Evolution of Antennapedia-Class Homeobox Genes

Jianzhi Zhang and Masatoshi Nei

Institute of Molecular Evolutionary Genetics and Department of Biology, The Pennsylvania State University, University Park, Pennsylvania 16802

Manuscript received June 13, 1995
Accepted for publication September 21, 1995

ABSTRACT

Antennapedia (*Antp*)-class homeobox genes are involved in the determination of pattern formation along the anterior-posterior axis of the animal embryo. A phylogenetic analysis of *Antp*-class homeodomains of the nematode, Drosophila, amphioxus, mouse, and human indicates that the 13 cognate group genes of this gene family can be divided into two major groups, *i.e.*, groups I and II. Group I genes can further be divided into subgroups A (cognate groups 1–2), B (cognate group 3), and C (cognate groups 4–8), and group II genes can be divided into subgroups D (cognate groups 9–10) and E (cognate groups 11–13), though this classification is somewhat ambiguous. Evolutionary distances among different amino acid sequences suggest that the divergence between group I and group II genes occurred ~1000 million years (MY) ago, and the five different subgroups were formed by ~600 MY ago, probably before the divergence of Pseudocoelomates (*e.g.*, nematodes) and Coelomates (*e.g.*, insects and chordates). Our results show that the genes that are phylogenetically close are also closely located in the chromosome, suggesting that the colinearity between the gene expression and gene arrangement was generated by successive tandem gene duplications and that the gene arrangement has been maintained by some sort of selection.

THE homeobox is a highly conserved sequence of ~180 nucleotides contained in many genes controlling development. It encodes the homeodomain that is capable of binding a DNA motif and regulating gene transcription (GEHRING *et al.* 1994). Homeobox genes are involved in the specification of the body plan, determination of cell fate, and several other basic developmental processes (MCGINNIS and KRUMLAUF 1992; LAWRENCE and MORATA 1994). More than 300 homeobox-containing genes and their relatives have been identified and sequenced in fungi, plants, and animals, and they can be classified into ≥30 different classes (KAPPEN *et al.* 1993; BÜRGLIN 1994).

The Antennapedia (*Antp*)-class homeobox genes are of special importance, because they specify the developmental patterning of the body segments along the anterior-posterior axis of the animal embryo. In all metazoans so far examined, this class of genes exists as one or more clusters of genes in the genome. The arrangement of the genes in the chromosome is identical or highly correlated to the order of the genes that specify the anterior-posterior body segments. In the developmental process, the 3′ end gene of the *Antp*-class homeobox gene complex is first transcribed, and the other genes are transcribed successively from the 3′ to the 5′ end of the complex (DUBOULE and MORATA 1994; KRUMLAUF 1994). Almost all *Antp*-class genes located in a chromosome have the same transcriptional direction with a few exceptions (see DISCUSSION).

Corresponding author: Masatoshi Nei, Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, 328 Mueller Laboratory, University Park, PA 16802. E-mail: nxm2@psuvm.psu.edu

In vertebrates, there are four clusters of *Antp*-class homeobox genes, and the genes within and between the clusters are evolutionarily related. They can be classified into 13 cognate gene groups (KAPPEN and RUDDLE 1993) (see Figure 1). However, none of the clusters has all 13 cognate genes. The amphioxus (*Branchiostoma floridae*), which is thought to be a sister group of vertebrates, has one cluster of ≥10 cognate genes (GARCIA-FERNÀNDEZ and HOLLAND 1994). There are eight *Antp*-class homeobox genes in *Drosophila melanogaster*, but they are located in two separate clusters on the same chromosome (LAWRENCE and MORATA 1994). The nematode (*Caenorhabditis elegans*) has one cluster of four *Antp*-class homeobox genes (BÜRGLIN and RUVKUN 1993; SALSER and KENYON 1994). The orthologous and paralogous relationships of these genes are believed to be as given in Figure 1 (KAPPEN and RUDDLE 1993; RUDDLE *et al.* 1994), though the relationships of cognate gene groups 6, 7, and 8 are somewhat ambiguous (see KRUMLAUF 1994).

There are studies of *Antp*-class homeobox genes in many other animals such as hydras (MURTHA *et al.* 1991; NAITO *et al.* 1993), flatworms (BARTELS *et al.* 1993), annelids (DICK and BUSS 1994; SNOW and BUSS 1994), crustaceans (CARTWRIGHT *et al.* 1993), acorn worms (PENDLETON *et al.* 1993), lampreys (PENDLETON *et al.* 1993), and others. These animals have very different body plans. However, the DNA sequences available from these studies are either truncated or their genomic organizations are unknown. Therefore, the orthologous or paralogous relationships of the genes are unclear.

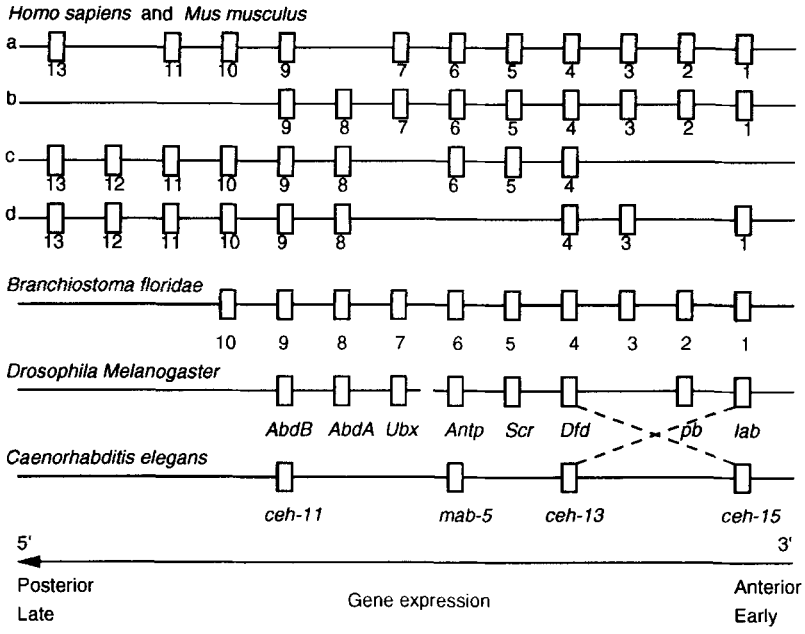


FIGURE 1.—Genomic organizations of Antennapedia-class homeobox genes in the human (*H. sapiens*), mouse (*M. musculus*), amphioxus (*B. floridae*), Drosophila (*D. melanogaster*) and nematode (*C. elegans*). There is an inversion of genes *ceh-13* and *ceh-15* in the nematode. The arrow indicates the order of gene expression. Moving along the clusters in a 3′ to 5′ direction, each successive gene expresses later in the developmental process and more posterior along the anterior-posterior axis of the animal embryo.

A number of authors have studied the evolutionary relationships of *Antp*-class homeobox genes with the aim of understanding the evolution of morphogenesis. However, many of these studies are qualitative (*e.g.*, KAPPEN *et al.* 1989; SCHUGHART *et al.* 1989; HOLLAND 1992). Schubert *et al.* (1993) conducted a phylogenetic analysis of human and Drosophila *Antp*-class homeobox genes, but the statistical accuracy of the phylogenetic tree is unclear. We have therefore conducted a detailed analysis of the evolutionary relationships of all *Antp*-class homeobox genes of which the genomic organization is known. We studied the evolutionary relationships of cognate gene groups as well as those of the four clusters of *Antp*-class homeobox genes.

MATERIAL AND METHODS

Amino acid sequence data: In the present study, we used amino acid sequences rather than nucleotide sequences, because synonymous nucleotide substitutions are apparently saturated in most gene comparisons and this would introduce noise in constructing phylogenetic trees (Russo *et al.*, 1996). We used altogether 98 *Antp*-class homeodomain sequences, four from the nematode (*C. elegans*), eight from Drosophila (*D. melanogaster*), 10 from the amphioxus (*B. floridae*), 38 from the mouse (*Mus musculus*) and 38 from the human (*Homo sapiens*). All the sequences were obtained from KAPPEN *et al.* (1993) except amphioxus sequences (GARCIA-FERNÀNDEZ and HOLLAND 1994), mouse *Hox-a-13* (HAACK and GRUSS 1993), and *Hox-c-12* and *Hox-c-13* sequences (PETERSON *et al.* 1994). There were no deletions and insertions in the amino acid sequences (60 amino acids) of these homeodomains, so alignment was straightforward.

Phylogenetic analysis: The evolutionary distance between two amino acid sequences was measured by the proportion of different amino acids between the sequences (*p*-distance). The reason for using *p*-distance rather than Poisson-correction distance or Dayhoff distance (see KUMAR *et al.* 1993) is that we are primarily interested in determining the topology of the phylogenetic tree, and for this purpose *p*-distance is often better than other distance measures in our experience (*e.g.*, HUGHES and NEI 1993). Similar results have been obtained about the *p*-distance and the Jukes-Cantor distance for nucleotide sequences (*e.g.*, SAITOU and NEI 1987; NEI 1991). The phylogenetic trees were constructed by using the neighbor-joining (NJ) method (SAITOU and NEI 1987) with *p*-distance. The NJ method is known to be quite efficient in obtaining reliable trees (NEI 1991; NEI *et al.* 1995). We did not use other commonly used methods such as maximum parsimony and maximum likelihood, because these methods could not handle the large number of sequences we analyzed. To root the phylogenetic tree for all the *Antp*-class homeobox genes, we used two non-*Antp*-class homeobox sequences from the nematode (*ceh-5* and *ceh-19*) (see KAPPEN *et al.* 1993) as outgroups. Computation of *p*-distances and construction of phylogenetic trees were conducted by using the computer software MEGA (KUMAR *et al.* 1993). In the construction of the phylogenetic tree of the four clusters of genes in vertebrates, we computed *p*-distances for all pairs of gene clusters from the human and the mouse (only cognate gene groups 1–10 were considered because the amphioxus cluster has only cognate genes 1–10). *p*-distances were computed by using all amino acids involved in each pairwise comparison (pairwise deletion option). The NJ tree was then constructed by using the amphioxus cluster as the outgroup.

The reliability of the trees obtained was tested by a

bootstrap method (FELSENSTEIN 1985) with 1000 replications and the confidence probability of each interior branch (RZHETSKY and NEI 1992; SITNIKOVA *et al.* 1995). The bootstrap test was conducted by using MEGA, whereas the confidence probability test of each interior branch was done by using a computer program developed by N. TAKEZAKI.

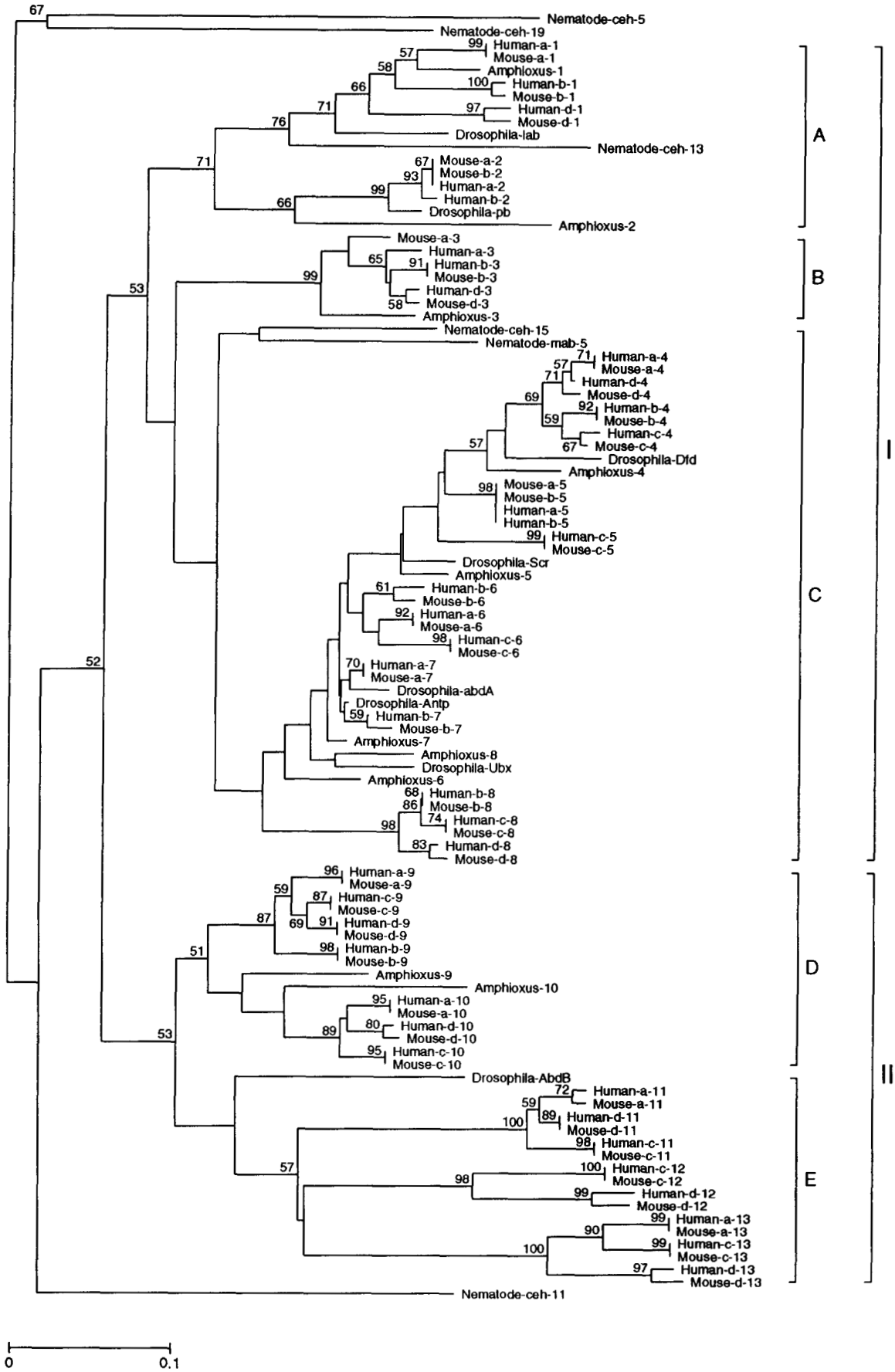
RESULTS

Evolutionary relationships of *Antp*-class homeobox genes: Figure 2 shows the phylogenetic tree of 98 *Antp*-class homeodomains from the five different species mentioned above. It is seen that the *Antp*-class homeobox genes can be classified into two major groups, group I and group II. Group I genes can further be divided into three subgroups, *i.e.*, subgroups A, B, and C. This subgrouping is not statistically well supported and thus provisional. Subgroup A includes cognate gene groups 1 and 2 of the mouse, human, and amphioxus; *pb* (proboscipedia) and *lab* (labial) of Drosophila; and *ceh-13* of the nematode. Inclusion of the gene *ceh-13* in this subgroup is reasonable, because *ceh-13* and *ceh-15* of the nematode are considered to be inverted and *ceh-13* seems to be orthologous with the cognate gene group 1 (BÜRGLIN and RUVKUN 1993). Subgroup B corresponds to cognate gene group 3 of the mouse, human, and amphioxus and is more closely related to subgroup C. However, this relationship is not statistically well supported. Subgroup C consists of cognate gene groups 4–8 of the mouse, human, and amphioxus; *Dfd* (Deformed), *Scr* (Sex combs reduced), *Antp* (Antennapedia), *abdA* (abdominal-A) and *Ubx* (Ultrabithorax) of Drosophila; and *mab-5* and *ceh-15* of the nematode. The nematode genes *ceh-15* and *mab-5* are supposed to be orthologous with cognate gene groups 4 and 6, respectively (KAPPEN and RUDDLE 1993), but they form a cluster separate from that of other genes in subgroup C. This cluster makes the current grouping of cognate genes questionable, and the function of the nematode genes might have differentiated from that of the other genes in subgroup C. However, it is also possible that this clustering pattern occurred by chance because the bootstrap values for the clusters within subgroup C are generally low. According to the tree in Figure 2, cognate group 5 genes do not form a monophyletic group, but this could be due to sampling error caused by the stochastic process of amino acid substitution. The homeobox has only 60 codons, so that it is difficult to obtain definitive conclusions from a phylogenetic analysis alone. Cognate group 6, 7, and 8 genes also do not necessarily form a monophyletic group. This could again be due to sampling error or to functional differentiation. Although the experimental results seem to support the orthologous relationships of Drosophila gene *Antp*, *Ubx*, and *abdA* with vertebrate cognate gene groups 6, 7, and 8, respectively (BACHILLER *et al.* 1994), our phylogenetic analysis gives no clear-cut support for these relationships. More extensive experimental study seems to be necessary.

Group II genes can be divided into two subgroups, *i.e.*, subgroups D and E. Subgroup D includes cognate gene groups 9 and 10 of the mouse, human, and amphioxus. The *Hox-9* gene of the amphioxus clusters with cognate group 10 genes, but this is probably due to sampling error because it is not supported by the bootstrap test. Subgroup E consists of cognate groups 11–13 of the mouse and human, and *AbdB* (Abdominal-B) of Drosophila. The Drosophila *AbdB* gene is supposed to belong to cognate group 9, and if we consider the low bootstrap values of the subgroups within group II, it seems that inclusion of this gene in subgroup E is due to sampling error. The nematode gene *ceh-11*, which was previously aligned with cognate group 9 genes (BÜRGLIN *et al.* 1991; WANG *et al.* 1993), appears to be a sister group of all other *Antp*-class genes, though this also could be due to sampling error.

Our bootstrap tests indicate that the clusters of major groups and subgroups of *Antp*-class homeobox genes are not as solid as they look. This is apparently because the number of amino acid sites used is small and the number of sequences analyzed is large. Particularly, the unexpected clustering of the nematode and Drosophila genes mentioned above are not statistically supported. By contrast, many clusters of human and mouse cognate genes are statistically significant. We have also conducted the confidence probability test of each interior branch. This test generally gave a higher probability of confidence than the bootstrap value, but since this test depends on a number of assumptions (SITNIKOVA *et al.* 1995), we have decided to rely on the bootstrap test, which is known to be quite conservative (*e.g.*, ZHARKIKH and LI 1992; HILLIS and BULL 1993; SITNIKOVA *et al.* 1995).

Because there are four clusters of *Antp*-class homeobox genes in the human and mouse and their orthologous and paralogous relationships seem to be more firmly established than those of other organisms, we constructed a phylogenetic tree for 76 genes from these two organisms. The tree obtained is presented in Figure 3. The topology of this tree is identical with that of the tree in Figure 2 for the part of human and mouse genes except the branching order of cognate gene groups belonging to subgroups C and D. However, the bootstrap tests show that the branching pattern of this tree is much more reliable than the previous one. The subdivision of the genes into the two major groups is now statistically significant. Cognate groups 11, 12, and 13 also form a monophyletic group with a bootstrap value of 90%. Many clusters representing different cognate groups are statistically significant. However, the branching order of cognate groups within subgroups C and D remains uncertain. Whether subgroup B is closer to subgroup A or C also remains statistically unresolved. Similarly, the phylogenetic position of cognate group


Nematode-ceh-5
Nematode-ceh-19
Human-a-1
Mouse-a-1
Amphioxus-1
Human-b-1
Mouse-b-1
Human-d-1
Mouse-d-1
Drosophila-lab
Nematode-ceh-13
Mouse-a-2
Mouse-b-2
Human-a-2
Human-b-2
Drosophila-pb
Amphioxus-2
Mouse-a-3
Human-a-3
Human-b-3
Mouse-b-3
Human-d-3
Mouse-d-3
Amphioxus-3
Nematode-ceh-15
Nematode-mab-5
Human-a-4
Mouse-a-4
Human-d-4
Mouse-d-4
Human-b-4
Mouse-b-4
Human-c-4
Mouse-c-4
Drosophila-Dfd
Amphioxus-4
Mouse-a-5
Mouse-b-5
Human-a-5
Human-b-5
Human-c-5
Mouse-c-5
Drosophila-Scr
Amphioxus-5
Human-b-6
Mouse-b-6
Human-a-6
Mouse-a-6
Human-c-6
Mouse-c-6
Human-a-7
Mouse-a-7
Drosophila-abdA
Drosophila-Antp
Human-b-7
Mouse-b-7
Amphioxus-7
Amphioxus-8
Drosophila-Ubx
Amphioxus-6
Human-b-8
Mouse-b-8
Human-c-8
Mouse-c-8
Human-d-8
Mouse-d-8
Human-a-9
Mouse-a-9
Human-c-9
Mouse-c-9
Human-d-9
Mouse-d-9
Human-b-9
Mouse-b-9
Amphioxus-9
Amphioxus-10
Human-a-10
Mouse-a-10
Human-d-10
Mouse-d-10
Human-c-10
Mouse-c-10
Drosophila-AbdB
Human-a-11
Mouse-a-11
Human-d-11
Mouse-d-11
Human-c-11
Mouse-c-11
Human-c-12
Mouse-c-12
Human-d-12
Mouse-d-12
Human-a-13
Mouse-a-13
Human-c-13
Mouse-c-13
Human-d-13
Mouse-d-13
Nematode-ceh-11
A
B
C
D
E
I
II
0
0.1

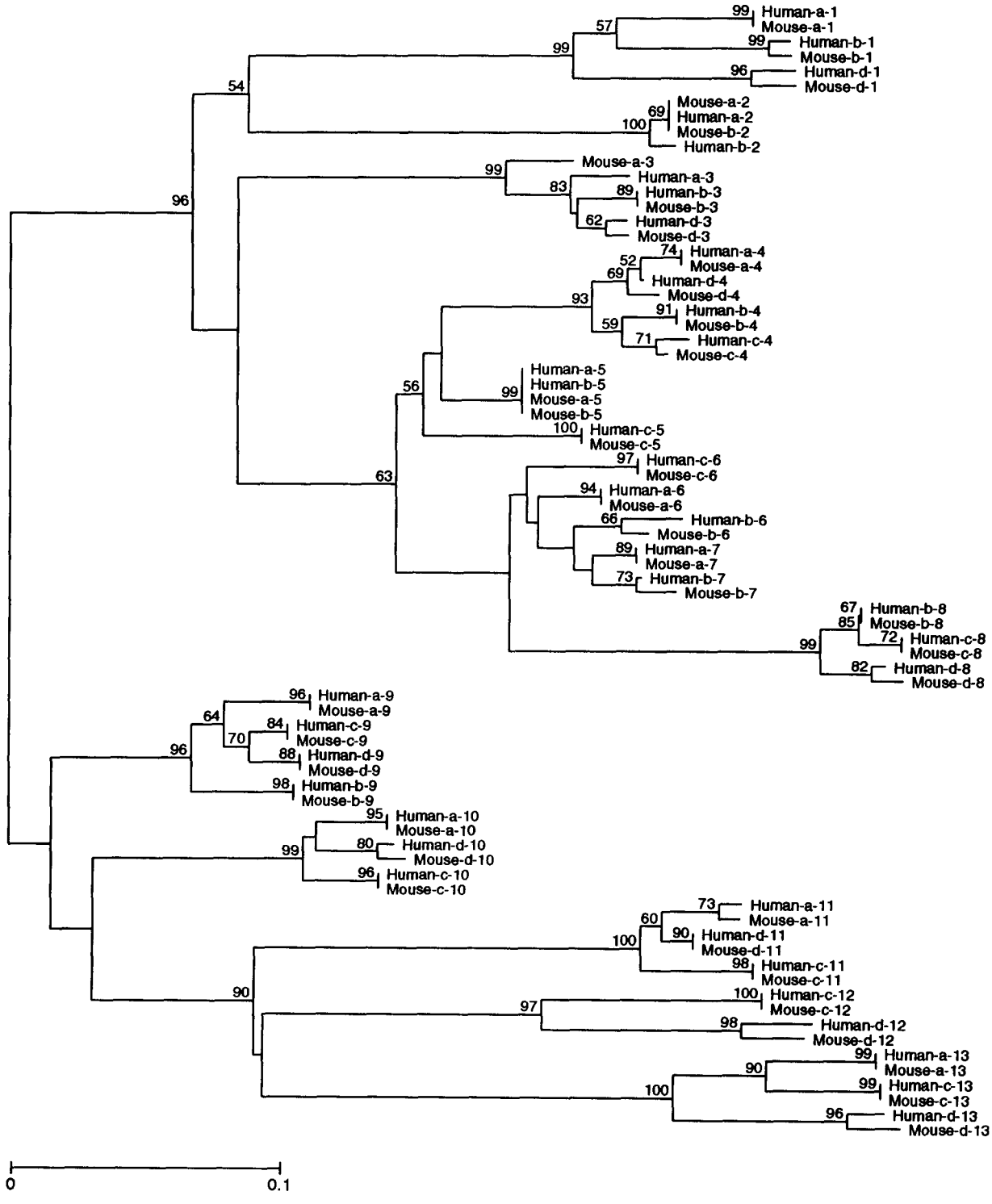


FIGURE 3.—Phylogenetic tree of 76 human and mouse *Antp*-class homeobox genes.

10 is unclear. In the present paper, we assume that cognate group 10 is closer to group 9, because Figure 2 shows a higher bootstrap value (51%) for the D subgroup than that (47%) for the branching order given in Figure 3.

Evolutionary relationships of the four clusters of cog-

FIGURE 2.—Phylogenetic tree of 98 *Antp*-class homeobox genes of the human, mouse, amphioxus, Drosophila, and nematode and two outgroup genes (nematode *ceh-5* and *ceh-19*). The tree is constructed by the NJ method. The numbers for interior branches are bootstrap percentages. Numbers <50 are not shown. The branches are measured in terms of the proportional difference of two amino acid sequences (*p*-distance) with the scale given below the tree.

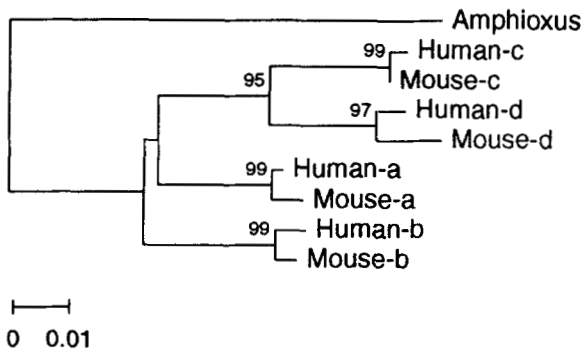


FIGURE 4.—Evolutionary relationships of four clusters of *Antp*-class homeobox genes of the human and mouse with the amphioxus cluster as an outgroup.

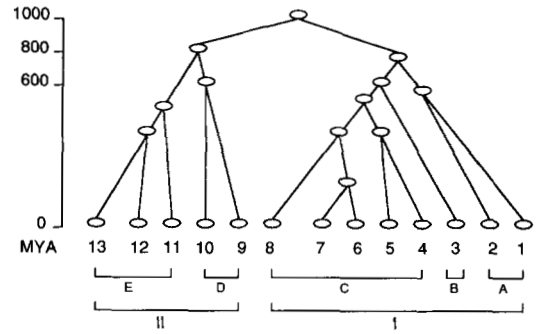


FIGURE 5.—Evolutionary scenario of 13 cognate groups of *Antp*-class homeobox gene family. This scenario is speculated according to the tree of Figure 3 except the phylogenetic position of cognate group 10 (see the text). Branch lengths are not proportional to evolutionary time. I and II are two major groups and A, B, C, D, and E are five subgroups (see the text).

nate genes in vertebrates: There are four clusters of *Antp*-class homeobox genes in vertebrates, but the evolutionary relationships of these clusters are not well established. Our NJ tree of the four clusters from the mouse and human is given in Figure 4, and the branching pattern is of the form (b(a(c, d))). However, we cannot rule out the branching pattern (a(b(c, d))) or ((a, b)(c, d)), because the interior branch between cluster b and the ancestor of clusters a, c, and d is not statistically significant. KAPPEN and RUDDLE (1993) constructed a tree of the four clusters of the human by minimizing the number of gene losses or gains. Their study has shown that the tree ((a, b)(c, d)) is one step shorter than (b(a(c, d))). If these four clusters evolved by genome duplication, the tree ((a, b)(c, d)) is more parsimonious than (a(b(c, d))) or (b(a(c, d))), because the tree ((a, b)(c, d)) can be explained by two events of genome duplication whereas the other trees require three genome duplications and three chromosome losses. At the present time, it is difficult to decide which of the three possible trees is correct.

DISCUSSION

Evolutionary scenario of *Antp*-class homeobox genes: Although the branching order of different cognate group genes is not firmly established in the trees of Figures 2 and 3, it is interesting to speculate the evolutionary history of *Antp*-class homeobox genes; it may give some useful information for planning future experimental studies. Theoretically it is possible to consider several scenarios, but the following one seems to be one of the most plausible ones. First, this gene family obviously evolved from a single ancestral gene through gene duplication, and it seems that the first duplication of this gene produced the ancestral genes of group I and II genes. Duplication of the ancestral group I gene subsequently generated subgroups A, B, and C (Figure 5). Subgroup A and B genes are involved in the segmentation of the anterior part, whereas subgroup C genes control the intermediate part along the anterior-posterior axis of the animal embryo. In the case of group II genes, the first gene duplication seems to have produced the ancestral genes of cognate groups 9–10 and of the others, and the subsequent duplication events generated the remaining cognate genes (Figure 5). A cluster of ≥ 10 cognate genes was formed before the amphioxus diverged from the vertebrate. Gene clusters a, b, c, and d in vertebrates were probably generated by two events of genome duplication in the early stage of vertebrate evolution, and later a few genes were lost by deletion events in each cluster. (The zebrafish, Xenopus, newt, chicken, mouse, and human are known to have the four clusters a, b, c, and d) (RUDDLE *et al.* 1994.) As speculated by KAPPEN and RUDDLE (1993), the generation of these clusters probably contributed to the evolution of more complex organisms.

Origin and evolution of *Antp*-class homeobox genes in invertebrates: The *Antp*-class homeobox genes exist in all metazoans so far surveyed including one of the most primitive metazoans, sponges (DEGNAN *et al.* 1995), but not in flagellates, amoeboids and ciliate protozoans, fungi, algae, and plants (DEGNAN *et al.* 1995). It seems that *Antp*-class homeobox genes originated at the very early stage of metazoan evolution.

Figure 2 suggests that the divergence of the nematode gene *ceh-13*, the Drosophila *lab* gene, and chordates cognate group 1 genes postdated the divergence of cognate gene groups 1 and 2, and that the last common ancestor of the nematode, Drosophila, and chordates already had cognate gene groups 1 and 2. Similarly, we can infer from Figure 2 that the last common ancestor of the nematode, Drosophila, and chordates also had the cognate group gene 3, the ancestral gene of subgroup C, and the ancestral gene of group II. In other words, there were at least five cognate genes in the last common ancestor of Pseudocoelomates (nematode in this paper) and Coelomates (insects and chordates in this papers). In the nematode, however, cognate group genes 2 and 3 seem to have been lost later, and the ancestral gene of groups 4–8

apparently gave rise to genes *mab-5* and *ceh-15*. It is interesting that the positions of *ceh-15* and *ceh-13* in the cluster are inverted according to the sequence comparison and functional analysis (see also BÜRGLIN and RUVKUN 1993). This could be the result of a chromosome inversion in the nematode.

Schubert *et al.* (1993) conducted a phylogenetic analysis of the human and Drosophila *Antp*-class genes. The topology of their tree (Figure 2 of SCHUBERT *et al.* 1993) is different from ours in Figure 3. In their tree, cognate gene group 8 is a sister group with cognate gene groups 1–7, so the genes of cognate groups 4–8 are not monophyletic as in our tree. To examine the reliability of SCHUBERT *et al.*'s tree, we reanalyzed their data and found that the bootstrap values of their tree are lower than ours, though the difference is not very large (data not shown). The difference between the two topologies apparently occurred because SCHUBERT *et al.* used only transversional nucleotide differences whereas we used amino acid sequences. Our analysis has shown that the transition/transversion ratio is not high (0.5–1.5) and that the nucleotide frequencies are nearly the same for all genes. Therefore, the use of transversion sites only would lose a considerable amount of phylogenetic information. When we used only nonsynonymous changes or nucleotide substitutions at first and second codon positions, the tree obtained was similar to our tree rather than to SCHUBERT *et al.*'s.

Origin of cognate gene groups 11–13: The tree in Figure 2 suggests that the divergence of the subgroups D and E occurred earlier than that of the amphioxus and vertebrate cognate group 10 genes. Therefore, if our tree is correct, the amphioxus must have had the ancestor of group 11–13 genes previously. The cognate group 11–13 genes have important functions in the limb formation in the mouse (HAACK and GRUSS 1993; MORGAN and TABIN 1993; DUBOULE 1994). Does the amphioxus have any organ that is homologous to the limbs of tetrapods? The amphioxus is thought to be a sister group of vertebrates and has some fin-like structures. However, the fin-like structures are neither paired nor separated as fish fins are. In the amphioxus, cartilage-like materials stiffen the dorsal fin, but no normal vertebrate skeleton is found. Therefore, it seems that the fin of the amphioxus, which may be homologous to the limbs of tetrapods, is poorly developed. For this reason, the cognate group 11–13 genes might have been lost during the evolution of the amphioxus, or the loss of these genes might have prevented the fin-like structures from developing into vertebrate-like skeletons. Of course, it is still possible that these genes actually exist in the amphioxus but have not been discovered, because the 5′ upstream region of the *Hox-10* gene has not been studied extensively (GARCIA-FERNÀNDEZ and HOLLAND 1994).

Colinearity of gene arrangement and phylogeny: Because of the remarkable conservation of the same order of gene arrangement on the chromosomes and the colinearity of gene expression pattern and gene arrangement, it is often stated that the arrangement order of *Antp*-class genes in the chromosome is important for their functions (*e.g.*, DUNCAN and LEWIS 1982). At least in Drosophila, however, the gene arrangement does not seem to be essential for their functions. In two experiments, splitting the bithorax complex into two pieces did not affect the development of the larva or adult (STRUHL 1984; TIONG *et al.* 1987). A translocation that separated the genes *lab* and *pb* from the rest of the complex also had no effect on the expression of these two genes (HAZELRIGG and KAUFMAN 1982). Moreover, there are some other types of homeobox genes or non-homeobox genes within the *Antp*-class homeobox gene cluster of Drosophila (see Figure 8 of BÜRGLIN 1994), and the existence of these genes does not seem to influence the expression pattern of *Antp*-class homeobox genes. Therefore, the colinearity of gene expression and gene arrangement looks to be unimportant.

We note that the phylogenetic tree of 13 cognate gene groups in Figure 3 can be described by the mathematical symbol (((1, 2)(3((4, 5)((6, 7)8))))(9(10(11-(12, 13))))). The genes that are phylogenetically close with each other are also closely located in the chromosome. This suggests that the colinearity of gene expression and gene order in each cluster partly reflects the events of tandem duplication of genes. In other words, if genes are tandemly duplicated without disturbing the transcription order, the genes that are closely located in the chromosome tend to have both sequence similarity and functional similarity. Therefore, the colinearity of gene arrangement and gene expression can be explained by the hypothesis of successive tandem gene duplication.

In practice, however, when the genes in a multigene family are duplicated, inversion of genes also often occurs so that the transcriptional direction varies from gene to gene as in the case of major histocompatibility complex genes of mammals (*e.g.*, HUGHES and NEI 1990) and immunoglobin heavy chain variable region genes of chicken (*e.g.*, REYNAUD *et al.* 1989). Yet, we do not see gene inversions very often in the *Antp*-class homeobox gene family. (Two exceptions are the *Dfd* gene in Drosophila and the *ceh-13* and *ceh-15* genes in the nematode). This is rather striking if we consider the long history of this gene family. It is therefore quite possible that the current gene arrangement is important for the function of the genes and the gene arrangement is maintained by some weak selection. If the selection is weak, experimental separation or translocation of genes would not disturb the development of an individual, yet the gene complex with inverted genes will have only a small chance of being fixed in the population. It is therefore likely that the current gene arrangement is maintained by some kind of purifying selection.

Coevolution of *Antp*-class homeobox genes and animal body plans: Biologists are interested in the evolution of *Antp*-class homeobox genes mainly because of the importance of this gene family in determining the animal body plan. It has been argued that the addition of new members of this gene family has been a driving force of morphological evolution in animals (*e.g.*, LEWIS 1978). Nevertheless, recent studies with insects and crustaceans suggest that changes in the number, size, and pattern of body structures are most likely to involve changes in the timing and spatial regulation of *Antp*-class homeobox genes or their target genes in the regulatory pathway (WARREN *et al.* 1994; AVEROF and AKAM 1995; CARROLL 1995; CARROLL *et al.* 1995). They also suggest that the *Antp*-class genes control only a very basic body plan on which the segmental diversity evolved. This is consistent with the fact that many organisms have the same organization of *Antp*-class homeobox genes even though their body plans are very different.

This suggests that the *Antp*-class genes evolved in the very early stage of metazoan evolution. To obtain some idea about the age of this class of genes, we tried to estimate the time of divergence between group I and group II genes. The average *p*-distance between the nematode gene *ceh-13* and the cognate group 1 genes of the other four organisms (Drosophila, amphioxus, mouse, and human) is 0.292, whereas the average *p*-distance between the nematode genes *ceh-15* and *mab-5* and the genes of other four organisms in subgroup C is 0.274. Therefore, the average divergence between the orthologous homeobox genes of the nematode and those of the other four organisms becomes 0.283. By contrast, the average *p*-distance of group I and II genes is 0.443. Therefore, if we assume that Pseudocoelomates (nematodes in this paper) and Coelomates (insects and chordates in this paper) diverged 550 million years (MY) ago (KNOLL 1992), the divergence between group I and group II genes is estimated to be 900 MY ago. In this computation we used *p*-distance, which is not proportional to evolutionary time. This can be rectified if we use Poisson-correction distance. Poisson-correction distance (d) can be obtained by $d = -\ln(1 - p)$, where p is the proportion of different amino acids between two sequences. This distance becomes 0.333 between the nematode and the chordates plus Drosophila and 0.585 for the divergence between group I and group II genes. Therefore, we obtain 1000 MY as the time of divergence between group I and II genes.

These estimates are certainly very crude, but if they are correct, the *Antp*-class homeobox gene family evolved in the very early stage of metazoan evolution (KNOLL 1992; CONWAY MORRIS 1993). Using similar calculations, we obtain a d distance of 0.440 for the divergence between subgroup A and subgroups B plus C, 0.365 between subgroups B and C, and 0.485 between subgroups D and E. In the above we have seen that a d distance of 0.333 corresponds to 550 MY. Therefore, the divergence between subgroup A and subgroups B plus C seems to have occurred $\sim 550 \times (0.440/0.333) = 730$ MY ago, whereas subgroups B and C apparently diverged ~600 MY ago. Similarly, subgroups D and E seem to have diverged ~800 MY ago. These computations suggest that the five subgroups A, B, C, D, and E were formed by 600 MY ago and that the evolution of these subgroups predated the divergence of Pseudocoelomates and Coelomates. Furthermore, following the tree topology of Figure 2, we earlier discussed the possibility that the cognate groups 1 and 2 had already diverged before the divergence of Pseudocoelomates and Coelomates. In other words, there were probably six *Antp*-class homeobox genes in the common ancestor of Pseudocoelomates and Coelomates. So it seems that *Antp*-class homeobox gene family of this ancestral organism was similar to that of the modern segmented animals though body segmentation probably did not exist at that time. Since the rate of amino acid substitution in this gene family varies from evolutionary lineage to evolutionary lineage and the paleontological estimate of the time of divergence between the nematode and the other four organisms is not accurate, the above estimates are very rough. Yet, they are not inconsistent with the argument that many new body plans evolved owing to the changes of the regulation of *Antp*-class homeobox genes or their target genes in the regulatory pathway rather than the acquisition of new members of this gene family (WARREN *et al.* 1994; AVEROF and AKAM 1995; CARROLL 1995; CARROLL *et al.* 1995).

We thank NAOKO TAKEZAKI for developing a computer program for computing the confidence probabilities of interior branches for amino acid sequence data and assistance in preparing Figures 2, 3, and 4. We are also grateful to CLAUDIA KAPPEN for her comments on an earlier version of this paper. This work was supported by research grants from the National Institutes of Health and the National Science Foundation to M.N.

LITERATURE CITED

AVEROF, M., and M. AKAM, 1995 *Hox* genes and the diversification of insect and crustacean body plans. Nature **376:** 420–423.

BACHILLER, D., A. MACIAS, D. DUBOULE and G. MORATA, 1994 Conservation of a functional hierarchy between mammalian and insect Hox/Hom genes. EMBO J. **13:** 1930–1941.

BARTELS, J. L., M. T. MURTHA and F. H. RUDDLE, 1993 Mutiple Hox/HOM-class homeoboxes in platyhelminthes. Mol. Phylogenet. Evol. **2:** 143–151.

BÜRGLIN, T. R., 1994 A comprehensive classification of homeobox genes, pp. 27–71 in *Guidebook to the Homeobox Genes,* edited by D. DUBOULE. Oxford University Press, New York.

BÜRGLIN, T. R., and G. RUVKUN, 1993 The *Caenorhabditis elegans* homeobox gene cluster. Curr. Opin. Gen. Dev. **3:** 615–620.

BÜRGLIN, T. R., G. RUVKUN, A. COULSON, N. C. HAWHINS, J. D. MCGHEE *et al.*, 1991 Nematode homeobox cluster. Nature **351:** 703.

CARROLL, S. B., 1995 Homeotic genes and the evolution of arthropods and chordates. Nature **376:** 479–485.

CARROLL, S. B., S. D. WEATHERBEE and J. A. LANGELAND, 1995 Homeotic genes and the regulation and evolution of insect wing number. Nature **375:** 58–61.

CARTWRIGHT, P., M. DICK and L. M. BUSS, 1993 HOM/Hox type homeoboxes in the chelicerate *Limulus polyphemus*. Mol. Phylogenet. Evol. **2:** 185–192.

CONWAY MORRIS, S., 1993 The fossil record and the early evolution of the Metazoa. Nature **361:** 219–225.

DEGNAN, B. M., S. M. DEGNAN, A. GIUSTI and D. E. MORSE, 1995 A *hox/hom* homeobox gene in sponges. Gene **155:** 175–177.

DICK, M. H., and L. M. BUSS, 1994 A PCR-based survey of homeobox genes in *Ctenodrilus serratus* (Annelida: polychaeta). Mol. Phylogenet. Evol. **3:** 146–158.

DUBOULE, D., 1994 How to make a limb? Science **268:** 575–576.

DUBOULE, D., and G. MORATA, 1994 Colinearity and functional hierarchy among genes of the homeotic complexes. Trends Genet. **10:** 358–364.

DUNCAN, I., and E. B. LEWIS, 1982 Genetic control of body segment differentiation in Drosophila, pp. 533–554 in *Developmental Order: Its Origin and Regulation,* edited by S. S. SUBTELNY. A. R. Liss, New York.

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

GARCIA-FERNÀNDEZ, J., and P. W. H. HOLLAND, 1994 Archetypal organization of the amphioxus Hox gene cluster. Nature **37:** 563–566.

GEHRING, W. J., Y. Q. QIAN, M. BILLETER, K. FURUKUBO-TOKUNAGA, A. F. SCHIER *et al.,* 1994 Homeodomain-DNA recognition. Cell **78:** 211–223.

HAACK, H., and P. GRUSS, 1993 The establishment of murine Hox-1 expression domains during patterning of the limb. Dev. Biol. **157:** 410–422.

HAZELRIGG, T., and T. C. KAUFMAN, 1982 Revertants of dominant mutations associated with the Antennapedia gene complex of *Drosophila melanogaster:* cytology and genetics. Genetics **105:** 581–600.

HILLIS, D. M., and J. J. BULL, 1993 An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. **42:** 182–192.

HOLLAND, P., 1982 Homeobox genes in vertebrate evolution. BioEssays **14:** 267–273.

HUGHES, A. L., and M. NEI, 1990 Evolutionary relationships of class II major-histocompatibility-complex genes in mammals. Mol. Biol. Evol. **7:** 491–514.

HUGHES, A. L., and M. NEI, 1993 Evolutionary relationships of the classes of major histocompatibility complex genes. Immunogenetics **37:** 337–346.

KAPPEN, C., and F. H. RUDDLE, 1993 Evolution of a regulatory gene family: HOM/HOX genes. Curr. Opin. Gen. Dev. **3:** 931–938.

KAPPEN, C., K. SCHUGHART and F. H. RUDDLE, 1989 Two steps in the evolution of Antennapetia-class vertebrate homeobox genes. Proc. Natl. Acad. Sci. USA **86:** 5459-5463.

KAPPEN, C., K. SCHUGHART and F. H. RUDDLE, 1993 Early evolutionary origin of major homeodomain sequence classes. Genomics **18:** 54–70.

KNOLL, A. H., 1992 The early evolution of eukaryotes: a geological perspective. Science **256:** 622–627.

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, version 1.01. Pennsylvania State University, University Park, PA.

LAWRENCE, P., and G. MORATA, 1994 Homeobox genes: their function in Drosophila segmentation and pattern formation. Cell **78:** 181–189.

LEWIS, E. B., 1978 A gene complex controlling segmentation in Drosophila. Nature **276:** 565-570.

MCGINNIS, W., and R. KRUMLAUF, 1992 Homeobox genes and axial patterning. Cell **66:** 283-302.

MORGAN, B. A., and C. J. TABIN, 1993 The role of homeobox genes in limb development. Curr. Opin. Gen. Dev. **3:** 668–674.

MURTHA, M. T., J. F. LECKMAN and F. H. RUDDLE, 1991 Detection of homeobox genes in development and evolution. Proc. Natl. Acad. USA **88:** 10711–10715.

NAITO, M., H. ISHIGURO, T. FUJISAWA and Y. KUROSAWA, 1993 Presence of eight distinct homeobox-containing genes in cnidarians. FEBS Lett. **333:** 271–274.

NEI, M., 1991 Relative efficiencies of different tree-making methods for molecular data, pp. 90–128 in *Phylogenetic Analysis of DNA Sequences,* edited by M. M. MIYAMOTO and J. CRACRAFT. Oxford University Press, New York.

NEI, M., N. TAKEZAKI and T. SITNIKOVA, 1995 Assessing molecular phylogenies. Science **267:** 253–255.

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.

PETERSON, R. L., T. PAPENBROCK, M. M. DAVDA and A. AWGULEWITSCH, 1994 The murine *Hoxc* cluster contains five neighboring *AbdB*-related Hox genes that show unique spatially coordinated expression in posterior embryonic subregions. Mech. Dev. **47:** 253-260.

REYNAUD, C.-A., A. DAHAN, V. ANQUEZ and J.-C. WEILL, 1989 Somatic hyperconvertion diversifies the single V_H gene of the chicken with a high incidence in the D region. Cell **59:** 171–183.

RUDDLE, F. H., J. L. BARTELS, K. L. BENTLEY, C. KAPPEN, M. T. MURTHA *et al.,* 1994 Evolution of *HOX* genes. Annu. Rev. Genet. **28:** 423–442.

RUSSO, C. A. M., N. TAKEZAKI, and M. NEI, 1996 Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol. Biol. Evol. (in press).

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

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

SALSER, S. J., and C. KENYON, 1994 Patterning *C. elegans:* homeotic cluster genes, cell fates and cell migrations. Trends Genet. **10:** 159–164.

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.

SCHUGHART, K., C. KAPPEN and F. H. RUDDLE, 1989 Duplication of large genomic regions during the evolution of vertebrate homeobox genes. Proc. Natl. Acad. Sci. USA **86:** 7067–7071.

SITNIKOVA, T., A. RZHETSKY and M. NEI, 1995 Interior-branch and bootstrap tests of phylogenetic trees. Mol. Biol. Evol. **12:** 319–333.

SNOW, P., and L. W. BUSS, 1994 HOM/Hox-Type homeoboxes from *Stylaria lacustris* (Annelida: oligochaeta). Mol. Phylogenet. Evol. **3:** 360–364.

STRUHL, G., 1984 Splitting the bithorax complex of Drosophila. Nature **308:** 454–457.

TIONG, S. Y., J. R. WHITTLE and M. C. GRIBBIN, 1987 Chromosomal continuity in the abdominal region of the bithorax complex of Drosophila is not essential for its contribution to metameric identity. Development **101:** 135–142.

WANG, B. B., M. M. MULLER-IMMERGLUCK, J. AUSTIN, N. T. ROBINSON, A. CHISHOLM *et al.,* 1993 A homeotic gene cluster patterns the anterioposterior body axis of *C. elegans.* Cell **74:** 29–42.

WARREN, R. W., L. NAGY, J. SELEGUE, J. GATES and S. CARROLL, 1994 Evolution of homeotic gene regulation and the function in flies and butterflies. Nature **372:** 458–461.

ZHARKIKH, A., and W.-H. LI, 1992 Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences: I. Four taxa with a molecular clock. Mol. Biol. Evol. **9:** 1119–1147.

Communicating editor: N. TAKAHATA