Dev Genes Evol (2008) 218:613–628
DOI 10.1007/s00427-008-0249-5

ORIGINAL ARTICLE

Conservation of linkage and evolution of developmental function within the *Tbx2/3/4/5* subfamily of T-box genes: implications for the origin of vertebrate limbs

Amy C. Horton · Navin R. Mahadevan · Carolina Minguillon · Kazutoyo Osoegawa · Daniel S. Rokhsar · Ilya Ruvinsky · Pieter J. de Jong · Malcolm P. Logan · Jeremy J. Gibson-Brown

Received: 8 May 2008 / Accepted: 5 September 2008 / Published online: 25 September 2008

Abstract T-box genes encode a family of DNA-binding transcription factors implicated in numerous developmental processes in all metazoans. The *Tbx2/3/4/5* subfamily genes are especially interesting because of their key roles in the evolution of vertebrate appendages, eyes, and the heart, and, like the Hox genes, the longevity of their chromosomal linkage. A BAC library derived from the single male amphioxus (*Branchiostoma floridae*) used to sequence the amphioxus genome was screened for *AmphiTbx2/3* and *AmphiTbx4/5*, yielding two independent clones containing both genes. Using comparative expression, genomic linkage, and phylogenetic analyses, we have reconstructed the evolutionary histories of these members of the T-box gene family. We find that the *Tbx2–Tbx4* and *Tbx3–Tbx5* gene pairs have maintained tight linkage in most animal lineages since their birth by tandem duplication, long before the divergence of protostomes and deuterostomes (e.g., arthropods and vertebrates) at least 600 million years ago, and possibly

Communicated by N. Satoh

Electronic supplementary material The online version of this article (doi:10.1007/s00427-008-0249-5) contains supplementary material, which is available to authorized users.

A. C. Horton · N. R. Mahadevan · J. J. Gibson-Brown (✉)
Department of Biology, Washington University in St. Louis,
1 Brookings Drive,
St. Louis, MO 63130, USA
e-mail: gibbroster@gmail.com

C. Minguillon · M. P. Logan
Division of Developmental Biology,
National Institute for Medical Research,
The Ridgeway, Mill Hill,
London NW7 1AA, UK

K. Osoegawa · P. J. de Jong
BACPAC Resources,
Children's Hospital Oakland Research Institute,
747 52nd Street,
Oakland, CA 94609, USA

D. S. Rokhsar
Department of Computational Genomics,
Joint Genome Institute of the US Department of Energy,
2800 Mitchell Avenue,
Walnut Creek, CA 94598, USA

I. Ruvinsky
Department of Ecology and Evolution, University of Chicago,
1101 East 57th Street,
Chicago, IL 60637, USA

Present addresses:
A. C. Horton
Genome Sequencing Center, Department of Genetics,
Washington University in St. Louis School of Medicine,
4444 Forest Park Ave.,
St. Louis, MO 63108, USA

C. Minguillon
CSIC–Institut de Biologia Molecular de Barcelona,
Parc Cientific de Barcelona,
C/ Josep Samitier 1-5,
08028 Barcelona, Spain

J. J. Gibson-Brown
Institute for Evolutionary Discovery,
909 Hiawatha Drive,
Mount Pleasant, MI 48858, USA

before the divergence of poriferans and cnidarians (e.g., sponges and jellyfish). Interestingly, we find that the gene linkage detected in all vertebrate genomes has been maintained in the primitively appendage-lacking, basal chordate, amphioxus. Although all four genes have been involved in the evolution of developmental programs regulating paired fin and (later) limb outgrowth and patterning, and most are also implicated in eye and heart development, linkage maintenance—often considered due to regulatory constraints imposed by limb, eye, and/or heart associated gene expression—is undoubtedly a consequence of other, much more ancient functional constraints.

Keywords T-box · Amphioxus · Limb · Evolution · Development

Introduction

T-box genes encode an ancient family of transcription factors involved in the development of multicellular animals as divergent as sponges and humans (Bollag et al. 1994; Adell and Muller 2005; Naiche et al. 2005). In vertebrates, the genes can be divided into eight subfamilies, all of which are represented in the stem chordate amphioxus (Ruvinsky et al. 2000b). The proteins all possess a highly conserved, sequence-specific, DNA-binding domain as well as divergent *trans*-regulatory domains characteristic of each subfamily. Among the first to be identified (Agulnik et al. 1996), the closely related *Tbx2/3* and *Tbx4/5* subfamily genes are especially interesting because of their role in the evolution of novel vertebrate structures including limbs (Ruvinsky and Gibson-Brown 2000).

Tbx2 and *Tbx3* are expressed in similar patterns at the anterior and posterior margins of both forelimbs and hindlimbs (Gibson-Brown et al. 1996, 1998a, b). Loss-of-function and misexpression studies suggest that both genes play a role in the anteroposterior patterning of digits (Suzuki et al. 2004), and *Tbx3* may be involved in limb positioning (Rallis et al. 2005). Because of their limb-specific expression in forelimbs and hindlimbs, *Tbx4* and *Tbx5* have been intensively studied for their roles in limb induction and the evolution of limb-specific morphologies. Evidence for their role in limb induction and patterning comes from four lines of evidence: (1) mutational studies, either experimentally induced or naturally underlying a congenital human birth defect, (2) ectopic limb induction experiments, (3) ectopic tissue grafting experiments, and (4) ectopic misexpression studies.

(1) In zebrafish, morpholino-induced knockdown of *tbx5* results in the loss of pectoral fins by inhibiting the migration of lateral plate mesoderm cells to the future site of bud eruption (Ahn et al. 2002). While *Tbx4* null mice show normal hindlimb field and bud initiation, hindlimbs fail to develop, possibly due to a failure of *Tbx4*-mediated *Fgf10* promotion and *dHand* repression (Naiche and Papaioannou 2003). Additionally, mutations in human *TBX4* and *TBX5* cause hindlimb- and forelimb-specific defects in small-patella (OMIM #147891) and Holt–Oram (OMIM #142900) syndromes, respectively (Basson et al. 1997; Bongers et al. 2004; Li et al. 1997). (2) Ectopic limb induction studies in the chick have shown that *Tbx4* and *Tbx5* expression correlates with the identity of the limb that subsequently develops. Ectopic limbs induced close to the wing predominantly express *Tbx5*, with *Tbx4* expression restricted to the posterior margin, and develop into wing-like limbs. Medially induced limbs express *Tbx5* anteriorly, *Tbx4* posteriorly, and become limbs of highly mosaic identity. Limbs induced near the level of the hindlimb primarily express *Tbx4*, with minimal *Tbx5* expression anteriorly, and develop into leg-like limbs (Gibson-Brown et al. 1998a; Isaac et al. 1998; Logan et al. 1998; Ohuchi et al. 1998). (3) Forelimb bud mesenchyme grafted to the hindlimb, and vice versa, develops structures (fingers or toes) appropriate to the tissue of origin, and is not regulated by the host tissue (Gibson-Brown et al. 1998a; Isaac et al. 1998; Logan et al. 1998; Ohuchi et al. 1998). (4) When coupled with FGF misexpression, ectopic misexpression of *Tbx5* in the hindlimb can result in the conversion of a leg to a wing-like structure. Conversely, ectopic *Tbx4* expression combined with FGF misexpression in the forelimb field leads to the development of a limb with a leg-like morphology (Rodriguez-Esteban et al. 1999; Takeuchi et al. 1999). It should be noted, however, that recent studies in mice have called into question the roles of these limb-specific genes in the determination of limb-specific morphologies. Minguillon et al. (2005), Naiche and Papaioannou (2006), and Hasson et al. (2007) have shown that *Tbx5* and *Tbx4*, while (respectively) necessary to initiate or maintain limb outgrowth, do not confer limb-specific morphologies on the appendages that subsequently develop, an observation in direct contrast to the experiments in chicks (Rodriguez-Esteban et al. 1999; Takeuchi et al. 1999). The reason for this discrepancy currently remains unclear, but might be related to the coincident reprogramming of axial identity in lateral plate mesoderm by ectopic expression of FGF in the chick flank (Cohn et al. 1997). Additional studies of Hox gene expression in response to ectopic FGF misexpression in chick lateral plate mesoderm might help resolve this enigma.

Phylogenetic analyses have indicated that *Tbx2*/*Tbx3* and *Tbx4*/*Tbx5* are paralogous gene pairs in vertebrates: each pair derives from the tandem duplication of an ancestral *Tbx2/3/4/5* locus (Agulnik et al. 1996). Linkage analysis in mice originally revealed that the *Tbx2–Tbx4* and *Tbx3–Tbx5* gene pairs occupy closely linked loci on different

chromosomes (Agulnik et al. 1996). These data, combined with the fact that paralogous flanking genes map to the same chromosomal regions (Ruvinsky and Silver 1997), indicated that the individuated *Tbx2*, *Tbx3*, *Tbx4*, and *Tbx5* genes arose from their linked *Tbx2/3* and *Tbx4/5* precursors by the duplication of a large chromosomal region within the vertebrate lineage, possibly during the first of two whole-genome duplications—long speculated, but now confirmed (Dehal and Boore 2005; Putnam et al. 2008)—at the base of the vertebrate lineage. The tandem duplication, cluster dispersion hypothesis is further supported by more recent phylogenetic analyses of the subfamily. The strongly supported placement of *Drosophila bifid* (*omb*) and *Caenorhabditis tbx2* within the *Tbx2/3* clade confirms that the initial *Tbx2/3/4/5* tandem duplication predated the divergence of protostomes and deuterostomes, with cluster duplication postdating this divergence within the chordate lineage. Discovery of a *Tbx2/3* gene in the urochordate, *Ciona intestinalis*, constrains the timing of cluster duplication to after the divergence of ascidians and vertebrates (Dehal et al. 2002; Delsuc et al. 2006). The characterization of distinct *Tbx4* and *Tbx5* genes in sharks (Tanaka et al. 2002), the most basally branching extant jawed vertebrates (gnathostomes), confirms that the *Tbx2/3*, *Tbx4/5* cluster duplication that gave birth to individual *Tbx2*, *Tbx3*, *Tbx4*, and *Tbx5* genes occurred before the origin of gnathostomes.

The cephalochordate, amphioxus, is the most informative extant organism with which to study the evolutionary origins of vertebrate genomes and developmental gene functions because its phylogenetic position and stem chordate-like morphology make it more similar to the last common invertebrate ancestor of the vertebrates than any other living organism (Bourlat et al. 2006; Putnam et al. 2008). Amphioxus closely resembles vertebrates, both morphologically and developmentally, possessing a notochord, a hollow dorsal neural tube, segmented muscles (embryologically derived from somites), and a perforated pharyngeal (branchial) region, making comparative gene expression/functional studies and homology assignments both rational and informative. Moreover, its genomic structure closely resembles that of vertebrates (Putnam et al. 2008), which allows the confident determination of orthology/paralogy relationships between chordate genes. Amphioxus contains representatives of each of the eight vertebrate T-box gene subfamilies, including a *Tbx2/3* and a *Tbx4/5* gene pair, definitively placing cluster duplication after the divergence of amphioxus and vertebrates (Ruvinsky et al. 2000a). By integrating the genomic, developmental expression and functional data of numerous *Tbx2/3/4/5* genes from a wide range of organisms with embryonic expression and genomic linkage data from amphioxus, we have attempted to reconstruct the evolutionary history of the *Tbx2/3* and *Tbx4/5* paralogous gene pairs since their basal metazoan origins.

Materials and methods

Amphioxus collection

Adult amphioxus (*Branchiostoma floridae*) were collected by shovel and sieve off the south shore of Courtney Campbell Causeway, or west shore of Picnic Island, in Old Tampa Bay (Tampa, FL) during the summer spawning season of 2003. Ripe adults were induced to spawn in the laboratory by electrical stimulation, their gametes collected, and *in vitro* fertilizations performed as described previously (Holland and Holland 1993). Embryos and larvae were raised in filtered seawater from the collection site at 23°C. Embryos were fixed at various stages from zygotes to 4-day larvae in 4% paraformaldehyde in MOPS buffer and stored at −20°C in 70% ethanol until use for whole-mount *in situ* hybridization. Several large (~6 cm), gravid males were live-shipped overnight in fresh seawater to BACPAC Resources, Children's Hospital Oakland Research Institute (CHORI) for genomic DNA extraction. The DNA from the gonads of one of these animals was used to construct BAC library CHORI-302 (see below) and sequence the amphioxus genome (Putnam et al. 2008).

RNA purification

Staged embryos were homogenized in 4 M guanidinium isothiocyanate, 25 mM Na citrate (pH 7.0), 0.5% sarkosyl, 0.1 M β-mercaptoethanol and stored at −80°C until use. For purification, 0.1 ml 2 M Na acetate (pH 4.0) was added to 1.0 ml embryo–guanidinium solution, followed by 1.0 ml H_2O-equilibrated phenol containing 0.1% hydroxyquinoline and 0.1 ml chloroform/isoamyl alcohol 49:1. The solution was chilled on ice for 15 min and spun at 10,000 ×*g* at 4°C for 20 min. Phenol/chloroform/isoamyl alcohol extraction of the supernatant was followed by isopropanol precipitation and resuspension of the pellet in water.

Construction of an amphioxus BAC library

BAC library CHORI-302, derived from sperm from a single adult male amphioxus, was constructed by partially digesting genomic DNA with EcoRI and EcoRI methylase. Size selection of the restriction fragments using pulsed-field gel electrophoresis was followed by ligation into pTARBAC2.1 vector. Subsequently, the products were transformed into DH10B (T1 phage-resistant) electro-competent cells (Invitrogen). This library has been arrayed into 384-

well microtiter dishes and gridded onto three, publicly available, high-density nylon filters (http://bacpac.chori.org/amphiox302.htm).

Screening of the BAC library and clone sequencing

Cloning of the amphioxus *Tbx2/3* and *Tbx4/5* genes has been described previously (Ruvinsky et al. 2000b). Gene-specific overgo probes based on variable sequences within the T-box-encoded DNA-binding domains were used to screen the CHORI-302 BAC library according to established protocols (Ross et al. 1999). *Tbx2/3* probe, 5′-CAA CGA CAT CAT GAA GCT TCC CTA CTG TCA C TT CCG CAC CTA-3′, *Tbx4/5* probe, 5′-CAG CGA GAA CAA CAA GTT TGA ACT GAA GAA GAC GTG TTT CAG-3′. BAC clones were shotgun sequenced by the Joint Genome Institute of the US Department of Energy.

Phylogenetic analysis

All metazoan T-box-encoded protein sequences retrieved through NCBI with lowest expect values on reciprocal BLASTP searches (query-human database and human-query organism database) for Tbx1, Tbx2, Tbx3, Tbx4, and Tbx5 were used in the initial alignment and analyses. Sequences with BLASTP expect values within 10^{-12} of the value for one of these genes were also included. Alignments were generated by CLUSTALW followed by manual editing to remove gaps and unalignable regions. The final alignment contained 188 amino acids spanning the T-box-encoded DNA-binding domain (the 'T-domain', Müller and Herrmann 1997) and its flanking regions (Fig. 1 of the Electronic supplementary material). Initial trees were reconstructed using default substitution matrices and other parameters in neighbor-joining (BIONJ, Gascuel 1997) and maximum likelihood (PhyML, Guindon and Gascuel 2003; Guindon et al. 2005; Animasova and Gascuel 2006) algorithms available through LIRRM (Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier) with 100 bootstrap replicates (BIONJ) or approximate likelihood-ratio test (PhyML). Sequences with multiple gaps in the final alignment were examined in initial trees, but removed from later analyses. Additionally, an echinoderm (XP_797010) and a ctenophore (comb jelly) sequence (ABL68081) were removed from further analyses as both methods, using a variety of substitution models, indicated long branches for these sequences: they always grouped together, but never grouped robustly within any of the Tbx1/10, Tbx2/3, or Tbx4/5 clades. Long Branch Attraction (Felsenstein 1978) has been definitively shown to cause the reconstruction of incorrect tree topologies across the phyla in question (Philippe et al. 2005). Vertebrate sequences that clearly fell within established groups were also pruned to just the few representatives necessary to establish the relationships of T-box genes of earlier diverging taxa. The final tree was based on 100 (PhyML) bootstrap replicates, with percentage support for each node of 500 (BIONJ) replicates shown beneath ML bootstrap support. Accession numbers: At *omb* (*Achaearanea tepidariorum*, common house spider, BAD16722), Bf *Tbx2/3* (AAG34888), Bf *Tbx4/5* (ABU50779, plus translation of the upstream flanking region in the genomic scaffold), Ce *tbx-2* (NP_498088), Ci *Tbx2/3* (NP_001027620), Dm *bi/omb* (NP_525070), Dr *tbx2b* (NP_571126), Dr *tbx3b* (NP_001095140), Dr *tbx4* (NP_570989), Dr *tbx5* (NP_570990), Gg *Tbx2* (XP_001235321), Gg *Tbx3* (AAC41297), Gg *Tbx4* (NP_001025708), Hs *TBX1* (NP_542377, NP_005983 and NP_542378), Hs *TBX10* (NP_005986), Lv *Tbx2/3* (*Lytechinus variegatus*, green sea urchin, AAM81744), Ml *Tbx2/3* (*Mnemiopsis leidyi*, ctenophore, ABL68080, Mm *Tbx2* (NP_033350), Mm *Tbx3* (NP_035665 and NP_932169), Mm *Tbx4* (NP_035666), Mm *Tbx5* (NP_035667), Mm *Tbx10* (NP_001001320), Nv *TbxA* (*Nematostella vectensis*, sea anemone, XP_001633951), Nv *TbxB* (XP_001633952), Pc *Tbx4/5* (*Podocoryne carnea*, ctenophore, CAE45765), Pp *Tbx2/3* (*Pleurobrachia pileus*, ctenophore, CAE45769), Sc *Tbx3* (*Scyliorhinus canicula*, shark, ABM89506), Sd *Tbx2* (*Suberites domuncula*, sponge, CAD66613), Sk *tbx2/3* (*Saccoglossus kowalevskii*, hemichordate, ABD97269), Sp *Tbx2/3* (*Strongylocentrotus purpuratus*, purple sea urchin, XP_779897), Ta *Tbx2/3* (*Trichoplax adhaerens*, placozoan, CAD70270). Statistical tests of alternative phylogenies—one-sided Kashino–Hasegawa test based on pairwise Shimodaira–Hasegawa tests (Shimodaira and Hasegawa 1999; Goldman et al. 2000; Kishino and Hasegawa 1989; Ota et al. 2000)—were conducted using Tree-Puzzle 5.2 (Schmidt et al. 2002) using Dayhoff substitution model and Gamma distributed rates in four categories estimated from the data set (as in defaults for PhyML). Probabilities reported in the text are normalized for multiple tests using the conservative Bonferroni correction.

Intron–exon structure determination

Data retrieved from NCBI MapViewer (http://www.ncbi.nlm.nih.gov/mapview/) and/or TBLASTN (http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html) versus genomic contig sequence was used to determine intron–exon structure (summarized in Fig. 1 and Table 1 of the Electronic supplementary material). Genomic sequences: Bf *Tbx2/3* and Bf *Tbx4/5* (CHORI302-78M15 contig55); Ci *Tbx2/3* (scaffold 212).

Retrieval of genomic linkage data

The *Tbx2*, *Tbx3*, *Tbx4*, and *Tbx5* genes, and their surrounding loci, were located within NCBI MapViewer

for the human, chimpanzee, rhesus macaque, mouse, rat, and chicken genomes using built-in search capabilities. Accession numbers for these, their intervening and flanking genes (out to adjacent loci conserved in all six species) were used to retrieve their sequence records and determine their orientation. Predicted splice sites were manually checked against consensus (nag/ggc..gt). Protein entries were then subjected to BLASTP searches to determine whether any sequence similarities exist between the adjacent genes across species.

Whole-mount *in situ* hybridization

Whole-mount *in situ* hybridization of fixed embryos and larvae were performed using the protocol of Holland et al. (1996). Digoxigenin-labeled riboprobes were synthesized using the Promega Riboprobe System kit. A full-length *AmphiTbx2/3* cDNA clone in pBluescript SK+ vector (Ruvinsky et al. 2000b) was linearized using XbaI and XhoI to make sense (T7) and antisense (T3) probes, respectively. An *AmphiTbx4/5* clone (#1A) was similarly linearized using SacI and EcoRV. Hybridized specimens were mounted in 80% glycerol and photographed with a Nikon DN100 digital camera on a Nikon E-600 microscope under DIC/Nomarski optics. Hybridized embryos were counterstained in Ponceau S, embedded in Spurr's resin, sectioned transversely (3 μm) using a LKB-Huxley ultramicrotome, and photographed using 40× and 100× (oil-immersion) objectives.

RT-PCR

Primers were designed to span an intron to distinguish between RT-amplified RNA and products derived from potentially contaminating genomic DNA. Primer sequences: Tbx2/3RT forward, GGG ATC AAT TCC ACA CGT ACG and reverse, AAA ACC GTG TTT GTC CGA GAT G (predicted product= 320 bp); Tbx4/5RT forward, GGG AAG GCT GAG CCC GCC and reverse, TCA AAC TTG TTG TTC TCG CTG G (predicted product=210 bp). Reactions were performed according to the directions of the OneStep RT-PCR Kit (QIAGEN) using 1 μl of RNA and 15 pmol of each primer under the following conditions: 30 min at 50°C, 15 min at 95°C, followed by 35 cycles of denaturation (30 s at 94°C), annealing (30 s at 62°C), and extension (30 s at 72°C), with a final extension period of 10 min.

Results

Phylogenetic analysis of the *Tbx2/3/4/5* subfamily

Since our previous analyses of the *Tbx2/3/4/5* subfamily (Agulnik et al. 1996; Ruvinsky et al. 2000a, b; Horton et al. 2003), additional sequences have become available from many phylogenetically informative, basally diverging taxa (Dehal et al. 2002; Adell et al. 2003; Gross et al. 2003; Martinelli and Spring 2003, 2005; Adell and Muller 2005; Akiyama-Oda and Oda unpublished [BAD16722], Lowe et al. 2006; Scholz and Technau unpublished [AAQ23383], O'Neill et al. 2007; Yamada et al. 2007). A reanalysis of the phylogenetic relationships within this subfamily was therefore performed (Fig. 1 and Fig. 1 of the Electronic supplementary material). Additionally, statistical tests comparing the tree topology in Fig. 1 to alternative hypotheses of biological significance—including lineage-specific duplications, alternate orthology assignments, and the relative timing of duplications—were also performed. Although this tree supports several non-biological groupings (e.g., placing zebrafish *tbx3b* together with chicken *tbx3* to the exclusion of mouse *Tbx3*), no tree with the established ordering of bilaterian sequences within these clades could be rejected as having significantly lower support (P=0.1112 to 0.1950). Additionally, we can exclude the possibility that *Ciona Tbx2/3*, both *Branchiostoma* genes, echinoderm and hemichordate *Tbx2/3*, and protostome *Tbx2/3* genes are actually members of the opposing subfamily (i.e., they lie within the *Tbx4/5* clade rather than the *Tbx2/3* clade, or vice versa; P<0.00005), or that any of these genes could equally well be assigned to a more basally divergent *Tbx2/3/4/5* clade (P≤0.0488 for amphioxus *Tbx2/3*, P≤0.0215 for all the other genes). This means not only that all the bilaterian orthology assignments are highly supported but also that statistical evaluation of the placements of *Drosophila* and spider *bi/omb* and *C. elegans tbx-2*, as well as *Ciona Tbx2/3*, and the absence of a *Tbx4/5* ortholog in the sequenced *Ciona*, *Drosophila*, and *Caenorhabditis* genomes strongly suggest that both arthropods and nematodes (i.e., possibly all ecdysozoans) and, independently, urochordates (tunicates) have lost their *Tbx4/5* orthologs. Why these animals can develop normally without their *Tbx4/5* genes, while these orthologs remain tightly linked in cephalochordates and vertebrates, remains a mystery.

Close linkage of *AmphiTbx2/3* and *AmphiTbx4/5* in amphioxus

BAC library CHORI-302, derived from a single male specimen of *B. floridae*, was screened with overgo probes (Cai et al. 1998; Han et al. 2000) against multiple developmental genes to evaluate the *X*-fold coverage of the genome by this library prior to whole-genome shotgun sequencing (Gibson-Brown et al. 2003; Putnam et al. 2008). Screens using overgo probes for *AmphiTbx2/3* and *AmphiTbx4/5* yielded multiple positive clones, including two (62D14 and 78M15) that hybridized to both probes

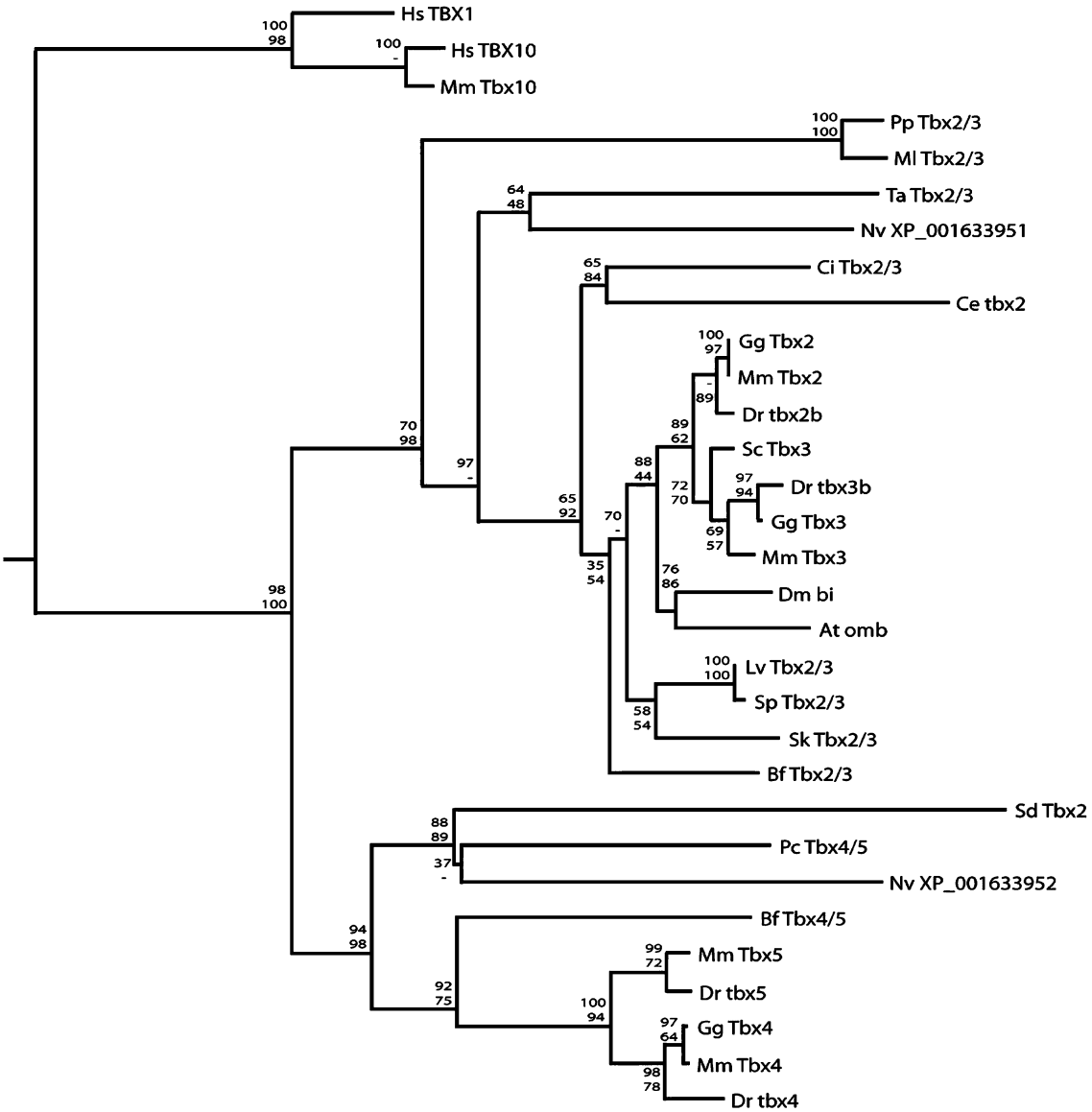

Fig. 1 Maximum likelihood (ML) tree of the *Tbx2/3/4/5* subfamily of T-box genes. *Numbers* indicate percentage of 100 ML bootstrap replicates supporting node (*top*); percentage of 500 neighbor-joining bootstrap replicates supporting node (*bottom*). Abbreviations: *At Achaearanea tepidariorum* (common house spider); *Bf*, *Branchiostoma floridae*; *Ce*, *Caenorhabditis elegans*; *Ci*, *Ciona intestinalis*; *Dm*, *Drosophila melanogaster*; *Dr*, *Danio rerio*; *Gg*, *Gallus gallus*; *Hs*, *Homo sapiens*; *Lv*, *Lytechinus variegatus* (green sea urchin); *Ml*, *Mnemiopsis leidyi* (ctenophore); *Mm*, *Mus musculus*; *Nv*, *Nematostella vectensis* (sea anemone); *Pc*, *Podocoryne carnea* (ctenophore); *Pp*, *Pleurobrachia pileus* (ctenophore); *Sc*, *Scyliorhinus canicula* (shark); *Sd*, *Suberites domuncula* (sponge); *Sk*, *Saccoglossus kowalevskii* (hemichordate); *Sp*, *Strongylocentrotus purpuratus* (purple sea urchin); *Ta*, *Trichoplax adhaerens* (placozoan)

(data not shown). This indicated that both genes are tightly linked within the amphioxus genome.

Sequencing and genome structure of *AmphiTbx2/3* and *AmphiTbx4/5*

As part of a pilot study for the Amphioxus Genome Sequencing Project, the Joint Genome Institute of the US Department of Energy sequenced both of the BAC clones (62D14 and 78M15) containing both *AmphiTbx2/3* and *AmphiTbx4/5*. The intron–exon structure of both genes was then determined. *AmphiTbx2/3* and *AmphiTbx4/5* are separated by 49 kb, compared to 9.7 to 270 kb for the corresponding vertebrate loci (Fig. 2, Tables 1 and 2 of the Electronic supplementary material). Early comparisons of the T-box genomic loci focused on the DNA-binding domain at the N-terminal end of the proteins (Wattler et al. 1998). The presence of a second intron (2b) in all of the *Tbx4/5* genes, which was absent from all *Tbx2/3* genes, was noted. We found the same second intron in *AmphiTbx4/5* that was absent from *AmphiTbx2/3*. We also detected an additional 5′ intron and exon conserved between *AmphiTbx4/5*

Fig. 2 Intron–exon structure of the chordate *Tbx2/3* and *Tbx4/5* genes and some of their representative vertebrate homologs. *Boxes* represent exons. *Lines* represent introns and intergenic regions. *Shaded areas* indicate T-box-encoded DNA-binding domains. Introns annotated according to Wattler et al. (1998)

and zebrafish, chicken, mouse, and human *Tbx4* and *Tbx5*, but absent from all of the *Tbx2/Tbx3* genes (Fig. 2). Interestingly, the C-terminal ends of the *Branchiostoma* and *Ciona Tbx2/3* genes differ from their vertebrate counterparts, containing a single large exon rather than two small exons (Fig. 2), suggesting that the intervening intron evolved within the vertebrate stem after the divergence of urochordates.

Conservation of linkage between the *Tbx2–Tbx4* and *Tbx3–Tbx5* loci despite constant local rearrangements

The *Tbx2–Tbx4* and *Tbx3–Tbx5* linked gene pairs are flanked by several conserved loci in all amniotes, including chickens, mice, rats, rhesus macaques, chimpanzees, and humans, confirming that they probably lie within paralogous regions (‘paralogons’, Coulier et al. 2000) derived from whole-genome duplications (Dehal and Boore 2005; Putnam et al. 2008) at the base of the vertebrate lineage (Fig. 3). Genes flanking the *Tbx2–Tbx4* cluster (*Bcas3* and *Brip1*) and the *Tbx3–Tbx5* cluster (*Thrap2* and *Rbm19*) all lie in the same position and orientation in all vertebrate species examined. Interestingly, several other genes and pseudogenes lie between these conserved flanking genes and the two T-box gene pairs. These intervening sequences maintain neither their orientation nor significant sequence similarity across all species, although some are conserved within a subset of closely related species (Fig. 3). Fourteen (of 20) linked loci have no apparent orthologs in this cluster. Of these, five (of eight) pseudogenes and nine (of 12) multiple-exon genes have no obvious orthologs. Of the six loci appearing in multiple amniote species, two are common to primates, three to apes (one of which, the *Naca2* pseudogene, has inverted its orientation), and one to murids (this appears to be a properly spliced, two-exon locus, with the second exon containing a large, but variable, number of AGG repeats, which are unlikely to actually code for proteins given the simplicity of the sequence). This observation is remarkable, as it suggests enormous activity within and adjacent to the T-box clusters in recent times—at least 17 gene cluster insertions/deletions, one inversion, and one local translocation

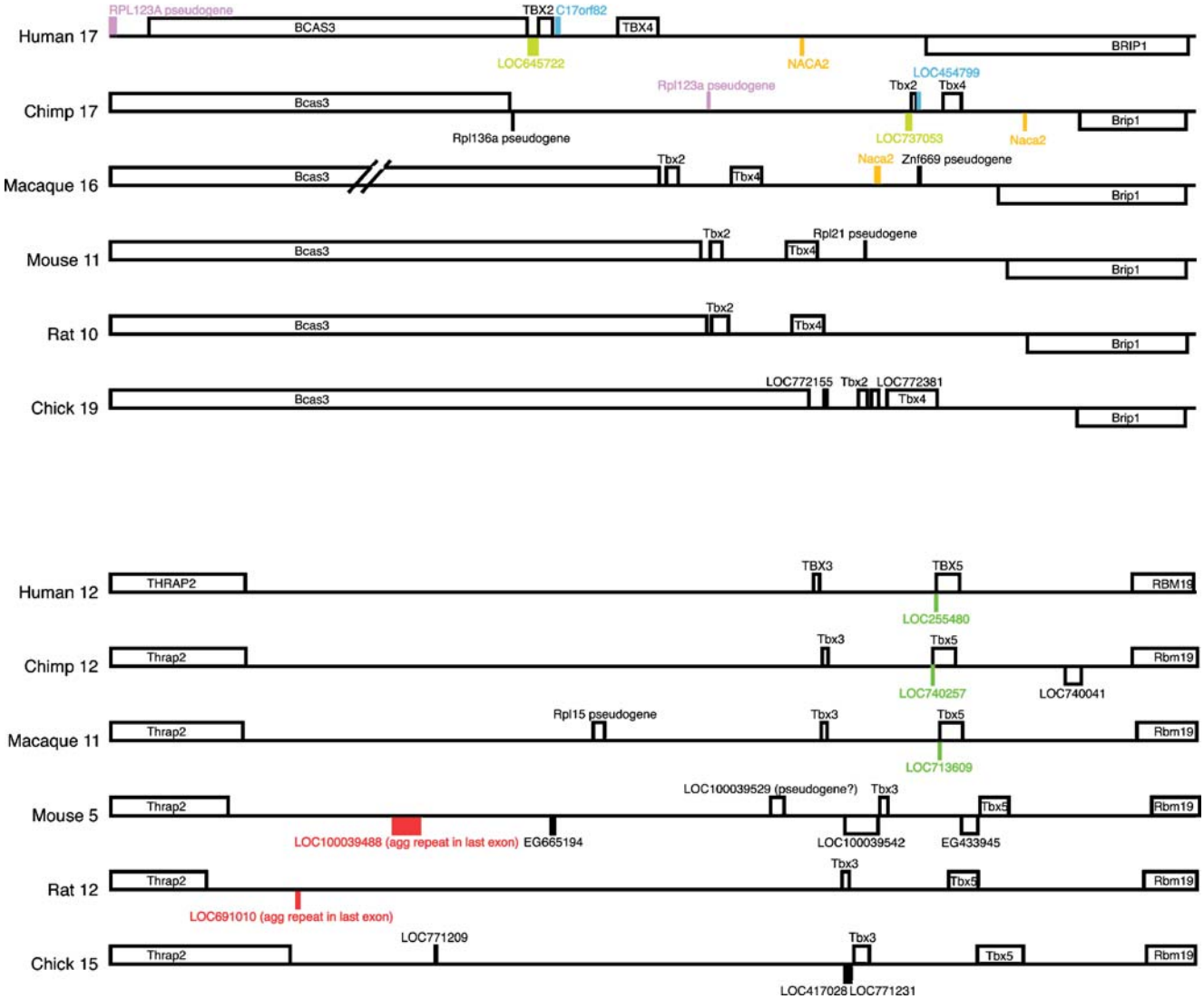

Fig. 3 Genomic map of loci surrounding the *Tbx2/Tbx4* and *Tbx3/Tbx5* loci in six representative amniote species. *Boxes* represent loci including introns (if any). Genes in same orientation as T-box genes are shown *above horizontal lines*, while genes in opposite orientation are placed *below the lines*. Spacing is relative (within cluster) rather than absolute (actual sizes vary between <460 kb and >2.5 Mb). Conserved loci (other than eight named anchoring genes) are displayed and labeled in *lavender*, *aqua*, *lime*, *orange*, *green*, and *red*

within this region since the divergence of humans and chickens, ~300 million years ago—without disrupting the close linkage or orientation of the two T-box gene pairs (or their flanking genes *Thrap2–Rbm19* and *Bcas3–Brip1*). Further, at least two insertions/deletions and one translocation have occurred within the past 6 million years, since the divergence of chimpanzees and humans, indicating that these regions have continued to be subject to local rearrangements over relatively short time scales.

Expression of *Tbx2/3* in amphioxus

Expression of *AmphiTbx2/3* is first detected around the lip of the blastopore during early gastrulation (Fig. 4a). This signal marks invaginating mesendoderm cells and is later seen in the posterior mesendoderm of early and 1-day larvae (Fig. 4b,c). *AmphiTbx2/3* is also expressed in the dorsal neural tube (in the presumptive cerebral vesicle) as well as two rostrocaudally restricted (branchial) regions of ventral gut endoderm (Fig. 4b,c,e–g) and in many, but not all, surface ectoderm cells (Fig. 4). Diffuse expression is also observed in a subset of notochord cells along the entire rostrocaudal axis (Fig. 4c,g). As neurulation progresses, signal persists in posterior mesendoderm, dorsal neural tube, gut endoderm, and surface ectoderm. By the 4-day larval stage, no expression can be detected in any tissue by *in situ* hybridization. RT-PCR data confirms and extends the expression profile revealed by *in situ* hybridization, showing persistent *AmphiTbx2/3* expression from 5 to 56 h post-fertilization (hpf) (Fig. 6).

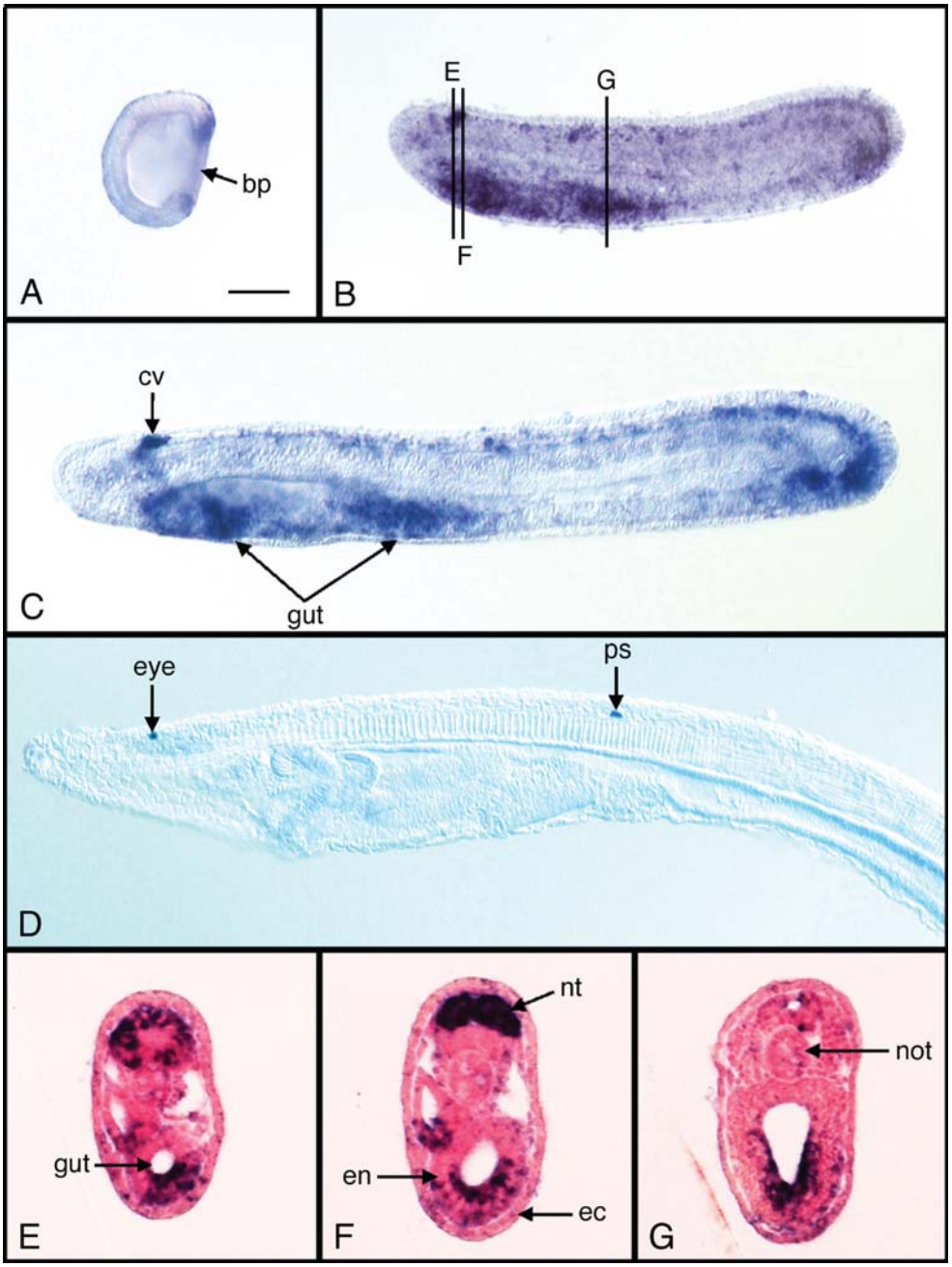

Fig. 4 Embryonic expression of amphioxus *Tbx2/3*. **a** Cap-shaped gastrula (6 h post-fertilization). **b** Lateral view of an early-stage larva (14 hpf). **c** Lateral view of 1-day larva (24 hpf). **d** Four-day larva, signal no longer present. From anterior to posterior, the *dark spots* (melanin granules) indicate positions of the eye and first pigment spot, respectively. **b–d** Anterior, *left*, posterior, *right*. Dorsal, *top*, ventral, *bottom*. **e–g** Transverse sections through an early larva. Planes of section as depicted in (**b**). Dorsal, *top*, ventral, *bottom*. Abbreviations: *bp*, blastopore; *cv*, cerebral vesicle; *ec*, ectoderm; *en*, endoderm; *eye*, eye spot; *not*, notochord; *nt*, neural tube; *ps*, pigment spot. *Scale bar*, **a–d** 50 μm, **e–g** 25 μm

Expression of *Tbx4/5* in amphioxus

Expression of *AmphiTbx4/5* is first detected by *in situ* hybridization in 2.5-day larvae (50–60 hpf) in medial, presumptive mesendoderm of the tail bud and in small bilateral patches of ventral, lateral mesoderm (Fig. 5a,c). Expression persists in the tail bud, while the ventrolateral domains progressively elongate (Fig. 5b), but is no longer seen in 3-day larvae. This very brief window of expression is supported by RT-PCR data, which reveal robust, yet transient, *AmphiTbx4/5* expression at 56 hpf (Fig. 6).

Expression is also seen in cells adjacent to the first pigment spot. This structure, also known as the Organ of Hesse, has a tripartite cell structure, in which two neuronal cells conduct signals from a central, melanin-containing, photoreceptor cell. The *AmphiTbx4/5* signal likely derives from the dendritic processes of the two neuronal cells neighboring the central photoreceptor (Fig. 5e,f).

Discussion

Evolution of the *Tbx2/3/4/5* subfamily genes

The most prominent feature within this subfamily of four vertebrate T-box genes is the conservation of linkage

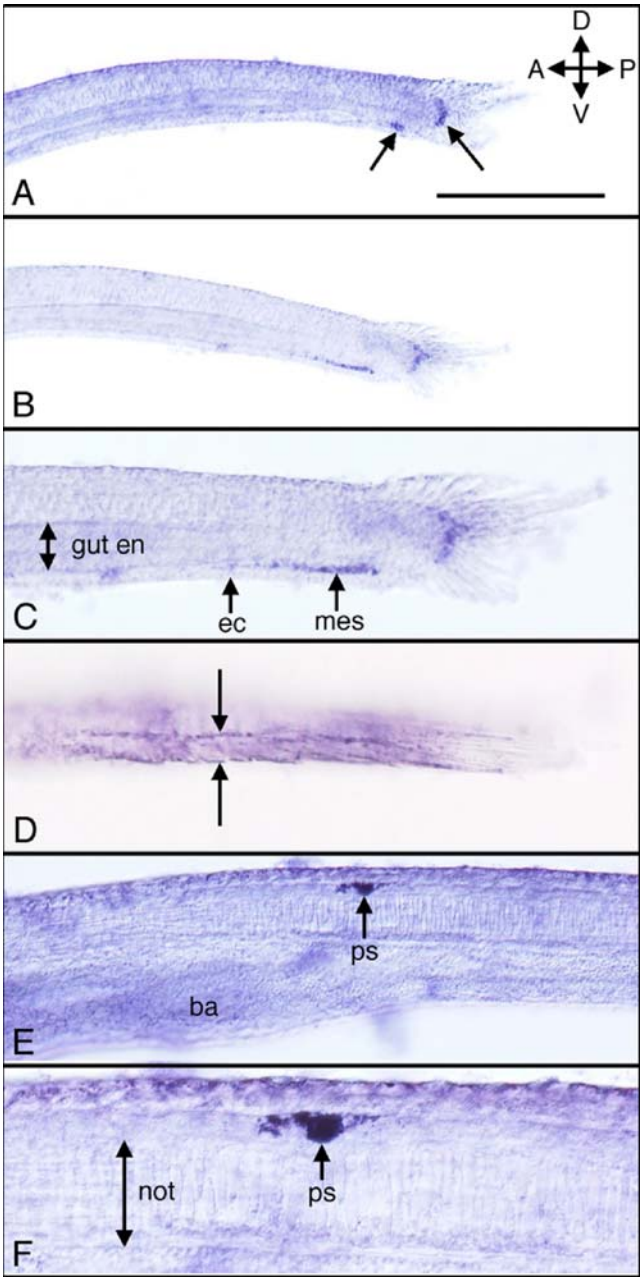

Fig. 5 Expression of amphioxus *Tbx4/5*. **a** *AmphiTbx4/5* is first detected in 2.5-day larvae in presumptive mesendoderm of the tail bud (*right arrow*) and in small bilateral patches of ventral, lateral mesoderm (*left arrow*). **b** Expression persists in presumptive mesendoderm of the tail bud, while the ventrolateral domains progressively elongate. **c** Higher magnification view of tail bud region shown in (**b**). **d** Ventral view of bilateral expression (indicated by *arrows*) within posterior ventrolateral mesoderm. **e** Expression in cellular processes adjacent to the first pigment spot (Organ of Hesse). **f** Higher magnification of region surrounding first pigment spot shown in (**e**). Anterior, *left*, posterior, *right*. Dorsal, *top*, ventral, *bottom*. *Scale bar*, **a–b** 1 mm, **c–e** 500 µm, **f** 250 µm. Abbreviations: *ba*, branchial anlagen; *ec*, ectoderm; *en*, endoderm; *mes*, mesoderm; *not*, notochord; *ps*, pigment spot

between the *Tbx2–Tbx4* and *Tbx3–Tbx5* subfamily members in all chordates and vertebrates except for the urochordate *Ciona*, which appears to have lost its *Tbx4/5* gene, independent of the loss of this same gene in protostomes such as *Drosophila* and *C. elegans* (Fig. 7). This preservation of linkage may result from the presence of shared *cis*-regulatory modules required for the coordinate expression of both loci located in inter- or intragenic regions. The large intergenic regions of the mouse and human *Tbx3* and *Tbx5* loci, in particular (Fig. 2), will certainly complicate characterization of the regulatory logic responsible for controlling the spatiotemporal expression of these genes.

The intron–exon structure of the *Tbx2* and *Tbx3* genes is remarkably conserved within chordates, with the exception of the large intron 1 in the *Tbx2/3* gene of *Ciona* (Fig. 2). Since *Drosophila Tbx2/3* (*omb/bi*) also possesses a large intron at this location, it is formally possible that the *Ciona* locus reflects the ancestral structure and that this intron was independently reduced in the lineages leading to cephalochordates and vertebrates. Although it is not clear whether the large intron 1 represents a shared-derived character (synapomorphy) of the ancestral *Tbx2/3/4/5* locus, it does appear to have been a character of the original *Tbx4/5* locus, as it is present in both amphioxus *Tbx4/5* and all vertebrate *Tbx4* loci. We therefore propose that this large intron was probably secondarily reduced in size in the *Tbx5* locus after the cluster duplication that gave birth to separate *Tbx4* and *Tbx5* genes during early vertebrate evolution, before the origin of gnathostomes. Similarly, the large intron 4a in amphioxus *Tbx4/5* and all vertebrate *Tbx5* genes is likely to be an ancestral feature. Conversely, the large intron 3 in *Tbx4* genes and large introns 5 and 6 in tetrapod *Tbx5* genes likely reflect more recent expansions.

Long-term maintenance of linkage between *Tbx2/3* and *Tbx4/5*

All members of the *Tbx2/3/4/5* subfamily have different functions during vertebrate limb development: *Tbx2* and *Tbx3* play roles in the anteroposterior patterning of both forelimbs and hindlimbs (Gibson-Brown 1996, 1998a, b; Suzuki et al. 2004), whereas *Tbx5* and *Tbx4* have roles in the initiation and/or maintenance of forelimb and hindlimb outgrowth, respectively (Gibson-Brown et al. 1998a; Minguillon et al. 2005; Naiche and Papaioannou 2003, 2006). Because of the closely related functions shared between *Tbx2*, *Tbx3*, *Tbx4*, and *Tbx5* during limb development, maintenance of linkage between these genes has often, by analogy with the *Hox* genes, been considered due to functional constraints imposed by the coordinate regulation of these genes during limb development. However, consistent with previous conclusions (Agulnik et al. 1996), we can

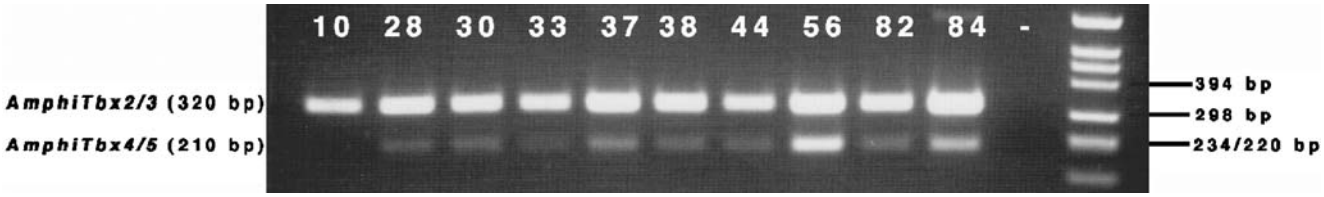

Fig. 6 RT-PCR of *AmphiTbx2/3* (*top band*) and *AmphiTbx4/5* (*bottom band*). *Numbers across top* indicate hours post-fertilization at which each RNA sample was collected

confirm that tandem duplication of the original *Tbx2/3/4/5* locus took place long before the divergence of protostomes and deuterostomes, and possibly before the divergence of ctenophores, cnidarians, and placozoans from bilaterians; despite reasonable claims that ctenophores and cnidarians possess definitive *Tbx2/3* genes (Martinelli and Spring 2005; Yamada et al. 2007), our statistical analyses indicate that, while we cannot reject these assignments, these genes could equally well be designated *Tbx2/3/4/5* genes ($P \leq 0.6048$ for ctenophore genes and $P \leq 0.0718$ for cnidarian genes). Whatever the specific timing, these events clearly predate the origin of vertebrate appendages by several hundred million years, indicating that linkage conservation is the consequence of some other, as yet unidentified selective constraint (e.g., the maintenance of distal regulatory elements controlling gene expression outside the limbs, see Menke et al. 2008).

Interestingly, most members of the *Tbx2/3/4/5* subfamily have also been implicated in vertebrate eye and heart development (reviewed by Naiche et al. 2005). Since evolution of the bilaterian "heart" (*sensu* an "Nkx/tinman-expressing contractile circulatory organ") predates the divergence of protostomes and deuterostomes (Olson 2006), but the evolution of eyes (*sensu* metazoan photoreceptive organs) is even more ancient, predating the divergence of diploblasts and triploblasts (Kozmik 2005; Fernald 2006), it is possible that this conservation of gene linkage is related to the more ancient process of eye development.

Since the relatively recent divergence of chicken and humans (Fig. 3), there have been numerous chromosomal

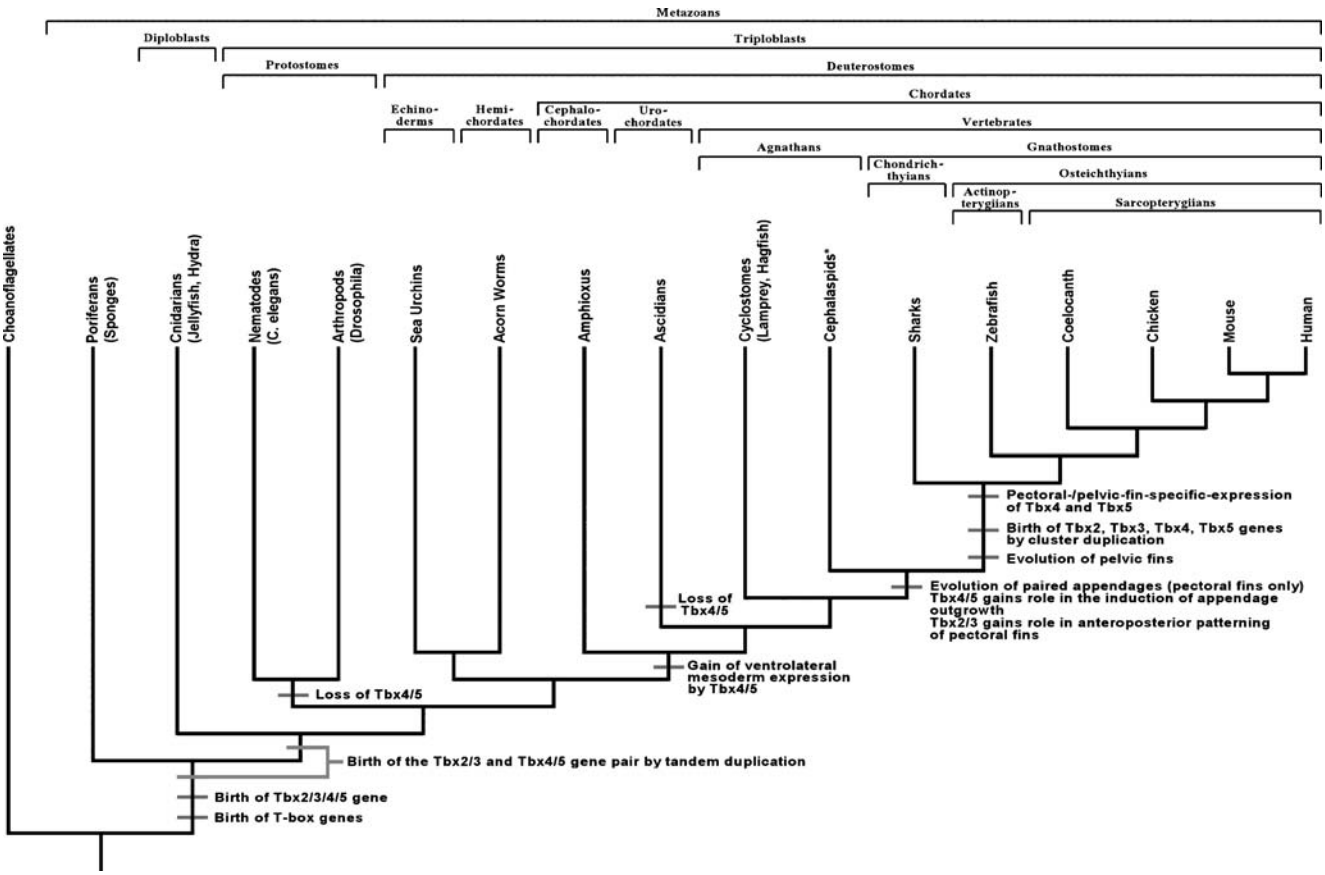

Fig. 7 Established phylogenetic species-tree upon which inferred character states for the evolution of developmental functions by the *Tbx2/3/4/5* subfamily of T-box genes have been mapped. Branch lengths not drawn to scale

rearrangements which have, nevertheless, disturbed neither the orientation nor linkage of the *Tbx2/Tbx4* and *Tbx3/Tbx5* genes or their flanking genes (*Bcas3–Brip1* and *Thrap2–Rbm19*), suggesting that, whatever selective pressures have maintained their linkage in amphioxus and vertebrates (including zebrafish, chickens, mice, and humans), the most important of these are still operating in all extant chordates except ascidians. The fact that this linkage has been maintained for over 600 million years, and possibly since before the divergence of sponges and vertebrates (Adell et al. 2003; Adell and Muller 2005), suggests that linkage within this gene subfamily must have been of fundamental importance to the development of most metazoan animals. Curiously, this linkage has been independently lost both in protostomes (e.g., *Drosophila* and *C. elegans*) and urochordates (e.g., *Ciona*), for all of which extensive genome sequences are available, yet none of which apparently possesses a *Tbx4/5* gene. This indicates that the archetypal function of these linked genes should be conserved among most, but not all, bilaterians. There are no obvious candidate functions that fit this criterion. Additional study of the longevity and function of linkage conservation between *Tbx2/3* and *Tbx4/5* subfamily members in diverse metazoan organisms is clearly warranted.

Evolution of limb-specific functions

During normal development in vertebrates, initiation of limb outgrowth takes place as a consequence of inductive interactions between rostrocaudally restricted fields of lateral plate mesoderm and the overlying ectoderm. A positive feedback loop between these two tissues maintains proximodistal outgrowth and determines the forelimb-/hindlimb-specific morphology of the appendage that subsequently develops (reviewed by Logan 2003). Expression of *Tbx4* and *Tbx5* in the presumptive hindlimb and forelimb fields, respectively, and during limb outgrowth, is conserved in all vertebrate taxa ranging from sharks to humans. Further, functional studies have implicated these genes as necessary for the induction, and possibly the patterning, of limbs. Here we report rostrocaudally restricted, bilateral expression of amphioxus *Tbx4/5* in ventrolateral mesoderm at the level of somites 15–20 (Fig. 5a–d). This appears very similar to the rostrocaudally restricted, limb-field-specific expression of *Tbx4* and *Tbx5* in vertebrates, which is confined to lateral plate mesoderm. While we have not yet been able to test the functions of *Tbx4/5* in amphioxus experimentally, our observations suggest that a *Tbx4/5* gene was probably expressed in an axially restricted manner in the ventrolateral mesoderm of the last common ancestor of amphioxus and vertebrates.

Importantly, the expression of *AmphiTbx4/5* also coincides with the posterior expression domain of *AmphiNk2-Tin*—the amphioxus ortholog of *Drosophila tinman* (Holland et al. 2003)—suggesting that an ancestral function of *AmphiTbx4/5* in chordates may well have involved heart cell specification. When paired appendages first evolved in vertebrates, *Tbx4/5* may have been co-opted from its original function in heart cell specification to play a role in the initiation of limb outgrowth (Fig. 7). This scenario would explain the evolutionary origins of the dual roles of vertebrate *Tbx5* in vertebrate heart and limb development, and of the so-called 'heart–hand syndromes' in humans, including Holt–Oram Syndrome (Basson et al. 1997; Li et al. 1997, OMIM #142900).

Tanaka et al. (2002), pursuing the long-held "fin-fold theory" of vertebrate appendage origins (reviewed by Goodrich 1958; Bowler 1996; Coates and Cohn 1999; Bemis and Grande 1999), proposed a model of fin evolution in which an ancestral *Tbx4/5* gene was co-opted to specify a single lateral fin fold in an ancient, limbless vertebrate. Modifying a previously proposed model (Gibson-Brown et al. 1998a; Ruvinsky and Gibson-Brown 2000) they suggested that duplication of the linked *Tbx2/3* and *Tbx4/5* loci allowed the evolution of discrete paired fins and the subfunctionalization of separate *Tbx5* and *Tbx4* genes in specifying the pectoral and pelvic fins, respectively (i.e., the "Genes before Limbs" model of Ruvinsky and Gibson-Brown 2000). However, the fin-fold theory of appendage evolution remains highly controversial, in part because of the paucity of fossil evidence supporting this hypothesis, but also conceptually, because the serial homology evident between forelimbs and hindlimbs (Owen 1849) would not logically derive from the subdivision of a previously established fin-fold field, but would logically follow from the homeotic transposition of a preexisting appendage (Tabin and Laufer 1993; Gibson-Brown et al. 1998a; Ruvinsky and Gibson-Brown 2000). We therefore suggest an alternative scenario (Fig. 8).

In an appendageless vertebrate ancestor, a *Tbx4/5* gene was expressed in ventrolateral mesoderm in a rostrocaudally restricted domain, as in amphioxus (Figs. 5 and 8). Co-option of this gene upstream of a lateral outgrowth program resulted in the evolution of paired pectoral fins in a jawless vertebrate ancestral to the, now extinct, cephalaspid fishes. These ancient agnathans only possessed pectoral fins and, despite extensive fossil data, show no evidence of ever having possessed pelvic appendages (Carroll 1987). This lateral expression domain of *Tbx4/5* was then homeotically transposed to an additional, more caudally positioned domain in lateral mesoderm, resulting in the induction of a pair of pelvic fins at a more caudal body position (see the "Fins before Genes" model, figure 5 in Ruvinsky and Gibson-Brown 2000). Induction of a common downstream fin outgrowth program at this new location would thus account for the serial homology evident between pectoral

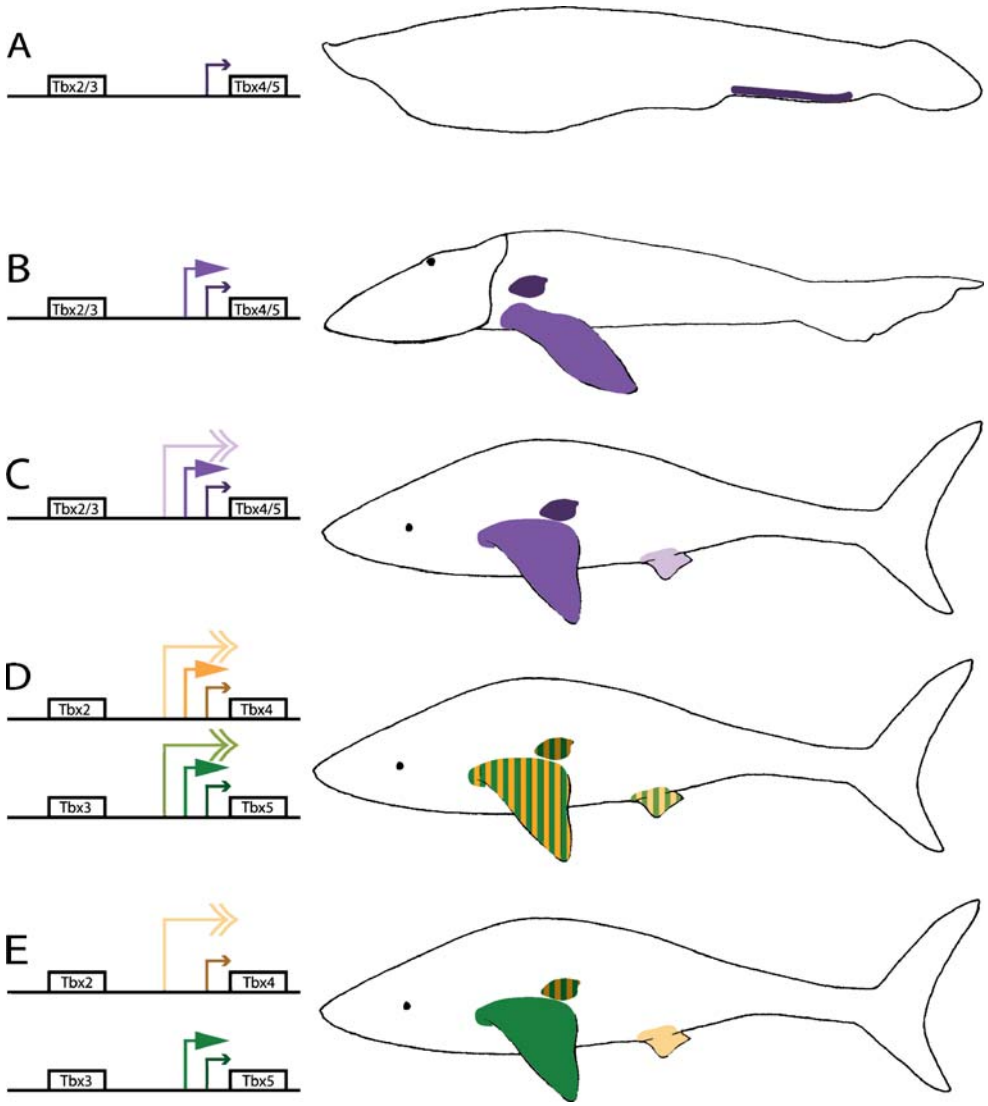

Fig. 8 Evolution of regulatory modules driving the expression of *Tbx4/5* genes in key chordates. T-box gene pairs depicted on *left*. Expression of *Tbx4/5* (*purple tones*), *Tbx4* (*solid or striped yellow tones*), and *Tbx5* (*solid or striped green tones*). *Dark tone* and *small arrowhead* represent ventral, lateral mesoderm expression (phylogenetically most widely conserved domain). *Intermediate tone* and *large arrowhead* represent pectoral fin/limb expression. *Light tone* and *double arrowhead* represent pelvic fin/limb expression. Animals from top to bottom: cephalochordate–vertebrate last common ancestor, osteostracan-like jawless vertebrate ancestor, early gnathostome ancestor, later gnathostome ancestor, gnathostome last common ancestor

fins and pelvic fins (later forelimbs and hindlimbs), because the developmental program switched on at each location would essentially be the same, although expressed within a different context (i.e., within the context of a more caudal Hox code domain in lateral plate mesoderm that also expressed another hindlimb-specific gene, *Pitx-1*; Lanctot et al. 1999; Szeto et al. 1999; Logan and Tabin 1999; Takeuchi et al. 1999). After the divergence of cephalochordates from vertebrates, but before the separation of chondricthyans (sharks) from osteichthyans (bony fish), a cluster duplication of the linked *Tbx2/3* and *Tbx4/5* loci occurred, resulting in the birth of *Tbx4* and *Tbx5* (as well as *Tbx2* and *Tbx3*). Since cluster duplication probably occurred as a result of a whole-genome duplication (Dehal and Boore 2005; Putnam et al. 2008), the complete *cis*-regulatory apparatus required for limb-specific gene expression would also have been duplicated during this event. Consequently, immediately after cluster duplication, both *Tbx4* and *Tbx5* would have been redundantly expressed in both the pectoral and pelvic appendages. Later, regulatory subfunctionalization (*sensu* Force et al. 1999) of *Tbx4* and *Tbx5* could have led to the limb-specific expression of these genes observed in all extant gnathostomes from sharks to tetrapods, including humans.

Conclusions

Close linkage of the *Tbx2/Tbx3* to the *Tbx4/Tbx5* genes has been maintained in all chordates—except the morphologically divergent, genomically degenerate ascidians—including the cephalochordate, amphioxus. This conservation of linkage suggests that essential *cis*-regulatory modules, located in intragenic or flanking regions, are likely to be found in the vicinity of these loci. Phylogenetic and phylogenomic analyses reaffirm previous subfamily assignments of these genes and, together with expression analyses during amphioxus development, allow their detailed evolutionary histories to be reconstructed. Statistical tests confirm that tandem

duplication of the ancestral *Tbx2/3/4/5* locus predated the divergence of protostomes and deuterostomes, and show that *Tbx4/5* orthologs were lost independently, at least twice, both in protostomes and urochordates. Furthermore, expression analyses in amphioxus suggest that *Tbx4/5* was co-opted from some previous function in ventral mesoderm specification (possibly heart field induction) to a new role in appendage induction after the divergence of cephalochordates and vertebrates, but before the separation of chondricthyans (cartilaginous fish) and osteichthyans (bony fish), during the Ordovician or Silurian periods, between 400 and 500 million years ago.

References

Adell T, Muller WE (2005) Expression pattern of the *Brachyury* and *Tbx2* homologues from the sponge *Suberites domuncula*. Biol Cell 97:641–650

Adell T, Grebenjuk VA, Wiens M, Muller WE (2003) Isolation and characterization of two T-box genes from sponges, the phylogenetically oldest metazoan taxon. Dev Genes Evol 213:421–434

Agulnik SI, Garvey N, Hancock S, Ruvinsky I, Chapman DL, Agulnik I, Bollag R, Papaioannou VE, Silver LM (1996) Evolution of mouse T-box genes by tandem duplication and cluster dispersion. Genetics 144:249–254

Ahn DG, Kourakis MJ, Rohde LA, Silver LM, Ho RK (2002) T-box gene tbx5 is essential for formation of the pectoral limb bud. Nature 417:754–758

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

Basson CT, Bachinsky DR, Lin RC, Levi T, Elkins JA, Soults J, Grayzel D, Kroumpouzou E, Traill TA, Leblanc-Straceski J, Renault B, Kucherlapati R, Seidman JG, Seidman CE (1997) Mutations in human cause limb and cardiac malformation in Holt–Oram syndrome. Nat Genet 15:30–35

Bemis WE, Grande L (1999) Development of the median fins of the North American paddlefish (*Polyodon spathula*), and a reevaluation of the lateral fin-fold hypothesis. In: Arratia G, Schulze HP (eds) Mesozoic fishes, vol. 2. Pfeil, Munich, pp 41–68

Bollag RJ, Siegfried Z, Cebra-Thomas J, Garvey N, Davison EM, Silver LM (1994) An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the *T*-locus. Nat Genet 7:383–389

Bongers EMHF, Duijf PHG, van Beersum SEM, Schoots J, van Kampen A, Burckhardt A, Hamel BCJ, Losan F, Hoefsloot LH, Yntema HG, Knoers NVAM, van Bokhoven H (2004) Mutations in the human TBX4 gene cause small patella syndrome. Am J Hum Genet 74:1239–1248

Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ (2006) Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 444:85–88

Bowler PJ (1996) Life's splendid drama. University of Chicago Press, Chicago

Cai WW, Reneker J, Chow CW, Vaishnav M, Bradley A (1998) An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. Genomics 54:387–397

Carroll RL (1987) Vertebrate paleontology and evolution. Freeman, New York

Coates MI, Cohn MJ (1999) Vertebrate axial and appendicular patterning: the early development of paired appendages. Am Zool 39:676–685

Cohn MJ, Patel K, Krumlauf R, Wilkinson DG, Clarke JDW, Tickle C (1997) Hox9 genes and vertebrate limb specification. Nature 387:97–101

Coulier F, Popovicci C, Villet R, Birnbaum D (2000) Meta*Hox* gene clusters. J Exp Zool 288:345–351

Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3:1700–1708

Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KE, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Necula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang HG, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee BI, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F, Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. Science 298:2157–2167

Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature 439:965–968

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27:401–410

Fernald RD (2006) Casting a genetic light on the evolution of eyes. Science 313:1914–1918

Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwaite J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545

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

Gibson-Brown JJ, Agulnik SI, Chapman DL, Alexiou M, Garvey N, Silver LM, Papaioannou VE (1996) Evidence of a role for T-box genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity. Mech Dev 56:93–101

Gibson-Brown JJ, Agulnik SI, Silver LM, Niswander L, Papaioannou VE (1998a) Involvement of T-box genes *Tbx2–Tbx5* in vertebrate limb specification and development. Development 125:2499–2509

Gibson-Brown JJ, Agulnik SI, Silver LM, Papaioannou VE (1998b) Expression of T-box genes *Tbx2–Tbx5* during chick organogenesis. Mech Dev 74:165–169

Gibson-Brown JJ, Osoegawa K, McPherson JD, Waterston RH, De Jong PJ, Rokhsar DS, Holland LZ (2003) A proposal to sequence the amphioxus genome submitted to the Joint Genome Institute of the US Department of Energy. J Exp Zool (Mol Dev Evol) 300B:5–22

Goldman N, Anderson JP, Rodrigo AG (2000) Likelihood-based tests of topologies in phylogenetics. Syst Biol 49:652–670

Goodrich ES (1958) Studies on the structure and development of vertebrates. Dover, New York

Gross JM, Peterson RE, Wu SY, McClay DR (2003) LvTbx2/3: a T-box family transcription factor involved in formation of the oral/aboral axis of the sea urchin embryo. Development 130:1989–1999

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

Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33:W557–W559

Han CS, Sutherland RD, Jewett PB, Campbell ML, Meincke LJ, Tesmer JG, Mundt MO, Fawcett JJ, Kim UJ, Deaven LL, Doggett NA (2000) Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. Genome Res 10:714–721

Hasson P, Del Buono J, Logan MPO (2007) *Tbx5* is dispensable for forelimb outgrowth. Development 134:85–92

Holland ND, Holland LZ (1993) Embryos and larvae of invertebrate deuterostomes. In: Stern CD, Holland PWH (eds) Essential developmental biology: a practical approach. IRL, Oxford, pp 21–32

Holland LZ, Holland PWH, Holland ND (1996) Revealing homologies of distantly related animals by in situ hybridization to developmental genes: amphioxus versus vertebrates. In: Ferraris JD, Palumbi SR (eds) Molecular zoology: advances, strategies and protocols. Wiley, New York, pp 267–282

Holland ND, Venkatesh TV, Holland LZ, Jacobs DK, Bodmer R (2003) *AmphiNk2-tin*, an amphioxus homeobox gene expressed in myocardial progenitors: insights into evolution of the vertebrate heart. Dev Biol 255:128–137

Horton AC, Mahadevan NR, Ruvinsky I, Gibson-Brown JJ (2003) Phylogenetic analyses alone are insufficient to determine whether genome duplication(s) occurred during early vertebrate evolution. J Exp Zool (Mol Dev Evol) 299B:41–53

Isaac A, Rodriguez-Esteban C, Ryan A, Altabef M, Tsukui T, Patel K, Tickle C, Izpisua-Belmonte JC (1998) Tbx genes and limb identity in chick embryo development. Development 125:1867–1875

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in *Hominoidea*. J Mol Evol 29:170–179

Kozmik Z (2005) Pax genes in eye development and evolution. Curr Opin Genet Dev 15:430–438

Lanctot C, Moreau A, Chamberland M, Tremblay ML, Drouin J (1999) Hindlimb patterning and mandible development require Ptx1 gene. Development 126:1805–1810

Li QY, Newbury-Ecob RA, Terrett JA, Wilson DI, Curtis AR, Yi CH, Gebuhr T, Bullen PJ, Robson SC, Strachan T, Bonnet D, Lyonnet S, Young ID, Raeburn JA, Buckler AJ, Law DJ, Brook JD (1997) Holt–Oram syndrome is caused by mutations in *TBX5*, a member of the *Brachyury* (*T*) gene family. Nat Genet 15:21–29

Logan M (2003) Finger or toe: the molecular basis of limb identity. Development 130:6401–6410

Logan M, Tabin CJ (1999) Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity. Science 283:1736–1739

Logan M, Simon HG, Tabin C (1998) Differential regulation of T-box and homeobox transcription factors suggests roles in controlling chick limb-type identity. Development 125:2825–2835

Lowe CJ, Terasaki M, Wu M, Freeman RM, Runft L, Kwan K, Haigo S, Aronowicz J, Lander E, Gruber C, Smith M, Kirschner M, Gerhart J (2006) Dorsoventral patterning in hemichordates: insights into early chordate evolution. PLoS Biol 4:e291 doi:10.1371/journal.pbio.0040291

Martinelli C, Spring J (2003) Distinct expression patterns of the two T-box homologues *Brachyury* and *Tbx2/3* in the placozoan *Trichoplax adhaerens*. Dev Genes Evol 213:492–429

Martinelli C, Spring J (2005) T-box and homeobox genes from the ctenophore *Pleurobrachia pileus*: comparison of *Brachyury*, *Tbx2/3* and *Tlx* in basal metazoans and bilaterians. FEBS Lett 579:5024–5028

Menke DB, Guenther C, Kingsley DM (2008) Dual hindlimb control elements in the *Tbx4* gene and region-specific control of bone size in vertebrate limbs. Development 135:2543–2553

Minguillon C, Del Buono J, Logan MP (2005) *Tbx5* and *Tbx4* are not sufficient to determine limb-specific morphologies but have common roles in initiating limb outgrowth. Dev Cell 8:75–84

Müller CW, Herrmann BG (1997) Crystallographic structure of the T domain–DNA complex of the *Brachyury* transcription factor. Nature 389:884–888

Naiche LA, Papaioannou VE (2003) Loss of *Tbx4* blocks hindlimb development and affects vascularization and fusion of the allantois. Development 130:2681–2693

Naiche LA, Papaioannou VE (2006) *Tbx4* is not required for hindlimb identity or post-bud hindlimb outgrowth. Development 134:93–103

Naiche LA, Harrelson Z, Kelly RG, Papaioannou VE (2005) T-box genes in vertebrate development. Annu Rev Genet 39:219–239

Ohuchi H, Takeuchi J, Yoshioka H, Ishimaru Y, Ogura K, Takahashi N, Ogura T, Noji S (1998) Correlation of wing-leg identity in ectopic FGF-induced chimeric limbs with the differential expression of chick *Tbx5* and *Tbx4*. Development 125:51–60

Olson EN (2006) Gene regulatory networks in the evolution and development of the heart. Science 313:1922–1927

O'Neill P, McCole RB, Baker CV (2007) A molecular analysis of neurogenic placode and cranial sensory ganglion development in the shark, *Scyliorhinus canicula*. Dev Biol 304:156–181

Ota R, Waddell PJ, Hasegawa M, Shimodaira H, Kishino H (2000) Appropriate likelihood ratio tests and marginal distributions for evolutionary tree models with constraints on parameters. Mol Biol Evol 17:798–803

Owen R (1849) On the nature of limbs. John Van Voorst, London

Phillipe H, Lartillot N, Brinkman H (2005) Multigene analyses of bilaterian animals corroborate the monophyly of ecdysozoa, lophotrochozoa, and protostomia. Mol Biol Evol 22:1246–1253

Putnam NH, Butts T, Ferrier DEK, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu J-K, Benito-Gutiérrez È, Dubchak I, Garcia-Fernàndez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Kapitonov V, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Sin-i T, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PWH, Satoh N, Rokhsar DS (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453:1064–1071

Rallis C, Del Buono J, Logan MPO (2005) *Tbx3* can alter limb position along the rostrocaudal axis of the developing embryo. Development 132:1961–1970

Rodriguez-Esteban C, Tsukui T, Yonei S, Magallon J, Tamura K, Izpisua-Belmonte JC (1999) The T-box genes *Tbx4* and *Tbx5* regulate limb outgrowth and identity. Nature 398:814–818

Ross R, Ross XL, Rueger B, Laengin T, Reske-Kunz AB (1999) Nonradioactive detection of differentially expressed genes using complex RNA or DNA hybridization probes. Biotechniques 26:150–155

Ruvinsky I, Gibson-Brown JJ (2000) Developmental bases of serial homology in vertebrate limb evolution. Development 127:5233–5244

Ruvinsky I, Silver LM (1997) Newly identified paralogous groups on mouse chromosomes 5 and 11 reveal the age of a T-box cluster duplication. Genomics 40:262–266

Ruvinsky I, Silver LM, Gibson-Brown JJ (2000a) Phylogenetic analysis of T-box genes demonstrates the importance of amphioxus for understanding evolution of the vertebrate genome. Genetics 156:1249–1257

Ruvinsky I, Oates AC, Silver LM, Ho RK (2000b) The evolution of paired appendages in vertebrates: T-box genes in the zebrafish. Dev Genes Evol 210:82–91

Schmidt HA, Strimmer K, Vingron M, von Haesseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114–1116

Suzuki T, Takeuchi J, Koshiba-Takeuchi K, Ogura T (2004) *Tbx* genes specify posterior digit identity through Shh and BMP signaling. Dev Cell 6:43–53

Szeto DP, Rodriguez-Esteban C, Ryan AK, O'Connell SM, Liu F, Kioussi C, Gleiberman AS, Izpisua-Belmonte JC, Rosenfeld MG (1999) Role of the bicoid-related homeodomain factor *Pitx1* in specifying hindlimb morphogenesis and pituitary development. Genes Dev 13:484–494

Tabin C, Laufer E (1993) *Hox* genes and serial homology. Nature 361:692–693

Takeuchi JK, Koshiba-Takeuchi K, Matsumoto K, Vogel-Hopker A, Naitoh-Matsuo M, Ogura K, Takahashi N, Yasuda K, Ogura T (1999) *Tbx5* and *Tbx4* genes determine the wing/leg identity of limb buds. Nature 398:810–814

Tanaka M, Münsterberg A, Anderson WG, Prescott AR, Hazon N, Tickle C (2002) Fin development in a cartilaginous fish and the origin of vertebrate limbs. Nature 416:527–531

Wattler S, Russ A, Evans M, Nehls M (1998) A combined analysis of genomic and primary protein structure defines the phylogenetic relationship of new members if the T-box family. Genomics 48:24–33

Yamada A, Pang K, Martindale MQ, Tochinai S (2007) Surprisingly complex T-box gene complement in diploblastic metazoans. Evol Dev 9:220–230