Evolution of Mouse *T-box* Genes by Tandem Duplication and Cluster Dispersion

Sergei I. Agulnik,* Nancy Garvey,* Sarah Hancock,* Ilya Ruvinsky,* Deborah L. Chapman,† Irina Agulnik,* Roni Bollag,*,[1] Virginia Papaioannou† and Lee M. Silver*

**Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544-1014 and †Department of Genetics and Development, College of Physicians and Surgeons of Columbia University, New York, New York 10032*

Manuscript received January 16, 1996
Accepted for publication May 22, 1996

ABSTRACT

The *T-box* genes comprise an ancient family of putative transcription factors conserved across species as divergent as *Mus musculus* and *Caenorhabditis elegans*. All T-box gene products are characterized by a novel 174–186-amino acid DNA binding domain called the *T-box* that was first discovered in the polypeptide products of the mouse *T* locus and the *Drosophila melanogaster optomotor-blind* gene. Earlier studies allowed the identification of five mouse *T-box* genes, *T*, *Tbx1–3*, and *Tbr1*, that all map to different chromosomal locations and are expressed in unique temporal and spatial patterns during embryogenesis. Here, we report the discovery of three new members of the mouse T-box gene family, named *Tbx4*, *Tbx5*, and *Tbx6*. Two of these newly discovered genes, *Tbx4* and *Tbx5*, were found to be tightly linked to previously identified *T-box* genes. Combined results from phylogenetic, linkage, and physical mapping studies provide a picture for the evolution of a *T-box* subfamily by unequal crossing over to form a two-gene cluster that was duplicated and dispersed to two chromosomal locations. This analysis suggests that *Tbx4* and *Tbx5* are cognate genes that diverged apart from a common ancestral gene during early vertebrate evolution.

THE *T-box* family of genes is defined by a homology domain called the *T-box* because it was first discovered within the mouse *T* locus (*Brachyury*) gene product (HERRMANN *et al.* 1990; BOLLAG *et al.* 1994). The canonical *T-box*, shared by all *T-box* polypeptides, is 174–186 amino acids in length. The *T-box* region, together with a small number of adjacent amino acid residues, has been shown to exhibit sequence-specific DNA binding activity *in vitro* (PFLUGFELDER *et al.* 1992; KISPERT and HERRMANN 1993). Recently, HERRMANN and colleagues demonstrated directly that the *T* locus product can function as a transcription factor, with an absolute requirement for the *T-box* domain (KISPERT *et al.* 1995). The *T* locus product plays a vital role in the differentiation of the notochord and induction of posterior mesoderm during early mouse embryogenesis. Results from studies of other species suggest that the *T* product plays a similar role in the development of all vertebrate organisms (HERRMANN and KISPERT 1994).

Before this report, five mouse *T-box* genes had been identified. These are the prototypical *T* locus, *Tbx1*, *Tbx2*, *Tbx3*, and *Tbr1* (BOLLAG *et al.* 1994; BULFONE *et al.* 1995). All are expressed with unique spatial and temporal patterns during embryogenesis (BOLLAG *et al.* 1994; BULFONE *et al.* 1995; CHAPMAN *et al.* 1996). In addition to the *T* locus product, two other *T-box*-containing polypeptides have been shown to exhibit sequence-specific DNA binding activity *in vitro* (N. GARVEY and R. BOLLAG, unpublished observations). *T-box* genes have also been uncovered in the genomes of other diverse metazoan organisms with six identified, to date, in the nematode *Caenorhabditis elegans* (AGULNIK *et al.* 1995 and unpublished data) and three in *Drosophila melanogaster* (PFLUGFELDER *et al.* 1992; KISPERT *et al.* 1994; unpublished data). It can be inferred from these data that T-box genes exist in all advanced metazoan species.

In this report, we describe the identification and analysis of three new *T-box* genes in the mouse, *Tbx4*, *Tbx5*, and *Tbx6*. Although linkage mapping of the original five mouse *T-box* genes suggested that members of this gene family might all be dispersed to different chromosomal locations, the new mapping data provides evidence for a cluster organization in one *T-box* subfamily. The results presented here, in combination with embryonic expression data, raise the possibility that two *T-box* genes, *Tbx4* and *Tbx5*, may have evolved apart originally to define unique characteristics of the fin precursors to the fore- and hindlimbs of all tetrapod vertebrates.

MATERIALS AND METHODS

cDNA library screening: A lambda gt10 cDNA library produced from day 8.5 mouse embryos was kindly provided by Dr. BRIGID HOGAN (Vanderbilt University). The library was screened with a ^{32}P-labeled probe from the *T-box* region of the *Tbx2* gene (BOLLAG *et al.* 1994). Low stringency filter hybridization was performed at 50° according to the protocol of

Corresponding author: Lee M. Silver, Department of Molecular Biology, Lewis Thomas Laboratory, Princeton University, Princeton, NJ 08544-1014. E-mail: lsilver@molbiol.princeton.edu

[1] *Present address:* Institute for Molecular Medicine and Genetics, Medical College of Georgia, Augusta, GA 30912.

CHURCH and GILBERT (1984), and an initial washing was performed at 50° in 2× SSC, 0.1% SDS followed by autoradiographic analysis. The same filters were rewashed at 65° in 0.1× SSC and analyzed again by autoradiography. A comparison of the two autoradiographs allowed the identification of clones that hybridize with *Tbx2* at low stringency but not high stringency. These clones were recovered through secondary and tertiary screens, and inserts were subcloned into Bluescript KS− (Strategene) for further analysis and sequencing with the Sequenase 2 kit (U.S. Biochemicals). This screen led to the identification of cDNA clones containing portions of three novel *T-box* genes: *Tbx4, Tbx5,* and *Tbx6.* Additional sequences from these three genes, as well as previously described *T-box* genes, were obtained by a combination of PCR-based and hybridization screening approaches with the same cDNA library as well as other embryonic libraries (CHAPMAN *et al.* 1996).

Linkage analysis: DNA samples from an interspecific backcross panel, BSS, developed at the JACKSON LABORATORY (ROWE *et al.* 1994) were used to determine the map positions of the *Tbx4* and *Tbx5* genes. To perform this analysis, uniquely hybridizing probes were obtained by PCR amplification of the *T-box* regions of *Tbx4* and *Tbx5* cDNA clones. Southern blots containing electrophoresed *Taq*I-digested genomic DNA from each backcross animal were hybridized sequentially to each probe at 65° (CHURCH and GILBERT 1984) and washed at 65° in 0.1× SSC, 0.1% SDS. The map positions of each gene were determined by comparison of the strain distribution patterns obtained in this experiment with the database held at the JACKSON LABORATORY (1995).

Phylogenetic analysis: In the present analysis, we compared amino acid sequences rather than nucleotide sequences because synonymous nucleotide substitutions are likely to be saturated over the range of metazoan evolution examined, which would lead to a high degree of noise in phylogenetic comparisons (ZHANG and NEI 1996). The PILEUP program from the GCG software package (GENETICS COMPUTER GROUP 1989) was used to obtain an initial alignment among *T-box* sequences. This alignment was adjusted manually with the ESEE sequence editor to maximize identity (CABOT and BECKENBACH 1989). Regions where alignment was uncertain due to extensive length and sequence variation are likely to have a low signal to noise ratio and were excluded from further analyses.

Phylogenetic trees were constructed based on Poisson-corrected distances using the neighbor-joining algorithm implemented with the METREE program (SAITOU and NEI 1987; RZHETSKY and NEI 1992). Confidence values for each interior-branch were also calculated with the METREE program (RZHETSKY and NEI 1994). For comparative purposes, we constructed a neighbor-joining tree based on nonsynonymous nucleotide substitutions computed by the method of NEI and GOJOBORI (1986) and conducted 2000 bootstrap replications (FELSENSTEIN 1985) as implemented with the MEGA program (KUMAR *et al.* 1993). All well-supported nodes of the amino acid sequence-based tree were confirmed with the nucleic acid tree, while some variability was observed for less well supported nodes.

YAC library screening and analysis: The Princeton (BURKE *et al.* 1991) and Whitehead (KUSUMI *et al.* 1993) mouse YAC libraries were screened for *T-box*-containing clones with a PCR-based protocol (GREEN and OLSON 1990). YAC clones were sized by pulsed field gel electrophoresis (PFGE) performed as described previously (BARLOW and LEHRACH 1987).

RESULTS

Identification of new *T-box* genes expressed in day 8.5 mouse embryos: Our original search for mouse *T-box* genes was based on the use of a pair of degenerate primers formulated to contain all possible codon permutations for two seven-residue peptides, YIHPDSP and AVTAYQN, that were conserved among all of the then-characterized *T* locus homologues as well as the *D. melanogaster T-box* homologue *omb* (BOLLAG *et al.* 1994; Figure 1). A PCR-based search with these primers led to the identification of three novel T-box genes, *Tbx1, Tbx2,* and *Tbx3,* that were expressed in day 12.5 embryos. Although the search for related genes with degenerate primer PCR is very sensitive, it has the drawback that single amino acid changes in the region assumed to be conserved can prevent the recovery of additional members of the gene family. In an attempt to overcome this limitation, we used a low stringency hybridization and washing protocol to search for novel *Tbx2*-cross-hybridizing members of the *T-box* gene family within a cDNA library prepared from day 8.5 embryos. Clones that hybridized to the *Tbx2* probe at low stringency, but did not hybridize to *Tbx1, Tbx2,* or *Tbx3* probes at high stringency, were selected as candidates for novel members of the *T-box* gene family.

Candidate clones were sorted into classes based on cross-hybridization to each other under conditions of high stringency. Multiple clones from each class were sequenced and regions with homology to the previously defined *T-box* domain (AGULNIK *et al.* 1995) were identified and compared to each other and all previously identified *T-box* genes. This strategy led to the identification of three novel *T-box* genes: *Tbx4, Tbx5,* and *Tbx6.*

The *T-box* domain is conserved in its entirety among *T-box* genes: Sequence analysis was performed across the *Tbx4, Tbx5,* and *Tbx6* coding regions. Translated regions that correspond to the previously defined *T-box* domain are presented in Figure 1, with additional sequence information deposited with GenBank. As shown in the figure, homology with all other characterized mouse T-box genes is maintained across the length of the 174–186-residue-long *T-box* domain. As is the case with other *T-box* genes, no extended homologies are observed outside the *T-box* region (data not shown).

Phylogenetic analysis: A phylogenetic analysis was performed on the conserved *T-box* domains of the *Tbx4, Tbx5,* and *Tbx6* genes in comparison with other *T-box* domains (Figure 2). A phylogenetic tree can provide several types of useful information. First, it can provide support for or against direct ortholog relationships between genes from different species. Second, it can provide information on the likely status of a gene family in organisms that are ancestral to groups of currently extant species. Finally, it provides an estimate of the relative time elapsed since the divergence of any two gene sequences from their most recent common ancestor.

With these points in mind, the phylogenetic tree presented in Figure 2 reveals several interesting features of the *T-box* gene family. First, the finding of direct vertebrate and invertebrate orthologs for both the *T*

```
          10        20        30        40        50        60        70        80        90
T     LWLRFKELTNEMIVTKNGRRMFPVLKVNVSGLDPNAMYSFLLDFVTADNHRWKYVNGEWVPGGKPEPQAP.SCVYIHPDSPNFGAHWMKAPVSFS
Tbx1  LWDEFNQLGTEMIVTKAGRRMFPTFQVKLFGMDPMADYMLLMDFVPVDDKRYAFHSSSWLVAGKRDPATPG.RVHYHPDSPAKGAQWMKQIVSFD
Tbx2  LWDQFHKLGTEMVITKSGRRMFPPFKVRVSGLDKKAKYILLMDIVAADDCRYKFHNSRWMVAGKADPEMP.KRMYIHPDSPATGEQWMAKPVAFH
Tbx3  LWDQFHKRGTEMVITKSGRRMFPPFKVRCSGLDKKAKYILLMTI.AADDCRYKFHNSRWMVAGKADPEMP.KRFYIHPDSPATGEQWMSKVVTFH
Tbx4  **********MIITKAGRRMFPSYKVKVTGMNPKTKYILLIDIVPADDHRYKFCDNKWMVAGKAEP.MPG.RLYVHPDSPATGAHWMRQLVSFQ
Tbx5  LWLKFRSGVTEMIITKAGGRMFPSYKVKVTGLNPKTKYILLMDIVPADDHRYKFADNKWSVTGKAEPAMPG.RLYVHPDSPATGAHWMRQLVSFQ
Tbx6  LWKEFSAVGTEMIITKAGRRMFPACRVSVTGLDPEARYLFLLDVVPVDGARYRWQGPDWEPSGKAEPRLP.DRVYIHPDSPATGAHWMRQPVSFH
Tbr1  LWLKFHRHQTEMIITKQGRRMFPFLSFNISGLDPTAHYNIFVDVILADPNHWRFQGGKWVPCGKADTNVQGNRVYMHPDSPNTGAHWMRQEISFG

          100       110       120       130       140       150       160       170       180       190
T     KVKLTNKLNGG...GQ.IMLNSLHKYEPRIHIVRVGGPQRM..........ITSHCFPETQFIAVTAYQNEEITALKIKYNPFAKAFLDAKERNDH
Tbx1  KLKLTNNLLDDN..GH.IILNSMHRYQPRFHVVYVAPRKDSEK....YEEENFKTFVFEETRFTAVTAYQNQQITQLKIASNPFAKGFRDCDPRLVS
Tbx2  KLKLTNNISDKH..GF.TILNSMHKYQPRFHIVRANDILKL.....PYST..FRTYVFPETDFIAVTAYQNDKITQLKIDNNPFAKGFRDTGNGRRE
Tbx3  KLKLTNNISDKH..GF.TILNSMHKYQPRFHIVRANDILKL.....PYST..FRTYLFPETEFIAVTAYQNDKITQLKIDNNPFAKGFRDIGHCRRE
Tbx4  KLKLTNNHLDPF..GH.IILNSMHKYQPRLHIVKADENNAFG....SKNTA.FCTHVFPETSFISVTSYQNHKITQLKIENNPFAKGFRGSDDSDLR
Tbx5  KLKLTNNHLDPF..GH.IILNSMHKYQPRLHIVKADENNGFG....SKNTA.FCTHVFPETAFIAVTSYQNHKITQLKIENNPFAKGFRGSDDLELH
Tbx6  RVKLTNSTLDPH..GH.LILHSMHKYQPRIHLVRATQLCSQHWGG.......VASFRFPETTFISVTAYQNPRITQLKIAANPFAKGFRENGRNCKR
Tbr1  KLKLTNNKGASNNNGQMVVLQSLHKYQPRLHVVEVNEDGTEDTSQPGR....VQTFTFPETQFIAVTAYQNTDITQLKIDHNPFAKGFRDNYDTIYT
```

FIGURE 1.—Comparison of *T-box* sequences from different mouse *T-box* polypeptides. The *T-box* domains from all eight characterized mouse *T-box* polypeptides, including the *Tbx4, Tbx5,* and *Tbx6* gene products described in this report, are aligned to maximize amino acid identity. The GenBank accession numbers for all sequences are as follows: *Tbx1* (U57327), *Tbx2* (U15566), *Tbx3* (U57328), *Tbx4* (U57329), *Tbx5* (U57330), *Tbx6* (U57331), *Tbr1* (U49250). The complete *T-box* homology domain varies in size from 174 to 186 amino acids in length. The two peptide regions that were used to develop degenerate oligonucleotides for the discovery of the *Tbx1, Tbx2,* and *Tbx3* genes are indicated in boxes. The sequence shown here for *Tbx1* is corrected from the version originally published by BOLLAG *et al.* (1994).

locus and *Tbx2/3* demonstrates the ancient origin of this gene family in the common ancestor to all animal species. With a confidence level of 99%, we can also postulate the existence of a third *T-box* gene, ancestral to *Tbx4/5,* within an early metazoan genome, even though an invertebrate ortholog of this gene has not yet been uncovered; this ortholog may be awaiting discovery or may have been lost during evolution. An additional *T-box* gene or genes, ancestral to *Tbx1, Tbx6,* and/or *Tbr1,* may also have existed in the common metazoan ancestor, with a degree of confidence below the 95% level required for significance. The phylogenetic data also suggest that two or more additional genes currently identified within the *C. elegans* genome (*tbx-7* and the ancestral *tbx-8/9* sequence) were also present in the common ancestor to nematodes and mice (Figure 2).

Another important conclusion that can be drawn from the phylogenetic tree concerns the evolutionary

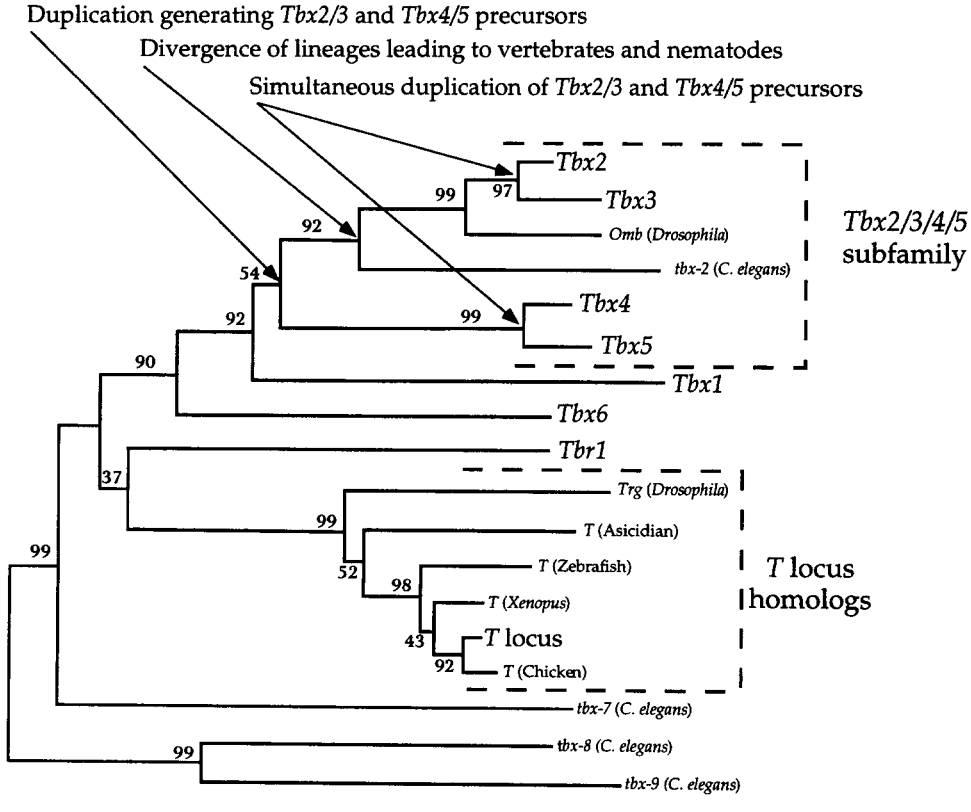

FIGURE 2.—Phylogenetic tree of *T-box* sequences. All mouse *T-box* genes characterized to date are shown in large type, and representative examples of *T-box* genes from other species are shown in smaller type. Branch lengths are proportional to estimates of evolutionary distances obtained from the phylogenetic analysis. Confidence values associated with the likelihood of selected branchpoints are indicated as percentile numbers. Branchpoints representing selected gene duplication or speciation events are indicated within the *Tbx2/3/4/5* subfamily.

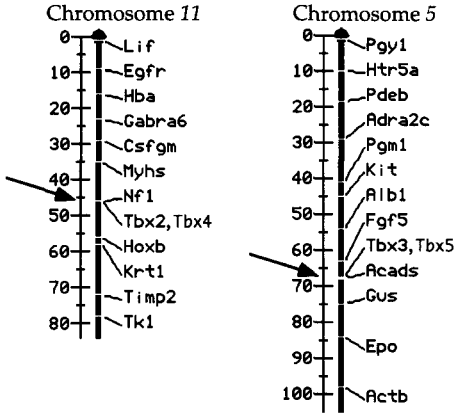

FIGURE 3.—Chromosomal map positions for *Tbx2, Tbx3, Tbx4,* and *Tbx5*. Linkage maps of chromosomes *11* and *5* are shown with reference loci, and *T-box* gene map positions are indicated with arrows. *Tbx2* and *Tbx4* are located at centimorgan position 46.25 on chromosome *11*. *Tbx3* and *Tbx5* are located at centimorgan position 67.5 on chromosome *5*. These reference maps were obtained from the Mouse Genome Database (MGD) World Wide Web site at the JACKSON LABORATORY (1995).

relationships among the four mouse genes *Tbx2, Tbx3, Tbx4,* and *Tbx5*. The data indicate that the *Tbx2* and *Tbx3* genes evolved apart from a common ancestral sequence through a duplication event that occurred subsequent to the divergence of mice and flies. As such, these two genes represent a cognate gene pair. Based on branch lengths and locations, it appears that *Tbx4* and *Tbx5* represent a second cognate gene pair, which can also be traced back to a duplication event that occured along the vertebrate lineage. Interestingly, these two cognate gene pairs were both formed at approximately the same point in evolutionary time.

***Tbx4* and *Tbx5* are each tightly linked to previously defined *T-box* genes:** The *Tbx4* and *Tbx5* genes were mapped with the use of DNA samples from the same panel of 94 backcross animals that had been used to map the *Tbx1, Tbx2,* and *Tbx3* genes to chromosomes *16, 11,* and *5,* respectively (BOLLAG *et al.* 1994). For this reason, it was possible to compare directly the data generated in this new analysis with the data obtained in our previous study. This comparison demonstrated identical map positions for the *Tbx2* and *Tbx4* genes on chromosome *11* and identical map positions for the *Tbx3* and *Tbx5* genes on chromosome *5* (Figure 3). Mapping and further analysis of *Tbx6* will be reported elsewhere (CHAPMAN *et al.* 1996).

These results were intriguing in that they suggest the possibility that these two linked *T-box* gene pairs might represent gene clusters formed by duplicative unequal crossing over events. However, with 100% concordance in segregation over a backcross panel of 94 animals, a 95% confidence interval for linkage distance still extends across 3.6 cM (SILVER 1995).

To determine the physical distance separating the *Tbx2* and *Tbx4* genes, we screened mouse YAC libraries with a PCR assay for *Tbx2* and recovered three independent YAC clones that contained this gene (FES.G4, B20.C11, C126.B9). Each of these clones was tested by Southern blot hybridization for the presence of *Tbx4,* and all exhibited a positive signal in a restriction fragment of the size expected from whole genome analysis. The smallest of these clones, B20.C11, has an insert size of 115-kb, as determined by PFGE, which represents the maximum distance that could separate *Tbx2* and *Tbx4* (data not shown). Thus, it appears likely that the *Tbx2* and *Tbx4* genes were duplicated from a single ancestral locus by a process of unequal crossing over. Attempts to recover YAC clones containing the *Tbx3* gene for similar analysis were unsuccessful.

DISCUSSION

Evolution of other well-conserved gene families that encode DNA-binding proteins: Two well-characterized families of transcription factors with critical roles in the development of both vertebrate and invertebrate organisms are the *Hox* genes and the *Pax* genes. When the genomic distribution of each of these gene families is examined in the mouse, very different patterns are observed. The classical antennapedia-like *Hox* genes are all located within one of four clusters that have been maintained throughout vertebrate evolution (KRUMLAUF 1994). In contrast, the *Pax* genes are all dispersed to different chromosomal locations (GRUSS and WALTHER 1992; WALLIN *et al.* 1993).

The order of genes within each *Hox* cluster is highly conserved, and members of each cluster often show maximal levels of relatedness to cognates within each of the other clusters. This implies a pathway of evolution in which unequal crossover events first caused the expansion of a single cluster of *Hox* genes in an early metazoan ancestor, with subsequent whole cluster duplication and dispersion events during vertebrate evolution (PENDLETON *et al.* 1993; SCHUBERT *et al.* 1993). The conservation of the *Hox* cluster organization in all vertebrate species, throughout 500 million years of evolution, is likely to be a consequence of one or more *cis*-acting regulatory elements that act across all members of each cluster. Disruption of cluster organization would then be strongly selected against. In contrast, it seems likely that the *Pax* genes originated with independent *cis*-acting regulatory elements that did not prevent their dispersion to distant chromosomal locations.

Linkage mapping of the five originally characterized mouse *T-box* genes suggested a *Pax*-like mode of evolution and regulation with dispersion of family members to different chromosomal locations (BOLLAG *et al.* 1994; BULFONE *et al.* 1995). However, the results obtained in the current study present a different picture for the *Tbx2/3/4/5* subfamily.

A model for the evolution of the *Tbx2, 3, 4,* and *5* genes: Phylogenetic analysis shows that *Tbx2* and *Tbx3*

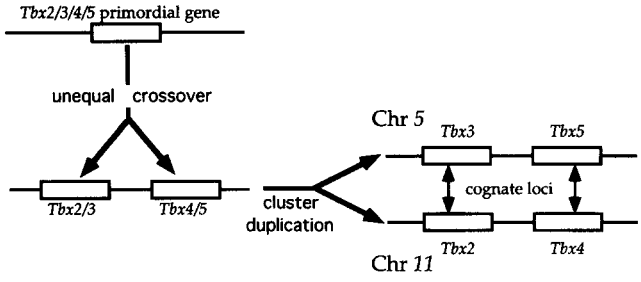

FIGURE 4.—A model for the evolution of the *Tbx2/3/4/5* subfamily. This model is based on an integration of the data shown in Figures 2 and 3. Details are described in the text.

form a cognate gene pair, as do *Tbx4* and *Tbx5* (Figure 2). However, the pairing of *T-box* genes according to chromosomal location yields a different result with tight linkage between *Tbx2* and *Tbx4,* and between *Tbx3* and *Tbx5* (Figure 3). The simplest interpretation of the combined phylogenetic and linkage data is presented in the form of the evolutionary model diagrammed in Figure 4. This model hypothesizes an initial duplication of a single ancestral gene by unequal crossing over to form a two-gene cluster that was later duplicated *en masse* with one copy dispersed to a different chromosomal location. This mode of evolution parallels that of the antennapedia class of *Hox* genes.

Phylogenetic analysis provides an estimate for the timing of the two duplication events shown in Figure 4. The initial duplication of an ancestral *Tbx2/3/4/5* gene by unequal crossing over must have occurred in a common metazoan ancestor to mice, flies, and nematodes, before the divergence of the vertebrate and nematode lineages over 600 million years ago (Figure 2). The subsequent duplication of the two-gene cluster appears to have occurred somewhere along the vertebrate lineage.

The persistence of tight linkage between two genes in *T-box* cluster through 600 million years of evolution suggests a selective advantage to this genomic arrangement that could be incurred by *cis*-regulatory elements that act upon both cluster members. A similar example of very tight linkage has also been found for two ancient *C. elegans T-box* genes, *tbx-8* and *tbx-9,* which have maintained a genomic separation of less than two kilobases (AGULNIK *et al.* 1995). In contrast, the other mouse *T-box* genes, the *T* locus, *Tbx1, Tbx6,* and *Tbr1,* appear to exist in genomic isolation like members of the *Pax* gene family (GRUSS and WALTHER 1992; CHAPMAN *et al.* 1996).

Duplication of *T-box* genes and the evolution of developmental complexity: The embryonic expression patterns of individual members of the *Tbx2/3/4/5* subfamily are consistent with their evolutionary relationships (BROWN *et al.* 1996; CHAPMAN *et al.* 1996). The *Tbx2* and *Tbx3* genes exhibit similar patterns of spatial and temporal expression as do the *Tbx4* and *Tbx5* genes. These conserved expression patterns are presumably a consequence of the recent divergence of both pairs of cognates from single ancestral genes within the vertebrate lineage (Figure 2). Together, the evolutionary and expression data suggest the possibility of a high degree of functional overlap between cognate genes. Nevertheless, important differences in cognate expression have been observed. The most striking of these occurs in the developing limb buds. *Tbx4* is expressed at much higher levels in the hindlimb bud, whereas *Tbx5* is expressed exclusively in the forelimb relative to the forelimb bud (BROWN *et al.* 1996; CHAPMAN *et al.* 1996).

The reciprocal patterns of limb bud expression exhibited by *Tbx4* and *Tbx5* suggests that each could be involved in the specification of unique features characteristic of the different limb types. This differential functionality could provide an evolutionary explanation for the maintenance of both the *Tbx2/4* and *Tbx3/5* clusters. In particular, phylogenetic data are consistent with a cluster duplication event that may have just preceded the developmental duplication of the pelvic fins into the pectoral fins in an ancient vertebrate ancestor (BROWN *et al.* 1996). Since, tetrapod fore- and hindlimbs evolved from the pectoral and pelvic fins, respectively, it is possible that *Tbx5* and *Tbx4* were divergently selected to play roles in the differential specification of fore- (pectoral) and hind- (pelvic) limb (fin) identity. Support for or against this hypothesis must await functional studies of the individual *T-box* genes.

This research was supported by a National Institutes of Health grant to L.M.S. (HD-20275) and by the Raymond and Beverly Sacker Foundation and the Alice Bohmfalk Charitable Trust (V.E.P.).

LITERATURE CITED

AGULNIK, S. I., R. J. BOLLAG and L. M. SILVER, 1995 Conservation of the T-box gene family from *M. musculus* to *C. elegans.* Genomics **25:** 214–219.

BARLOW, D. P., and H. LEHRACH, 1987 Genetics by gel electrophoresis: the impact of pulsed field gel electrophoresis on mammalian genetics. Trend. Genet. **3:** 167–171.

BOLLAG, R. J., Z. SIEGFRIED, J. CEBRA-THOMAS, N. GARVEY, E. M. DAVISON *et al.,* 1994 An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the T-locus. Nature Genet. **7:** 383–389.

BULFONE, A., S. M. SIMGA, K. SHIMAMURA, A. PETERSON, L. PUELLES *et al.,* 1995 T-brain-1: a homolog of Brachyury whose expression defines molecularly distinct domains within the cerebral cortex. Neuron **15:** 63–78.

BURKE, D. T., J. M. ROSSI, D. S. KOOS and S. M. TILGHMAN, 1991 A mouse genomic library of yeast artificial chromosome clones. Mammal. Genome **1:** 65.

CABOT, E. L., and A. T. BECKENBACH, 1989 Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comput. Appl. Biosci **5:** 233–234.

CHAPMAN, D. L., N. GARVEY, S. HANCOCK, M. ALEXIOU, S. AGULNIK *et al.,* 1996 Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. Dev. Dynamics (in press).

CHURCH, G. M., and W. GILBERT, 1984 Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:** 1991–1995.

FELSENSTEIN, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

GENETICS COMPUTER GROUP, 1989 *The Wisconsin GCG Package.* Maryland Biotechnology Institute, Madison, WI.

GIBSON-BROWN, J. J. G., S. AGULNIK, D. L. CHAPMAN, M. ALEXIOU, N.

GARVEY *et al.*, 1996 Evidence of a role for T-box genes in the evolution of limb morphogenesis and the specification of forelimb/hindlimb identity. Mech. Dev. (in press).

GREEN, E. D., and M. V. OLSON, 1990 Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc. Natl. Acad. Sci. USA **87:** 1213–1217.

GRUSS, P., and C. WALTHER, 1992 Pax in development. Cell **69:** 719–722.

HERRMANN, B. G., and A. KISPERT, 1994 The T genes in embryogenesis. Trends Genet. **10:** 280–286.

HERRMANN, B. G., S. LABEIT, A. POUSTKA, T. KING and H. LEHRACH, 1990 Cloning of the T gene required in mesoderm formation in the mouse. Nature **343:** 617–622.

JACKSON LABORATORY, 1995 MGD: The Mouse Genome Database (http://www.informatics.jax.org/mgd.html).

KISPERT, A., and B. G. HERRMANN, 1993 The Brachyury gene encodes a novel DNA binding protein. EMBO J. **12:** 3211–3220.

KISPERT, A., B. G. HERRMANN, M. LEPTIN and R. REUTER, 1994 Homologs of the mouse Brachyury gene are involved in the specification of posterior terminal structures in Drosophila, Tribolium, and Locusta. Genes Dev. **8:** 2137–2150.

KISPERT, A., B. KOSCHORZ and B. HERRMANN, 1995 The T protein encoded by Brachyury is a tissue-specific transcription factor. EMBO J. **14:** 4763–4772.

KRUMLAUF, R., 1994 Hox genes in vertebrate development. Cell **78:** 191–201.

KUMAR, S., K. TAMURA and M. NEI, 1993 *MEGA: Molecular Evolutionary Genetics Analysis.* The Pennsylvania State University, University Park, PA.

KUSUMI, K., J. S. SMITH, J. A. SEGRE, D. S. KOOS and E. S. LANDER, 1993 Construction of a large-insert yeast artificial chromosome library of the mouse genome. Mammal. Genome **4:** 391–392.

NEI, M., and T. GOJOBORI, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

PENDLETON, J. W., B. K. NAGAI, M. T. MURTHA and F. H. RUDDLE, 1993 Expansion of the *Hox* gene family and the evolution of chordates. Proc. Natl. Acad. Sci. USA **90:** 6300–6304.

PFLUGFELDER, G. O., H. ROTH and B. POECK, 1992 A homology domain shared between Drosophila optomotor-blind and mouse Brachyury is involved in DNA binding. Biochem. Biophys. Res. Commun. **186:** 918–925.

ROWE, L. B., J. H. NADEAU, R. TURNER, W. N. FRANKEL, V. A. LETTS *et al.*, 1994 Maps from two interspecific backcross panels available as a community genetic mapping resource. Mammal. Genome **5:** 253–274.

RZHETSKY, A., and M. NEI, 1992 A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. **9:** 945–967.

RZHETSKY, A., and M. NEI, 1994 METREE: a program package for inferring and testing minimum-evolution trees. Comput. Appl. Biosci. **10:** 409–412.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

SCHUBERT, F. R., K. NIESELT-STRUWE and P. GRUSS, 1993 The antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution. Proc. Natl. Acad. Sci. USA **90:** 143–147.

SILVER, L. M., 1995 *Mouse Genetics: Concepts and Application.* Oxford University Press, New York.

WALLIN, J., Y. MITZUTANI, K. IMAI, N. MIYASHITA, K. MORIWAKI *et al.*, 1993 A new Pax gene, Pax-9, maps to mouse chromosome *12.* Mammal. Genome **4:** 354–358.

ZHANG, J., and M. NEI, 1996 Evolution of Antennapedia-class homeobox genes. Genetics **142:** 295–303.

Communicating editor: K. ARTZT