Heredity (2013) 111, 57–65
© 2013 Macmillan Publishers Limited All rights reserved 0018-067X/13
www.nature.com/hdy

ORIGINAL ARTICLE

# The discovery of *Foxl2* paralogs in chondrichthyan, coelacanth and tetrapod genomes reveals an ancient duplication in vertebrates

MT Geraldo[1], GT Valente[1], ASK Braz[2] and C Martins[1]

The *Foxl2* (*forkhead box L2*) gene is an important member of the forkhead domain family, primarily responsible for the development of ovaries during female sex differentiation. The evolutionary studies conducted previously considered the presence of paralog *Foxl2* copies only in teleosts. However, to search for possible paralog copies in other groups of vertebrates and ensure that all predicted copies were homolog to the *Foxl2* gene, a broad evolutionary analysis was performed, based on the forkhead domain family. A total of 2464 sequences for the forkhead domain were recovered, and subsequently, 64 representative sequences for *Foxl2* were used in the evolutionary analysis of this gene. The most important contribution of this study was the discovery of a new subgroup of *Foxl2* copies (ortholog to *Foxl2B*) present in the chondrichthyan *Callorhinchus milii,* in the coelacanth *Latimeria chalumnae,* in the avian *Taeniopygia guttata* and in the marsupial *Monodelphis domestica*. This new scenario indicates a gene duplication event in an ancestor of gnathostomes. Furthermore, based on the analysis of the syntenic regions of both *Foxl2* copies, the duplication event was not exclusive to *Foxl2*. Moreover, the duplicated copy distribution was shown to be complex across vertebrates, especially in tetrapods, and the results strongly support a loss of this copy in eutherian species. Finally, the scenario observed in this study suggests an update for *Foxl2* gene nomenclature, extending the actual suggested teleost naming of *Foxl2A* and *Foxl2B* to all vertebrate sequences and contributing to the establishment of a new evolutionary context for the *Foxl2* gene.

*Heredity* (2013) **111,** 57–65; doi:10.1038/hdy.2013.19; published online 3 April 2013

**Keywords:** *forkhead box L2*; vertebrate genome; paralog copies; gene duplication

## INTRODUCTION

The forkhead domain, also called the winged helix domain, is present in a large number of proteins that constitute a family of transcription factors that activate important pathways of embryogenesis (Weigel *et al.*, 1989; Mahlapuu *et al.*, 2001; Lehmann *et al.*, 2003) and cell differentiation in eukaryotes (Brissette *et al.*, 1996; Dottori *et al.*, 2001; Nakae *et al.*, 2003). This domain, determined by the Pfam database (Finn *et al.*, 2010) as including 96 amino acids, was named after the description of the crystal structure of the hepatocyte nuclear factor-3$\gamma$ (HNF-3$\gamma$)/FOXA family, which is involved in pancreas and liver tissue development (Clark *et al.*, 1993). Many proteins of the forkhead family are considered to be tissue-specific regulators of development (Bravieri *et al.*, 1997). Examples include craniopharyngeal development—FOXE1 (Lehmann *et al.*, 2003), cell growth and insulin responsiveness—FOXO1 (Gross *et al.*, 2008), hair formation and keratinocyte differentiation—FOXN1 (Nehls *et al.*, 1994) and ovarian formation and function—FOXL2 (Cocquet *et al.*, 2002; Baron *et al.*, 2004; Uhlenhaut and Treier, 2006; Veitia, 2010; Jaubert *et al.*, 2011). The role of each protein member of the domain family was discovered primarily by studies of congenital defects associated with mutations, often observed within the forkhead domain (Benayoun *et al.*, 2011).

The *FOXL2* (*forkhead box L2*) gene is an important member of this extensive family and is primarily responsible for ovarian development and maturation (Uhlenhaut *et al.*, 2009). The majority of studies involving *FOXL2* describe functional differences of mutations associated with blepharophimosis–ptosis–epicanthus inversus syndrome (BPES), which is characterized by eyelid malformations (BPES type II) or, in some cases, premature ovarian failure (BPES type I) (Zlotogora *et al.*, 1983; Crisponi *et al.*, 2001; De Baere *et al.*, 2001). An exclusive feature of the protein, observed only in mammals, is a tract of 14 alanines, which has been suggested to be a region with strong functional constraints (Cocquet *et al.*, 2003). For instance, expansions of alanines represent ∼30% of the causes of BPES type II (De Baere *et al.*, 2003).

Although there are vast amounts of data available from clinical and expression analysis of *Foxl2*, only a few studies, restricted to a reduced set of vertebrate species from teleosts to mammals, are focused on the evolution of the gene. To date, the most interesting observation in the evolutionary analysis is the presence of paralogs in the fish group, the origin of which was suggested to be in accordance with the fish-specific whole genome duplication (Christoffels *et al.*, 2004; Jaillon *et al.*, 2004). An extra copy of the *Foxl2* gene has been reported in *Oncorhynchus mykiss* (rainbow trout), *Takifugu rubripes* (fugu), *Tetraodon nigroviridis* (pufferfish) (Baron *et al.*, 2004), *Danio rerio* (zebrafish), *Gasterosteus aculeatus* (stickleback), *Oryzias latipes* (medaka) (Jiang *et al.*, 2011) and *Salmo salar* (Atlantic salmon) (von Schalburg *et al.*, 2011).

[1]Integrative Genomics Laboratory, Department of Morphology, Institute of Biosciences, Sao Paulo State University–UNESP, Botucatu, Sao Paulo, Brazil and [2]Laboratory of Computational Biology and Bioinformatics, Center of Natural and Human Sciences, Federal University of ABC–UFABC, Santo Andre, Sao Paulo, Brazil
Correspondence: Professor C Martins, Department of Morphology, Institute of Biosciences, UNESP–Sao Paulo State University, Botucatu, Sao Paulo 18618-970, Brazil.
E-mail: cmartins@ibb.unesp.br

Received 20 September 2012; revised 29 January 2013; accepted 20 February 2013; published online 3 April 2013

Motivated by the previous evolutionary context of *Foxl2*, the present work was conducted to improve the robustness of *Foxl2* evolutionary history, thus including a large number of sequences and also searching for new copies of *Foxl2* in vertebrates. Evolutionary descriptions of new *Foxl2* single and duplicated copies were obtained for different groups of vertebrates, including the chondrichthyan *Callorhinchus milii* (elephant shark), the coelacanth *Latimeria chalumnae* (Comoros coelacanth), the bird *Taeniopygia guttata* (zebra finch) and the marsupial *Monodelphis domestica* (opossum). Moreover, single copies of *Foxl2* from the neotropical cichlid species *Cichla monoculus* (peacock bass) and the chondrichthyan species *Rhizoprionodon lalandei* (Brazilian sharpnose shark) and *Callorhinchus callorynchus* (Plownose chimaera) were sequenced. Furthermore, the syntenic region analyses of both *Foxl2* copies in different vertebrate species were discussed together with the phylogenetic results to support the present conclusions about the duplication event of *Foxl2*.

Contrary to the previous report on the origin of paralog copies of *Foxl2*, this work shows an ancient origin of the paralog copies in ancestors of gnathostomes, leading to a suggestion for a new nomenclature for those genes, namely, *Foxl2A* (the most studied gene form, generally known as *Foxl2*) and *Foxl2B* (for the diverged duplicated form) in vertebrates.

## MATERIALS AND METHODS

The methodology adopted in the present work is summarized in Figure 1.

### Sequence acquisition and alignment procedures

*Forkhead domain sequences.* Protein sequences of the forkhead family were collected from a large number of eukaryotes including fungi, plants and animals. A complete list of Uniprot identifiers (converted to GenBank accession numbers in the Uniprot browser, http://www.uniprot.org), available in the Pfam database (Finn *et al.*, 2010) for the forkhead family, was used to retrieve the corresponding nucleotide and amino-acid sequences utilizing the GBreader software (Razente HL, Braz ASK, Scott LPB (unpublished)). This procedure allowed for the acquisition of 2464 sequences of all available forkhead family proteins in GenBank (Supplementary Material 1).

To avoid redundancy due to highly similar sequences, and thus ensure the use of only a set of representative data, a clustering methodology was performed using the CD-HIT software (Li and Godzik, 2006). This program clustered proteins with identity $\geqslant$90% (the threshold used in the present work) and retrieved only one representative protein from each cluster. This threshold value was applied because the forkhead family exhibits a high conservation level in the domain region and, thus, lower values would indicate the loss of important representative data. The previous analysis resulted in a set of sequences (1023 amino-acid sequences relative to 370 eukaryotic species) that were aligned with the HMMER v. 3 software (Finn *et al.*, 2011) using a Hidden Markov Model (HMM) profile of the forkhead domain. The final alignment was edited in GeneDoc (Nicholas *et al.*, 1997) software to select the conserved regions of the domain and avoid poorly aligned portions (final alignment length: 111 sites; see Supplementary Material 2).

*Foxl2 sequences.* Because our focus is on Foxl2, besides the sequences presented in the aforementioned data set, other Foxl2 sequences were obtained from the Elephant Shark Genome Project (http://esharkgenome.imcb.a-star.edu.sg/), Ensembl (Flicek *et al.*, 2011), BouillaBase (http://BouillaBase.org) and JGI (http://genome.jgi.doe.gov/) databases (access from December 2011 to March 2012). Only full-length sequences annotated as FOXL2 were retrieved in the first search; however, deep searches were also conducted in all metazoan genomes available in these databases to detect additional, unreported duplicated copies. A tblastn search was performed using Foxl2 sequences from species closely related to the target species as queries. Because the resulting hits were centered only in the forkhead domain, the retrieval was expanded across the hit to obtain the complete *Foxl2* coding sequence. These sequences were submitted to an ORF (open reading frame) search and protein translation using Geneious 4.8.5 software (Drummond *et al.*, 2009). Moreover, other Foxl2 sequences were retrieved by a blastp search directly from the GenBank database, available at the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/). A complete list of Foxl2-retrieved sequences and details of the database searches are shown in Supplementary Material 3.

Experimental procedures were also employed to retrieve *Foxl2* sequences for three organisms, the chondrichthyan species *R. lalandei* and *C. calloryncus*, and the neotropical cichlid *C. monoculus*. The animals were collected from Brazilian rivers and marine areas according to Brazilian laws for environmental protection (wild collection permit, SISBIO 12337-1, 15729-1). The experimental research on the animals was conducted according to the international guidelines of Sao Paulo State University (Protocol no. 34/08—CEEA/IBB/UNESP). The tissue samples were available at the Laboratory of Integrative Genomics of Sao Paulo State University, and the genomic DNA was extracted from muscle and liver tissues using the phenol–chloroform method (Sambrook and Russel, 2001). The *Foxl2* sequences were amplified by PCR using degenerate primers (forward 5′-GTNGCNYTNATHGCNATGGC-3′ and reverse 5′-CCARTANSWRCARTGCATCAT-3′) constructed with the Primer3-Plus software (Untergasser *et al.*, 2007). For the construction of these primers, several Foxl2 protein sequences of vertebrates (Supplementary Material 4) were retrieved based on a blastp search in the Expasy Proteomic Server (http://ca.expasy.org/) using a Foxl2 sequence from *Oreochromis niloticus* (accession number Q6JA05) as the query. These sequences were aligned using ClustalW (Thompson *et al.*, 1994), and the primers were constructed for the most conserved regions. The PCR cycling sequence was as follows: 1: 95 °C for 2 min; 2: 95 °C for 1 min; 3: 55 °C for 30 s; 4: 72 °C for 1 min; 5: steps 2–4 for 30 cycles; and 6: 72 °C for 5 min. The amplicons were ~250 bp, and the full gene sequences were obtained by the genome walking technique using Genome Walker kit (Clontech Laboratories, Mountain View, CA, USA) according to the manufacturer's protocol. All amplicons were cloned in p-GEM-T plasmid vectors (Promega Corporation, Madison, WI, USA), and the corresponding clones were sequenced with an ABI Prism 3100 DNA sequencer (Perkin-Elmer, Waltham, MA, USA) using ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction kits (Perkin-Elmer). Both strands from different clones were sequenced at least three times to ensure the sequencing quality. Finally, each of these sequences had their amino-acid sequences predicted by an ORF search performed using the Geneious 4.8.5 program (Drummond *et al.*, 2009). These sequences were also included in the data set for forkhead phylogeny.

The full Foxl2 amino-acid sequences present in the forkhead data set used in the domain phylogeny (Supplementary Material 5), and their clustered full

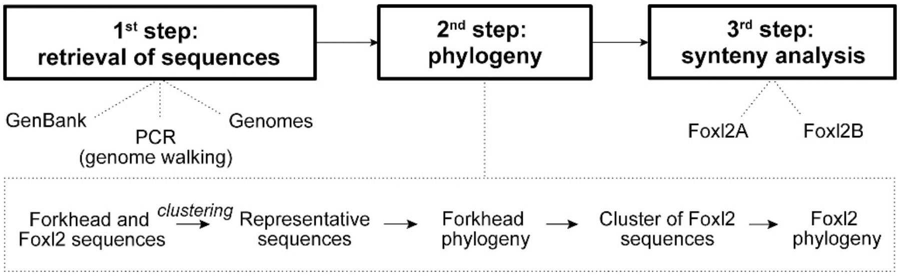

**Figure 1** Fluxogram of the methodology adopted in the present work. The general steps performed for the evolutionary analysis of *Foxl2* gene.

sequences, were used for the subsequent Foxl2 phylogeny reconstruction. The amino-acid sequences were aligned using the Muscle algorithm (Edgar, 2004), and obviously misplaced aligned regions were corrected manually. Poor-quality regions observed in the N-terminal portion were excluded from the analysis using the Seaview software (Supplementary Material 6). At the end, the reference data set for Foxl2 phylogeny was composed of 64 amino-acid sequences relevant to 50 metazoan species, and the alignment length included 392 sites.

### Forkhead and Foxl2 phylogenetic analysis

The choice of the best-fit model of evolution was performed with ProtTest3 (Darriba *et al*., 2011) for the forkhead and Foxl2 sequences using the Akaike information criterion (Akaike, 1974) for the best model selection.

The phylogenetic reconstruction was determined by maximum likelihood and Bayesian methods implemented in the Phyml v3.0.1 (Guindon and Gascuel, 2003) and Beast v1.7.0 (Drummond and Rambaut, 2007) software, respectively. The Bayesian analyses were conducted with Beast software allocated in CIPRES Science Gateway (Miller *et al*., 2010). For maximum likelihood analysis, the approximate likelihood ratio test (Shimodaira-Hasegawa-like) reliability test (Anisimova and Gascuel, 2006) was adopted, and the values were supported by posterior probabilities obtained by Bayesian analysis. For Bayesian method generations, the burn-in was determined in Tracer (Rambaut and Drummond, 2007) through log likelihood scores, and data were summarized in TreeAnnotator (Drummond and Rambaut, 2007) after trees that were out of the convergence area had been discarded. The visualization and the final tree edition were performed using FigTree v1.3.1 (Drummond and Rambaut, 2007) and the software package Mesquite (Maddison and Maddison, 2001). In all phylogenetic analyses, the proportion of invariable sites and $\gamma$-distributed rate variation across sites were estimated, and the substitution rate categories were set in four categories. The random local clock model for Bayesian analysis was chosen because it allows for rate variation among lineages, and the model can run a series of local molecular clocks as described in Drummond and Suchard (2010) (see parameters in Tables 1 and 2).

### Evaluation of the *Foxl2* syntenic region

In an attempt to provide additional support for the orthology for most of the newly described *Foxl2* sequences, especially the coelacanth *L. chalumnae*, the avian *T. guttata* and the marsupial *M. domestica* copies, and to understand the composition and organization of the genes in the proximity of *Foxl2*, the syntenic regions were analyzed in *D. rerio*, *G. aculeatus*, *G. morhua*, *O. latipes*, *O. niloticus*, *T. nigroviridis*, *T. rubripes*, *L. chalumnae*, *T. guttata* and *M. domestica*. Unfortunately, it was not possible to obtain the syntenic regions of *Foxl2* for *C. milii* because the genomic data in the blast hits are still not organized in scaffolds and the reads are short (http://esharkgenome.imcb.a-star.edu.sg/). Furthermore, for comparison, some ortholog sequences from representative species such as *Xenopus tropicalis* (frog) and *Homo sapiens* were included in the analysis of the syntenic regions.

Most of the synteny analysis was conducted using Ensembl with the genes identified using its genome browser. Because there were only partial transcript predictions for *O. niloticus* and *L. chalumnae*, without gene identification, each partial transcript sequence in the genomic region of *Foxl2A* and *Foxl2B* was retrieved and then queried by a blastp search against the NCBI database, with the identities and the E-values annotated from the highest similar hits from blast results (Supplementary Material 7). For each partial transcript, the retrieval was expanded to 50 kb of the flanking regions. The sequences acquired were inserted into the online program Softberry FGENESH (http://www.softberry.com/) using *T. rubripes* as the reference genome to recover the whole gene transcriptional region and confirm the identification of genes in these syntenic regions (Supplementary Material 7). However, for the gene *Pik3cd* (*phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit $\delta$ isoform*), there was no transcript annotation in the *L. chalumnae Foxl2B* syntenic region in Ensembl. Because this gene appeared in our analysis as an important marker for *Foxl2B*, a tblastn search was conducted in Ensembl to identify the localization of *Pik3cd* in *L. chalumnae* using a sequence of the Pik3cd protein from *D. rerio* (accession number AAH54896) as the query. Because the resulting hit from the blast search was positioned to be syntenic with *Foxl2B*, 50 kb of the flanking region was retrieved, and gene prediction was performed using Softberry FGENESH to exclude a possible pseudogene and to confirm the correct identification of Pik3cd in *L. chalumnae*. For the final predicted gene, its corresponding protein sequence (Supplementary Material 7) was used as the query for a blastp search in NCBI to check the protein prediction reliability and integrity.

**Table 1 Maximum likelihood reconstruction tree parameters**

| *Data* | *Substitution model* | *Tree searching operations* | *Starting tree* | *Branch support* |
|---|---|---|---|---|
| Forkhead | LG | Best of NNI | Neighbor-joining | aLRT (SH-like) |
| Foxl2 | Dayhoff | Best of SPR | Neighbor-joining | aLRT (SH-like) |

Abbreviations: aLRT, approximate likelihood ratio test; Foxl2, forkhead box L2; LG, Le and Gascuel; NNI, nearest neighbor interchange; SH, Shimodaira-Hasegawa; SPR, subtree pruning and regrafting.

### Exhaustive search for the *Foxl2* duplicated copy

To assure that *Foxl2* duplicated copies were not missed during the blast search, we proceeded to perform an exhaustive search in the Ensembl genomes, isolating the flanking regions in which *Foxl2B* was located and conducting a gene name search of important *Foxl2B* vicinity markers identified in our work, including *Pik3cd*, *Tmem201* (*transmembrane protein 201*), *Clstn1* (*calsyntenin-1*), *Ctnnbip1* (*β-catenin-interacting protein 1*) and *Lzic* (*leucine zipper and Ctnnbip1 domain-containing protein*). We restricted our searches to annotated genomes that exhibited at least one of those markers. Once a syntenic region was identified, the nucleotide sequences of that region were retrieved. These sequences ranged from 15 to 1000 kb, depending upon the size of the available region. A gene prediction was performed in Softberry FGENESH (http://www.softberry.com/) using *T. rubripes*, *X. tropicalis* and *H. sapiens* as reference genomes. For all predicted genes, their corresponding amino-acid sequences were used as queries for a blastp search in NCBI to identify results with similarity to the forkhead protein family. When any similarity was observed, the corresponding sequence was inserted in the domain phylogeny to verify the homology with Foxl2.

### Nomenclature reference to systematic relationships

All taxonomic groups and systematic relationships mentioned in this work were based on the iTOL (Interactive Tree of Life) (Letunic and Bork, 2011).

## RESULTS AND DISCUSSION

### *Foxl2* gene search and sequencing

During the genome searches of annotated sequences, blast and exhaustive searches for *Foxl2*, diverged duplicated copies were found in the chondrichthyan *C. milii*, in the teleosts *A. burtoni*, *G. morhua*,

**Table 2 Bayesian method reconstruction tree parameters**

| *Data* | *Substitution model* | *Base frequencies* | *Starting tree* | *Generations/burn-in* | *Sample frequency* | *Branch support* |
|---|---|---|---|---|---|---|
| Forkhead | WAG | Estimated | Randomly generated | 50 000 000/17 000 | 1000 | Posterior probability |
| Foxl2 | Dayhoff | Estimated | Randomly generated | 30 000 000/10 000 | 1000 | Posterior probability |

Abbreviations: Foxl2, forkhead box L2; WAG, Whelan And Goldman.

*M. zebra* and *O. niloticus*, in the coelacanth *L. chalumnae*, in the bird *T. guttata* and in the mammal *M. domestica* (one copy ortholog to *Foxl2A* and the other to *Foxl2B*, as will be discussed). On the other hand, the *Foxl2* sequences obtained by genome walking exhibited just one copy of the gene in the chondrichthyans *R. lalandei* (accession number JX012129) and *C. callorynchus* (accession number JX012130), and in the teleost *C. monoculus* (accession number JX012131). However, the degenerate primers used probably anneal only to *Foxl2A* sequences, precluding the detection of paralog copies. The same outcome may have occurred in the other teleost species in which the *Foxl2* sequences were obtained by PCR procedures in previous studies.

All the sequences obtained for *Foxl2* from genome database searches and experimental procedures, which were subsequently used for phylogenetic analysis, exhibited a single exon (as expected for the *Foxl2* gene) and a translated protein with variable sizes (Supplementary Material 3). Furthermore, the present gene searches revealed the presence and absence of *Foxl2* copies.

## Forkhead phylogeny

The domain phylogeny combined sequences from all forkhead family groups (A–R) (Figure 2a). As our goal with this approach was to demonstrate that all Foxl2 analyzed sequences were monophyletic (especially the ones predicted in the present study), additional analyses on the domain phylogeny have not been implemented. Besides, there are well-established studies that already address the evolution of the forkhead family (Kaestner *et al.*, 2000; Mazet *et al.*, 2003; Katoh, 2004; Tu *et al.*, 2006; Hannenhalli and Kaestner, 2009).

The Foxl2 branch was clearly identified from the rest of the forkhead proteins, and all the Foxl2 sequences obtained were placed in a monophyletic group, thus characterizing them as members of the Foxl2 subfamily (Figure 2a and Supplementary Material 5). Besides, the domain phylogeny showed the following: (1) the sequence of the sponge *Suberites domuncula* seems to be the proper outgroup for Foxl2 phylogeny and (2) there is evidence of orthology among the extra Foxl2 copies. For instance, all the new duplicated copies obtained in this study (see the topic '*Foxl2* gene search and sequencing'), as well as the paralogs reported in other papers, were placed within the Foxl2 branch, including the diverged duplicated copies in the tetrapods *T. guttata* and *M. domestica* within a clade with the duplicated copies in *T. nigroviridis* and *T. rubripes*. Moreover, the sequence of the agnathan *P. marinus* that could not be inserted in

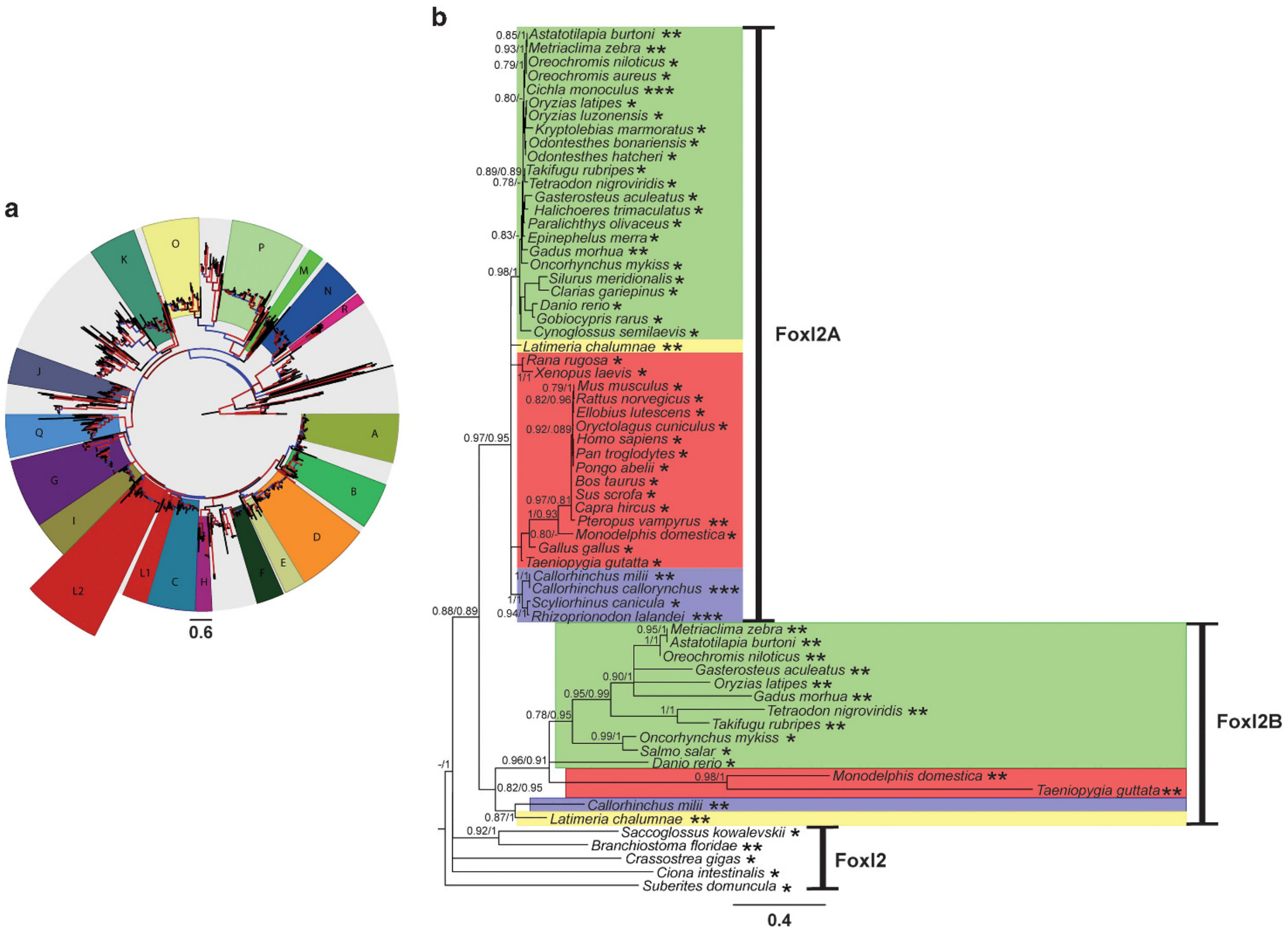
a
K
O
P
M
N
R
J
Q
A
G
B
I
D
E
F
H
C
L1
L2
0.6
b
0.85/1 Astatotilapia burtoni **
0.93/1 Metriaclima zebra **
0.79/1 Oreochromis niloticus *
Oreochromis aureus *
Cichla monoculus ***
0.80/- Oryzias latipes *
Oryzias luzonensis *
Kryptolebias marmoratus *
Odontesthes bonariensis *
Odontesthes hatcheri *
0.89/0.89 Takifugu rubripes *
0.78/- Tetraodon nigroviridis *
Gasterosteus aculeatus *
Halichoeres trimaculatus *
Paralichthys olivaceus *
0.83/- Epinephelus merra *
Gadus morhua **
Oncorhynchus mykiss *
0.98/1 Silurus meridionalis *
Clarias gariepinus *
Danio rerio *
Gobiocypris rarus *
Cynoglossus semilaevis *
Latimeria chalumnae **
Rana rugosa *
1/1 Xenopus laevis *
0.79/1 Mus musculus *
0.82/0.96 Rattus norvegicus *
Ellobius lutescens *
0.92/.089 Oryctolagus cuniculus *
Homo sapiens *
Pan troglodytes *
Pongo abelii *
Bos taurus *
Sus scrofa *
0.97/0.81 Capra hircus *
1/0.93 Pteropus vampyrus **
0.80/- Monodelphis domestica *
Gallus gallus *
Taeniopygia gutatta *
1/1 Callorhinchus milii **
Callorhinchus callorynchus ***
1/1 Scyliorhinus canicula *
0.94/1 Rhizoprionodon lalandei ***
0.97/0.95
0.88/0.89
Foxl2A
0.95/1 Metriaclima zebra **
1/1 Astatotilapia burtoni **
Oreochromis niloticus **
Gasterosteus aculeatus **
0.90/1 Oryzias latipes **
0.95/0.99 Gadus morhua **
1/1 Tetraodon nigroviridis **
0.78/0.95 Takifugu rubripes **
0.99/1 Oncorhynchus mykiss *
Salmo salar *
0.96/0.91 Danio rerio *
0.98/1 Monodelphis domestica **
0.82/0.95 Taeniopygia guttata **
Callorhinchus milii **
0.87/1 Latimeria chalumnae **
Foxl2B
-/1
0.92/1 Saccoglossus kowalevskii *
Branchiostoma floridae **
Crassostrea gigas *
Ciona intestinalis *
Suberites domuncula *
Foxl2
0.4

**Figure 2** Phylogenetic relationship of the forkhead domain and Foxl2 sequences. The letters in the forkhead phylogeny (**a**) represent each different member of the domain family (FoxA to FoxR). The branch colors indicate the consensus statistical support from PhyML and BEAST analysis: blue (<70%) and red (⩾70%). The scale bars indicate the average number of amino acid substitutions per site. In the Foxl2 phylogeny (**b**), the first and second branch values represent the approximate likelihood ratio test (aLRT; SH-like) and the posterior probability from the PhyML and BEAST programs, respectively. The clades of each Foxl2 form are indicated by right side bars. The highlighted clades represent chondrichthyans (blue), teleosts (green), coelacanth (yellow) and tetrapods (red). The asterisks represent the procedure used to obtain each sequence in the phylogeny: accession number from GenBank (*), genome search (**) and genome walking (***). The scale bars indicate the average number of amino-acid substitutions per site.

the posterior Foxl2-specific phylogeny exhibited Foxl2A orthology in the domain approach (Supplementary Material 5).

### A new view of *Foxl2* evolution

The Foxl2 phylogeny based only on the amino-acid sequences of the forkhead domain (data not shown) did not exhibit a total resolute tree, probably because of the high conservation observed in this region, and was thus characterized by a low phylogenetic signal. In this case, based only on an almost full-length amino-acid sequence alignment, a tree with well-supported branches was generated (Figure 2b).

The first thing that is evident in the phylogeny of Foxl2 is that it is possible to split the tree into three main clades (evidenced by a black bar in the right side of Figure 2b). Based on this result, together with other findings of the present work (that is, results from syntenic region analysis), a new nomenclature for the *Foxl2* gene in vertebrates is proposed. In summary, one clade includes the *Foxl2* gene of nonvertebrate taxa; another includes the two distinct copies of *Foxl2* from vertebrates, which we propose have evolved from the same duplication event. One copy is known in tetrapods as *Foxl2* and is considered the most thoroughly studied form of the gene (here named *Foxl2A*), and the other copy includes the duplicated diverged form (here named *Foxl2B*).

The Foxl2A sequences obtained from genomes and experimental procedures, together with other sequences already known, were grouped within the Foxl2A clade of vertebrates, thus confirming their orthology (Figure 2b). Interestingly, Foxl2B copies were placed in a monophyletic group, including the newly described ones in this study (*C. milii*, *A. burtoni*, *G. morhua*, *M. zebra*, *O. niloticus*, *L. chalumnae*, *T. guttata* and *M. domestica*; Figure 2b). The Foxl2B sequences were represented as a separate clade from the Foxl2A sequences. The main point of this finding is the fact that the *C. milli*, *L. chalumnae*, *T. guttata* and *M. domestica* duplicated copies were positioned within the Foxl2B clade, which refutes the hypothesis, proposed by Jiang *et al.* (2011), that the paralog Foxl2 copies are a product of the duplication event exclusive of teleosts. Thus, we hypothesized that the Foxl2B copies arose from an old duplication event; however, it was necessary to perform an analysis in the syntenic region of the predicted paralog copies and compare them with the teleost copies to reinforce the orthology of the duplicated sequences.

In some cases, orthology among sequences can be inferred from synteny analysis because small syntenic blocks of the genome are frequently maintained and inherited throughout distant lineages, such as the *H. sapiens* and *D. rerio* genomes (Postlethwait *et al.*, 2000). The present synteny analysis clearly showed that the flanking gene composition for *Foxl2* copies was similar in the analyzed species. For a clearer presentation, only the common syntenic genes among the analyzed species are shown in Figure 3.

In the *Foxl2A* synteny analysis (Figure 3a), a common general composition as well as gene orientation was observed in all species analyzed with emphasis on *Mrps22* (*mitochondrial 28S ribosomal protein S22*) and *Pik3cb* (*β isoform of the phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit*) that were observed near the *Foxl2A* copies (except *Mrps22* for *O. niloticus* and *T. nigroviridis*; however, this result can be related to the limited scaffold size in the case of *O. niloticus*). The *Foxl2A* syntenic region for the fish O*ryzias latipes* was not analyzed here because syntenic information for this region is currently unavailable in Ensembl.

A conservation pattern was also displayed in the syntenic region of the *Foxl2B* copy (Figure 3b). In this case, *Pik3cd* and *Tmem201* were both close to *Foxl2B* in all analyzed species, except for *G. morhua*, which does not show the copy of *Tmem201*; however, this result can also be related to the limited scaffold size. It was possible to observe some rearrangements in the orientation and repositioning of some genes. For instance, the *Tmem201* gene was located upstream to *Foxl2B* in *D. rerio*, *L. chalumnae*, *T. guttata* and *M. domestica*, whereas it was downstream in all other analyzed species, suggesting the occurrence of rearrangements in this genomic region. Another feature observed in the *Foxl2B* cluster of *L. chalumnae*, *T. guttata* and *M. domestica* was the inversion rearrangement involving the *Clstn1* and *Ctnnbip1* genes.

The syntenic regions of all described *Foxl2B* copies, including the new ones in this study, exhibited high composition conservation, which supports the hypothesis of the orthology of these sequences, also shown by the phylogeny of Foxl2. Based on this scenario of orthology among all *Foxl2B* genes, the hypotheses of the duplicated copies being generated in an exclusive duplication event of teleosts (Jiang *et al.*, 2011) can be discarded. The duplication event must have occurred in an ancestor of gnathostomes, for the *Foxl2B* copy to appear in chondrichthyans, teleosts, coelacanth and tetrapods (Figure 4). Posteriorly, the *Foxl2A* and *Foxl2B* copies would have suffered an additional duplication event exclusive of teleosts. In this context, the generated copies were possibly lost, as the presence of new duplicated copies, besides *Foxl2A* and *Foxl2B*, was not reported in teleosts. Indeed, the presence of a possible *Foxl2A* pseudogene described in *S. salar*, with short identity with the forkhead domain, and exhibiting a divergent N and C terminal regions when compared with *Foxl2A* and *Foxl2B* (von Schalburg *et al.*, 2011), may represent a remnant of the *Foxl2* copies generated in the exclusive whole genome duplication occurred in teleosts.

A significant observation was that the *Pik3cb* and *Pik3cd* genes in the *Foxl2A* and *Foxl2B* syntenic regions, respectively, have been reported to have evolved from a duplication that gave rise to the two forms (Brown and Auger, 2011).The same study verified that both *Pi3k* copies remained in teleosts and tetrapods; therefore, the duplication event also occurred in an ancient ancestor in vertebrates. Taken together with our present analysis, this scenario indicates that *Foxl2A* and *Foxl2B* were not products of a single gene duplication, and considering the timing of the duplication as placed somewhere in the early lineages of vertebrates, we may speculate about this event as an outcome from one of the rounds of whole genome duplication occurred in the ancestor of vertebrates (1R) and gnathostomes (2R) (Donoghue and Purnell, 2005; Meyer and Van de Peer, 2005; Kasahara, 2007). Searches of other genes observed in the syntenic region of *Foxl2*, such as *Clstn1* and *Clstn2* (*Calsyntenin-2*); *Rbp2* (*Retinol-binding protein 2*) and *Rbp7* (*Retinol-binding protein 7*); and *Nmnat1* (*Nicotinamide nucleotide adenylyltransferase 1*) and *Nmnat3* (*Nicotinamide nucleotide adenylyltransferase 3*), did not reveal any related evidence associating them with a duplication event. In this context, it would be very informative to perform a phylogenetic analysis of those genes to understand their evolutionary history and their relationship to the *Foxl2* and *Pi3k* duplication event.

An additional important observation was that no duplicated copy of *Foxl2* was detected in any of the 23 eutherian genomes searched, even in the exhaustive search in 18 of those species, thus suggesting that the *Foxl2B* copy was lost in this group (Figure 4 and Supplementary Material 8). *Foxl2B* was also not detected in the amphibian species *X. tropicalis*, in the sauropsidian *Anolis carolinensis* (anole lizard), *Gallus gallus* (chicken) and *Meleagris gallopavo* (turkey), in the monotreme *Ornithorhynchus anatinus* (platypus) or in the methatherian species *Sarcophilus harrisii.* These results indicate an evolutionary process characterized by events of independent losses

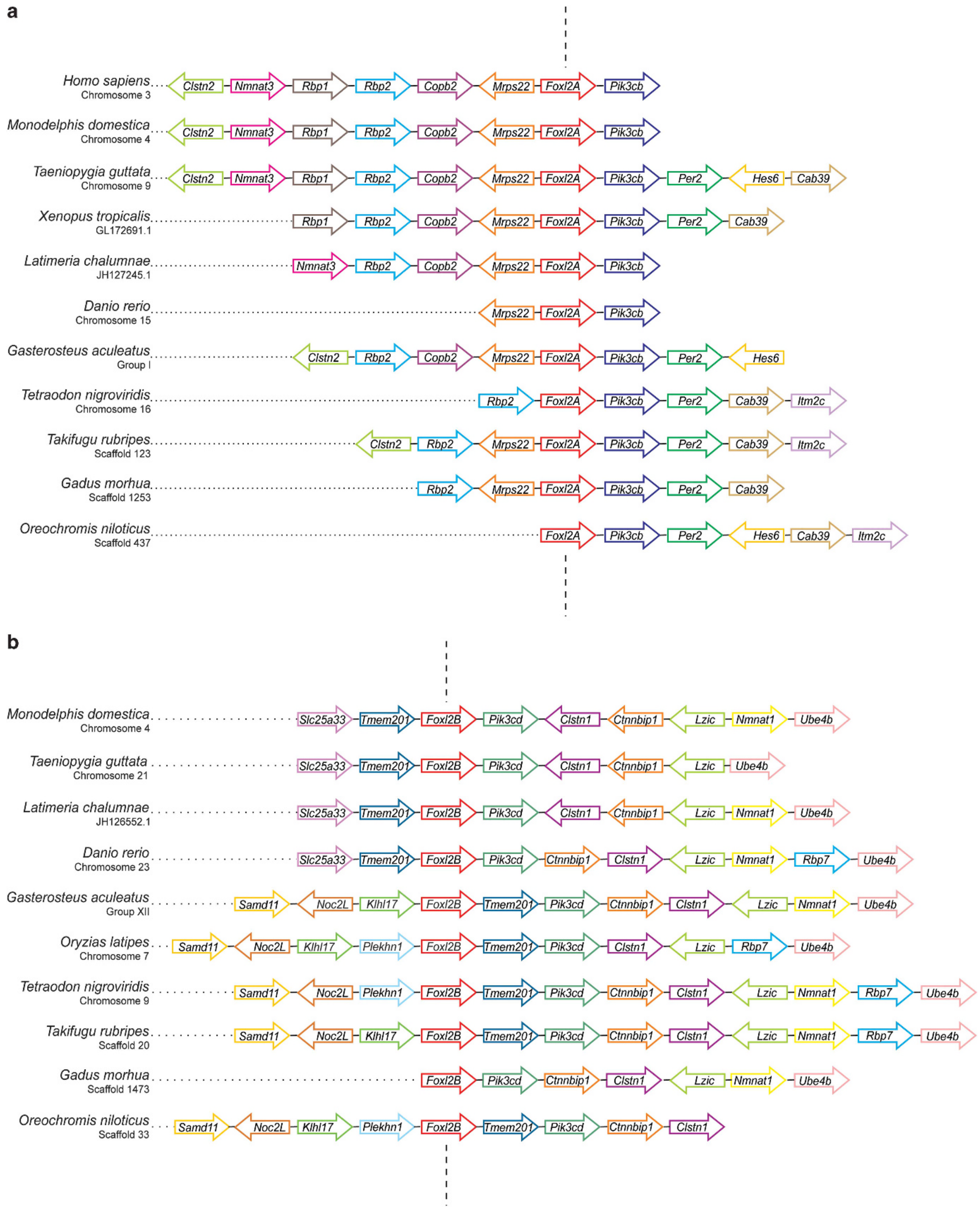
a
Homo sapiens Chromosome 3
Clstn2 Nmnat3 Rbp1 Rbp2 Copb2 Mrps22 Foxl2A Pik3cb
Monodelphis domestica Chromosome 4
Clstn2 Nmnat3 Rbp1 Rbp2 Copb2 Mrps22 Foxl2A Pik3cb
Taeniopygia guttata Chromosome 9
Clstn2 Nmnat3 Rbp1 Rbp2 Copb2 Mrps22 Foxl2A Pik3cb Per2 Hes6 Cab39
Xenopus tropicalis GL172691.1
Rbp1 Rbp2 Copb2 Mrps22 Foxl2A Pik3cb Per2 Cab39
Latimeria chalumnae JH127245.1
Nmnat3 Rbp2 Copb2 Mrps22 Foxl2A Pik3cb
Danio rerio Chromosome 15
Mrps22 Foxl2A Pik3cb
Gasterosteus aculeatus Group I
Clstn2 Rbp2 Copb2 Mrps22 Foxl2A Pik3cb Per2 Hes6
Tetraodon nigroviridis Chromosome 16
Rbp2 Foxl2A Pik3cb Per2 Cab39 Itm2c
Takifugu rubripes Scaffold 123
Clstn2 Rbp2 Mrps22 Foxl2A Pik3cb Per2 Cab39 Itm2c
Gadus morhua Scaffold 1253
Rbp2 Mrps22 Foxl2A Pik3cb Per2 Cab39
Oreochromis niloticus Scaffold 437
Foxl2A Pik3cb Per2 Hes6 Cab39 Itm2c
b
Monodelphis domestica Chromosome 4
Slc25a33 Tmem201 Foxl2B Pik3cd Clstn1 Ctnnbip1 Lzic Nmnat1 Ube4b
Taeniopygia guttata Chromosome 21
Slc25a33 Tmem201 Foxl2B Pik3cd Clstn1 Ctnnbip1 Lzic Ube4b
Latimeria chalumnae JH126552.1
Slc25a33 Tmem201 Foxl2B Pik3cd Clstn1 Ctnnbip1 Lzic Nmnat1 Ube4b
Danio rerio Chromosome 23
Slc25a33 Tmem201 Foxl2B Pik3cd Ctnnbip1 Clstn1 Lzic Nmnat1 Rbp7 Ube4b
Gasterosteus aculeatus Group XII
Samd11 Noc2L Klhl17 Foxl2B Tmem201 Pik3cd Ctnnbip1 Clstn1 Lzic Nmnat1 Ube4b
Oryzias latipes Chromosome 7
Samd11 Noc2L Klhl17 Plekhn1 Foxl2B Tmem201 Pik3cd Clstn1 Lzic Rbp7 Ube4b
Tetraodon nigroviridis Chromosome 9
Samd11 Noc2L Plekhn1 Foxl2B Tmem201 Pik3cd Ctnnbip1 Clstn1 Lzic Nmnat1 Rbp7 Ube4b
Takifugu rubripes Scaffold 20
Samd11 Noc2L Klhl17 Foxl2B Tmem201 Pik3cd Ctnnbip1 Clstn1 Lzic Nmnat1 Rbp7 Ube4b
Gadus morhua Scaffold 1473
Foxl2B Pik3cd Ctnnbip1 Clstn1 Lzic Nmnat1 Ube4b
Oreochromis niloticus Scaffold 33
Samd11 Noc2L Klhl17 Plekhn1 Foxl2B Tmem201 Pik3cd Ctnnbip1 Clstn1

**Figure 3** Schematic representation of the syntenic regions of *Foxl2* copies. *Foxl2A* (**a**) and *Foxl2B* (**b**) are positioned as the reference genes. The transcriptional orientation of the genes is depicted by the direction of the arrows. The observed genes represent only the common markers for the analyzed species and not the complete syntenic block composition. *Cab39*, *calcium-binding protein 39*; *Clstn1*, *calsyntenin-1*; *Clstn2*, *calsyntenin-2*; *Copb2*, *coatomer protein complex subunit β2*; *Ctnnbip1*, *β-catenin-interacting protein 1*; *Foxl2A*, *isoform A of the forkhead box L2*; *Foxl2B*, *isoform B of the forkhead box L2*; *Hes6*, *hairy and enhancer of split 6*; *Itm2c*, *integral membrane protein 2C*; *Klhl17*, *Kelch-like protein 17*; *Lzic*, *leucine zipper and Ctnnbip1 domain-containing protein*; *Mrps22*, *mitochondrial 28S ribosomal protein S22*; *Nmnat1*, *nicotinamide nucleotide adenylyltransferase 1*; *Nmnat3*, *nicotinamide nucleotide adenylyltransferase 3*; *Noc2l*, *nucleolar complex protein 2 homolog*; *Per2*, *period circadian protein homolog 2*; *Pik3cb*, *β-isoform of the phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit*; *Pik3cd*, *δ-isoform of the phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit*; *Plekhn1*, *pleckstrin homology domain-containing family N member 1*; *Rbp2*, *retinol-binding protein 2*; *Rbp7*, *retinol-binding protein 7*; *Samd11*, *sterile α-motif domain-containing protein 11*; *Slc25a33*, *solute carrier family 25 member 33*; *Tmem201*, *transmembrane protein 201*; *Ube4b*, *ubiquitin conjugation factor E4 B*.

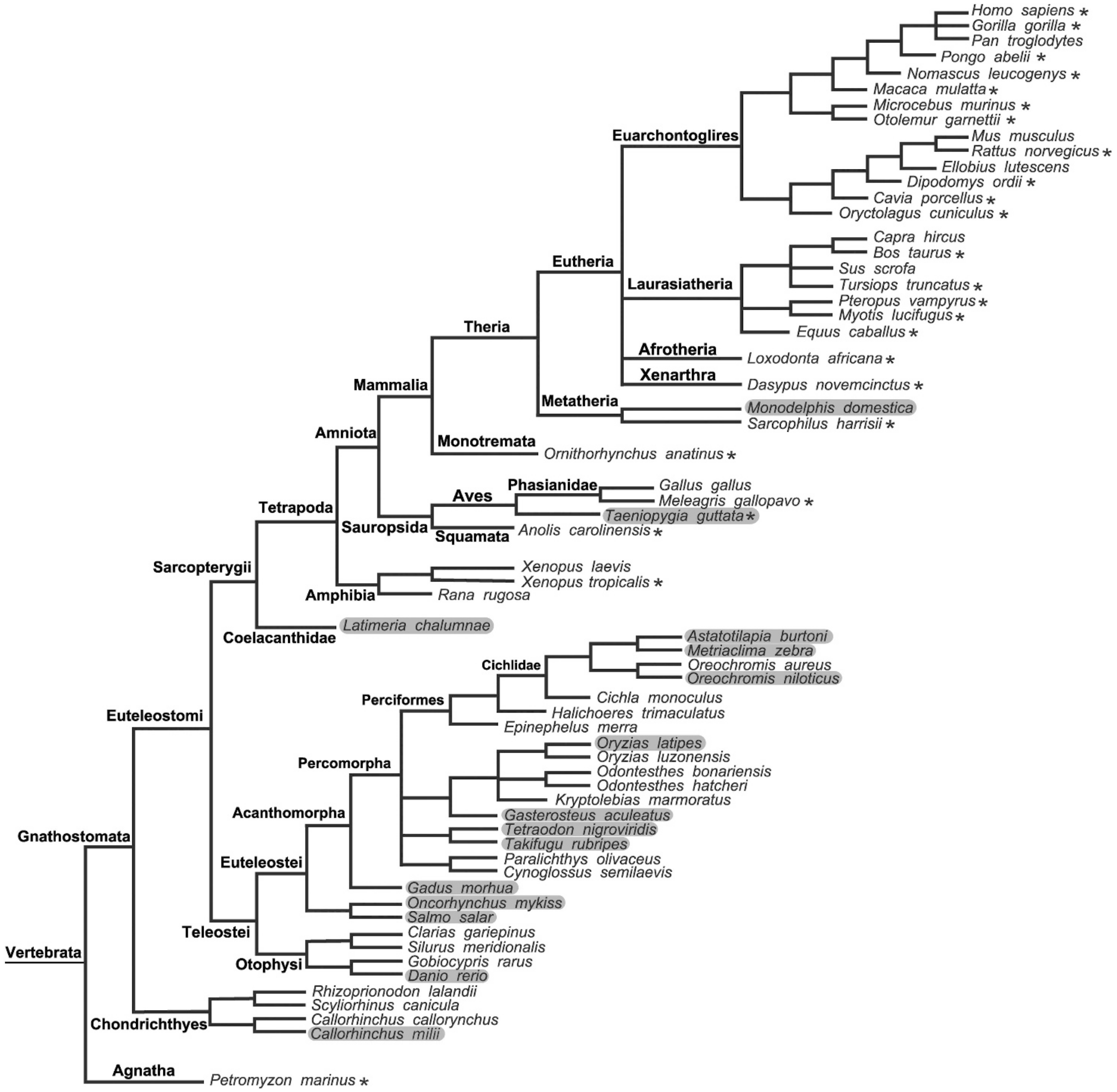

**Figure 4** Summary of the evolutionary scenario for *Foxl2* copies plotted in the classical phylogeny of vertebrates. Taxa are labeled as carrying (highlighted in gray) or missing the *Foxl2B* copy. The asterisks (*) represent the species in which an exhaustive search for the *Foxl2B* copy was performed.

of *Foxl2* copies during the evolutionary history of gnathostomes. However, except for eutherians, few species were analyzed, and because most of these genomes are still in an assembly process, this scenario of independent losses will only be clearly understood and confirmed with future studies.

### Foxl2B activity

The scenario described in the current work for paralog copies also raises an important issue about the role of Foxl2B, especially with regard to sexual differentiation. In an expression study involving the fish *O. mykiss* (Baron *et al.*, 2004), *Foxl2B* was shown to be expressed later than *Foxl2A* in ovaries. The authors suggested that this difference indicated a neofunctionalization process for Foxl2B. In another study with the fish *S. salar* (von Schalburg *et al.*, 2011), *Foxl2B* exhibited a similar but also very distinctive pattern of expression compared with *Foxl2A*. This difference was observed especially in males, in which *Foxl2B* was expressed at higher levels in testes and extragonadal tissues when compared with *Foxl2A*. In females, however, Foxl2B was restricted to the gills, skin and ovary. Although the authors did not draw any conclusions about the roles of Foxl2B, the different patterns of expression in different types of tissues could indicate a subfunctionalization or a neofunctionalization process for the duplicated copies. However, only analyzing the expression pattern of the single ancestral gene, we can clearly understand the outcome of the duplicated copies of Foxl2.

Another interesting observation was based on the analysis of the polyalanine tract that is exclusive to mammals. In *M. domestica*, the Foxl2A protein showed 16 alanines (Supplementary Material 6), unlike the strict value of 14 alanines suggested by Cocquet *et al.* (2003) for all mammals. In addition, the Foxl2B protein in *M. domestica* did not exhibit the polyalanine tract, which may indicate a change in the protein function compared with the Foxl2A protein. However, how Foxl2B adapted its function during the evolution of gnathostomes remains unknown, pending additional functional studies.

## CONCLUSION

The broad study with the domain approach and the incorporation of new Foxl2 duplicated sequences in the inferred phylogeny provided the basis for a novel discussion of evolutionary studies of *Foxl2*. The current study identified duplicated forms of *Foxl2* in different vertebrate groups, which suggests the need for a revision of the nomenclature for *Foxl2A* and *Foxl2B*. Notably, the established evolutionary analysis strongly supported the conclusion that the predicted *Foxl2B* copies in chondrichthyans, coelacanth and tetrapods are indeed orthologous to all *Foxl2B* copies of teleosts. Furthermore, the duplication event that gave rise to the paralog copies was not unique to teleosts and was a common event that occurred in an ancestor gnathostome. Our results also suggest that this duplication was not exclusive to *Foxl2*. From the description of new *Foxl2B* copies in more fish species, obtained by genome searches, it is reasonable to believe that there are more teleost species, not yet reported, that carry a *Foxl2B* copy. In this context, it remains unknown whether most species in this group carry this additional copy, whereas others have lost it, or if all teleost species have the duplicated version of *Foxl2*. The distribution of *Foxl2B* was shown to be diversified in tetrapods, with a strong indication of loss of this copy in eutherians. In addition to this newly reported scenario, more information is required for understanding the role of the *Foxl2B* copy in vertebrates, especially in relation to sexual differentiation.

## DATA ARCHIVING

Sequence data have been submitted to GenBank: accession numbers JX012129, JX012130 and JX012131.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

This study was supported by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) from Brazil. We are grateful to Dr OBF Gadig for providing tissue samples of *C. callorynchus*.

Akaike H (1974). A new look at the statistical model identification. *IEEE Trans Automat Contr* **19**: 716–723.

Anisimova M, Gascuel O (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol* **55**: 539–552.

Baron D, Cocquet J, Xia X, Fellous M, Guiguen Y, Veitia RA (2004). An evolutionary and functional analysis of FoxL2 in rainbow trout gonad differentiation. *J Mol Endocrinol* **33**: 705–715.

Benayoun BA, Caburet S, Veitia RA (2011). Forkhead transcription factors: key players in health and disease. *Trends Genet* **27**: 224–232.

Bravieri R, Shiyanova T, Chen TH, Liao XB (1997). Different DNA contact schemes are used by two winged helix proteins to recognize a DNA binding sequence. *Nucleic Acids Res* **25**: 2888–2896.

Brissette JL, Li J, Kamimura J, Lee D, Dotto GP (1996). The product of the mouse nude locus Whn, regulates the balance between epithelial cell growth and differentiation. *Genes Dev* **10**: 2212–2221.

Brown JR, Auger KR (2011). Phylogenomics of phosphoinositide lipid kinases: perspectives on the evolution of second messenger signaling and drug discovery. *BMC Evol Biol* **11**: 4.

Christoffels A, Koh EG, Chia JM, Brenner S, Aparicio S, Venkatesh B (2004). Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol Biol Evol* **21**: 1146–1151.

Clark KL, Halay ED, Lai ES, Burley SK (1993). Co-crystal structure of the Hnf-3/fork head DNA-recognition motif resembles histone-H5. *Nature* **364**: 412–420.

Cocquet J, De Baere E, Gareil M, Pannetier M, Xia X, Fellous M *et al.* (2003). Structure, evolution and expression of the FOXL2 transcription unit. *Cytogenet Genome Res* **101**: 206–211.

Cocquet J, Pailhoux E, Jaubert F, Servel N, Xia X, Pannetier M *et al.* (2002). Evolution and expression of FOXL2. *J Med Genet* **39**: 916–921.

Crisponi L, Deiana M, Loi A, Chiappe F, Uda M, Amati P *et al.* (2001). The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nat Genet* **27**: 159–166.

Darriba D, Taboada GL, Doallo R, Posada D (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**: 1164–1165.

De Baere E, Beysen D, Oley C, Lorenz B, Cocquet J, De Sutter P *et al.* (2003). FOXL2 and BPES: mutational hotspots, phenotypic variability, and revision of the genotype-phenotype correlation. *Am J Hum Genet* **72**: 478–487.

De Baere E, Dixon MJ, Small KW, Jabs EW, Leroy BP, Devriendt K *et al.* (2001). Spectrum of FOXL2 gene mutations in blepharophimosis-ptosis-epicanthus inversus (BPES) families demonstrates a genotype-phenotype correlation. *Hum Mol Genet* **10**: 1591–1600.

Donoghue PC, Purnell MA (2005). Genome duplication, extinction and vertebrate evolution. *Trends Ecol Evol* **20**: 312–319.

Dottori M, Gross MK, Labosky P, Goulding M (2001). The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. *Development* **128**: 4127–4138.

Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R *et al.* (2009). Geneious v4.8.5, Available from. http://www.geneious.com

Drummond AJ, Rambaut A (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**: 214.

Drummond AJ, Suchard MA (2010). Bayesian random local clocks, or one rate to rule them all. *BMC Biol* **8**: 114.

Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Finn RD, Clements J, Eddy SR (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29–W37.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE *et al.* (2010). The Pfam protein families database. *Nucleic Acids Res* **38**:Database issue D211–D222.

Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y *et al.* (2011). Ensembl 2011. *Nucleic Acids Res* **39**: D800–D806.

Gross DN, van den Heuvel APJ, Birnbaum MJ (2008). The role of FoxO in the regulation of metabolism. *Oncogene* **27**: 2320–2336.

Guindon S, Gascuel O (2003). PhyML: a simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**: 696–704.

Hannenhalli S, Kaestner KH (2009). The evolution of Fox genes and their role in development and disease. *Nat Rev Genet* **10**: 233–240.

Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E *et al.* (2004). Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **431**: 946–957.

Jaubert F, Galmiche L, Lortat-Jacob S, Fournet JC, Fellous M (2011). Foxl-2 in gonad development and pathology. *Arkh Patol* **73**: 10–13.

Jiang W, Yang Y, Zhao D, Liu X, Duan J, Xie S *et al.* (2011). Effects of sexual steroids on the expression of foxl2 in Gobiocypris rarus. *Comp Biochem Physiol B Biochem Mol Biol* **160**: 187–193.

Kaestner KH, Knochel W, Martinez DE (2000). Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev* **14**: 142–146.

Kasahara M (2007). The 2R hypothesis: an update. *Curr Opin Immunol* **19**: 547–552.

Katoh M (2004). Human FOX gene family. *Int J Oncol* **25**: 1495–1500.

Lehmann OJ, Sowden JC, Carlsson P, Jordan T, Bhattacharya SS (2003). Fox's in development and disease. *Trends Genet* **19**: 339–344.

Letunic I, Bork P (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**:Web Server issue W475–W478.

Li W, Godzik A (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

Maddison WP, Maddison DR (2001). Mesquite: a modular system for evolutionary analysis. Version 2.75. http://mesquiteproject.org

Mahlapuu M, Ormestad M, Enerback S, Carlsson P (2001). The forkhead transcription factor Foxf1 is required for differentiation of extra-embryonic and lateral plate mesoderm. *Development* **128**: 155–166.

Mazet F, Yu JK, Liberles DA, Holland LZ, Shimeld SM (2003). Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. *Gene* **316**: 79–89.

Meyer A, Van de Peer Y (2005). From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* **27**: 937–945.

Miller MA, Pfeiffer W, Schwartz T (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proc Gateway Computing Environments Workshop (GCE).* New Orleans, LA, USA, pp 1–8.

Nakae J, Kitamura T, Kitamura Y, Biggs WH 3rd, Arden KC, Accili D (2003). The forkhead transcription factor Foxo1 regulates adipocyte differentiation. *Dev Cell* **4**: 119–129.

Nehls M, Pfeifer D, Schorpp M, Hedrich H, Boehm T (1994). New member of the winged-helix protein family disrupted in mouse and rat nude mutations. *Nature* **372**: 103–107.

Nicholas KB, Nicholas HB, Deerfield DW (1997). GeneDoc: Analysis and visualization of genetic variation. *EMBnet News* **4**: 14.

Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, Kelly PD, Chu F *et al.* (2000). Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* **10**: 1890–1902.

Rambaut A, Drummond AJ (2007). Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer

Sambrook J, Russel DW (2001). *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory Press: New York.

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.

Tu Q, Brown CT, Davidson EH, Oliveri P (2006). Sea urchin Forkhead gene family: phylogeny and embryonic expression. *Dev Biol* **300**: 49–62.

Uhlenhaut NH, Jakob S, Anlag K, Eisenberger T, Sekido R, Kress J *et al.* (2009). Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell* **139**: 1130–1142.

Uhlenhaut NH, Treier M (2006). Foxl2 function in ovarian development. *Mol Genet Metab* **88**: 225–234.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* **35**: W71–W74.

Veitia RA (2010). FOXL2 versus SOX9: a lifelong ‘battle of the sexes’. *Bioessays* **32**: 375–380.

von Schalburg KR, Yasuike M, Yazawa R, de Boer JG, Reid L, So S *et al.* (2011). Regulation and expression of sexual differentiation factors in embryonic and extragonadal tissues of Atlantic salmon. *BMC Genomics* **12**: 31.

Weigel D, Jurgens G, Kuttner F, Seifert E, Jackle H (1989). The homeotic gene fork head encodes a nuclear-protein and is expressed in the terminal regions of the Drosophila embryo. *Cell* **57**: 645–658.

Zlotogora J, Sagi M, Cohen T (1983). The blepharophimosis, ptosis, and epicanthus inversus syndrome: delineation of two types. *Am J Hum Genet* **35**: 1020–1027.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)