spleen.


A
Relative expression
2.5
2.0
1.5
1.0
0.5
0.0
Female
Male
B
H
I
K
L
S
G
GATA2a for females and males

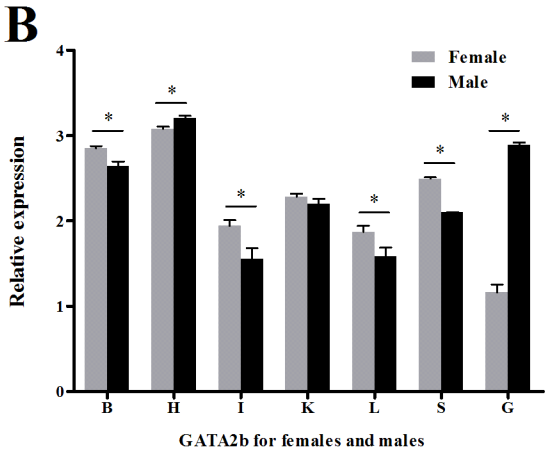

B
Relative expression
4
3
2
1
0
Female
Male
B
H
I
K
L
S
G
GATA2b for females and males

Table 1(on next page)

Results of sites model analyses on the teleost GATA2 Bayesian gene tree.

Table 1 Results of sites model analyses on the teleost GATA2 Bayesian gene tree.

Tree	Model	ln*L*	κ	Null	LRT	df	*P*-value	site	BEB
GATA2a	M0	-4832.177	2.463	NA					
	M1a	-4775.943	2.642	NA					
	M2a	-4775.943	2.642	M1a	0	2	1.000		
	M3	-4732.818	2.523	M0	198.718	4	0.000		
	M7	-4733.231	2.524	NA					
	M8a	-4735.152	2.540	NA					
	M8	-4733.231	2.524	M7	0	2	1.000		
				M8a	3.842	1	0.050		
GATA2b	M0	-4083.986	2.472	NA					
	M1a	-4053.831	2.560	NA					
	M2a	-4053.831	2.560	M1a	0	2	1.000		
	M3	-4034.159	2.488	M0	99.654	4	0.000		
	M7	-4034.060	2.485	NA					
	M8a	-4037.538	2.506	NA					
	M8	-4030.944	2.490	M7	6.232	2	**0.044**	356(P)	0.95*
				M8a	13.188	1	**0.00028**		

Note: Abbreviations: ln*L*, *ln* likelihood; κ, transition/transversion ratio; df, degrees of freedom; NA, not applicable

21