Evolution of the *Tbx6/16* sub-family genes in vertebrates: insights from zebrafish

Supplementary Materials II

Daegwon Ahn, Kwan-Hee You, and Cheol-Hee Kim

Fig. S3

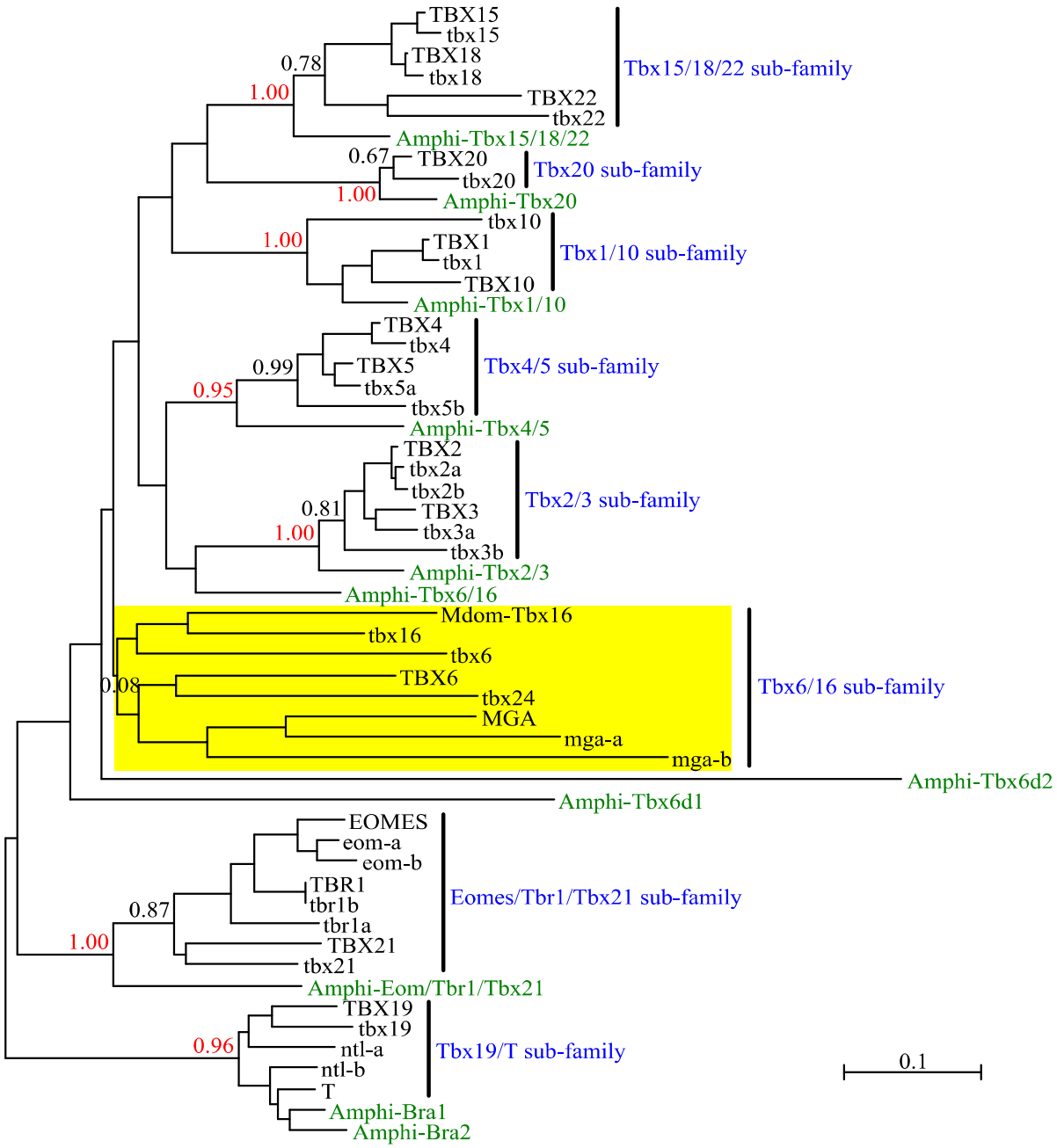


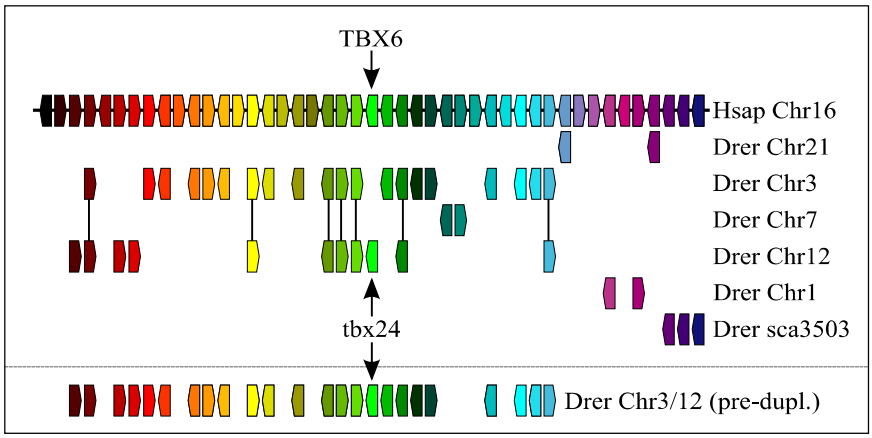
Figure S3. Neighbor-joining tree showing the phylogeny of human (*Homo sapiens*), zebrafish (*Danio rerio*), and amphioxus (*Branchiostoma floridae*) T-box genes.

Neighbor-joining tree was constructed by BioNJ algorithm using the amino acid distances with Poisson correction (Gascuel 1997). Bootstrap values (expressed in proportions) are given for

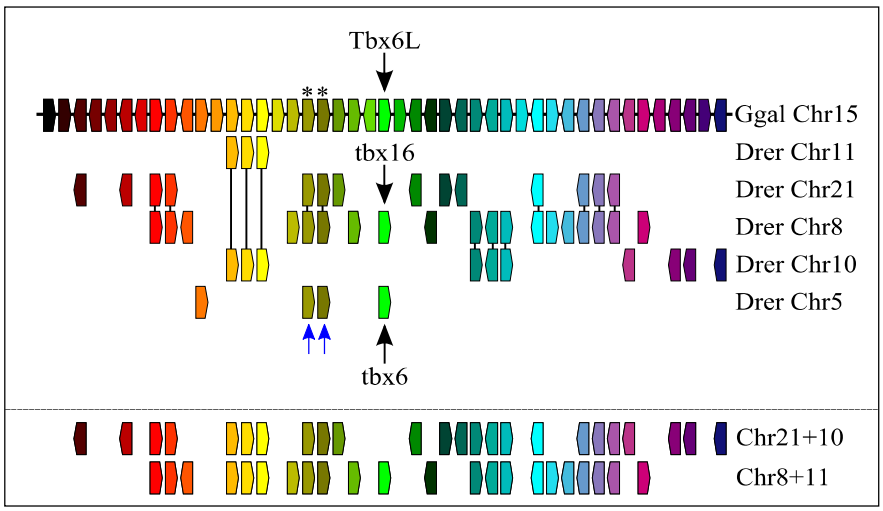
the nodes representing individual vertebrate T-box gene sub-family (numbers written in black) or vertebrate sub-family plus its amphioxus ortholog(s) (numbers written in red). Scale bar represents 0.1 amino acid substitutions per site. Note that the *Tbx6/16* sub-family (boxed in yellow rectangle) is monophyletic in this tree (albeit with a very weak statistical support) just like any other sub-families of vertebrate T-box genes. For Genbank accession numbers and full-length amino acid sequences of the genes included in this tree, see Fig. S1. For the alignment of amino acid sequences used in the analysis, see Fig. S2. Mdom: *Monodelphis domestica* (grey short-tailed opossum).

Fig. S5

A


TBX6
Hsap Chr16
Drer Chr21
Drer Chr3
Drer Chr7
Drer Chr12
Drer Chr1
Drer sca3503
tbx24
Drer Chr3/12 (pre-dupl.)

B


Tbx6L
Ggal Chr15
Drer Chr11
tbx16
Drer Chr21
Drer Chr8
Drer Chr10
Drer Chr5
tbx6
Chr21+10
Chr8+11

C

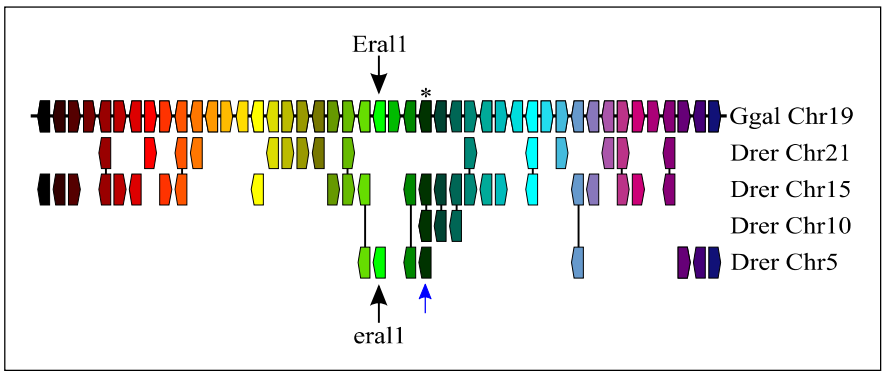

Eral1
Ggal Chr19
Drer Chr21
Drer Chr15
Drer Chr10
Drer Chr5
eral1

Figure S5. Alignview diagrams[1)] of *Genomicus Browser* showing the chromosomal distribution of zebrafish genes orthologous to the genes on the selected segments of tetrapod chromosomes.

In each diagram, tetrapod genes are represented by colored symbols, which are arranged in the same order as they appear on the chromosome. The pointed side of each symbol denotes the direction of transcription for each gene. Beneath tetrapod genes, presence of their fish orthologs in the zebrafish genome is indicated by the presence of matching symbols, which are arranged in rows corresponding to individual zebrafish chromosomes harboring these genes. In Alignview diagrams, like the dot plot diagrams of *Synteny Database*, zebrafish genes are not arranged in their natural order of appearance on each chromosome. Therefore, the presence of zebrafish gene symbols beneath the tetrapod genes signifies only the presence of zebrafish orthologs for tetrapod genes, not the presence of such genes at the same relative positions on the respective zebrafish chromosome[2)]. In *Genomicus Browser*, orthology between tetrapod and zebrafish genes is established using the information from the *EnsemblCompara GeneTree* dataset (Vilella et al. 2009), which uses phylogenetic tree-building algorithms to determine the evolutionary relationships among genes from a diverse range of species available in Ensembl database[3)].

(A) Alignment of the orthologous genes between *TBX6*-containing region of the human genome and the whole genome of zebrafish. Zebrafish genes orthologous to the 45 human genes (*TBX6* plus 22 neighboring genes on either side) in the *TBX6*-containing region of human chromosome 16 are distributed on 6 different chromosomes, with the majority of genes located on chromosomes 3 and 12. Note that, on chromosome 12, *tbx24* is syntenic to several of the zebrafish orthologs of the human genes in the neighborhood of *TBX6*. From the distribution of paralogous pairs (genes connected by vertical lines) between zebrafish chromosomes, it is also evident that, with respect to the *TBX6*-containing region of the human genome, zebrafish chromosomes 3 and 12 form a pair of duplicated regions resulting from the teleost-specific whole genome duplication event (note that, over the *TBX6*-containing region of the human genome, zebrafish chromosomes 3 and 12 share 7 sets of paralogs while other chromosomes share none). Notice that, if we take this into consideration and extend the comparisons to the

"reconstructed" zebrafish chromosome that might have been present before the genome duplication (shown in the bottom row), *tbx24* becomes syntenic to a much larger number of zebrafish orthologs of the *TBX6* gene's genomic neighbors.

(B) Alignment of the orthologous genes between *Tbx6L*-containing region of the chicken genome and the whole genome of zebrafish. Zebrafish genes orthologous to the 45 genes in the *Tbx6L*-containing region of the chicken chromosome 15 (*Tbx6L* plus 22 neighboring genes on either side) are distributed on 5 different chromosomes, with the majority of the orthologs located on chromosomes 21 and 8. Note that, compared to *tbx6*, *tbx16* is syntenic to a much larger number of zebrafish genes orthologous to the *Tbx6L*'s genomic neighbors. From the distribution of paralogous pairs (genes connected by vertical lines) between zebrafish chromosomes, it is also clear that, with respect to the *Tbx6L*-containing region of the chicken genome, zebrafish chromosomes 8 and 21 would constitute a pair of genomic regions that had been produced by the teleost-specific whole genome duplication event (note that, over the *Tbx6L*-containing region of the chicken genome, zebrafish chromosomes 8 and 21 have 8 sets of paralogs between them while other pairs of chromosomes have 3 sets at most). Presence of zebrafish orthologs on other zebrafish chromosomes over this region might then be interpreted as resulting from the post-duplication inter-chromosomal translocation events, which had moved the genes to several other parts of the zebrafish genome (see Kasahara et al. 2007). Notice that, using this assumption, zebrafish orthologs on chromosomes 10 and 11 could be "rolled back" into their hypothetical pre-translocation positions on chromosomes 21 and 8 to reproduce the pre-translocation syntenies for these chromosomes (shown in bottom rows), but the genes on chromosome 5 do not seem to fit into this overall picture of 'duplication followed by inter-chromosomal translocation events', suggesting that they might derive from a different part of the ancestral genome[4)] (see below).

(C) Alignment of the orthologous genes between *Eral1*-containing region of the chicken genome and the whole genome of zebrafish. Zebrafish genes orthologous to the 45 genes in the *Eral1*-containing region of the chicken genome (*Eral1* plus 22 neighboring genes on either side) are distributed on 4 different chromosomes, with the majority of the orthologs located on chromosomes 21 and 15. Note that, in contrast to the prior comparison with the *Tbx6L*-

containing region of the chicken genome (see B), zebrafish chromosome 5 is now seen to have many more orthologs of the chicken genes in this comparison. Also, it is now much easier to explain the presence of the orthologs on chromosome 5 by the 'duplication followed by inter-chromosomal translocation event' model (see B), since most of the chromosome 5 genes[5)] can now easily be "rolled back" into their hypothetical pre-translocation positions on chromosome 21. This indicates that these chromosome 5 genes, which include the zebrafish *eral1* gene and its genomic neighbors *c17orf63a* and *flot2a* (the two symbols on either side of *eral1*), must have been on chromosome 21 before they were displaced from their original positions.

Notes on Figure S5:

1) For simplicity and ease of presentation, original diagrams have been edited to remove those parts that are not needed for our present purpose. Hsap: *Homo sapiens* (human). Drer: *Danio rerio* (zebrafish). Ggal: *Gallus gallus* (chicken).

2) *Genomicus Browser* also allows drawing of genes in their natural order of presence in its "Phyloview" diagrams, which generate views of genes comparable to the local synteny trace diagrams. However, this is more difficult to manage on screen, since the zebrafish genes are often separated by a large number of other genes in the zebrafish genome, even when their tetrapod orthologs are organized in tight clusters.

3) In contrast, *Synteny Database* uses a modified reciprocal best hit (mRBH) method (based on BLAST patterns) coupled with the single-linkage clustering algorithm to determine the orthology and paralogy of genes from a selected pair of species (e.g., human and zebrafish). Because *Synteny Database* employs a strictly pairwise comparison of the species in its determination of the orthology and paralogy, it often mis-identifies orthologs or paralogs when one or more of the relevant genes are missing from one of the species (see the case study for the *ARNTL* family genes in Catchen et al. 2009). This is less of a problem in *Genomicus Browser* since the routine inclusion of multiple species in its orthology determination pipeline in *EnsemblCompara* database usually ensures that the deleterious effect of the "missing" gene from one species would be compensated away by the presence of comparable genes from other species. Because of this,

we based all our local synteny trace diagrams on the Phyloview diagrams of *Genomicus Browser*.

4) Over the *Tbx6L*-containing region of the chicken genome, there are four zebrafish genes on chromosome 5 that are judged to be orthologous to the chicken genes in this region. However, two of these (blue arrows) are deemed orthologous to the chicken genes which already have the requisite number of zebrafish orthologs (i.e., 2) on zebrafish chromosomes 21 and 8. Closer inspection of the symbols reveals that these symbols actually represent a single zebrafish gene (*zgc:66350*) that is drawn twice in the diagram to denote its orthology to a tandem pair of chicken *Gstt1* genes (marked with *'s). This gene (*zgc:66350*) has an amino acid sequence similar to zebrafish genes *gstt1a* (on chromosome 8) and *gstt1b* (on chromosome 21), but the phylogenetic position of this gene on *EnsemblCompara GeneTree* indicates that although *zgc:66350* is related to the vertebrate *Gstt1* genes (including tetrapod *Gstt1-Gstt2* and fish *gstt1a-gstt1b* genes) it does not belong to the *Gstt1* gene cluster itself (data not shown). This indicates that *zgc:66350* is a product of a duplication event more ancient than the teleost-specific whole genome duplication event that had produced the *gstt1a* and *gstt1b* genes, suggesting that this gene might provide yet another example of an anciently duplicated gene that had been lost in tetrapods but had been retained in fishes. We therefore tentatively re-name this gene as *gstt1L* (Fig. 3C) and will treat it as a gene with a sufficiently distinct identity rather than as a paralog of another zebrafish gene.

5) The exception to this would be the zebrafish *si:dkey-174n20.1* gene (blue arrow). This gene provides the third example of zebrafish chromosome 5 genes (the other two being *gstt1L* and *tbx6*) that had been produced by a whole genome duplication event pre-dating the fish-specific whole genome duplication (data not shown). This gene is related to the *Dhrs13* gene (*) of chickens.

Fig. S6

A

```
Drer-tbx6       LWDKFSSIGTEMLITKSGRRMFPSCKVTVTGLNPKVKYVVIMDMVPFDNHKYKWNKDCWE
Drer-tbx16      LWRSFHEIGTEMIITKPGRRMFPHCKISLSGLVPYAKYILLVDMVPEDGLRYKWNKDKWE
Olat-tbx16a     LWRTFHEIGTEMIITKPGRRMFPHCKVSLSGLIPCAKYILLVDMVPEDGFRYKWNKDKWE
Olat-tbx16b     LWKSFHDIGTEMIITKHGRRMFPHCSIALTGLQPLTDYVIMVDMIPVDGFKYKWKKEKWE
Gacu-tbx16a     LWKTFCEIGTEMIITKPGRRMFPHCKINLSGLIPCAKYILLVDMVPEDGFRYKWNKEKWE
Gacu-tbx16b     LWKSFHTFGTEMIITKHGRRMFPHCSISLSGLQPFANYVVMIDMVPADGFKYKWKKEQWE
Mdom-Tbx16      LWTEFYRVGTEMVITKSGRRMFPQCKIRLSGLVPYLKYVVLADFVSVDNFRYKWAKDQWE
Ggal-Tbx6L      LWMKFHQIGTEMIITKSGRRMFPQCKIKVSGLIPYAKYLMLVDFVPVDNFRYKWNKDQWE
Hsap-TBX6       LWKEFSSVGTEMIITKAGRRMFPACRVSVTGLDPEARYLFLLDVIPVDGARYRWQGRRWE

Drer-tbx6       VNGSSDPHLPNRFFIHPDSPAPGQKWMQYPISFHKLKLTNNTLNSNGLVVLHSMHKYQPR
Drer-tbx16      VAGKAEPQPPYRTYLHPDSPAPGSHWMKQPVSFLKLKLTNNALDQHGHIILHSMHRYHPR
Olat-tbx16a     VAGKAEPQPPCRTYVHPDSPAPGSHWMKQSISFLKLKLTNNTLDQHGHIILHSMHRYHPR
Olat-tbx16b     VAGTAEPQPPWRTYTHPDSPAPGRHWMKHPLSFLKMKLTNNTLDQHGHIILHSMHRYYPR
Gacu-tbx16a     VAGKAEPQPPCRTYLHPESPAPGSHWMKQSISFLKLKLTNNTLDQHGHIILHSMHRYHPR
Gacu-tbx16b     AAGKAEPQPPCRTYVHPDSPAPGSHWMKQPLSFLRMKLTNNTLDQHGHIVLHSMHRYYPR
Mdom-Tbx16      VAGKAEPQLPGRSYIHPDSPAYGSHWMKEPVSFHKMKLTNNTLDQHGHIILHSMHRYQPR
Ggal-Tbx6L      VAGKAEPQLPCRTYVHPDSPAPGSHWMKEPVSFQKLKLTNNTLDQHGHIILHSMHRYKPR
Hsap-TBX6       PSGKAEPRLPDRVYIHPDSPATGAHWMRQPVSFHRVKLTNSTLDPHGHLILHSMHKYQPR

Drer-tbx6       LHIVQSPDPCTPHNPGAYLRFTFPEAAFIAVTAYQNQEITKLKIDNNPFAKGFRD
Drer-tbx16      FHIVQADDLYSVR-WSVFQTFTFPETSFTAVTAYQNTKITKLKIDHNPFAKGFRD
Olat-tbx16a     FHVVQADDLFSVR-WSVFQMFTFPETSFTAVTAYQNTKITKLKIDHNPFAKGFRD
Olat-tbx16b     FHVIQTDSSHTVR-WGSFQTFSFPETVFTAVTAYQNPKITKLKIDHNPFAKGFRE
Gacu-tbx16a     FHIVQADDLFSVR-WSVFQTFTFPETSFTAVTAYQNTKITKLKIDHNPFAKGFRD
Gacu-tbx16b     FYVTQADSPYTIC-WAPFQTFSFPETTFTAVTAYQNPRITKLKIHHNPFAKGFRE
Mdom-Tbx16      FLVAQADDLFNVC-WNLFQVFSFPQTVFISVTAYQNEQITKLKIDNNPFAKGFRE
Ggal-Tbx6L      FHIVQADDLFSVR-WSIFQVFSFPETVFTSVTAYQNEQITKLKIDNNPFAKGFRE
Hsap-TBX6       IHLVRAAQLCSQH-WGGMASFRFPETTFISVTAYQNPQITQLKIAANPFAKGFRE
```

B

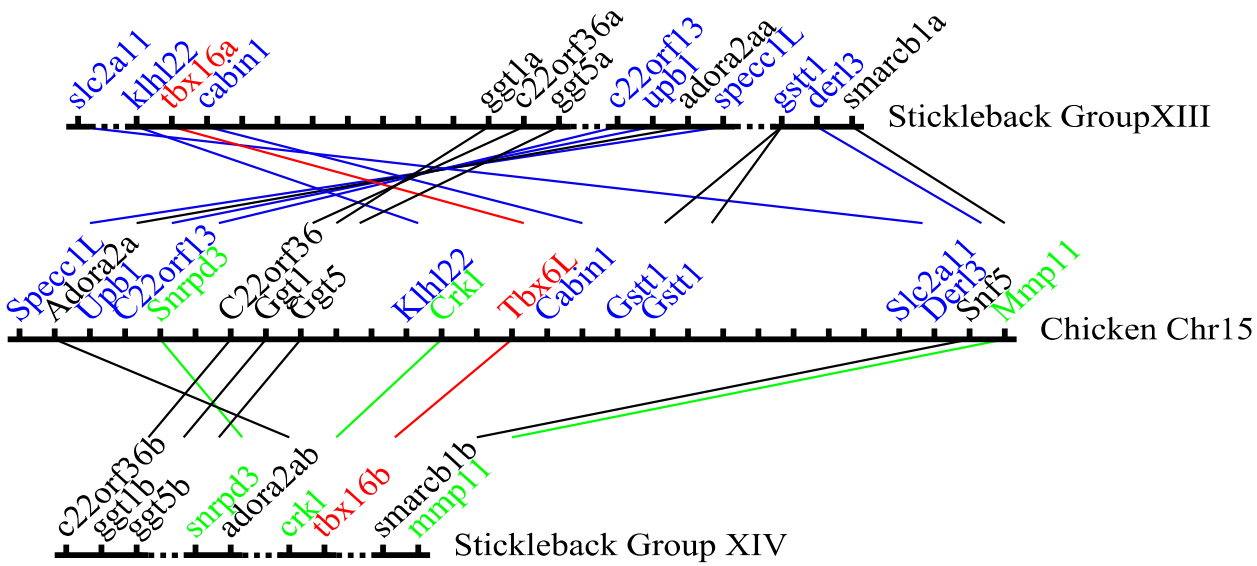


Fig. S6 (cont.)

C

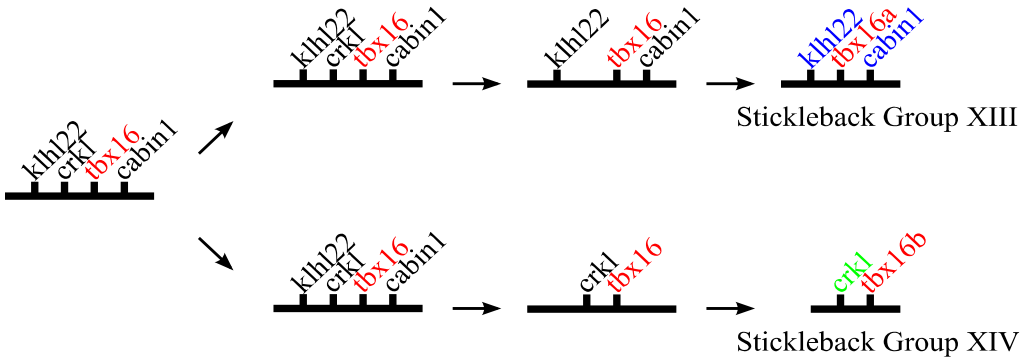


Figure S6. Sequence, structure, and evolutionary origin of teleost *tbx16b* genes.

(A) Alignment of the amino acid sequences of T-domains from the *Tbx16* orthologs of zebrafish (Drer), medaka (Olat), stickleback (Gacu), opossum (Mdom), and chicken (Ggal). For comparison, human (Hsap) *TBX6* gene (NM_004608) is included as an outgroup. For the phylogenetic analysis, the sole unalignable region containing a gap (shaded dark grey) was excluded. Portions of amino acid sequences encoded by different exons are color-coded as in Figure S4. Note that *tbx16b* genes from medaka and stickleback have the same exon-intron structure as the *Tbx16* genes from other animals. Sequences are taken from: NM_131052 (Drer-tbx6); NM_131058 (Drer-tbx16); ENSORLG00000014344 (Olat-tbx16a); ENSGACG00000012454 (Gacu-tbx16a); conceptual translation of the genomic DNA (Olat-tbx16b and Gacu-tbx16b); XM_001377868 (Mdom-Tbx16); and NM_001030367 (Ggal-Tbx6L).

(B) Local synteny trace diagram showing the paralogy between *tbx16a* and *tbx16b* genes of the stickleback. Genes and chromosomes are represented by short vertical lines and long horizontal lines in the diagram, respectively. Only those genes that are orthologous to at least one other gene on another chromosome are identified by names. Orthologous correspondences between chicken and stickleback genes are represented by thin lines. Chicken and stickleback genes showing a 1-to-1 rather than the expected 1-to-2 ratio of correspondences (indicating the loss of one of the duplicates in the stickleback genome) are indicated by colors. Hatched lines in stickleback chromosomes represent the presence of numerous other genes in the corresponding

intervals. Presence of two *Gstt1* genes in chicken chromosome 15 is likely to be caused by a duplication of *Gstt1* gene in the chicken lineage. Distances between genes are not drawn in scale. Note that in terms of the local synteny, *tbx16a* and *tbx16b* genes are occupying equivalent positions in the stickleback genome.

(C) Scenario for the evolution of *tbx16*-containing regions of the stickleback (and medaka) genome by a balanced gene loss. After the duplication of the whole genome, Groups XIII and XIV had lost a copy of *crkl* and copies of *klhl22* and *cabin1* genes, respectively. Present-day configuration of genes results from the rearrangement of genes within each region after the loss.

Fig. S7

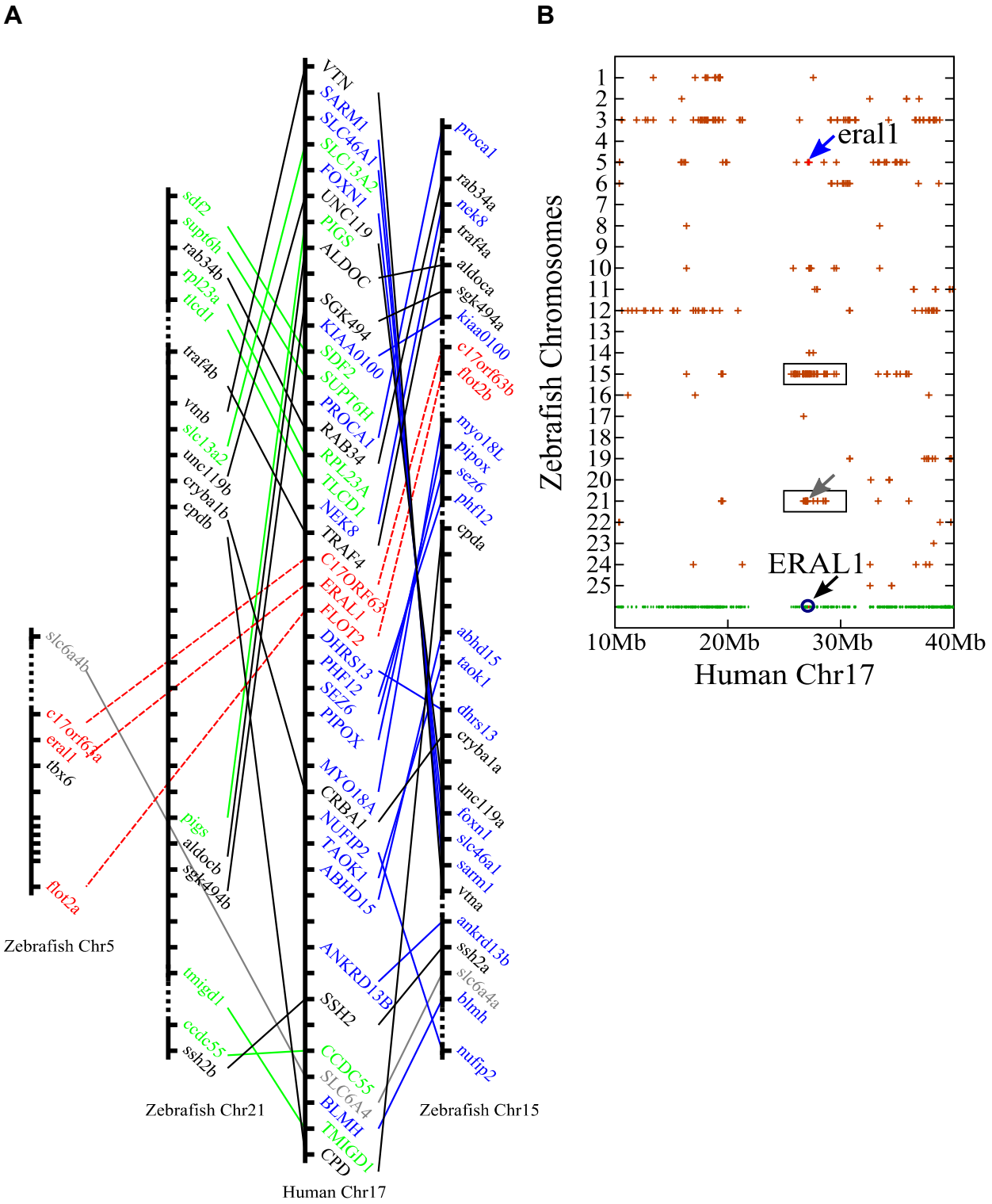


Figure S7. Translocation of the *tbx6*-containing chromosomal region in zebrafish.

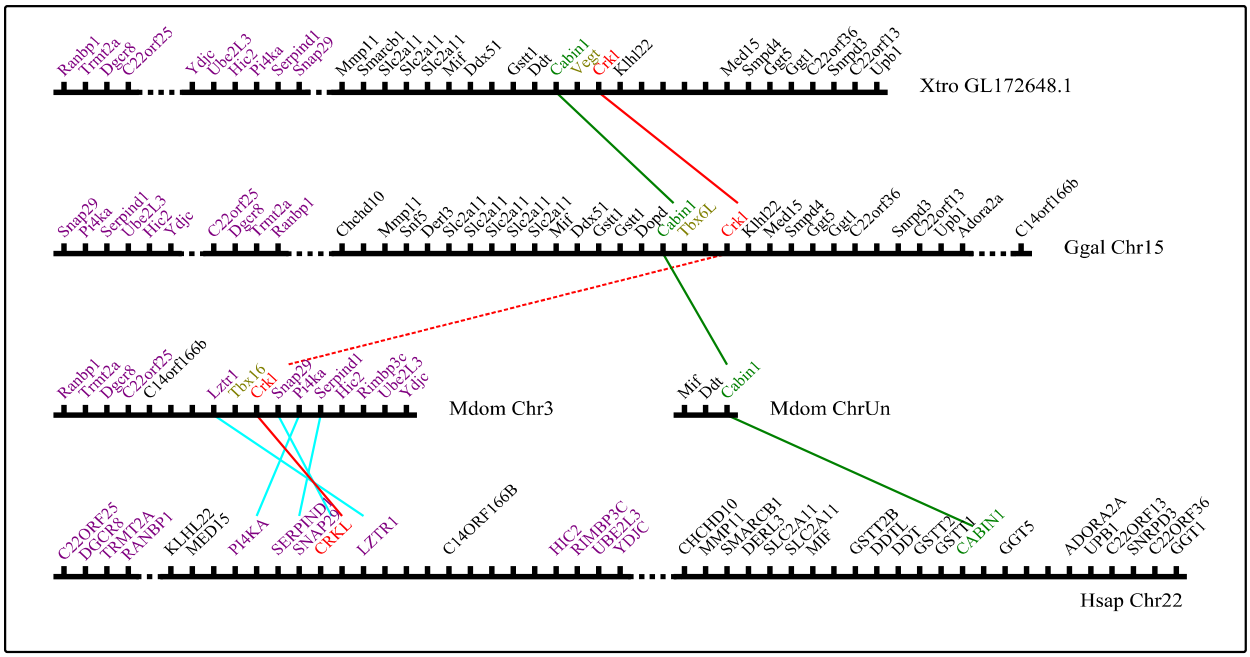
(A) Local synteny trace diagram showing the orthology between human genes neighboring the *C17ORF63-ERAL1-FLOT2* gene cluster (red) and zebrafish genes on the corresponding regions of zebrafish genome. Genes and chromosomes are represented by short vertical and long horizontal lines in the diagram (for ease of presentation, the whole diagram has been rotated 90° in a clockwise direction). Only the genes that are orthologous to at least one other gene on another chromosome are identified by names. Orthologous correspondences between human and zebrafish genes are represented by thin lines. Human and zebrafish genes showing a 1-to-1 rather than the 1-to-2 ratio of correspondences (indicating the loss of one of the duplicates in zebrafish genome) are marked by colors. Hatched lines in zebrafish chromosomes represent the presence of a large number of other genes in the corresponding intervals. Short hatched lines in zebrafish chromosome 15 denote the intervals containing less than 10 genes. Distances between genes are not drawn in scale. Tight packing of 6 genes on zebrafish chromosome 5 represents the tandem duplication of a *pim3*-like gene in that region. Note that in the genomic regions shown here zebrafish chromosomes 15 and 21 effectively form a paralogous pair. Notice that, over the interval containing *C17ORF63*, *ERAL1*, and *FLOT2* genes, orthologous correspondences between human and zebrafish genes suddenly shift from between human chromosome 17 and zebrafish chromosome 21 (solid thin lines) to between human chromosome 17 and zebrafish chromosome 5 (hatched thin lines).

(B) Dot plot diagram showing the chromosomal distribution of zebrafish genes related to the human chromosome 17 genes. Human genes are represented by green dots at the bottom row which denote the chromosomal locations of human genes in terms of the distances from the telomere. Only the genes in 10-40Mb region are shown. Zebrafish genes considered orthologous to the human chromosome 17 genes are represented by crosses drawn above the corresponding human genes on a series of invisible parallel lines representing individual zebrafish chromosomes. Note that crosses represent just the presence of zebrafish orthologs, not the actual locations of those orthologs, on a particular zebrafish chromosome. Zebrafish *eral1* gene is represented by a red cross on chromosome 5. Human *ERAL1* gene is shown at the bottom row as a red dot surrounded by a blue circle. The apparent absence of human genes in 22-26Mb region is due to the presence of centromeric sequences in that region. Note that

human genes neighboring *ERAL1* have their zebrafish orthologs mostly on chromosomes 15 and 21 (boxed regions), although *ERAL1* itself and its immediate neighbors *C17ORF63* and *FLOT2* have their zebrafish orthologs on chromosome 5 (blue arrow). Notice that zebrafish chromosome 21 has a small break in synteny (grey arrow) indicating the absence of the orthologs of *C17ORF63*, *ERAL1*, and *FLOT2* genes on this chromosome.

Fig. S8

A



B

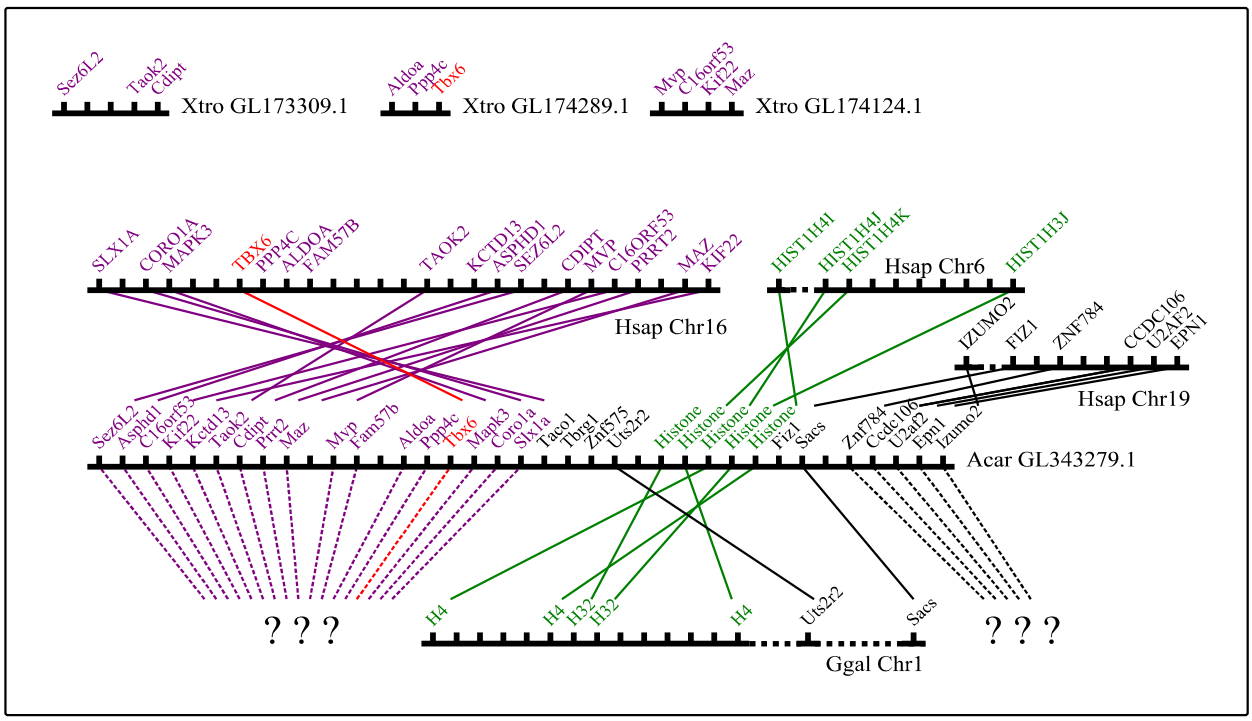


Figure S8. Putative locations of the missing *Tbx6/16* genes in human and chicken genomes.

(A) Location of the *Tbx16* ortholog in the human genome. In non-mammal tetrapods, such as frog (Xtro) and chicken (Ggal), *Tbx16* genes (*Vegt* in frog and *Tbx6L* in chicken: olive) are found between *Cabin1* (green) and *Crkl* (red) genes. These three genes (*Cabin1*, *Tbx16*, and *Crkl*) are in turn nestled within a larger syntenic cluster (genes written in black color) showing a conserved gene order between frog and chicken. In mammals, the *Cabin1-Tbx16-Crkl* gene cluster is broken into two, such that *Tbx16-Crkl* genes are now located in another genomic region (genes written in purple color), which is well-separated from *Cabin1* that has remained in its original genomic position (surrounded by genes written in black color). In humans (Hsap), this new, "mammalian" genomic neighborhood of *Tbx16* (as exemplified by the *Tbx16*-containing region of opossum (Mdom) genome) is preserved largely intact, albeit with some scrambling of gene orders (thin cyan lines). This indicates that the remnant (or the "ghost") of the *Tbx16* gene in humans would be found in the region between *CRKL* and *LZTR1* genes. Note that even though *CRKL* is now separated from *CABIN1* in humans (*CRKL* is located about 3.1Mb away from *CABIN1* in the human genome), these genes are still on the same chromosome (chromosome 22) and therefore the putative chromosomal location of the *Tbx16* gene in humans does not change in a search if *CABIN1* instead of *CRKL* is assumed to be its nearest genomic neighbor (as is the case in zebrafish).

(B) Location of the *Tbx6* ortholog in the chicken genome. In the genome of green anole lizard (Acar), which, among tetrapods with sequenced genomes, is the closest living relative of birds, *Tbx6* is located next to *Ppp4c* (which is also true in human (Hsap), opossum/wallaby (data not shown), and frog (Xtro) genomes). These two genes in turn form an integral part of a large gene cluster (genes written in purple color) which has a conserved synteny (thin purple lines) with the genes located in the *TBX6*-containing region of the human chromosome 16. Curiously, in the current assembly of the chicken genome, *Tbx6* as well as a large number of other genes in its genomic neighborhood are not be found (represented by thin hatched lines leading to the question marks), indicating that—since it is unlikely that chicken genome is really devoid of that many genes—*Tbx6* and its genomic neighbors must be located in one of the genomic regions that are failed to be sequenced (see International Chicken Genome Sequencing Consortium 2004). Moreover, since these genes seem to be missing in the genomes of other birds as well (data not

shown), it is likely that the apparent absence of these genes in the chicken genome is not due to any stochastic failure in sequencing or assembly of clones but rather due to the physical property of the *Tbx6*-containing region of the bird genome that makes the cloning and sequencing particularly challenging (see International Chicken Genome Sequencing Consortium 2004).

Unfortunately, at present, we cannot even be certain of the chromosome on which *Tbx6* may reside in the chicken genome. In the genome of green anole lizard, *Tbx6* and its neighbors appear to be syntenic to the histone gene complex (green), which, in the chicken genome, is located on the chromosome 1. However, we found that, on Acar scaffold GL343279.1, there is a large sequencing gap in the intergenic region between *Slx1a* and *Taco1*, suggesting a possible mis-assembly of sequences in the lizard genome. Because of this, it cannot be assumed with confidence that the *Tbx6*-containing region would also be syntenic to the histone gene complex in the chicken genome. Note that another chromosome in the chicken genome that may more likely harbor *Tbx6* and its genomic neighbors is chromosome 14, which contains the vast majority (about 87%) of the chicken orthologs of human genes from the p-arm of human chromosome 16 (*TBX6* is located on the p-arm, in a region close to the centromere).

Fig. S9

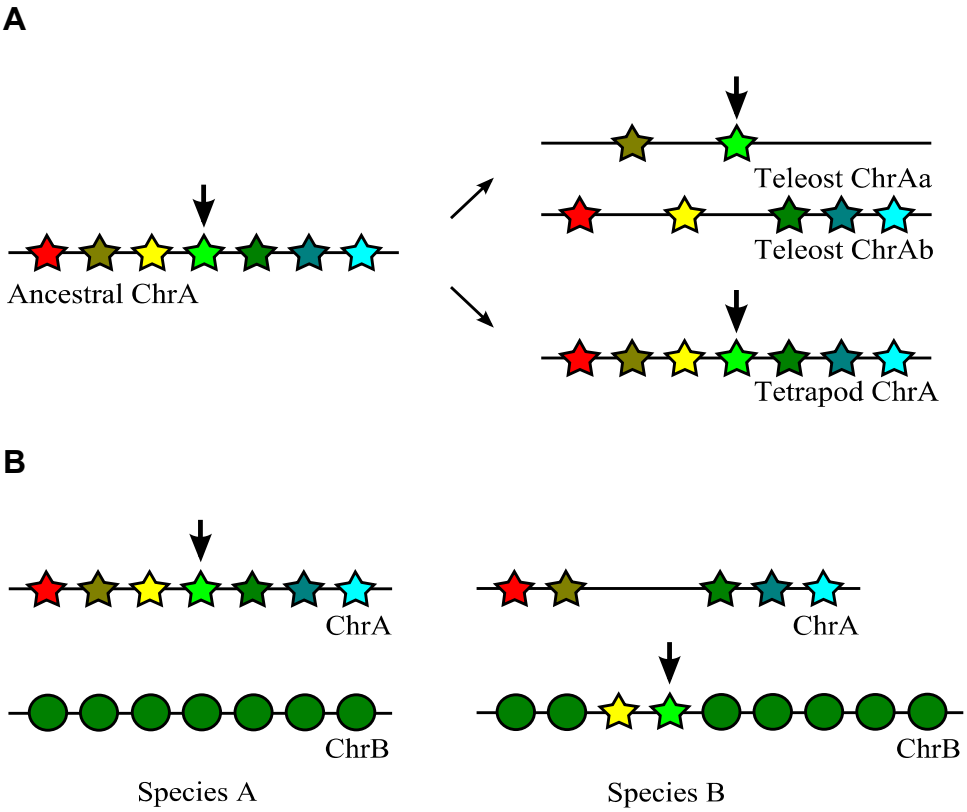


Figure S9. Reduction of syntenies by loss or translocation of genes.

Assignment of orthologies by synteny criteria may become difficult in cases of gene loss or translocation of genes if the evolutionary history of the corresponding genomic regions is not taken into consideration.

(A) Reduction of synteny by a balanced but unequal loss of genes after the duplication of a genome. In this scenario, duplicated chromosomes (or regions of chromosomes) lose genes in a complementary manner but with heavier losses in one chromosome. In this case, focusing only on the regions containing the genes of interest (arrows) may lead to a correct but misleading conclusion that there is only a weak conservation of synteny between the regions surrounding the genes of interest.

(B) Reduction of synteny by a small-scale translocation of genes. Again, focusing only on the

genes of interest (arrows) and surrounding regions may lead to a correct but still misleading conclusion based on the comparison between chromosome A of species A and chromosome B of species B that these genes share few common neighborhood genes.

Supplementary Materials References

Catchen JM, Conery JS, Postlethwait JH. 2009. Automated identification of conserved synteny after whole-genome duplication. *Genome Res* 19: 1497-1505.

Gascuel O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol* 14: 685-695.

International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695-777.

Kasahara M, Naruse K, Sasaki S, et al. (38 co-authors). 2007. The medaka draft genome and insights into vertebrate genome evolution. *Nature* 447: 714-719.

Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. *Mol Biol Evol* 25: 1307–1320.

Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. 2009. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* 19: 327-335.