OPEN ACCESS Freely available online

PLOS ONE

Rapid Diversification of *FoxP2* in Teleosts through Gene Duplication in the Teleost-Specific Whole Genome Duplication Event

Xiaowei Song[1,2,3], Yajun Wang[2*], Yezhong Tang[1*]

1 Department of Herpetology, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, Sichuan, China, **2** College of Life Science, Sichuan University, Chengdu, Sichuan, China, **3** University of Chinese Academy of Sciences, Beijing, China

Abstract

As one of the most conserved genes in vertebrates, *FoxP2* is widely involved in a number of important physiological and developmental processes. We systematically studied the evolutionary history and functional adaptations of *FoxP2* in teleosts. The duplicated *FoxP2* genes (*FoxP2a* and *FoxP2b*), which were identified in teleosts using synteny and paralogon analysis on genome databases of eight organisms, were probably generated in the teleost-specific whole genome duplication event. A credible classification with *FoxP2*, *FoxP2a* and *FoxP2b* in phylogenetic reconstructions confirmed the teleost-specific *FoxP2* duplication. The unavailability of *FoxP2b* in *Danio rerio* suggests that the gene was deleted through nonfunctionalization of the redundant copy after the Otocephala-Euteleostei split. Heterogeneity in evolutionary rates among clusters consisting of *FoxP2* in Sarcopterygii (Cluster 1), *FoxP2a* in Teleostei (Cluster 2) and *FoxP2b* in Teleostei (Cluster 3), particularly between Clusters 2 and 3, reveals asymmetric functional divergence after the gene duplication. Hierarchical cluster analyses of hydrophobicity profiles demonstrated significant structural divergence among the three clusters with verification of subsequent stepwise discriminant analysis, in which FoxP2 of *Leucoraja erinacea* and *Lepisosteus oculatus* were classified into Cluster 1, whereas FoxP2b of *Salmo salar* was grouped into Cluster 2 rather than Cluster 3. The simulated thermodynamic stability variations of the forkhead box domain (monomer and homodimer) showed remarkable divergence in FoxP2, FoxP2a and FoxP2b clusters. Relaxed purifying selection and positive Darwinian selection probably were complementary driving forces for the accelerated evolution of *FoxP2* in ray-finned fishes, especially for the adaptive evolution of *FoxP2a* and *FoxP2b* in teleosts subsequent to the teleost-specific gene duplication.

Citation: Song X, Wang Y, Tang Y (2013) Rapid Diversification of *FoxP2* in Teleosts through Gene Duplication in the Teleost-Specific Whole Genome Duplication Event. PLoS ONE 8(12): e83858. doi:10.1371/journal.pone.0083858

Editor: Pierre Boudinot, INRA, France

Received August 6, 2013; **Accepted** November 18, 2013; **Published** December 9, 2013

Funding: Funding came from the Chinese Academy of Sciences Bairen Jihua (KSCX2-YW-R-077) and the National Natural Science Foundation of China (NSFC 31272304) to YZT and NSFC 31172202 to YJW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* Email: cdwyjhk@gmail.com (YW); tangyz@cib.ac.cn (YT)

Introduction

FoxP2 is a key transcription factor gene in the *FoxP* subfamily [1]. The gene possesses multiple functionally important domains, including the zinc-finger, leucine-zipper and forkhead box (Figure 1A). Genetic studies show that when *FOXP2* is rendered dysfunctional in humans, seen in cases of pathogenic translocation, missense mutation (R553H) and nonsense mutation (R328X), severe disorders of speech and language can result [2,3]. The gene is also likely involved in vocal control in zebra finches, mice and bats [4-7]. In addition, it has been found that *Foxp2* plays a nonessential role in the production of innate emotional vocalizations in mouse pups [8]. Based on these findings, it is reasonable to infer that the role of *FoxP2* in vocal communication is a derived function as opposed to an ancestral one. Aside from its role in vocal communication, *FoxP2* also plays a pleiotropic role in cell differentiation, signal transduction, organogenesis (e.g. lungs), and neural circuit plasticity in the central nervous system (CNS) [3-6,9-16].

Nevertheless, basic functions of the gene remain unclear due to limited studies. Functional studies of *FoxP2* have been mainly performed on mammals and birds but have seldom been conducted on other vertebrates (i.e. reptiles, amphibians and fishes) [11,17-21]. The conservative expression patterns of *FoxP2* seen during embryonic development in homologous brain regions of mammals, birds, frogs, and fishes [17] imply that *FoxP2* is associated with similar functions in the CNS of vertebrates. In mice, *Foxp2* is essential for lung development, especially for postnatal lung alveolarization [12,22]. *FoxP2* of

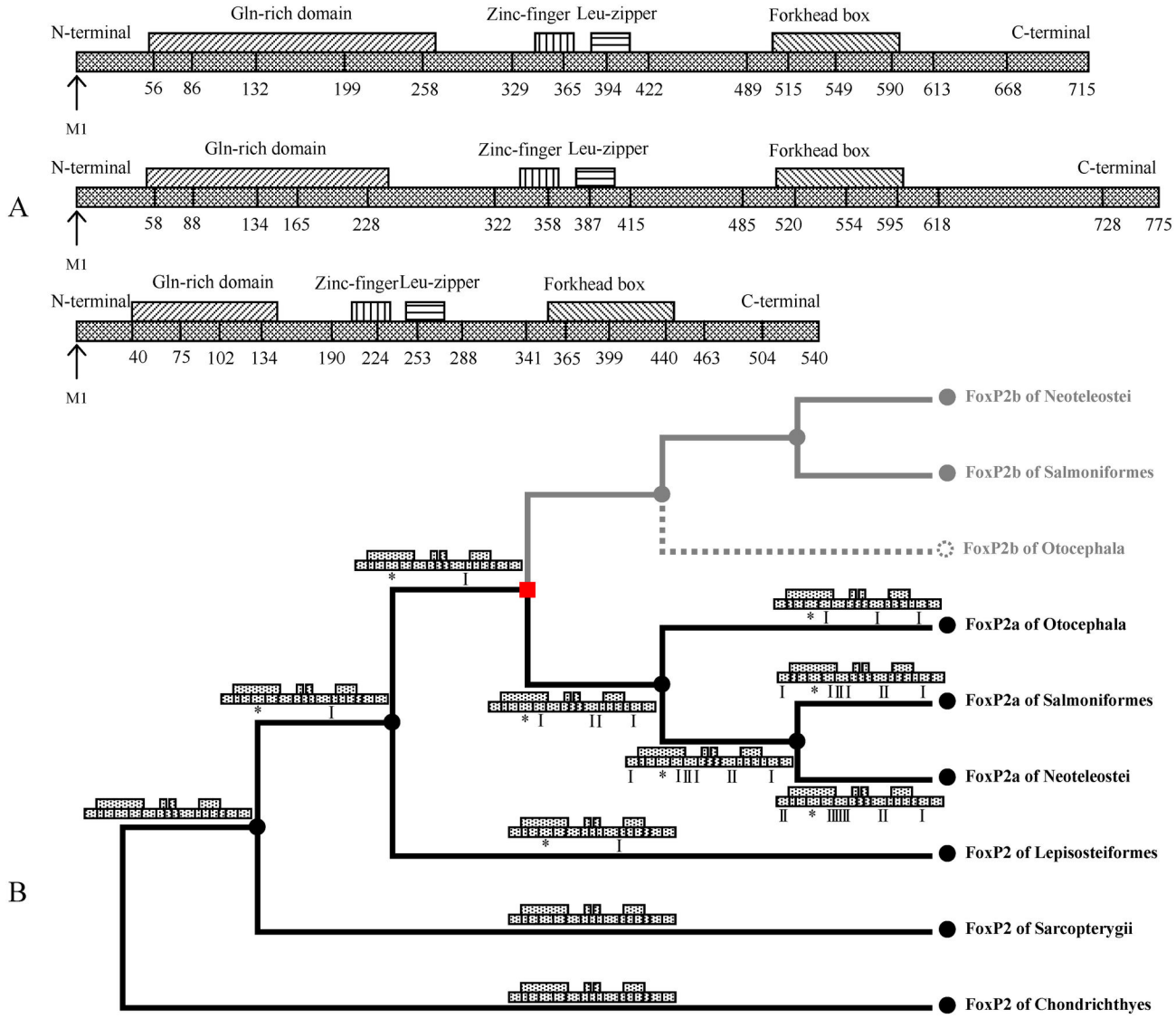

Figure 1

Figure 1. Schematic drawings and predicted evolutionary changes of FoxP2, FoxP2a and FoxP2b. (A) FOXP2 in *Homo sapiens* (top), FoxP2a in *Oryzias latipes* (middle) and FoxP2b in *Oryzias latipes* (bottom). M1 – Methionine coded by initiator; Numbers – Amino acid sites linking exons. (B) Branches of FoxP2b are shown in grey. Red square node indicates teleost-specific FoxP2 duplication. Symbols of '*' and 'I' under schematic drawings represent losses of poly-Glutamine and inserts, respectively. The square dotted branch indicates lineage-specific deletion of FoxP2b in Otocephala. Despite a comparatively short length, exon-compositions of FoxP2b in euteleosts are still obscure.

doi: 10.1371/journal.pone.0083858.g001

medaka shows weak transcriptional activity, *in vitro*, for regulation of the lung-specific *CC10* gene of mice[19].

Despite the fact that *FoxP2* is one of the most conserved genes in vertebrates [23-25], structures and functions of the gene in teleosts probably show substantial divergence [18,19,21]. It is still an open question about how and why such substantial divergence occurred. Either positive selection or relaxation of purifying selection may have forced the divergence of the gene. In addition, gene duplication often results in diversifying natural selection pressures on the duplicated genes, further forcing their fast evolution [26]. Since the fish-specific whole genome duplication, which is also known as the 3rd round of whole genome duplication (3R-WGD), in the common ancestor of teleosts seems to be correlated with species diversification of teleosts [27], we hypothesize that *FoxP2* possibly has been duplicated in teleosts. Therefore, the logical existence of teleost-specific *FoxP2* paralogs should be investigated further. In addition, the teleost swim bladder has been considered to be a homologous organ to the lung on the basis of the consistency of both developmental blastema and molecular pathways, despite the fact that its main function is buoyancy regulation rather than respiration [28,29]. Thus, the divergence of *FoxP2* in teleosts may coincide with the adaptation of the swim bladder.

The present study aims to shed light on the underlying mechanisms from which the remarkable divergence of *FoxP2* in ray-finned fishes, particularly teleosts, resulted. We systematically investigated the evolutionary history and functional adaptations of *FoxP2* through a series of analyses on phylogenetic relationships, natural selection pressures, and divergence of structures and functions. The study concerning rapid evolution of *FoxP2* in ray-finned fishes will prompt further explorations of *FoxP2*'s functional adaptations.

Materials and Methods

Molecular biological methods for data collection

The experimental animals included *Colisa lalia*, *Pachytriton labiatus*, *Rana daunchina*, *Phrynocephalus vlangalii* and *Gekko gecko* (Information S1). All procedures used were approved by the Animal Care Committee of Chengdu Institute of Biology, Chinese Academy of Sciences. Animals were first anesthetized with pelltobarbitalum natricum before sacrifice. Total RNA was extracted from the brains of the animals by using the TRIzol reagent (Invitrogen). The cDNA templates were synthesized by RevertAid First Strand cDNA Synthesis Kit (Fermentas). The primers for the subsequent polymerase chain reaction (PCR) were designed by utilizing Primer Premier 5 and Oligo 6 softwares in conjunction with Primer 3 plugged into Biology Workbench of San Diego Supercomputer Center (http://workbench.sdsc.edu) and synthesized by Sangon Biotech (Shanghai) Co., Ltd (Information S1). PCRs were performed using EasyTaq DNA polymerase or TransTaq HiFi DNA polymerase (Beijing TransGen Biotech Co., Ltd.). PCR products were purified by Axyprep DNA Gel Extraction Kit (AxyGEN) and cloned into the pTA2 vector (Toyobo). Gene sequencing was carried out by Invitrogen Company (Shanghai).

In silico methods for data collection

We downloaded 18 coding sequences or fragments from GenBank or Ensembl directly (Information S1). We obtained 19 additional coding sequences or fragments through screening genomes and assessed the results using the Basic Local Alignment Search Tool (BLAST) in the National Center for Biotechnology Information (NCBI) [30] (Information S1). The data bases for screening *FoxP2* included the NCBI, the USCS Genome Bioinformatics website (http://genome.ucsc.edu/) [31] and the Elephant Shark Genome Project (http://esharkgenome.imcb.a-star.edu.sg/) [32,33] (Information S1). Since amino acid sequences of *FoxP2* in vertebrates are highly conserved, we used protein queries (*Danio rerio* [DQ061052.1], *Xenopus laevis* [BC170268.1], *Taeniopygia guttata* [AY395709.1] and *Mus musculus* [NM_053242.4]) to predict *FoxP2* coding sequences through genome screenings. According to the results of phylogenetic reconstruction as well as synteny and paralogon analysis, the formerly named *FoxP2* of teleosts was identified to be *FoxP2a*, as one of two paralogs (i.e. *FoxP2a* and *FoxP2b*). The nomenclature for these genes was therefore corrected in the present study to avoid confusion. Due to the drastic divergence and unclear characteristics of *FoxP2b*, we predicted *FoxP2b* coding sequences by a two step approach: first, searching contig locations with queries of FoxP2b in *Gadus morhua* (GW852741.1) and *Oryzias latipes* (XM_004069881.1); second, screening out the coding sequences with GENSCAN (http://genes.mit.edu/GENSCAN.html) [34].

Synteny and paralogon analysis

To investigate whether *FoxP2* paralogs exclusively exist in teleosts, we conducted synteny and paralogon analysis to determine homologous (orthologous or paralogous) relationships of sequences involved [35,36]. The synteny and paralogon analysis of *FoxP2* loci was performed on genome databases of eight representative species in Ensembl: *Homo sapiens* (GRCh37), *Taeniopygia guttata* (taeGut3.2.4), *Anolis carolinensis* (AnoCar2.0), *Xenopus tropicalis* (JGI_4.2), *Danio rerio* (Zv9), *Oryzias latipes* (MEDAKA1), *Oreochromis niloticus* (Orenil1.0) and *Tetraodon nigroviridis* (TETRAODON8). Twenty upstream and 20 downstream protein-coding genes flanking *FoxP2*, *FoxP2a* and *FoxP2b* were selected for the synteny and paralogon analysis. The homologous relationships among sequences were determined using the prediction method in Ensembl (http://www.ensembl.org) [37]. Results of pairwise region comparisons were integrated into one figure (Figure 2 and Information S2).

Phylogenetic reconstructions

In order to facilitate comparison and discussion, all site and exon numbers of FoxP2, FoxP2a and FoxP2b were made to correspond to the numbers of the counterparts in human FOXP2 (Figure 1A). Multiple sequence alignments were executed with clustalW in the BioEdit software (version 7.0.5.3) [38,39] (Information S3). Since exon composition and the functions of *FoxP2b* were unknown, we only phylogenetically analyzed the conserved forkhead box domain, i.e. exons 12, 13, 14, 15 and partial exon11, in Data set 1 (consisting of 41

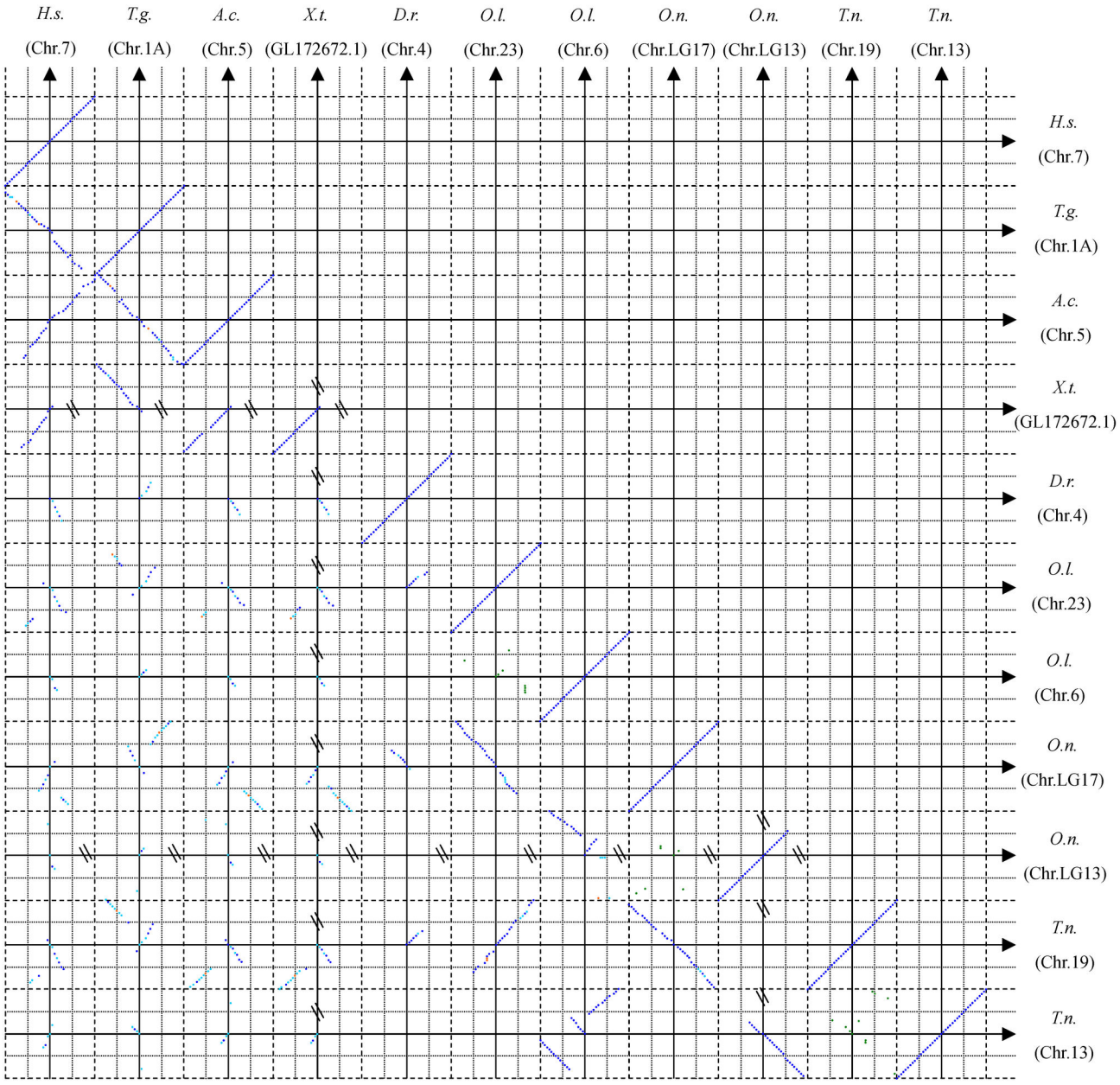

Figure 2

Figure 2. Pairwise comparisons of 40 (20 upstream and 20 downstream) protein-coding genes flanking FoxP2, *FoxP2a* or *FoxP2b* among eight vertebrates. The solid lines represent chromosomes (or scaffold) of species, i.e. Chr. 7 of *Homo sapiens* (H.s.), Chr. 1A of *Taeniopygia guttata* (T.g.), Chr. 5 of *Anolis carolinensis* (A.c.), GL172672.1 of *Xenopus tropicalis* (X.t.), Chr. 4 of *Danio rerio* (D.r.), Chr. 23 and 6 of *Oryzias latipes* (O.l.), Chr. LG17 and LG13 of *Oreochromis niloticus* (O.n.) and Chr. 19 and 13 of *Tetraodon nigroviridis* (T.n.). The arrow head of each solid line points toward the forward direction of the corresponding chromosome. Crossing points of the solid lines indicate gene loci for FoxP2, *FoxP2a* or *FoxP2b*. Short dashed lines and dotted lines show the borders of pairwise comparisons and the measuring scale of gene loci, respectively. The points with blue, sky blue, orange and green indicate that the corresponding gene pairs are one-to-one orthologous, one-to-many or many-to-many orthologous, possibly orthologous and intraspecific paralogous, respectively.

doi: 10.1371/journal.pone.0083858.g002

FoxP2, *FoxP2a* and *FoxP2b* fragments) (Figure 3 and Information S4). Phylogenetic trees generated with a single intraspecific paralogous gene were able to avoid polytomy. In view of the occurrence of excessive divergence in teleost *FoxP2b*, we reconstructed gene trees using the *FoxP2* and *FoxP2a* coding sequences (i.e. Data set 2) to infer the species tree (Figure 4 and Information S4). Since exons 6, 16 and partial exon8 were unavailable in some species, we deleted these exons and the low complexity poly-glutamine (poly-Q) region (i.e. exon5) in Data set 2. We assessed substitution saturation of these data sets before the phylogenetic analyses using DAMBE software (version 5.2.57) [40]. Maximum Likelihood (ML) and Neighbor-Joining (NJ) trees were constructed based on nucleotide or amino acid sequences using MEGA software (version 5.10) [41,42]. We used the general time reversible (GTR) model with gamma distributed plus invariant sites (G+I) and the maximum composite likelihood (MCL) method [43] for ML and NJ trees based on nucleotide sequences, respectively. The models/methods for both the ML and NJ trees based on amino acid sequences were the Jones-Taylor-Thornton (JTT) model with G [44]. The bootstrap method (1000 replications) was used in phylogenetic reconstructions for the ML and NJ trees [45]. Bayesian phylogenetic analysis was performed with MrBayes software (version 3.1.2) [46]. We partitioned nucleotide coding sequences into three datasets (first, second and third codon position sites) which had some unlinking parameters (statefreq, revmat, shape and pinvar) across the partitions. The analyses by MrBayes were based on GTR+G+I with generations (10,000,000), sample frequency (1000) and temperature (0.05). All other options were left at default settings.

Analysis of functional divergence

The coefficient of functional divergence (θ_λ) is an efficient parameter for testing type I functional divergence after gene duplication [47]. Type I functional divergence refers to different functional constraints in paralogous gene clusters after gene duplication. The method used here is based on the principle that the coefficient of evolutionary rate correlation (r_λ) reflects functional divergence in the gene clusters. Specifically, the coefficient of functional divergence is calculated with the formula, $\theta_\lambda = 1 - r_\lambda$, using DIVERGE software (version 1.04) [47]. A statistical significance level of $\theta_\lambda > 0$ is computed using a likelihood ratio test (LRT) through comparing whether $\delta = -2 * \ln(LR)$ is greater than the value of the $\chi^2_{[1]}$ distribution at a significance level of 0.05. LR represents the ratio of the maximum likelihood that $\theta_\lambda > 0$ to that of $\theta_\lambda = 0$. In addition, the critical amino acids that are responsible for the type I functional divergence can be identified using a reasonable cutoff value (> 1) for the posterior probability ratio based on the hidden Markov model.

We tested the functional divergence in Data set 1 among different gene clusters, i.e. Cluster 1 (*FoxP2* of Sarcopterygii), Cluster 2 (*FoxP2a* of Teleostei) and Cluster 3 (*FoxP2b* of Teleostei) based on the teleost lineage specific *FoxP2* duplication in the ancestral 3R-WGD event (Figure 5 and Information S5). A stringent cutoff value (> 2.33) for the posterior probability ratio was used, i.e. a posterior probability of a site-specific rate difference (> 0.7). Since the teleost *FoxP2a* exons surrounding functionally important domains (zinc-finger, leucine-zipper and forkhead box) show greater divergence than these functional domains, we also tested the functional divergence between Cluster 1 and Cluster 2 in Data set 2 to determine if the two clusters exhibit homogeneous evolutionary rates in an extensive region (Figure 6 and Information S5).

Analysis of structural divergence

To investigate the functional adaptations of a protein, the evolutionary changes of its molecular structures must first be studied. High correlation between sequence similarity and structure similarity [48] indicates that structural traits (e.g. charge, polarity and hydropathy) of amino acids in a protein determine its higher level structures. Thus, the hydrophobicity values of amino acids calculated with the Kyte-Doolittle method were utilized to evaluate the structure of FoxP2, FoxP2a and FoxP2b [49]. The amino acid sequences in Data sets 1 and 2 were transformed into hydrophobicity profiles composed of normalized hydrophobicity values from 0 to 1. In addition, we defined 'gap' sites in sequence alignments as a normalized hydrophobicity value of 0.5. For the sake of capturing general differences between protein structures, we executed hierarchical cluster analyses based on four kinds of pairwise distances (i.e. Cityblock, Correlation, Cosine and Euclidean) using a Matlab program (MathWorks) (Information S6). The average-linkage cluster method was chosen for the cluster analyses. The cophenetic correlation coefficients between the cophenetic distances obtained from a cluster tree and the original distances used to construct the tree can be used to assess how faithfully the tree represents the original distances between sequences. In addition, we used the inconsistency coefficient as a measure of dissimilarity between downward links (branches) or clusters connected by a link. Although an inconsistency coefficient threshold could be arbitrarily selected to classify the sequences, we set a flexible cutoff value (0.8) greater than the baseline value (0.707) to avoid an exhaustive classification. The depth denoting the number of levels of the cluster tree for calculating the inconsistency coefficient was set to 2.

Furthermore, we performed a stepwise discriminant analysis using SPSS software (version 20) to evaluate the consistency of the classification in the four cluster trees constructed with each data set. The variable selection method used in the discriminant analysis was Mahalanobis distance. The prior probability for groups was computed according to size. We kept the F values for entry or removal of variables and the covariance matrix for classification as defaults. Coefficients for both the canonical discriminant function and Fisher's linear discriminant function were computed. Finally we evaluated the classification results by using the percentage of original grouped cases and cross-validated grouped cases correctly classified. In Data set 1, we focused on the rationalization of groupings for FoxP2 of *Leucoraja erinacea*, *Lepisosteus oculatus* and *Rana daunchina*, as well as FoxP2b of *Salmo salar* and *Gadus morhua*. The previous three clusters (i.e. Cluster 1, Cluster 2 and Cluster 3) were used to discriminate

Figure 3

Figure 3. NJ tree based on nucleotide sequences in Data Set 1. Bootstrap values lower than 50 on branches are not shown.
doi: 10.1371/journal.pone.0083858.g003

these sequences. With Data set 2, we investigated whether FoxP2 of *Leucoraja erinacea* and *Lepisosteus oculatus* were grouped together with Cluster 1 or with Cluster 2. In order to study whether the classification was a remaining trace after a long-term evolutionary history, we also used the three clusters to discriminate hydrophobicity profiles of ancestral sequences

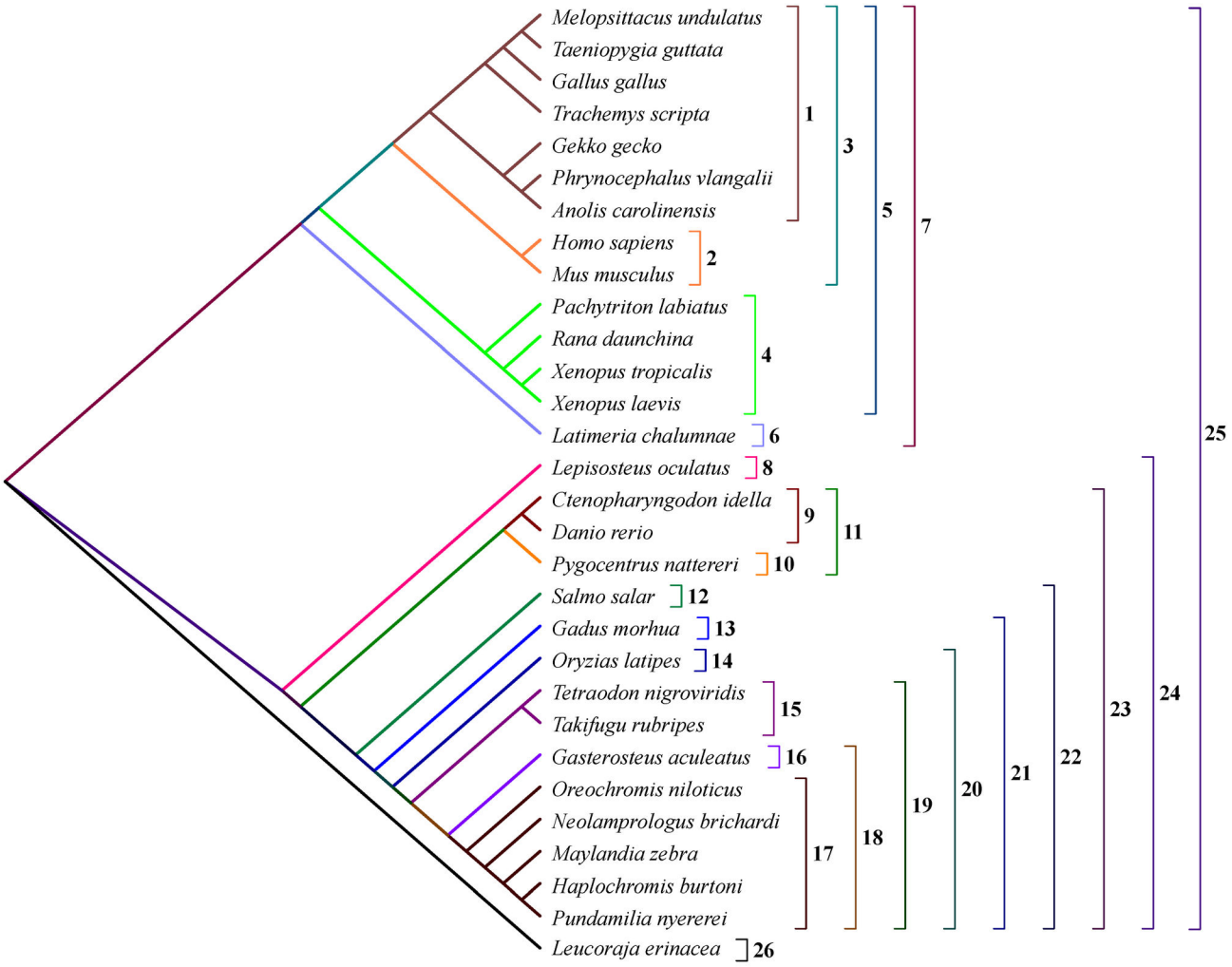

Figure 4

Figure 4. Phylogenetic topology tree of 30 species. 1, Sauropsida; 2, Mammalia; 3, Amniota; 4, Amphibia; 5, Tetrapoda; 6, Coelacanthiformes; 7, Sarcopterygii; 8, Lepisosteiformes; 9, Cypriniformes; 10, Characiformes; 11, Otocephala; 12, Salmoniformes; 13, Gadiformes; 14, Beloniformes; 15, Tetraodontiformes; 16, Gasterosteiformes; 17, Perciformes; 18, unnamed; 19, Percomorpha; 20, Acanthopterygii; 21, Neoteleostei; 22, Euteleostei; 23, Teleostei; 24, Actinopterygii; 25, Osteichthyes; 26, Chondrichthyes.
doi: 10.1371/journal.pone.0083858.g004

from Data sets 1 and 2, which were inferred with a Dayhoff substitution rate matrix using the codeml program of PAML software [50].

Evaluation of thermodynamic stability

The three-dimensional (3D) structure of the forkhead domain of human FOXP2 (PDB: 2A07) provides a model to investigate changes in the stability of the local 3D structure in different evolutionary scenarios. We applied FoldX (version 3.0 b5.1), a fast and effective computer algorithm for estimating the effect of mutations on stability [51], in order to predict the differences in total Gibbs free energy (DTE) from wild-type (i.e. human FOXP2) to mutants (i.e. extant sequences and predicted ancestral ones as described above). Amino acid changes in the forkhead domain between two neighboring nodes are shown in Figure 5. We repaired the model file to optimize the structure prior to these energy calculations. In addition to using chains A, B and J as a model for predicting the DTE of monomers, we extracted chains A, B, F and G to use as a model for evaluating the DTE of homodimers. For our calculations, we ignored water

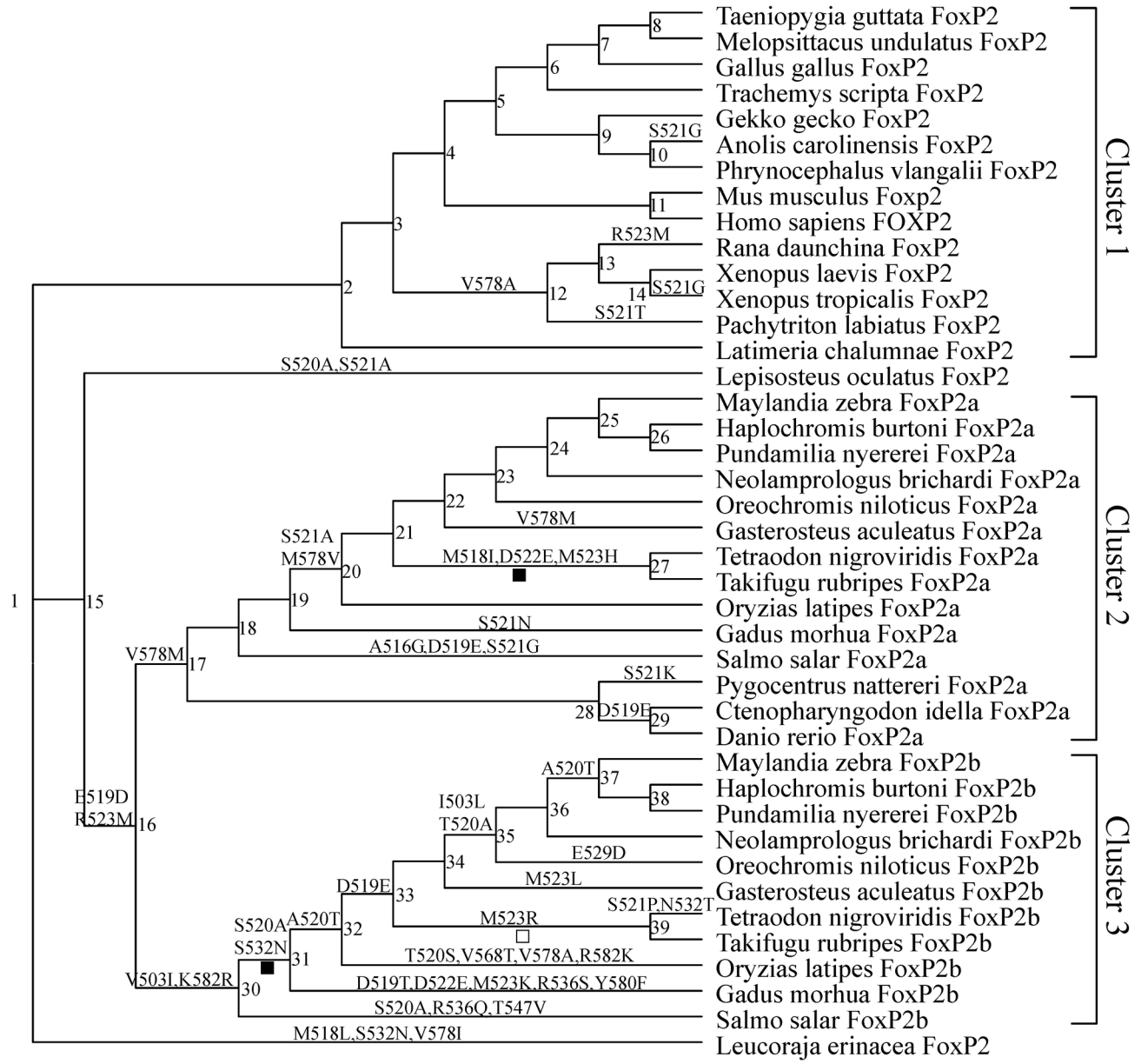

Figure 5

Figure 5. The phylogenetic topology tree for functional divergence and natural selection pressures tests in Data Set 1. Numbers adjacent to nodes represent ancestral FoxP2, FoxP2a and FoxP2b. Forkhead domain evolutionary changes between neighboring nodes are listed above the branches. The filled square under a branch indicates that the gene has undergone positive selection in the evolutionary status; the blank square represents that the gene has experienced accelerated evolution.

doi: 10.1371/journal.pone.0083858.g005

bridges and retained binding metals in crystals. We executed 20 energy prediction runs for each mutant to get an average difference in total Gibbs free energy (ADTE). If the ADTE of a mutant was greater than 0, the mutant was destabilizing; alternatively, if the ADTE was found to be less than 0, the mutant was stabilizing. According to the ADTE value of each

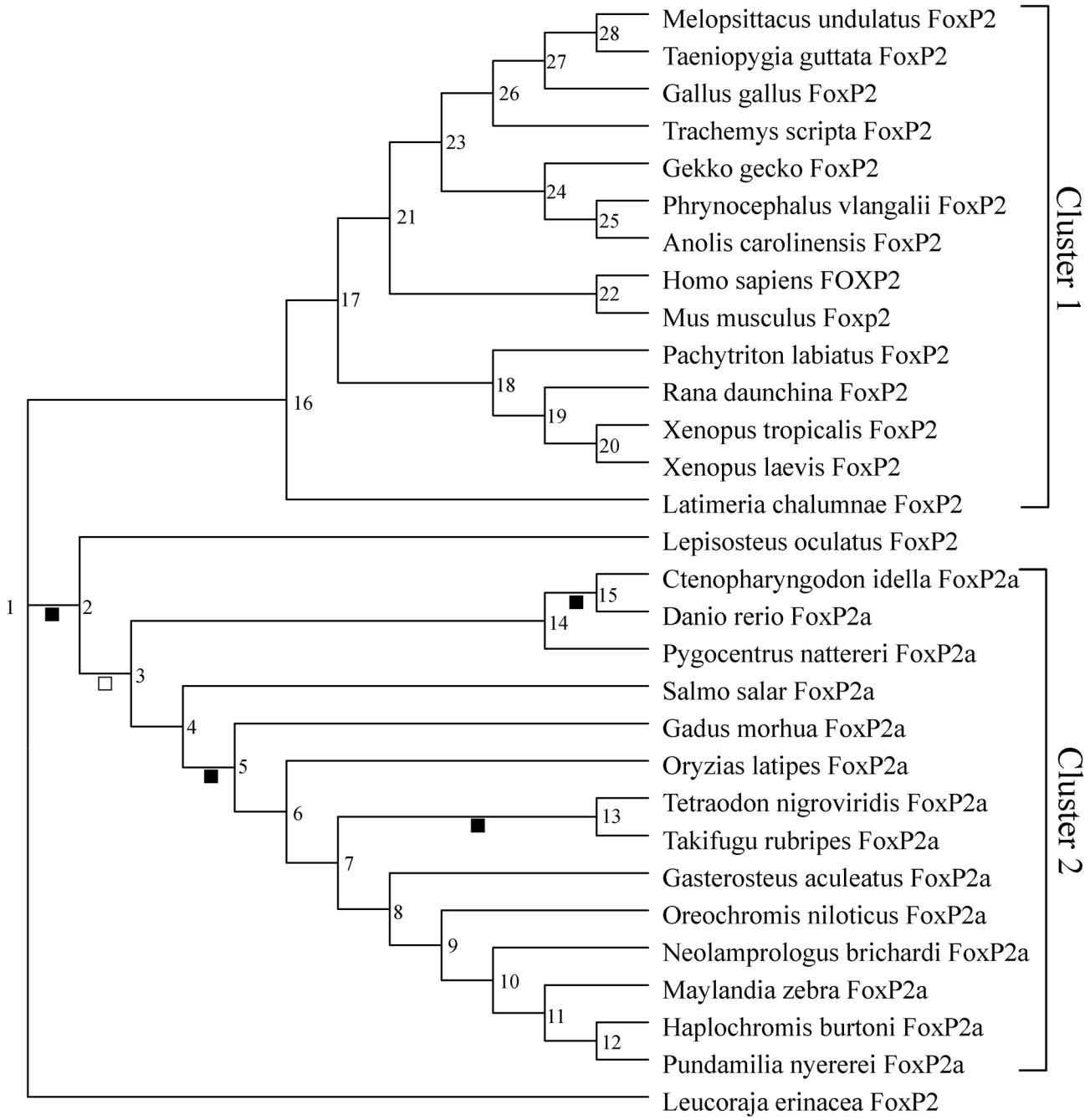

Figure 6

Figure 6. The phylogenetic topology tree for functional divergence and natural selection pressures tests in Data Set 2. Numbers adjacent to nodes represent ancestral FoxP2 and FoxP2a. The filled square under a branch indicates that the gene has undergone positive selection in the evolutionary status; the blank square indicates that the gene has experienced accelerated evolution.

doi: 10.1371/journal.pone.0083858.g006

Table 1. Estimates of the coefficient of functional divergence (θ_λ) among clusters.

	Clusters 1 and 2	Clusters 1 and 3	Clusters 2 and 3	Clusters 1 and 2 (')
θ_λ	0.267	0.578	0.706	0.402
α	0.071	0.401	0.321	0.220
SE of θ_λ	0.196	0.322	0.174	0.106
δ	1.845	3.215	16.395*	14.261*

The rightmost column with an apostrophe lists results from Data set 2; the other three columns list results from Data set 1. The starred δ value indicates the column is statistically significant. α – Gamma shape parameter of rate variation among sites; SE – Standard error; δ – LRT value of θ_λ; Cluster 1 – *FoxP2* of Sarcopterygii; Cluster 2 – *FoxP2a* of Teleostei; Cluster 3 – *FoxP2b* of Teleostei.

doi: 10.1371/journal.pone.0083858.t001

mutant calculated with the same wild model, we determined the stability changes of monomers or homodimers in each evolutionary scenario using ADTE variations. Since the error margin in FoldX was approximately 0.5 kcal/mol [52], ADTE changes within this range were deemed insignificant.

Test of natural selection pressures

The analyses of functional and structural divergence suggest that *FoxP2a* and *FoxP2b* in teleosts experienced dramatic evolutionary changes after gene duplication. Therefore, we tested natural selection pressures with Data set 1 based on the corresponding phylogenetic topology tree (Figure 5) using the codeml program implemented in PAML software (version 4.6) [50]. In view of the occurrence of heterogeneous evolutionary rates between conserved functional domains (e.g. the forkhead box) and the exons flanking those functional domains, we also tested natural selection pressures with Data set 2 to investigate whether the regions without Data set 1 underwent divergent natural selection pressures (Figure 6). Codon substitution models including site models, branch specific models and branch-site model A were utilized. The LRT was used to determine the fitness of the models to the data sets analyzed [50]. For the LRT, $2\Delta l$ (twice the log likelihood difference) between a model and its nested model was compared to the value of the χ^2 distribution with degrees of freedom equal to the numerical difference between the free parameters of the two models. The nested model was rejected at a significance level of $P < 0.05$.

First, three pairs of site models were tested on the data sets and compared based on the LRT (Information S7). Comparison of Model 3 (discrete) to Model 0 (basic or one ratio) was used to test whether ω varied among sites. Comparisons of Model 2 (positive selection) to Model 1 (nearly neutral), and Model 8 (beta and ω > 1) to Model 7 (beta) were conducted in order to detect positively selected sites with the Bayes empirical Bayes (BEB) method [53]. The test of variable ω among different branches was performed by comparing the free ratios model (independent ω ratio for each branch) to Model 0.

We used branch specific models to detect whether Cluster 1, Cluster 2 and Cluster 3 experienced different natural selection pressures (Information S7). Due to a significantly variable ω ratio among sites and branches, the branch-site model A, which allowed ω to vary among sites and across branches, was used to detect positive selection affecting a few sites along specific foreground lineages (Figure 5, Figure 6 and Information S7). Test 1 and Test 2 were performed for these branch-site models. The null models in Test 1 and Test 2 are Model 1 and branch-site model A with ω2 = 1 fixed, respectively [53]. Although Test 1 is unable to provide direct evidence of the existence of positive Darwinian selection, it can detect accelerated evolution attributed to potential positive selection or relaxed purifying selection. In contrast, Test 2 is a powerful test of positive natural selection. Positively selected sites inferred with the BEB procedure were replaced by the corresponding sites of human FOXP2. If *p* (one tail probability for χ^2 distribution) value was not the statistically significant ($P < 0.05$), we neglected Test 2 and positively selected sites.

Results

Phylogenetic relationships between FoxP2, *FoxP2a* and *FoxP2b*

Homologous relationships between each pair of genes were sketched out in Figure 2. The 20 upstream and 20 downstream protein-coding genes flanking *FoxP2* shared a highly conserved synteny in four tetrapod species: *Homo sapiens*, *Taeniopygia guttata*, *Anolis carolinensis* and *Xenopus tropicalis* (Figure 2 and Information S2). In comparison, many co-localized genes around the *FoxP2a* and *FoxP2b* gene loci in *Danio rerio*, *Oryzias latipes*, *Oreochromis niloticus* and *Tetraodon nigroviridis* were orthologs of the one-to-two type. The 40 genes surrounding *FoxP2a* in *Oryzias latipes*, *Oreochromis niloticus* and *Tetraodon nigroviridis* resembled each other in synteny. Similarly, the co-localized genes flanking *FoxP2b* in *Oryzias latipes*, *Oreochromis niloticus* and *Tetraodon nigroviridis* shared a conserved synteny. In addition, *FoxP2a* in *Danio rerio* possessed a chromosomal location similar to that of *FoxP2a* in *Oryzias latipes*, *Oreochromis niloticus* and *Tetraodon nigroviridis*. However, a candidate ortholog of *FoxP2b* in these three teleosts could not be screened out in the genome of *Danio rerio*. In addition, *FoxP2a* and *FoxP2b* in *Oryzias latipes*, *Oreochromis niloticus* and *Tetraodon nigroviridis* resided in paralogous chromosome segments, i.e. paralogons (Figure 2). The linked paralogous gene pairs shared between *FoxP2a* and *FoxP2b* in *Oryzias latipes*, *Oreochromis niloticus* and *Tetraodon nigroviridis* were eight, six and nine, respectively.

Multiple evolutionary changes (inversion, deletion and translocation) could be pinpointed in the chromosomal segments in which *FoxP2* was located during the long-term differentiation of these vertebrates. For instance, inversion of the chromosomal segment of *Taeniopygia guttata* occurred after its separation with the other three tetrapods. In comparison with chromosomal segments containing *FoxP2a* in *Danio rerio*, *Oryzias latipes* and *Tetraodon nigroviridis*, an inversion occurred in the corresponding location of *Oreochromis niloticus*. A microinversion of chromosomal segments containing *FoxP2a* occurred in *Oryzias latipes*, *Oreochromis niloticus* and *Tetraodon nigroviridis* relative to

their tetrapod counterparts. The microinversion was also seen in chromosomal segments containing *FoxP2b* of *Oryzias latipes*, *Oreochromis niloticus* and *Tetraodon nigroviridis*.

Multiple sequence alignments showed that *FoxP2a* and particularly *FoxP2b* were both greatly divergent from their *FoxP2* orthologs. *FoxP2a* in teleosts possessed a 16-exon composition resembling that in human *FOXP2* (Figure 1A). Exon compositions of *FoxP2b* and *FoxP2a* in *Oryzias latipes* were similar to each other (Figure 1A). Specifically, *FoxP2a* of teleosts contained variable amino acid or fragment inserts (3–11), yet lacked a poly-Q repeat between site 154 and site 192 (Figure 1B and Information S3). All of the inserts were located in regions (i.e. exons 2, 6, 7, 11, 12 and 16) flanking three functional domains (zinc-finger, leucine-zipper and forkhead box). In contrast, these corresponding regions were likely truncated in *FoxP2b* of *Oryzias latipes* (Figure 1A).

All phylogenetic trees constructed with nucleotide or amino acid sequences in Data set 1 strongly supported a classification of *FoxP2*, *FoxP2a* and *FoxP2b* (Figure 3 and Information S4). *FoxP2a* of *Danio rerio*, *Ctenopharyngodon idella* and *Pygocentrus nattereri* were identified with a high degree of credibility. The gene trees reconstructed with the *FoxP2* and *FoxP2a* coding nucleotide sequences in Data set 2 (Information S4) were able to credibly reflect the phylogenetic relationships of the organisms involved [54] (Figure 4). The topological gene trees applied to subsequent tests in Data sets 1 and 2 (Figures 5 and 6) were revised ones conforming to the predicted species tree (Figure 4).

Functional divergence among FoxP2, FoxP2a and FoxP2b

The coefficient of functional divergence (θ_λ) between Cluster 2 and Cluster 3 in Data set 1 was significantly greater than 0 (Table 1 and Figure 5). In spite of the statistical insignificance of θ_λ between Clusters 1 and 2 (or 3), the θ_λ between Clusters 1 and 3 was greater than that between Clusters 1 and 2. Moreover, a significant functional divergence between Clusters 1 and 2 was detected in Data set 2 (Table 1 and Figure 6), indicating that sites in regions flanking the functional domain (i.e. the forkhead box) have evolved at heterogeneous rates after the split of Cluster1 and Cluster 2 (Information S5).

We plotted the posterior probability profiles of site-specific rate differences among clusters in Data set 1 and Data set 2 (Information S5). In Data set 1, the baseline posterior probability (i.e. posterior probability of the majority of sites) between Cluster 2 and Cluster 3 was the largest (0.683); the second largest was between Cluster 1 and Cluster 3 (0.565); the smallest was between Cluster 1 and Cluster 2 (0.260). Furthermore, there were 53, 2 and 0 amino acid sites with posterior probabilities above the cutoff value (0.70) when comparing Clusters 2 and 3, 1 and 3, and 1 and 2, respectively. The baseline posterior probability (0.384) when comparing Cluster 1 and Cluster 2 in Data set 2 was lower than 0.5, indicating no significant changes in the evolutionary rates of these sites occurred after the divergence of these two clusters. Ten amino acid sites with posterior probabilities greater than 0.7 were all located in exons (i.e. 3, 7, 11 and 17) surrounding functional domains, i.e. the zinc-finger, leucine-zipper and forkhead box.

Structural divergence among FoxP2, FoxP2a and FoxP2b

Analyses on both data sets showed that hierarchical dendrograms, based on four kinds of pairwise distances (i.e. City-block, Correlation, Cosine and Euclidean), yielded highly consistent inconsistency coefficients and clusters (Information S6). All cophenetic correlation coefficients in these dendrograms were greater than 0.95, indicating that the trees faithfully represented original pairwise distances among sequences.

According to the cutoff value (0.8) for inconsistency coefficients, the three clusters (Cluster 1, Cluster 2 and Cluster 3) were consistently separated from one another in the four hierarchical dendrograms from Data set 1 (Information S6), although the detailed classification was somewhat different. Specifically, FoxP2b of *Salmo salar* and *Gadus morhua* were singly classified in these dendrograms. FoxP2 of *Leucoraja erinacea* and *Lepisosteus oculatus* were grouped into Cluster 1, however FoxP2 of *Rana daunchina* was misclassified into Cluster 2 in the trees based on Correlation, Cosine and Euclidean pairwise distances. In addition, some links connecting dissimilar members (i.e. with inconsistency coefficients > 0.8) existed in these three clusters, indicating possible structural heterogeneity in the clusters. Notably, all dendrograms revealed that FoxP2a of Acanthopterygii were dissimilar to FoxP2a of *Salmo salar*, *Gadus morhua* and Otocephala. FoxP2a of *Takifugu rubripes* and *Tetraodon nigroviridis* showed significant dissimilarity from FoxP2a of other seven acanthopterygian fishes. Amphibian FoxP2 were possibly divergent from other tetrapod FoxP2. Subsequent stepwise discriminant analysis, which correctly classified high percentages of the original grouped cases (100%) and cross-validated grouped cases (97.2%), strongly demonstrated that FoxP2 of *Leucoraja erinacea*, *Lepisosteus oculatus* and *Rana daunchina* belonged to Cluster 1. Additionally, FoxP2b of *Salmo salar* and *Gadus morhua* were classified into Cluster 2 and Cluster 3, respectively. Moreover, FoxP2 in the ancestor of Osteichthyes and the ancestor of Actinopterygii, as well as all the ancestral FoxP2 in Sarcopterygii, were grouped into Cluster 1. FoxP2 in the ancestor of Teleostei, FoxP2b in the ancestor of Euteleostei and all the ancestral FoxP2a in Teleostei were classified as members of Cluster 2. All ancestral FoxP2b of Neoteleostei were classified into Cluster 3.

Similarly, analysis of structural divergence on Data set 2 produced consistent results (Information S6), classifying two clusters (Cluster 1 and Cluster 2). FoxP2 of *Leucoraja erinacea* and *Lepisosteus oculatus* were grouped into Cluster 1. The existence of a substantial inconsistency coefficient (> 0.8) in Cluster 2 indicated significant divergence between FoxP2a of Acanthopterygii and that of *Salmo salar*, *Gadus morhua* and Otocephala. FoxP2 of *Leucoraja erinacea* and *Lepisosteus oculatus* were confirmed to be members of Cluster 1 in subsequent stepwise discriminant analysis. In addition, FoxP2 in the ancestor of Osteichthyes and the ancestor of Actinopterygii were grouped together with all ancestral FoxP2

in Sarcopterygii into Cluster 1. All ancestral FoxP2a in Teleostei were classified into Cluster 2. Furthermore, 100% of the original grouped cases and cross-validated grouped cases were correctly classified.

Variations of thermodynamic stability in FoxP2, FoxP2a and FoxP2b

The simulated ADTE variations of monomers and homodimers in different evolutionary scenarios showed that such ADTE variations were not always in unison (Figure 7). In 53 of 78 pieces of evolutionary trajectories, both monomers and homodimers underwent insignificant ADTE changes or maintained stable ADTE values. Different ADTE variations between monomers and homodimers occurred in more than half of the rest (i.e. 14/25). Moreover, dramatic divergence of ADTE fluctuations in the three clusters (i.e. Cluster 1, Cluster 2 and Cluster 3) indicated that the thermodynamic stability of the forkhead domain experienced divergent variations in these clusters. In Cluster 1, the numbers of significant ADTE changes in monomers and homodimers were 0 and 4, respectively. Likewise in Cluster 2, the numbers were 5 and 6; in Cluster 3, the numbers were 9 and 9.

Variation of natural selection pressures on *FoxP2*, *FoxP2a* and *FoxP2b*

The comparison of Model 3 to Model 0 in both data sets revealed variable natural selection pressures across sites (Information S7). Notwithstanding the fact that no positively selected sites were detected in comparison of Model 2 with Model 1, 521S was found using a more robust comparison between Model 8 and Model 7. In addition, comparison of the free ratios model to Model 0 revealed significant heterogeneity of the ω ratio in different branches.

Subsequent analysis of branch specific models in Data set 1 demonstrated that Cluster 3 experienced divergent natural selection pressures against both Cluster 1 and Cluster 2 (Information S7). Specifically, ω ratios of Cluster 1 and Cluster 3 in the independent two-ratio model were 0.022 and 0.139, respectively. Similarly, the ω ratio of Cluster 3 (0.139) was greater than that of Cluster 2 (0.030) in their two-ratio model. The insignificantly better fit of the independent two-ratio model (ω ratios: 0.022 in Cluster 1, 0.029 in Cluster 2) than the interdependent one-ratio model (ω ratio: 0.026 in Cluster 1 and Cluster 2) showed no differences of selection pressures in Cluster 1 and Cluster 2. However, a significant difference in selection pressures between Cluster 1 and Cluster 2 was demonstrated in Data set 2 (Information S7). The ω ratio of Cluster 1 (0.046) was slightly smaller than that of Cluster 2 (0.063) in the independent two-ratio model.

Three foreground lineages (i.e. *FoxP2a* in ancestor of Tetraodontiformes, *FoxP2b* in ancestor Neoteleostei and *FoxP2b* in the ancestor of Tetraodontiformes) were found to have experienced accelerated evolution using Test 1 for branch-site models A in Data set 1 (Figure 5 and Information S7). Moreover, *FoxP2a* in the ancestor of Tetraodontiformes and *FoxP2b* in the ancestor of Neoteleostei probably underwent positive Darwinian selection. Most of the positively selected sites were situated surrounding the forkhead domain, specifically in exons 11, 12 and 15.

Likewise, five ancestral branches in Data set 2 (i.e. *FoxP2* in the ancestor of Actinopterygii, *FoxP2a* in the ancestor of Teleostei, *FoxP2a* in the ancestor of Cypriniformes, *FoxP2a* in the ancestor of Neoteleostei and *FoxP2a* in the ancestor of Tetraodontiformes) likely experienced accelerated evolution (Figure 6 and Information S7). Positive natural selection was detected in four ancestral branches, i.e. *FoxP2* in the ancestor of Actinopterygii, *FoxP2a* in the ancestor of Cypriniformes, *FoxP2a* in the ancestor of Neoteleostei and *FoxP2a* in the ancestor of Tetraodontiformes. The positively selected sites were mostly located in exons 7, 11, 15 and 17, flanking important functional domains (i.e. the zinc-finger, leucine-zipper and forkhead box). The test of selection pressures on 27 selected sequences in Data set 2, containing exons 6 and 16, produced similar results (not shown). These two exons with teleost-specific inserts were also demonstrated targets of positive natural selection and relaxed purifying selection.

Discussion

Duplication of FoxP2 in the 3R-WGD event

In this report, we identify the teleost-specific *FoxP2* duplication using synteny and paralogon analyses of related chromosomal segments. The subsequent phylogenetic reconstructions of *FoxP2*, *FoxP2a* and *FoxP2b* definitively confirm their homologous relationships. The orthologous one-to-two type relationships between *FoxP2* in tetrapods and *FoxP2a* and *FoxP2b* in teleosts are now established. The comparatively variable linearity of genes flanking *FoxP2a* and *FoxP2b* in teleosts probably results from multiple evolutionary changes (inversion, deletion and translocation) in these chromosomal segments after the tetrapod-teleost split. Although *FoxP2a* and *FoxP2b* show remarkable divergence in teleosts, the phylogenetic trees reconstructed with four member genes (i.e. *FoxP1*, *FoxP2*, *FoxP3* and *FoxP4*) of the *FoxP* subfamily clearly demonstrate that *FoxP2a* and *FoxP2b* in teleosts are members of *FoxP2* lineage (not shown). In comparison with *FoxP2*, greater divergence (e.g. exon-composition) of *FoxP2b* than *FoxP2a* suggests that *FoxP2b* evolves faster than *FoxP2a* after the gene duplication. Although dimerization of FoxP1, FoxP2 and FoxP4 has been demonstrated, which is important for their DNA binding and transcriptional activity [55], it is not clear whether FoxP2a and FoxP2b interact with one another in teleosts, leading to peculiar characteristics of exons flanking important functional domains in FoxP2a and FoxP2b.

The absence of *FoxP2b* in *Danio rerio* probably results from pseudogenization [26] of the redundant copy after the split of Otocephala and Euteleostei (Figure 1B). In addition to *FoxP2a* of *Danio rerio*, *FoxP2a* of two other species in Otocephala (*Ctenopharyngodon idella* and *Pygocentrus nattereri*) are phylogenetically determined. However, deletion of *FoxP2b* in *Ctenopharyngodon idella* and *Pygocentrus nattereri* is uncertain because their genomes are not available. The retention of *FoxP2b* in Euteleostei implies that the gene may conserve some functions of the parental *FoxP2* through

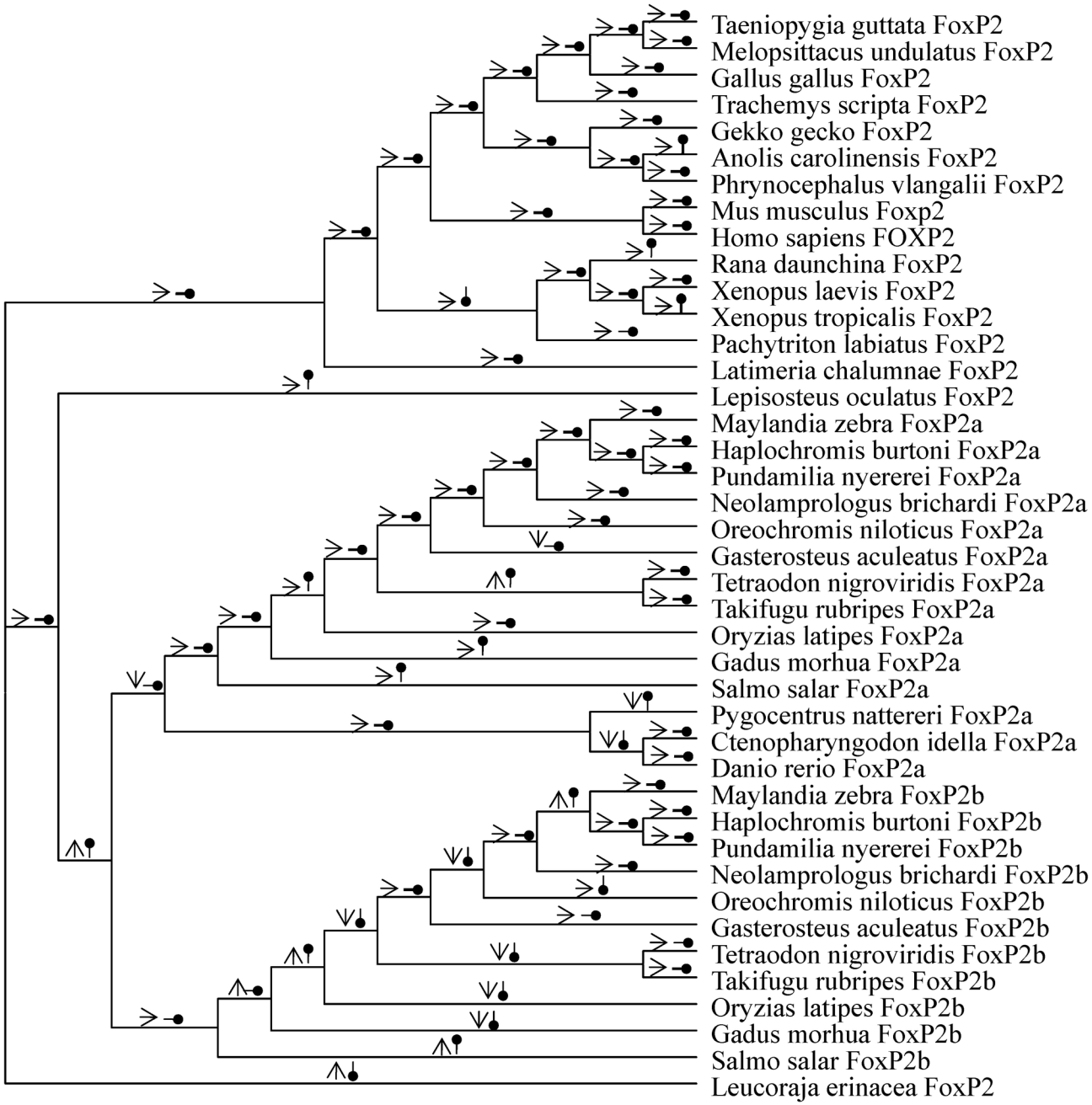

Figure 7

Figure 7. The thermodynamic stability changes of monomers (arrows with open arrowheads) and homodimers (arrows with circled arrowheads) based on average differences in total Gibbs free energy (ADTE). Rightward pointing arrowheads indicate insignificant thermodynamic stability changes. Upward and downward pointing arrowheads represent significant destabilizing and stabilizing thermodynamic stability changes, respectively.

doi: 10.1371/journal.pone.0083858.g007

subfunctionalization or may gain novel functions through neofunctionalization. The linked paralogon shared by *FoxP2a* and *FoxP2b* in teleosts indicates that these two duplicated genes were generated in a large-scale genome duplication event, i.e. 3R-WGD event [27,36] (Figure 2B). In addition, it should be noted that extra duplication of *FoxP2* may occur in polyploid vertebrates, such as *Xenopus laevis* and *Salmo salar* [56].

Rapid evolution of *FoxP2a* and *FoxP2b* after the gene duplication

The significant functional divergence between FoxP2a and FoxP2b in Data set 1 implies that FoxP2a and FoxP2b probably execute divergent functions in teleosts. The θ_λ between Cluster 1 and Cluster 3 is greater than that between Cluster 1 and Cluster 2, suggesting that FoxP2b has evolved at a faster rate than FoxP2a after the gene duplication event. In addition, significant functional divergence of Cluster 1 and Cluster 2 in Data set 2 shows that the regions surrounding the functional domain (i.e. the forkhead box) rather than the functional domain itself in FoxP2a and FoxP2 are targets of divergent evolutionary rates. Therefore, the functions of *FoxP2a* may be more conservative than those of *FoxP2b* in teleosts.

Hierarchical cluster analysis and stepwise discriminant analysis of hydrophobicity profiles in Data sets 1 and 2 suggest that the structures of FoxP2, FoxP2a and FoxP2b are remarkably divergent from one another. In the stepwise discriminant analysis, FoxP2b of *Salmo salar* groups together with FoxP2a, as well as FoxP2 in the ancestor of Teleostei and FoxP2b in the ancestor of Euteleostei together with all ancestral FoxP2a in Teleostei, implying that structural divergence between FoxP2a and FoxP2b was probably very weak at the early stage of post-duplication. Perhaps the weak structural divergence of FoxP2a and FoxP2b generated functional redundancy of FoxP2b in Otocephala, further leading to the lineage-specific loss of FoxP2b. The significant structural divergence between FoxP2a and FoxP2b seems to coincide with the split of Neoteleostei from Euteleostei, implying more divergent functions between the genes of Neoteleostei than Salmoniformes. Additionally, some teleost lineages show significant FoxP2a structural divergence, likely indicating divergent functional adaptations in the lineages.

The significantly divergent variations of thermodynamic stability among FoxP2a, FoxP2b and FoxP2 clusters imply divergent functional adaptations of the forkhead domains in FoxP2a, FoxP2b and FoxP2. Intriguingly, the significant destabilizing effects of mutations (E519D and R523M) on both monomers and homodimers suggest that functional divergence of FoxP2 in teleosts possibly began in the pre-duplication phase. It has been demonstrated that the two mutations, as well as 521S, are responsible for the weak repressive activity of the medaka FoxP2a on the lung-specific gene of *CC10* [19]. In addition, predicted mutations in FoxP2b of the ancestor of Neoteleostei (S520A and S532N) and those (M518I, D522E and M523H) in FoxP2a of the ancestor of Tetraodontiformes may have led to significant destabilizing effects. Although the correlation between stability variation and molecular divergence is not clear, the simulated fluctuation of thermodynamic stability in each evolutionary scenario theoretically reveals a history of buffering and compensatory trade-offs in protein stability [57]. The theoretical history, therefore, suggest the nature of evolutionary changes in the forkhead box which have stability effects thereby aiding our investigation of the functional divergence of related mutations.

Our tests of natural selection pressures based on branch specific models strongly suggest the *FoxP2b* lineage has undergone a generally relaxed purifying selection pressure, which is significantly greater than that of the *FoxP2a* and *FoxP2* lineages. The regions flanking the functional domain of the forkhead box, rather than the forkhead domain per se, in the *FoxP2a* lineage has experienced a milder purifying selection pressure than those in the *FoxP2* lineage. Thus the asymmetrical natural selection pressures acting on the *FoxP2a*, *FoxP2b* and *FoxP2* clusters, particularly between *FoxP2a* and *FoxP2b*, likely led to asymmetrical functional divergence among these clusters. Furthermore, natural selection pressures on branches and sites in each cluster show significant heterogeneity, especially positive natural selection acting on several ancestral branches. Notably, positive selection pressure mainly acts on regions flanking functional domains. Therefore, positive selection pressure possibly was the driving force in forming the teleost-specific inserts in FoxP2a and the truncated regions in FoxP2b. It is also notable that *FoxP2* in the ancestor of Actinopterygii probably underwent positive selection, suggesting that positive selection played a role in variation during the pre-duplication phase [58]. The potential pseudogenization and subsequent loss of *FoxP2b* in Otocephala possibly resulted from excessively relaxed purifying selection on the redundant copy in the fixation phase [26,58]. Positive Darwinian selection acts on *FoxP2a* and *FoxP2b* after Neoteleostei branched off from Euteleostei, implying adaptive evolution of these two genes coincided with diversification of Neoteleostei.

Insights into the functional adaptations of *FoxP2a* and *FoxP2b*

Studies of FoxP2 in both animal models and humans have shown that the pleiotropic gene is widely involved in a series of important physiological activities and developmental processes [1,4,17]. In addition to the neural control of vocalization, it has been demonstrated that Foxp2 is a crucial regulator in *Mus musculus* for lung and esophageal development, especially for postnatal lung alveolarization [12,22,59]. These findings indicate that *FoxP2* is critically involved in the development of foregut derived pulmonary organ systems through regulating lung-specific genes, e.g. *CC10* and surfactant protein C (SPC).

The teleost swim bladder has been considered homologous to the lung on the basis of consistency of developmental blastema and molecular pathways [28,29]. It has been suggested that the surfactant systems essential for air filled organs are also homologous in gas bladders and lungs [60]. However, the functions of the swim bladder and lung are markedly different from each other. The swim bladder has evolved to be an organ mainly controlling buoyancy regulation not oxygen respiration [29]. Studies of *FoxP2a* in teleosts have

provided only limited information concerning its functions for organ development [18,19,21]. Fishes share a conservative expression pattern of *FoxP2a* in development of CNS with other vertebrates. No *FoxP2a* expression has been detected in the swim bladder of adult *Danio rerio* [29], whereas the area from which the swim bladder is derived shows high *FoxP2a* expression in early (7 days post-fertilization) embryonic development in *Maylandia zebra* [20]. Expression of *FoxP2a* decreases to a light but visible level in the swim bladder of late larval (20 days post-fertilization) embryos of *Mchenga conophorus* [20] and possibly disappears before adulthood, as is the case in *Danio rerio* [29].

Given the essential functions of *FoxP2* in foregut derived air-breathing organ, rapid evolution of *FoxP2* (*FoxP2a* and *FoxP2b*) in teleosts is possibly correlated with adaptive evolution of the swim bladder [19]. Notably *Lepisosteus oculatus* retains a swim bladder to breathe air and possesses *FoxP2* resembling that of tetrapods. The three species of Otocephala (*Danio rerio*, *Ctenopharyngodon idella* and *Pygocentrus nattereri*) and one species of Protacanthopterygii (*Salmo salar*) are physostomous fishes with a pneumatic duct connecting the swim bladder to the gastrointestine. It has been suggested that the Otocephala-specific loss of *FoxP2b* possibly results from excessively relaxed purifying selection on the functionally redundant gene. In spite of retention of *FoxP2b* in *Salmo salar*, the divergence of *FoxP2a* and *FoxP2b* is noticeably weak. Positive natural selection probably acts on *FoxP2a* and *FoxP2b* to favor the divergence of the genes in Neoteleostei, which are physoclistous fishes (i.e. swim bladders of these fishes have lost a pneumatic duct). Taken together, the present study provides a useful introduction to studying the functional involvement of *FoxP2a* and *FoxP2b* in teleosts.

Supporting Information

Information S1. Species coverage, PCR primers and predicted FoxP2 sequences. (PDF)

Information S2. Protein-coding genes around *FoxP2*, *FoxP2a* or *FoxP2b* in eight species. Twenty upstream (with negative order number) and 20 downstream (with positive order number) protein-coding genes.
(PDF)

Information S3. Amino acid or fragment inserts in FoxP2 and FoxP2a of ray-finned fishes. (PDF)

information S4. Phylogenetic reconstructions based on Data sets 1 and 2. (PDF)

Information S5. Posterior probability profiles of site-specific rate difference among clusters in Data sets 1 and 2. (PDF)

Information S6. Hierarchical dendrograms of genes in Data sets 1 and 2. (PDF)

Information S7. Test of natural selection pressures on Data sets 1 and 2. (PDF)

Acknowledgements

We thank Dr. Steven E. Brauth of Department of Psychology, University of Maryland and Ms. Megan N. Li for helpful comments and critical reading of the manuscript; All referees for their thoroughgoing reviews and improving the quality of the paper; Drs. Jianguo Cui, Li Ding and Yin Qi for providing experimental animals; Ting Zhou, Guian Huang and Jianhui Yue for help with molecular biological experiments.

Author Contributions

Conceived and designed the experiments: XWS. Performed the experiments: XWS. Analyzed the data: XWS. Wrote the manuscript: XWS YJW YZT.

References

1. Takahashi H, Takahashi K, Liu FC (2009) FOXP genes, neural development, speech and language disorders. Adv Exp Med Biol 665: 117-129. PubMed: 20429420.
2. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413: 519-523. doi:10.1038/35097076. PubMed: 11586359.
3. Fisher SE, Marcus GF (2006) The eloquent ape: genes, brains and the evolution of language. Nat Rev Genet 7: 9-20. doi:10.1038/nrn1835. PubMed: 16369568.
4. Fisher SE, Scharff C (2009) FOXP2 as a molecular window into speech and language. Trends Genet 25: 166-177. doi:10.1016/j.tig.2009.03.002. PubMed: 19304338.
5. Haesler S, Rochefort C, Georgi B, Licznerski P, Osten P et al. (2007) Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. PLoS Biol 5: 2885-2897. PubMed: 18052609.
6. Groszer M, Keays DA, Deacon RM, de Bono JP, Prasad-Mulcare S et al. (2008) Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. Curr Biol 18: 354-362. doi:10.1016/j.cub.2008.01.060. PubMed: 18328704.
7. Li G, Wang J, Rossiter SJ, Jones G, Zhang S (2007) Accelerated FoxP2 evolution in echolocating bats. PLOS ONE 2: e900. doi:10.1371/journal.pone.0000900. PubMed: 17878935.
8. Gaub S, Groszer M, Fisher SE, Ehret G (2010) The structure of innate vocalizations in Foxp2-deficient mouse pups. Genes Brain Behav 9: 390-401. doi:10.1111/j.1601-183X.2010.00570.x. PubMed: 20132318.
9. Lai CS, Gerrelli D, Monaco AP, Fisher SE, Copp AJ (2003) FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. Brain 126: 2455-2462. doi:10.1093/brain/awg247. PubMed: 12876151.
10. Haesler S, Wada K, Nshdejan A, Morrisey EE, Lints T et al. (2004) FoxP2 expression in avian vocal learners and non-learners. J Neurosci 24: 3164-3175. doi:10.1523/JNEUROSCI.4369-03.2004. PubMed: 15056696.
11. Schön C, Wochnik A, Rössner A, Donow C, Knöchel W (2006) The FoxP subclass in Xenopus laevis development. Dev Genes Evol 216: 641-646. doi:10.1007/s00427-006-0073-8. PubMed: 16609867.

12. Shu W, Lu MM, Zhang Y, Tucker PW, Zhou D et al. (2007) Foxp2 and Foxp1 cooperatively regulate lung and esophagus development. Development 134: 1991-2000. doi:10.1242/dev.02846. PubMed: 17428829.
13. Spiteri E, Konopka G, Coppola G, Bomar J, Oldham M et al. (2007) Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. Am J Hum Genet 81: 1144-1157. doi:10.1086/522237. PubMed: 17999357.
14. Bonkowsky JL, Wang X, Fujimoto E, Lee JE, Chien CB et al. (2008) Domain-specific regulation of foxP2 CNS expression by lef1. BMC Dev Biol 8: 103. doi:10.1186/1471-213X-8-103. PubMed: 18950487.
15. Schulz SB, Haesler S, Scharff C, Rochefort C (2010) Knockdown of FoxP2 alters spine density in Area X of the zebra finch. Genes Brain Behav 9: 732-740. doi:10.1111/j.1601-183X.2010.00607.x. PubMed: 20528955.
16. Enard W, Gehre S, Hammerschmidt K, Hölter SM, Blass T et al. (2009) A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell 137: 961-971. doi:10.1016/j.cell.2009.03.041. PubMed: 19490899.
17. Scharff C, Petri J (2011) Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. Philos Trans R Soc Lond B Biol Sci 366: 2124-2140. doi:10.1098/rstb.2011.0001. PubMed: 21690130.
18. Bonkowsky JL, Chien CB (2005) Molecular cloning and developmental expression of foxP2 in zebrafish. Dev Dyn 234: 740-746. doi:10.1002/dvdy.20504. PubMed: 16028276.
19. Itakura T, Chandra A, Yang Z, Xue XD, Wang B et al. (2008) The medaka FoxP2, a homologue of human language gene FOXP2, has a diverged structure and function. J Biochem 143: 407-416. PubMed: 18079162.
20. Norsworthy M (2011) Novel expression sites and genetic diversity of FoxP2 in Lake Malawi cichlids.
21. Shah R, Medina-Martinez O, Chu LF, Samaco RC, Jamrich M (2006) Expression of FoxP2 during zebrafish development and in the adult brain. Int J Dev Biol 50: 435-438. doi:10.1387/ijdb.052065rs. PubMed: 16525940.
22. Shu W, Yang H, Zhang L, Lu MM, Morrisey EE (2001) Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. J Biol Chem 276: 27488-27497. doi:10.1074/jbc.M100636200. PubMed: 11358962.
23. Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V et al. (2002) Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418: 869-872. doi:10.1038/nature01025. PubMed: 12192408.
24. Zhang J, Webb DM, Podlaha O (2002) Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. Genetics 162: 1825-1835. PubMed: 12524352.
25. Webb DM, Zhang J (2005) FoxP2 in song-learning birds and vocal-learning mammals. J Hered 96: 212-216. doi:10.1093/jhered/esi025. PubMed: 15618302.
26. Zhang J (2003) Evolution by gene duplication: an update. Trends Ecol Evol 18: 292-298. doi:10.1016/S0169-5347(03)00033-8.
27. Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. Nat Rev Genet 10: 725-732. doi:10.1038/nrg2600. PubMed: 19652647.
28. Yin A, Korzh S, Winata CL, Korzh V, Gong Z (2011) Wnt Signaling Is Required for Early Development of Zebrafish Swimbladder. PLOS ONE 6: e18431. doi:10.1371/journal.pone.0018431. PubMed: 21479192.
29. Zheng W, Wang Z, Collins JE, Andrews RM, Stemple D et al. (2011) Comparative Transcriptome Analyses Indicate Molecular Homology of Zebrafish Swimbladder and Mammalian. Lung - PLOS ONE 6: e24019. doi:10.1371/journal.pone.0024019.
30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410. doi:10.1016/S0022-2836(05)80360-2. PubMed: 2231712.
31. Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12: 656-664. doi:10.1101/gr.229202. Article published online before March 2002 PubMed: 11932250
32. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402. doi:10.1093/nar/25.17.3389. PubMed: 9254694.
33. Venkatesh B, Kirkness EF, Loh YH, Halpern AL, Lee AP et al. (2007) Survey sequencing and comparative analysis of the elephant shark (Callorhinchus milii). Genome - PLOS Biol 5: e101.
34. Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic. DNA - J Mol Biol 268: 78-94. doi:10.1006/jmbi.1997.0951.
35. Frazer KA, Elnitski L, Church DM, Dubchak I, Hardison RC (2003) Cross-species sequence comparisons: a review of methods and available resources. Genome Res 13: 1-12. doi:10.1101/gr.222003. PubMed: 12529301.
36. Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3: e314. doi:10.1371/journal.pbio.0030314. PubMed: 16128622.
37. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R et al. (2009) EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res 19: 327-335. PubMed: 19029536.
38. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680. doi:10.1093/nar/22.22.4673. PubMed: 7984417.
39. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95-98.
40. Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. J Hered 92: 371-373. doi:10.1093/jhered/92.4.371. PubMed: 11535656.
41. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425. PubMed: 3447015.
42. Tamura K, Peterson D, Peterson N, Stecher G, Nei M et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739. doi:10.1093/molbev/msr121. PubMed: 21546353.
43. Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci U S A 101: 11030-11035. doi:10.1073/pnas.0404206101. PubMed: 15258291.
44. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8: 275-282. PubMed: 1633570.
45. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution: 783-791.
46. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574. doi:10.1093/bioinformatics/btg180. PubMed: 12912839.
47. Gu X (1999) Statistical methods for testing functional divergence after gene duplication. Mol Biol Evol 16: 1664-1674. doi:10.1093/oxfordjournals.molbev.a026080. PubMed: 10605109.
48. Wood TC, Pearson WR (1999) Evolution of protein sequences and structures. J Mol Biol 291: 977-995. doi:10.1006/jmbi.1999.2972. PubMed: 10452901.
49. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105-132. doi:10.1016/0022-2836(82)90515-0. PubMed: 7108955.
50. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586-1591. doi:10.1093/molbev/msm088. PubMed: 17483113.
51. Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320: 369-387. doi:10.1016/S0022-2836(02)00442-4. PubMed: 12079393.
52. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F et al. (2005) The FoldX web server: an online force field. Nucleic Acids Res 33: W382-W388. doi:10.1093/nar/gki387. PubMed: 15980494.
53. Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22: 1107-1118. doi:10.1093/molbev/msi097. PubMed: 15689528.
54. Nelson JS (2006) Fishes of the World. Wiley.
55. Li S, Weidenfeld J, Morrisey EE (2004) Transcriptional and DNA binding activity of the Foxp1/2/4 family is modulated by heterotypic and homotypic protein interactions. Mol Cell Biol 24: 809-822. doi:10.1128/MCB.24.2.809-822.2004. PubMed: 14701752.
56. Gregory TR (2005) The evolution of the genome. Elsevier Academic Press.
57. Tokuriki N, Tawfik DS (2009) Stability effects of mutations and protein evolvability. Curr Opin Struct Biol 19: 596-604. doi:10.1016/j.sbi.2009.08.003. PubMed: 19765975.
58. Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11: 97-108. PubMed: 20051986.
59. Yang Z, Hikosaka K, Sharkar MT, Tamakoshi T, Chandra A et al. (2010) The mouse forkhead gene Foxp2 modulates expression of the

lung genes. Life Sci 87: 17-25. doi:10.1016/j.lfs.2010.05.009. PubMed: 20553735.
60. Daniels CB, Orgeig S, Sullivan LC, Ling N, Bennett MB et al. (2004) The Origin and Evolution of the Surfactant System in Fish: Insights into the Evolution of Lungs and Swim Bladders. Physiol Biochem Zool 77: 732-749. doi:10.1086/422058. PubMed: 15547792.