Three novel *T*-box genes in *Caenorhabditis elegans*

Sergei I. Agulnik, Ilya Ruvinsky, and Lee M. Silver

Abstract: The *T*-box gene family consists of members that share a unique DNA binding domain. The best characterized *T*-box gene, *Brachyury* or *T*, encodes a transcription factor that plays an important role in early vertebrate development. Seven other recently described mouse *T*-box genes are also expressed during development. In the nematode *Caenorhabditis elegans*, four *T*-box genes have been characterized to date. In this study, we describe three new *C. elegans T*-box genes, named *Ce-tbx-11*, *Ce-tbx-12*, and *Ce-tbx-17*. *Ce-tbx-11* and *Ce-tbx-17* were uncovered through the sequencing efforts of the *C. elegans* Genome Project. *Ce-tbx-12* was uncovered through degenerate PCR analysis of *C. elegans* genomic DNA. *Ce-tbx-11* and *Ce-tbx-17* are located in close proximity to the four other previously described *T*-box genes in the central region of chromosome III. In contrast, Ce-tbx-12 maps alone to chromosome II. Phylogenetic analysis of all known *T*-box domain sequences provides evidence of an ancient origin for this gene family.

Key words: transcription factor, *T*-box genes, evolution, *Brachyury*.

Résumé : La famille de gènes *T*-box comprend des membres qui partagent un domaine de liaison à l'ADN unique. Le mieux caractérisé de ces gènes, *Brachyury* ou *T*, code pour un facteur de transcription qui joue un rôle important tôt lors du développement des vertébrés. Sept autres gènes *T*-box récemment décrits chez la souris sont également exprimés lors du développement. Chez le nématode *Caenorhabditis elegans*, quatre gènes *T*-box ont été caractérisés à ce jour. La présente étude a permis d'identifier trois nouveaux gènes *T*-box chez le *C. elegans*: *Ce-tbx-11*, *Ce-tbx-12* et *Ce-tbx-17*. Les gènes *Ce-tbx-11* et *Ce-tbx-17* ont été identifiés grâce au projet génome en cours chez le *C. elegans*. Le gène *Ce-tbx-12* a été obtenu par amplification PCR de l'ADN génomique du *C. elegans* à l'aide d'amorces dégénérées. Les gènes *Ce-tbx-11* et *Ce-tbx-17* sont situés très proches des quatre gènes *T*-box rapportés précédemment, soit dans la région centrale du chromosome III. Quant au gène *Ce-tbx-12*, il est situé plutôt sur le chromosome II. Une analyse phylogénétique de toutes les séquences connues de domaines *T*-box suggère que cette famille de gènes est d'origine ancienne.

Mots clés : facteur de transcription, gènes *T*-box, évolution, *Brachyury*.

[Traduit par la Rédaction]

Introduction

The *T*-box gene family shares a homologous domain, named the *T*-box, that was first described within the mouse *Brachyury* (*T*) locus (Herrmann et al. 1990; Bollag et al. 1994). The *T*-box domain is 174–186 amino acids in length and has been shown to be associated with sequence specific DNA binding activity (Pflugfelder et al. 1992; Kispert and Herrmann 1993). In recent work described by Kispert et al. (1995), the product of the prototypical member of the *T*-box gene family, *Brachyury*, was shown to behave as a transcription factor with both activation and repression domains when assayed within tissue-culture cells. This study provided the biochemical basis for the action of the *Brachyury*-gene product in the differentiation of the notochord and the induction of posterior mesoderm during mammalian embyrogenesis. In studies of the *Brachyury* homolog in other species, a similar role in mesodermal formation was demonstrated (Herrmann and Kispert 1994).

To date, eight mouse *T*-box genes have been identified (Bollag et al. 1994; Bulfone et al. 1995; Agulnik et al. 1996). All these genes have highly specific patterns of expression during embryogenesis (Bollag et al. 1994; Bulfone et al. 1995; Chapman et al. 1996). Members of the *T*-box gene family have also been found in other metazoan species, with four having been identified in *Drosophila melanogaster* (Pflugfelder et al. 1992; Kispert et al. 1994; Brook and Cohen 1996; unpublished data) and four in the nematode *Caenorhabditis elegans* (Agulnik et al. 1995).

In this report, we describe the identification and analysis of three new *T*-box genes in the genome of the nematode *C. elegans*. This brings the total number of *C. elegans T*-box genes to seven. Surprisingly, it appears likely that all seven genes were already established at the outset of metazoan evolution.

Corresponding Editor: B. Golding.

Received November 12, 1996. Accepted March 18, 1997.

S.I. Agulnik,[1] I. Ruvinsky, and L.M. Silver. Department of Molecular Biology, Princeton University, Princeton, NJ 08544-1014, U.S.A.

[1] Author to whom all correspondence should be addressed (e-mail: SAGULNIK@WATSON.PRINCETON.EDU).

Materials and methods

PCR amplification, cloning, and sequencing

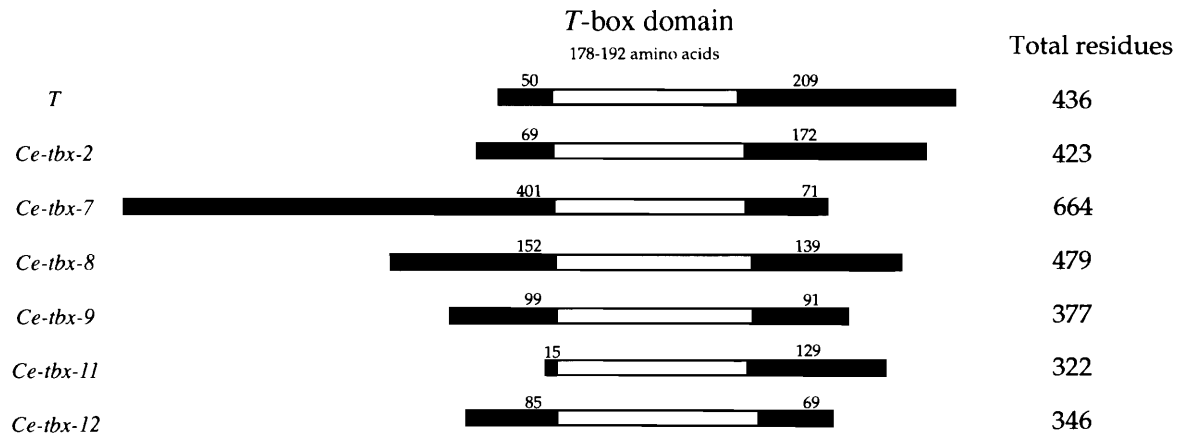
A 200-ng aliquot of *C. elegans* genomic DNA was subjected to PCR amplification with degenerate forward (TA(CT) AT(ACT) CA(CT) CCN GA(CT) (AT)(CG)N CC) and reverse (GTN ACN GCN

Fig. 1. cDNA sequence and predicted amino acid sequence of the *Ce-tbx-12* gene. The in-frame terminator codon in the 5′-UTR is shown in bold. The polyadenylation signal (AATAAA) at position 1201 is underlined. The cDNA sequence has been deposited in GenBank (see Fig. 2).

```
      TCAGTATCGGCGGAATTCGAGGATCGGGTACCATGGCTCTAACGTATTCTTCTTTTAATTCGGGGAAAATATCTG    -80
      AAAATTGTCAGAAAAATAACTGGAAATTTTTATAGTCCCGGAGCTTCTTATTCCATAAGATACTAAATTTTGAAAAAAA   -1

      ATG TCA AAA AGA TGT GCA AGT GAT AGA GAT GAT AAA GAT TCG GAG CCC GTT AAG AAA CCC    60
1     M   S   K   R   C   A   S   D   R   D   D   K   D   S   E   P   V   K   K   P

      CGT TTT TCG ATT GCC AAC ATT CTT GAT GAA GTA GAA GAC GAG GAA GAT GTG GAA GTT GAT   120
21    R   F   S   I   A   N   I   L   D   E   V   E   D   E   E   D   V   E   V   D

      GTT GAA GAC GTG GAT GAC GTG GaT TTA TCA TCA ATT CCA TCG AAA AGT CCT GAA AGA TCA   180
41    V   E   D   V   D   D   V   D   L   S   S   I   P   S   K   S   P   E   R   S

      AGG GGT CGG CCG AAG ATT GGC TTG AAA ATG AAG GAA GGA AAT CTA CCA ATC GAG TGC AAA   240
61    R   G   R   P   K   I   G   L   K   M   K   E   G   N   L   P   I   E   C   K

      TTA GAG GGG TCC GAG CTA TGG GCG AAA TTC TTC GAT TTG GGC ACA GAG ATG ATT ATC ACA   300
81    L   E   G   S   E   L   W   A   K   F   F   D   L   G   T   E   M   I   I   T

      AAA AGT GGA AGG CGA ATG TTC CCA ACA GTT AAA GTT TCC TTC ACA AAT GTA ATA CTA GAT   360
101   K   S   G   R   R   M   F   P   T   V   K   V   S   F   T   N   V   I   L   D

      GCT CTA TAT TAT ATC TTC CTA GAT GTA GTC CCA GTG GAC TCC AAA AGA TAC CGT TAC ATC   420
121   A   L   Y   Y   I   F   L   D   V   V   P   V   D   S   K   R   Y   R   Y   I

      TAC AAC AAA TCA GCT TGG CTC ACA GCC GGT AAA GCG GAG CCT GTG CCA AAA AAT CGA TAT   480
141   Y   N   K   S   A   W   L   T   A   G   K   A   E   P   V   P   K   N   R   Y

      TAC CTT CAC CCG GAC TCC CCA TTT ACC GGT GAT CAA CTG CTG AAG CAC GTG ATT TCG TTT   540
161   Y   L   H   P   D   S   P   F   T   G   D   Q   L   L   K   H   V   I   S   F

      GAA AAG ACT AAA TTG ACA AAT AAC GAG GTG GAT AAG ACG GGT CAT CTA ATC CTA AAC TCA   600
181   E   K   T   K   L   T   N   N   E   V   D   K   T   G   H   L   I   L   N   S

      ATG CAC AAA TAC CAG CCA CGT ATT CAC ATT GTC CAA CGC CAG AAA GCC AAC CCC TTG GAC   660
201   M   H   K   Y   Q   P   R   I   H   I   V   Q   R   Q   K   A   N   P   L   D

      CCC AAC AAA GTA GTG ATG AGT GAA GAA AAA CAC TGC ACC TAC ACA TTT CCA GAA ACA CAA   720
221   P   N   K   V   V   M   S   E   E   K   H   C   T   Y   T   F   P   E   T   Q

      TTT ATG GCA GTG ACG GCT TAT CAA AAT CAG TTG ATA ACA AAG CTG AAA ATC GAG AAA AAC   780
241   F   M   A   V   T   A   Y   Q   N   Q   L   I   T   K   L   K   I   E   K   N

      CCT TTT GCA AAG GGG TTT AGG GAT CCT ACA GGA AGG TCG CCG GAT GAA ATG GAA AGG TCC   840
261   P   F   A   K   G   F   R   D   P   T   G   R   S   P   D   E   M   E   R   S

      CCA GGC GAC ATG ATG CTC TCC AAC TTC TAC CAC TCC TCT GCT CTG CGA CAG GCA ATG TTT   900
281   P   G   D   M   M   L   S   N   F   Y   H   S   S   A   L   R   Q   A   M   F

      CAG CAG TGC CTG AGC AAG ACT CTT CAG TTG AAT CCA TCA ATT ATG ATG CTT TAT CAG AAT   960
301   Q   Q   C   L   S   K   T   L   Q   L   N   P   S   I   M   M   L   Y   Q   N

      GTT TTT CCA ACG GGG AAT TCG TTG CCA GCA GGA CCG ACG GTA CCC GGC AAT CCT GCA GAA  1020
321   V   F   P   T   G   N   S   L   P   A   G   P   T   V   P   G   N   P   A   E

      ATT TCA ATA AAA TCA GAA TAG                                                    1041
341   I   S   I   K   S   E   AMB

      CTCGCTGCAAATCGTCTTCCACGTGTAAATAACGCAATTGGCCATTTTTTTGCAAAATCTTTAATTTTTTTAGTGTAATTT  1120
      CTGTAATTTTTTTGTGGTGGTATTCATATGATATCCGGTTTTAAATTTGTGTGCATGCATGTGTCTGAATATTTGAATTT  1199
      GAATAAAAAAATTGATGTAAAAAAAAAAA                                                  1229
```

TA(CT) CA(AG) AA(CT) GA) primers. Amplification conditions were 94°C for 1 min, 50°C for 1 min, and 72°C for 2 min for 35 cycles. A single resulting product was purified with the Qiagen kit, cloned using the TA cloning kit (Invitrogen), and sequenced using the Sequenase kit (US Biochemicals). Sequence information was used to devise a pair of gene-specific primers. PCR amplification of 1 μL of a cDNA library (5×10^9 phu/mL) was performed with the gene-specific primers for 35 cycles (94°C for 1 min, 55°C for 1 min, and 72°C for 1 min). The resulting 170-bp product was used as a hybridization probe to obtain a complete cDNA clone. In each case of PCR

Fig. 2. Position of *T*-box domains within predicted protein products of the *T*-locus gene and *C. elegans T*-box genes. Sizes of the N- and C-terminal regions that flank the *T*-box domain are indicated by numbers. Total sizes of predicted products are shown to the right. The GenBank accession numbers for *C. elegans T*-box genes are *Ce-tbx-2*, U11279; *Ce-tbx-7*, U50193; *Ce-tbx-8* and *Ce-tbx-9*, Z29443; *Ce-tbx-11*, U21310; and *Ce-tbx-12*, U56081.



cloning, three independent products were sequenced to eliminate PCR-derived mutations.

Screening of cDNA and genomic libraries

We screened 2×10^5 clones from both embryonic cDNA and genomic libraries with a specific probe under stringent conditions. Positive clones were isolated, and cDNA inserts were subcloned in BlueScript KS (-) and sequenced from both strands using specific primers.

Evolutionary analysis

We have used amino acid sequences to reconstruct the phylogenetic relationships within the *T*-box gene family because they provide a more robust approach to examining the distant evolutionary relationships as nucleotide changes tend to become saturated. The initial alignment was done using the PILEUP program of the GCG package (Genetics Computer Group 1989) and later improved manually. Two regions of the alignment were omitted from further study, owing to extensive length and sequence variation (Agulnik et al. 1996). Phylogenetic trees were constructed using the neighbor-joining algorithm of Saitou and Nei (1987) as implemented in the METREE program (Rzhetsky and Nei 1994). We used Poisson-corrected distances to account for multiple hits, which are likely when distant comparisons are performed (for discussion see Kumar et al. (1993)). The reliability of internal nodes was assessed by computing the confidence–probability values (Rzhetsky and Nei 1992) and by 1000 bootstrap replications (Felsenstein 1985). We have used the MEGA program (Kumar et al. 1993) to conduct the bootstrap analysis as well as to compute distances.

Results

Identification of three new *C. elegans T*-box genes

To uncover novel a *T*-box gene(s) in the *C. elegans* genome, we performed PCR analysis on total genomic DNA, using a pair of degenerate primers encoding two conserved regions within the previously characterized *T*-domain sequence: YIHPDSP and VTAYQNE. This same primer pair was used successfully in the initial discovery of the first three mouse *T*-box genes: *Tbx1*, *Tbx2*, and *Tbx3* (Bollag et al. 1994).

A single product, 0.7 kb in size, was observed as the result of PCR amplification from the *C. elegans* genome. Sequence analysis of this product revealed two regions of homology to the *T*-box consensus sequence flanking a possible intron between them. Based on this sequence, we designed two specific internal primers and used this pair for PCR amplification from an aliquot of an embryonic cDNA library. The sequence of the resulting PCR product coincided completely with the two regions of the genomic sequence that showed evidence of similarity to the conserved *T*-box domain. This result confirmed the existence of an intron in the middle of the genomic sequence and established the locations of splice donor and acceptor sites.

The cloned PCR-amplified cDNA product was used as a hybridization probe to screen the embryonic cDNA library. Two independent clones of 1.4 and 0.8 kb were recovered. The larger clone contained the full open reading frame of a novel gene (Fig. 1), along with a 3′-untranslated region (3′-UTR) with a polyadenylation signal and a poly(A) tail. Based on our previously established system of *T*-box nomenclature (Agulnik et al. 1995; Agulnik et al. 1996), we are calling this novel gene *Ce-tbx-12*. The deduced product of *Ce-tbx-12* consists of 346 residues, with the position of the *T*-box domain extending from amino acid 86 to amino acid 276 (Figs. 1 and 2).

To establish the positions of other introns within the *Ce-tbx-12 T*-box domain, we used specific primers to amplify different parts of the *T*-box from genomic DNA. The resulting analysis shows, that along with the above described intron, there are three others. Their locations are indicated in Fig. 3.

Since the publication of our first report describing the identification of four *T*-box genes in the *C. elegans* genome, *Ce-tbx-2*, *Ce-tbx-7*, *Ce-tbx-8*, and *Ce-tbx-9* (Agulnik et al. 1995), two additional members of this gene family have been identified through the *C. elegans* Genome Sequencing Project. The first of these was designated F40H6.4 (accession No. U21310). We have given the name *Ce-tbx-11* to this gene. The second gene was found within the cosmid H14A12 (Wilson et al. 1994) and we have named it *Ce-tbx-17*.

Phylogenetic analysis of the *C. elegans T*-box genes

Figure 3 shows an alignment of the *T* domains of all known *C. elegans T*-box genes. This alignment acted as the basis for building the phylogenetic tree that has 21 *T*-box domains shown in Fig. 4. As there is no unequivocal way of building a rooted tree in the absence of an established outgroup sequence, we have used the unweighted pair-group method with arithmetic averaging (UPGMA) (as implemented in the MEGA program) to locate the most likely root position. This analysis placed the root between the *Ce-tbx-11* sequence and the rest of the gene family, since the sequence of *Ce-tbx-11* showed the most dramatic divergence from the "consensus" *T* domain. In particular, the *Ce-tbx-11* sequence has amino acid substitutions in 14 residues (shown as boldfaced letters in Fig. 3) that are otherwise conserved in all other *C. elegans* members of the *T*-box gene family. This provides evidence for a very early duplication event that may have occurred prior to the onset of metazoan evolution.

In contrast, the *Ce-tbx-12* and *Ce-tbx-17* sequences lie securely within the realm of previously described *T*-box genes in mice and other organisms. In particular, *Ce-tbx-12* shows greatest similarity to the set of non-*Brachyury T*-box genes defined in the mouse genome, and is distinct from the *C. elegans* genes *Ce-tbx-7*, *Ce-tbx-8*, *Ce-tbx-9*, and *Ce-tbx-17*. The most closely related sequence to *Ce-tbx-12* is found in the recently described *Drosophila H15* gene.

Map locations of the *C. elegans T*-box genes

The *Ce-tbx-12* gene was mapped with the use of two independent protocols. First, the original PCR-derived cDNA fragment was hybridized to an ordered array of YAC (yeast artificial chromosome) clones (kindly provided by Dr. A. Coulson, Cambridge, England) spotted onto a nylon filter. Two overlapping YACs from chromosome II showed a positive signal after hybridization, Y54A7 and Y49C6. Second, lambda clones containing *Ce-tbx-12* that had been isolated from a genomic library were subjected to fingerprint data analysis by Dr. Coulson. The results demonstrate the localization of the gene to two neighboring cosmid clones, W07E6 and C55G12, which are themselves contained within YAC Y54A7 (Fig. 5). This result suggests that Y49C6 hybridization is probably an artifact.

Ce-tbx-11 and *Ce-tbx-17* have been localized by the *C. elegans* Genome Sequencing Project to the same region of chromosome III that contains four other previously characterized *T*-box genes (Agulnik et al. 1995; Fig. 5). The closest of these to *Ce-tbx-11* is *Ce-tbx-7*, which is located in the neighboring cosmid, ZK328, about 15 kb away.

Discussion

In this report, we describe the characterization of three new members of the *C. elegans T*-box gene family, which we have named *Ce-tbx-11*, *Ce-tbx-12*, and *Ce-tbx-17*. In total, seven distinct *C. elegans T*-box genes have now been characterized. Phylogenetic analysis indicates an ancient origin for all seven genes. The topology of the tree shown in Fig. 4 allows us to mark the divergence of the nematode and vertebrate lineages at the branchpoint between *Ce-tbx-2* and the closest mouse ortholog pair *Tbx2/Tbx3*. Surprisingly, based on topology alone, it appears that *Ce-tbx-12*, *Ce-tbx-7*,

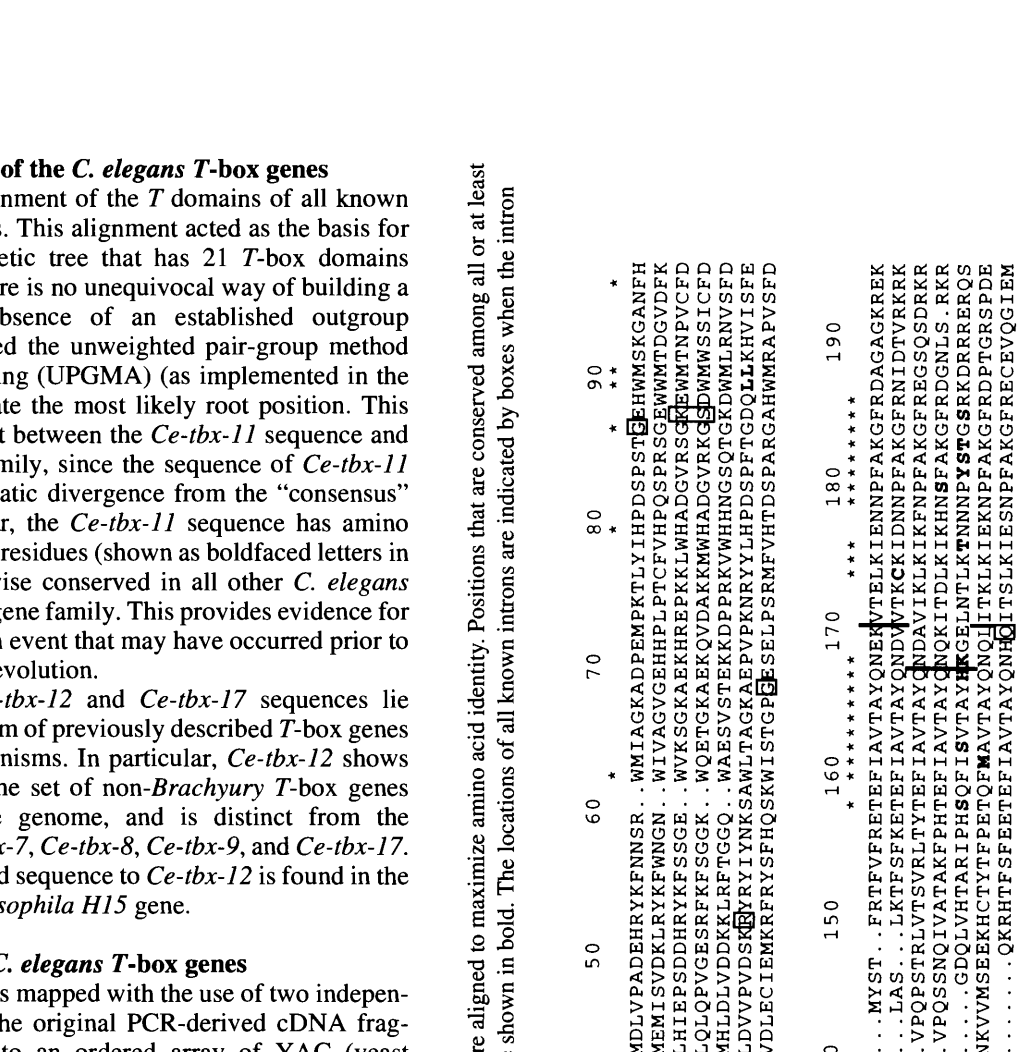


Fig. 3. Alignment of *T*-box domains from seven *C. elegans T*-box genes. The sequences are aligned to maximize amino acid identity. Positions that are conserved among all or at least six sequences are marked by a star. Amino acid residue substitutions in these positions are shown in bold. The locations of all known introns are indicated by boxes when the intron occurs within a codon, and by lines when the intron occurs between codons.

Fig. 4. Phylogenetic tree of *T*-box sequences. All *C. elegans T*-box genes characterized to date are shown in large type. Confidence–probability values, indicated as percentile numbers, are shown above a particular branch (only values higher than 70% are presented). The percent frequency of occurrence of the given branch in 1000 bootstrap replicates is indicated under the branch. Mm-*T* (*Brachyury*), *Tbr-1*, and *Tbx1-6* are the mouse genes; *Trg*, *H15*, and *omb* are the *Drosophila* genes; and Ac-*T*, Zf-*T*, and Xl-*T* are *Brachyury* homologs from ascidians, zebrafish, and *Xenopus laevis*, respectively.

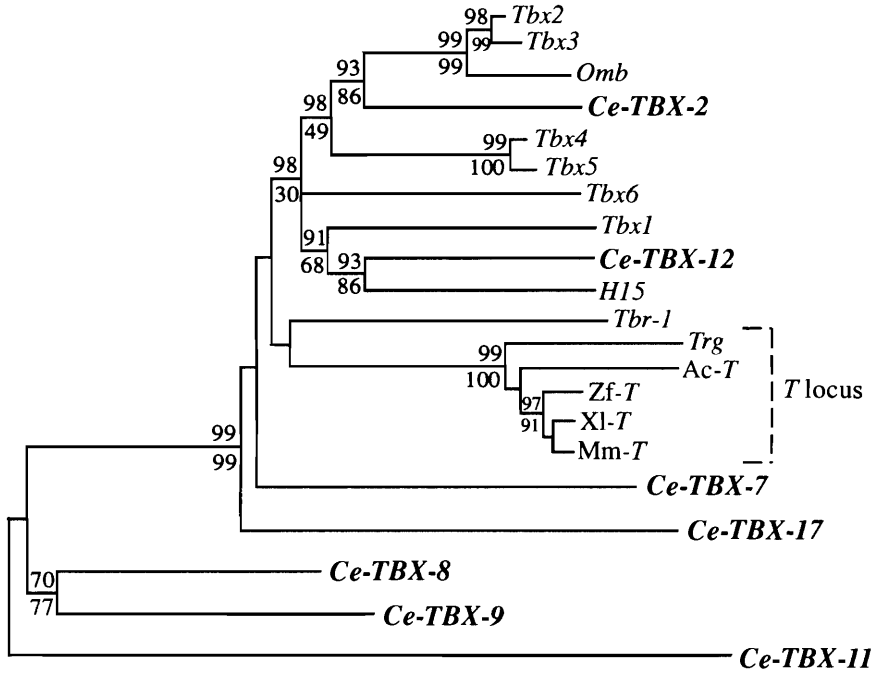
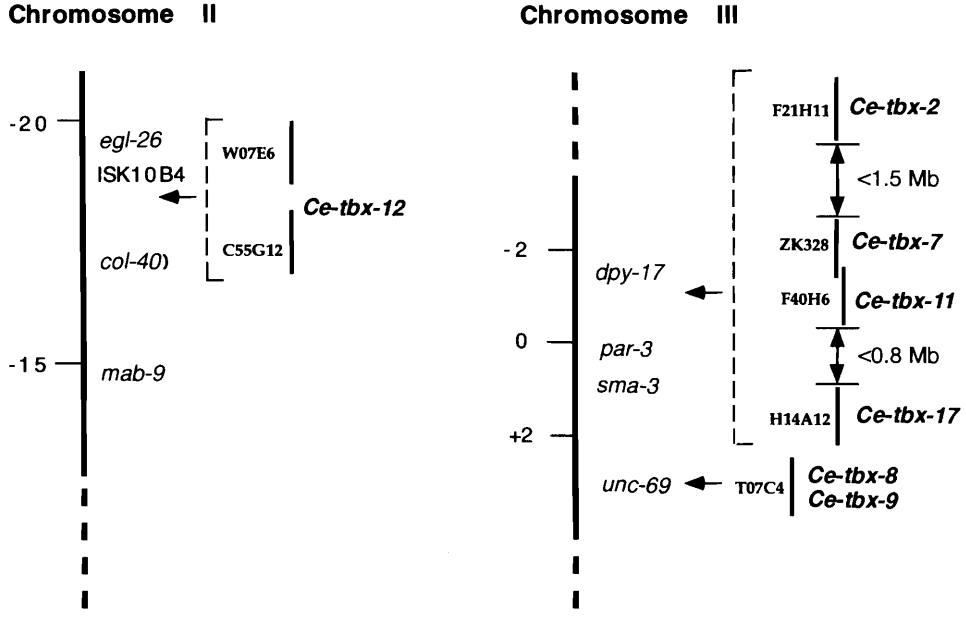


Fig. 5. Chromosome map positions of *C. elegans T*-box genes. Some marker genes closely linked to *T*-box genes are shown for orientation. *T*-box gene containing cosmids are shown in boldfaced letters. The *Ce-tbx-12* gene was located in the terminal region of chromosome II. Six other *T*-box genes are clustered within 5 map units of the central part of chromosome III.



Ce-tbx-11, and *Ce-tbx-17*, and the ancestral sequence to *Ce-tbx-8/Ce-tbx-9* all existed as separate genes prior to this very early metazoan branch point. This result underscores the very ancient nature of the *T*-box gene family. It also suggests that each of these *C. elegans* genes might have mammalian homologs.

The phylogenetic tree can be used to infer the existence of potential homologous relationships between individual *T*-box

genes from different species. Thus, *Ce-tbx-2* is more closely related to the mouse genes *Tbx2* and *Tbx3* than it is to any of the other *C. elegans T*-box genes. This suggests the possibility of a homologous relationship. A second homologous relationship is suggested by the presence of *Ce-tbx-12* and the *Drosophila T*-box gene *H15* together on an isolated branch of the tree. However, none of the other *C. elegans T*-box genes display obvious relationships to particular *T*-box genes present in other species. And, in those cases of suggested homology, future functional studies will be required to prove or reject interspecies relationships.

An analysis of intron positions within the *C. elegans T*-box domains generally confirms the established tree topology (Fig. 3). The most divergent, *Ce-tbx-11*, possesses only one intron, which is also present in all other *C. elegans* genes. A three amino acid shift in the position of this intron separates the more closely related *Ce-tbx-2*, *Ce-tbx-7*, *Ce-tbx-12*, and *Ce-tbx-17* genes from the *Ce-tbx-8*, *Ce-tbx-9*, and *Ce-tbx-11* genes. An exon–intron boundary shift occurred in the common ancestor of *Ce-tbx-2*, *Ce-tbx-7*, *Ce-tbx-12*, and *Ce-tbx-17*; a shift also occurred in the position of the common middle intron, after separation from the ancestral gene of *Ce-tbx-8/Ce-tbx-9*.

Characterization of the chromosomal locations of the seven *C. elegans T*-box genes shows that six of them map to a small region of less then 5 map units on chromosome III. Within this group, two genes, *Ce-tbx-8* and *Ce-tbx-9*, are located within 2 kb of each other, and two others, *Ce-tbx-7* and *Ce-tbx-11*, are located within 15 kb of each other. The fact that no *T*-box genes have been uncovered within sequenced regions of equal or greater length from chromosomes II, IV, V, and X suggests that the clustering on chromosome III is not a chance event. Rather, it appears likely that all five *T*-box genes are derived from ancient duplications caused by unequal crossing over. However, in the absence of selection, it seems likely that dispersion would have taken place over the long span of evolutionary time since these events occurred. The fact that dispersion has not occurred provides support for a model in which *T*-box genes share common regulatory elements that prevent their separation from each other.

To date, none of the *C. elegans T*-box genes has been subjected to functional analysis. However, it is possible to infer certain functional attributes based on a comparison of the presumed protein products. All seven genes show sequence similarity in the DNA binding domain defined by the *T*-locus product. DNA binding activity has now been associated with products derived from four different *T*-box genes (Pflugfelder et al. 1992; Kispert and Herrmann 1993; N. Garvey, R.J. Bollag, and L.M. Silver, unpublished data), and a role as a transcriptional regulator has been demonstrated for the *T*-locus product (Kispert et al. 1995). These results strongly suggest that all members of the *T*-box gene family will be DNA binding proteins, and further, that they may all be involved in transcriptional regulation.

Kispert et al. (1995) have demonstrated the existence of two transactivation and two repression domains in the 209 residue long C-terminal region of the prototypical mouse *Brachyury* gene product. The observation that each *C. elegans* gene product shows no detectable sequence similarity in the N-terminal and C-terminal regions that flank the *T*-box, and the observation that these non-*T*-box regions vary greatly in length, lead to the likelihood that each will have different functional specificity. However, issues of functionality will only be resolved with further genetic and biochemical studies.

Note added in proof: While the manuscript was in the editorial office, a new *T*-box gene (accession No. AF000261) was uncovered through the efforts of the *C. elegans* Genome Project. In accordance with our nomenclature it was named *Ce-tbx-18*. It is localized between map units 5 and 6 on the left arm of chromosome II. Evolutionary analysis indicates that *Ce-tbx-18* is the most divergent *T*-box gene in *C. elegans*, as it lacks some elements conserved between *Ce-tbx-11* and the rest of the family.

Acknowledgments

This research was supported by a grant from the National Institutes of Health (HD20275). We thank Alan Coulson of the Medical Research Council Research Centre in Cambridge, England for help with the chromosomal mapping of the *Ce-tbx-12* gene, and Richard Wilson from Washington University in St. Louis for unpubished sequence information.

References

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

Agulnik, S.I., Garvey, N., Hancock, S., Ruvinsky, I., Chapman, D.L., Agulnik, I., Bollag, R., Papaioannou, V.E., and Silver, L.M. 1996. Evolution of *T-box* genes by tandem duplication and cluster dispersion. Genetics, **144**: 249–254.

Bollag, R.J., Siegfried, Z., Cebra-Thomas, J., Garvey, N., Davison, E.M., and Silver, L.M. 1994. An ancient family of embryonically expressed mouse genes sharing a conserved protein motif with the *T* locus. Nature Genet. **7**: 383–389.

Brook, W.J., and Cohen, S.M. 1996. Antogonistic interection between Wingless and Decapentaplegic responsible for dorsal-ventral pattern in the *Drosophila* leg. Science (Washington, D.C.), **273**: 1373–1377.

Bulfone, A., Simga, S.M., Shimamura, K., Peterson, A., Puelles, L., and Rubenstein, J.L.R. 1995. *T-brain-1*: a homolog of *Brachyury* whose expression defines molecularly distinct domains within the cerebral cortex. Neuron, **15**: 63–78.

Chapman, D.L., Garvey, N., Hancock, S., Alexiou, M., Agulnik, S., Gibson-Brown, J.J., Cebra-Thomas, J., Bollag, R.J., Silver, L.M., and Papaioannou, V.E. 1996. Expression of the T-box family genes, *Tbx1–Tbx5*, during early mouse development. Dev. Dyn. **206**: 379–390.

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

Genetics Computer Group. 1989. The Wisconsin Genetics Computer Group package. Maryland Biotechnology Institute, Madison, Wis.

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

Herrmann, B.G., Labeit, S., Poustka, A., King, T., and Lehrach, H. 1990. Cloning of the *T* gene required in mesoderm formation in the mouse. Nature (London), **343**: 617–622.

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

Kispert, A., Herrmann, B.G., Leptin, M., and Reuter, R. 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., Koschorz, B., and Herrmann, B. 1995. The T protein encoded by *Brachyury* is a tissue-specific transcription factor. EMBO J. **14**: 4763–4772.

Kumar, S., Tamura, K., and Nei, M. 1993. MEGA: molecular evolutionary genetics analysis. Pennsylvania State University, University Park, Pa.

Pflugfelder, G.O., Roth, H., and Poeck, B. 1992. A homology domain shared between *Drosophila optomotor-blind* and mouse *Brachyury* is involved in DNA binding. Biochem. Biophys. Res. Commun. **186**: 918–25.

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

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

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

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Shownkeen, R., Smaldon, N., Smith, A., Sonnhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wohldman, P. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. Nature (London), **368**: 32–38.

This article has been cited by:

1. S Sinha. 2000. Differential DNA binding and transcription modulation by three T-box proteins, T, TBX1 and TBX2. *Gene* **258**:1-2, 15-29. [CrossRef]
2. Christian T Carson, Eric R Kinzler, Brian A Parr. 2000. Tbx12, a novel T-box gene, is expressed during early stages of heart and retinal development. *Mechanisms of Development* **96**:1, 137-140. [CrossRef]
3. Moritz Meins, Deborah J. Henderson, Shomi S. Bhattacharya, Jane C. Sowden. 2000. Characterization of the Human TBX20 Gene, a New Member of the T-Box Gene Family Closely Related to the Drosophila H15 Gene. *Genomics* **67**:3, 317-332. [CrossRef]
4. Dae-gwon Ahn, Ilya Ruvinsky, Andrew C. Oates, Lee M. Silver, Robert K. Ho. 2000. tbx20, a new vertebrate T-box gene expressed in the cranial motor neurons and developing cardiovascular structures in zebrafish. *Mechanisms of Development* **95**:1-2, 253-258. [CrossRef]
5. Sergei I. Agulnik, Virginia E. Papaioannou, Lee M. Silver. 1998. Cloning, Mapping, and Expression Analysis ofTBX15,a New Member of the T-Box Gene Family. *Genomics* **51**:1, 68-75. [CrossRef]
6. Virginia E. Papaioannou, Lee M. Silver. 1998. The T-box gene family. *BioEssays* **20**:1, 9-19. [CrossRef]