Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 143–147, January 1993
Evolution

The Antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution

FRANK R. SCHUBERT*, KATJA NIESELT-STRUWE†, AND PETER GRUSS*‡

Abteilungun für *Molekulare Zellbiologie and †Biochemische Kinetik, Max-Planck-Institut für Biophysikalische Chemie, W-3400 Göttingen, Federal Republic of Germany

Communicated by Manfred Eigen, September 23, 1992

ABSTRACT **The developmental control genes containing an Antennapedia-type homeobox are clustered in insects and vertebrates. The evolution of these genes was studied by the construction of evolutionary trees and by statistical geometry in sequence space. The comparative analysis of the homeobox sequences reveals the subdivision of the Antennapedia-type homeobox genes into three classes early in metazoan evolution. This observation suggests an important function of these genes even in the most primitive metazoans. Subsequent duplication events generated a cluster of at least five homeobox genes in the last common ancestor of insects and vertebrates. These genes later independently gave rise to the 13 groups of paralogous genes in vertebrates and to the 11 Antennapedia-type genes in the *Drosophila* complexes.**

The homeobox consists of a 183-base-pair sequence encoding a trihelical DNA binding motif. This domain has been found in several genes in a wide variety of eukaryotic organisms (for reviews see refs. 1 and 2). The protein products of some of the homeobox genes act as ubiquitous transcription factors, whereas most are involved in the control of embryonic development. In the fruit fly *Drosophila* as well as in the house mouse, the segment identity along the anterior–posterior axis of the embryo is specified by homeotic selector genes belonging to the Antennapedia-type of homeobox genes (reviewed in ref. 3). Both in insects (HOM genes) and in vertebrates (Hox genes) these genes are located in clusters. The beetle *Tribolium* (4, 5) has its homeotic genes arranged in one complex (HOM-C), whereas *Drosophila* has them split into two complexes [Bithorax complex (BX-C) and Antennapedia complex (ANT-C)]. In mammals, the four clusters (Hox-1 to Hox-4) can be aligned according to sequence similarities of their genes, thus defining 13 groups of paralogous genes. Kappen *et al.* (6) have therefore suggested that cluster duplication events underlie the evolution of the vertebrate Hox genes.

A closer examination of the homeobox gene clusters in insects and vertebrates revealed remarkable resemblances (7, 8). The *Drosophila* HOM genes are arranged in a similar order on the chromosome as their vertebrate counterparts. Moreover, in both organisms, these genes exhibit a correlation between their expression boundaries and their position in the cluster: generally the further downstream (3′) a gene is in its cluster, the more anterior its expression boundary is in the embryo (for reviews see refs. 9 and 10). A common ancestral cluster for insects and vertebrates therefore appears to be highly probable (7, 8, 11). Reports of a related homeobox gene cluster in the nematode *Caenorhabditis* (12) and of Antennapedia-type homeobox genes in cnidarians (13, 14) support this hypothesis.

To gain insight into the evolution of developmental control mechanisms, as well as the phylogenetic relationships between the metazoans, several studies of the evolution of the Antennapedia-type homeobox genes were done by comparative sequence analyses (1, 6, 11, 15–17). While these investigations could reveal some aspects of homeobox gene evolution, many of the evolutionary steps which led to the linear order of the genes in the ancestral cluster of insects and vertebrates and, more recently, to the final organization of the homeobox gene clusters still remain obscure. To understand these events, we used the neighbor-joining method (18) combined with bootstrapping resampling (19) to build an evolutionary tree for the human and *Drosophila* homeobox genes. In addition, statistical geometry (20) was applied to confirm the major branching points of the resulting tree and to evaluate doubtful nodes. In contrast to tree construction programs that are based on the overall distance only, the method of statistical geometry in sequence space uses also positional information, thus obtaining a higher sensitivity with respect to the assignment of topologies for sequence sets. Our results suggest an early trichotomy of the Antennapedia-type homeobox genes arising from a common progenitor. At the time of the insect/vertebrate divergence two of the resulting genes (the Antennapedia–Deformed precursor and the proboscipedia–labial precursor) had already undergone a further duplication, while the Abdominal-B precursor remained singular. Later in the evolution of vertebrates and insects, gene duplications in all three of the original classes occurred.

MATERIALS AND METHODS

Nucleotide Sequences. An alignment of homeobox sequences of 38 published human HOX genes (21–24) and 11 Antennapedia-type genes of the *Drosophila* clusters (25–31, 40) was generated; no gaps were introduced. A total of 183 nucleotides for each gene were analyzed, of which 40 were identical in all 49 sequences.

Statistical Geometry. The method of statistical geometry provides a complementary tool for constructing an evolutionary tree for a given sequence set. Based on the analysis of quartets of sequences, this method allows one to distinguish cases of tree-like, bush-like, or randomized topologies. For binary (i.e., purine/pyrimidine, R/Y) sequences there are eight distinguishable classes of positions in three categories describing their entry distribution (Fig. 1). For a set of n sequences, the values of the position classes are calculated for each of the possible $\binom{n}{4}$ quartets. The average values of these parameters are then computed, giving rise to the mean quartet geometry of the set.

Before an evolutionary tree was constructed for a sequence set, the underlying topology was evaluated by statistical geometry for all three codon positions. The largest ($\bar{l}$), middle ($\bar{m}$), and smallest ($\bar{s}$) of the three box dimensions were averaged separately. While a tree-like geometry then is characterized by one large box dimension ($\bar{l} >> \bar{m}, \bar{s}$), a

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡To whom reprint requests should be addressed.

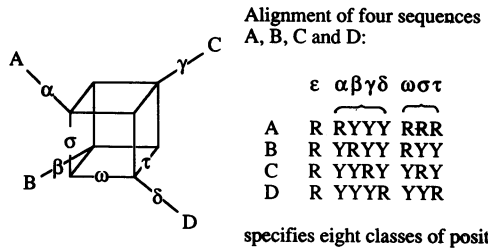

FIG. 1. Schematic diagram representing the distance correlation among four sequences (A–D) in purine/pyrimidine (R/Y) sequence space. Eight position classes falling into three categories are defined: d_0 (=ε) sums up all positions with equal entries in all four sequences, d_1 (=$\alpha + \beta + \gamma + \delta$) is the sum of the positions where one sequence differs from the other three (like α with A differing from B, C, and D), and d_2 (=$\omega + \sigma + \tau$) are the positions where two sequences differ pairwise from the other two. The representative geometry is then a three-dimensional box with the box dimensions ω, σ, and τ and the protrusions α, β, γ, and δ.

bush-like geometry is indicated by three box dimensions of comparable length ($\bar{l} + \bar{s} \approx 2\bar{m}$). If, in addition, the average protrusions are of similar magnitude as the box dimensions ($\bar{d}_1/4 \approx \bar{m}$), then the set is completely randomized. For every set of sequences analyzed, the statistical geometry showed a randomized third codon position. This is not surprising, considering the long history of homeobox gene evolution, with selection pressure working predominantly on the amino acid sequence.

The method was also applied in evaluating the significance of particular branchings in the evolutionary trees. The relevant sequence set was partitioned into four subsets based on the clustering of the tree. Sequence space geometries were computed for all quartets containing one member of each of the four subsets. The average values of these quartet geometries were used to describe the topology of the common node of the four subsets. Also in all these calculations, the third codon position appeared highly randomized. It was therefore judged to be noninformative in describing the topology and was omitted from the computations.

Construction of Evolutionary Trees. For each pair of sequences, the distances were computed for the purine/pyrimidine sequences by summing up the positions occupied by different symbols. These absolute distances were divided by the number of positions considered. Since the third codon position had been shown by statistical geometry to be randomized, the evolutionary trees were based on the first and second codon positions only.

When two groups of paralogous genes were compared, group distances between these two groups were computed by first calculating the distance for each pairwise combination of the individual members of the two groups and then taking the average of these distances.

Unrooted evolutionary trees from the distance matrices were constructed by using the NJTREE program, kindly provided by N. Saitou (University of Tokyo), as well as the CLUSTAL V package (32). Both programs are based on the neighbor-joining algorithm; in the latter case the tree construction was combined with bootstrapping resampling. The trees shown are built on transversions, which may avoid the problems caused by the different frequencies of transversions and transitions. However, similar results were obtained using AGCT or amino acid sequences. Trees were printed by using the NJDRAW program, kindly provided by J. Ferguson (University of Texas, Houston).

RESULTS AND DISCUSSION

The Vertebrate Hox Gene Clusters Have Evolved from a Common Ancestor by Duplication. The Hox clusters appear to

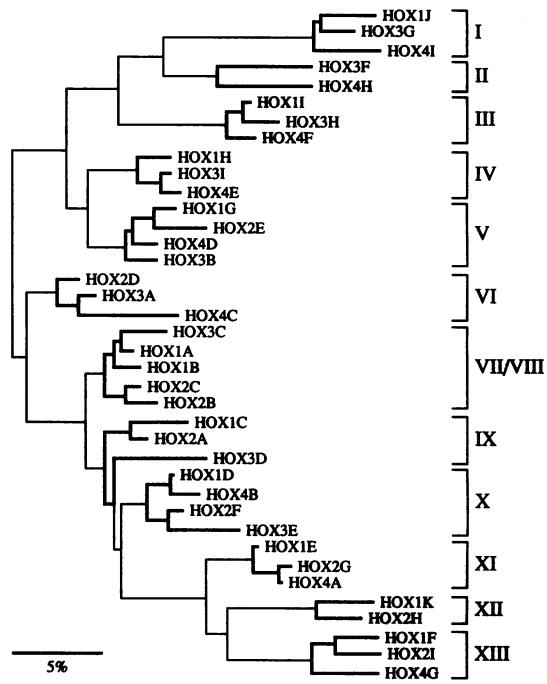

FIG. 2. Unrooted evolutionary tree for the 38 published human HOX genes. The nomenclature follows the guidelines given in (ref. 33). The distance matrix resulted from the calculation of the pairwise Hamming distances of the R/Y sequences for the first and second codon position. The unrooted evolutionary tree was constructed by applying the NJTREE program to the distance matrix. Thick lines indicate the branching within the paralog groups. The branch lengths of the tree represent percent divergence, not evolutionary distances. A scale showing 5% divergence is drawn (as in Figs. 3–5).

be similarly organized in all vertebrates, with the genes of the human HOX clusters covering the homologs of all other published vertebrate Hox genes. This allows the study of vertebrate Hox gene evolution using sequences of only one species. Therefore, the homeobox sequences of all 38 published HOX genes were analyzed by the neighbor-joining method. The resulting evolutionary tree clearly separates most groups of paralogous genes from one another (Fig. 2). Only the paralog groups VII–IX in the middle of the cluster show so few differences that they are not resolved in the homeobox-based tree. However, analyses of sequences flanking the homeobox subdivide these genes into the paralog groups (ref. 17 and data not shown).

The results argue for the monophyletic origin of each paralog group, suggesting a single ancestral cluster with 13 genes before the radiation of the vertebrates, 350–450 million years ago. This is consistent with the hypothesis that the four vertebrate Hox clusters result from duplications of the ancestral cluster (6, 17), followed by the deletion of certain genes in some of the clusters.

The Vertebrate and Insect Homeobox Gene Clusters Are Derived from an Ancestral Cluster Containing At Least Five Genes. The similar structure of the *Drosophila* and vertebrate homeobox gene clusters has led to the hypothesis of a common homeobox gene cluster in the insect/vertebrate ancestor. If this were true, the tree program should arrange the human paralog groups on branches together with those *Drosophila* genes that share the same precursor in the ancestral cluster.

To test this hypothesis, an evolutionary tree was constructed for the human paralog groups and the genes of the *Drosophila* HOM complexes (Fig. 3). The branching order of the tree not only confirms the common ancestral cluster for the human and *Drosophila* genes but also allows the precur-

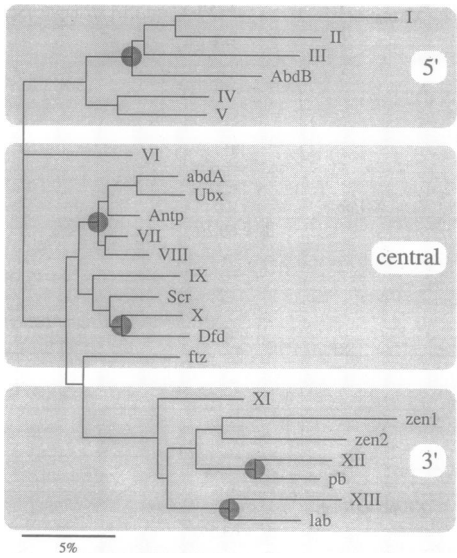

FIG. 3. Unrooted evolutionary tree for the human paralog groups and the *Drosophila* Antennapedia-type homeobox genes. In the case of the human genes, the group distances for the paralog groups (I–XIII) were calculated as described in *Materials and Methods*. As in Fig. 2, the tree is based on the first and second codon position in R/Y space. Circles mark the five precursors in the ancestral cluster at the insect/vertebrate divergence. The three original classes are underlaid by shaded boxes.

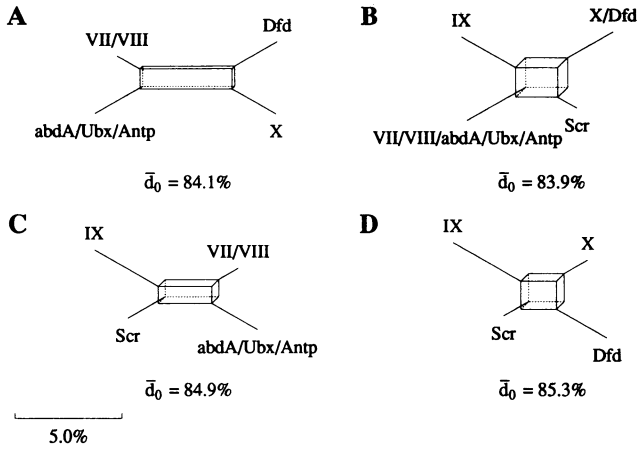

FIG. 4. Statistical geometry analysis for the central genes. The mean geometry of the node between the paralog groups VII and VIII; the *Drosophila* genes *abdA*, *Ubx*, and *Antp*, paralog group X; and *Drosophila Dfd* was calculated for the first and second codon position in R/Y space as described in *Materials and Methods*. (*A*) In addition, the mean geometries of the nodes between *Scr* and paralog group IX and the *Antp*-like and the *Dfd*-like genes (*B*), paralog groups VII and VIII and *abdA*, *Ubx*, and *Antp* (*C*), and paralog groups X and *Dfd* (*D*) are shown. The average percentage of identical positions, $\bar{d}_0$, between the four groups connected by the node is given.

sors connecting certain paralog groups with certain *Drosophila* genes to be inferred.

On the 5′ side the paralog groups I–V are clearly set apart from the other vertebrate paralog groups. All the groups, in particular paralog groups I and II, are separated by long branches. This suggests high evolutionary rates and, thus, a low selection pressure for these genes. Presumably, the genes were redundant shortly after the radiation of the 5′ class, whereas later on they obtained specific roles in controlling the development of newly acquired structures such as limbs (34). All the five vertebrate paralog groups correspond to one *Drosophila* gene only, Abdominal-B (*AbdB*), suggesting one common ancestor for these 5′-located genes in insects and vertebrates.

In the central part of the clusters, two groups can be distinguished. According to the tree, the genes abdominal-A (*abdA*), Ultrabithorax (*Ubx*), and Antennapedia (*Antp*) and the paralog groups VII and VIII evolved independently in the vertebrate and insect lines from a common precursor. Despite its more diverged homeobox, paralog group VI is placed close to the paralog groups VII and VIII based on the analysis of the flanking sequences.

From these *Antp*-like genes, paralog group X and Deformed (*Dfd*) are relatively easily discriminated as their own group by both the tree analysis (Fig. 3) and the statistical geometry of the node (Fig. 4*A*). The relationships between the other genes are more ambiguous. Paralog group IX and Sex combs reduced (*Scr*) cannot be grouped together nor assigned to the *Antp*-like or the *Dfd*-like genes unequivocally, as indicated by the roughly equal box dimensions of the nodes between these groups (Fig. 4 *B–D*), although they appear to be slightly more related to the *Dfd*-like genes. Similarly, the origin of the pair-rule gene fushi tarazu (*ftz*) could not be resolved.

At the 3′ end of the clusters, the evolutionary tree assembles labial (*lab*) and paralog group XIII together on one branch, suggesting a common precursor for these genes.

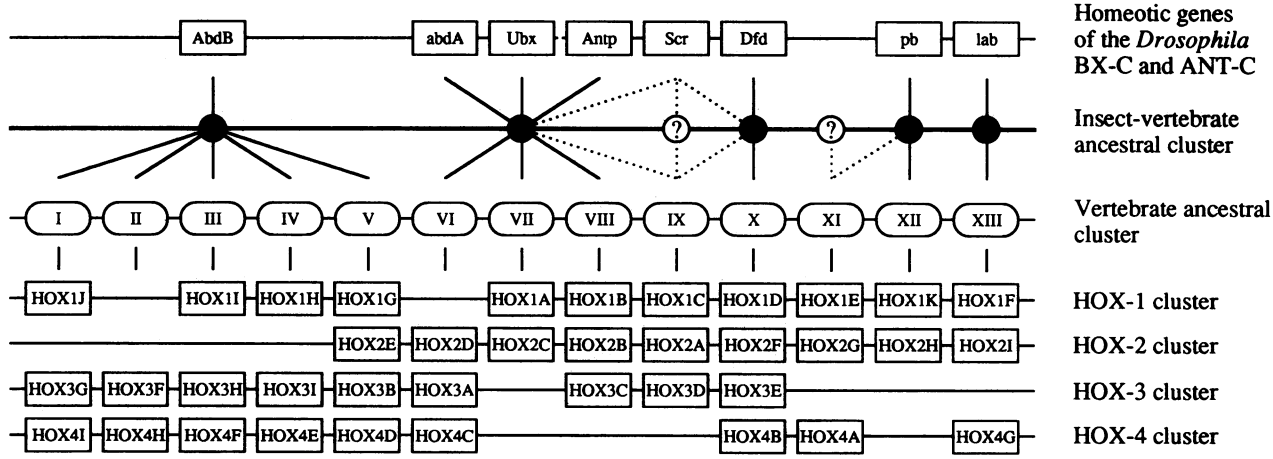

FIG. 5. Evolution of homeobox gene clusters. The proposed homeobox gene cluster of the insect/vertebrate ancestor is shown with the five precursors for the *AbdB*-, *Antp*-, *Dfd*-, *pb*-, and *lab*-like genes (●) and the questionable precursors for *Scr*/paralog group IX and for paralog group XI (?). Lines connect the vertebrate and insect genes with their putative precursors; dashed lines indicate uncertain relations. For the vertebrate line the proposed ancestral 13-gene cluster as well as the four derived human HOX clusters are shown.

Another branch point connects the *Drosophila* gene proboscipedia (*pb*) with the vertebrate paralog group XII. Paralog group XI shows some structural similarities to paralog group XII in the flanking sequences. However, the exact relation to the *pb*- or *lab*-like genes cannot be defined. A comparable situation is found for the zygotic *zerknüllt* genes *zen1* and *zen2* of *Drosophila*, for which no connection to one particular other 3'-located gene can be drawn.

These analyses strongly argue for a common ancestral cluster at the root of the insect and vertebrate homeobox gene clusters. From the branching order of the evolutionary tree for the human HOX genes and the *Drosophila* HOM genes, we can deduce the probable structure of this ancient cluster (Fig. 5). At the insect/vertebrate divergence it must have contained at least five genes, the precursors of the *AbdB*-like, *Antp*-like, *Dfd*-like, *pb*-like, and *lab*-like genes. Subsequent independent duplications led to the 13 paralog groups of vertebrates and the 11 Antennapedia-type genes of *Drosophila*.

The Genes in the Homeobox Gene Clusters Can Be Subdivided into Three Classes Suggesting an Ancient Three-Gene Cluster. The evolutionary tree for the human and *Drosophila* homeobox genes, having already revealed the precursors in the insect/vertebrate ancestor, allows a look further into the past. The five genes of the proposed insect/vertebrate ancestor are divided into three classes (Fig. 3).

The genes on the 5' end of the cluster are clearly distinguishable as a class and were probably represented by only one gene up to the divergence of insects and vertebrates. As in *Drosophila*, only one member of the 5' class, *TgHBox4* (35), is known in echinoderms. Duplications of this ancestral gene occurred later in the line leading to the vertebrates.

In the central part of the cluster, the *Dfd*-like and the *Antp*-like genes are linked by a common branch point. Thus, a joint precursor for these central genes can be deduced from the tree, defining a class that differs from the 5' and the 3' gene classes.

The evolutionary tree also suggests the 3'-located *pb*- and *lab*-like genes to form a separate class. The geometry of the node in Fig. 6*A* supports a common precursor for these genes despite the size of the two smaller box dimensions. The latter are reduced even when *lab* is not included in the computations. Still, as the homolog of paralog group XIII, *lab* is clearly placed into the 3' class. In addition, the geometry of the questionable node classifies paralog group XI as a member of the 3' class (Fig. 6*B*). The bootstrapping analysis supports the grouping of the 3' genes versus the central and 5' genes with about 70% confidence. However, the remaining uncertainties indicate the high number of parallel and reverse mutations accumulated by these genes. Contradictory results concerning the root of the 3' class are obtained by the tree analyses: in some cases the genes of the 3' class are arranged as a unique group (Fig. 3), in others as subsequent divergents of the *Dfd*-like genes in the central class (Fig. 2). The box geometry of the terminal genes toward the central genes argues for the separation of the 3' class prior to the duplication of the central class progenitor (Fig. 6*C*), supporting the branching shown in Fig. 3.

It is interesting that the splits between the three original classes progressively seem to become more indistinct. In particular, the paralog groups VI and XI show features both of their own and of the neighboring class. This characteristic, however, is restricted to the homeobox only. It seems to have been advantageous to develop a cluster showing a continuum with respect to the homeodomain sequence.

The identification of the three classes of Antennapedia-type genes suggests that a three-gene cluster existed early in the evolution of metazoans. Further support for this hypothesis comes from the characterization of Antennapedia-type genes in organisms that branched off before the separation of insects and vertebrates. In the *Caenorhabditis* cluster, a *lab*-like

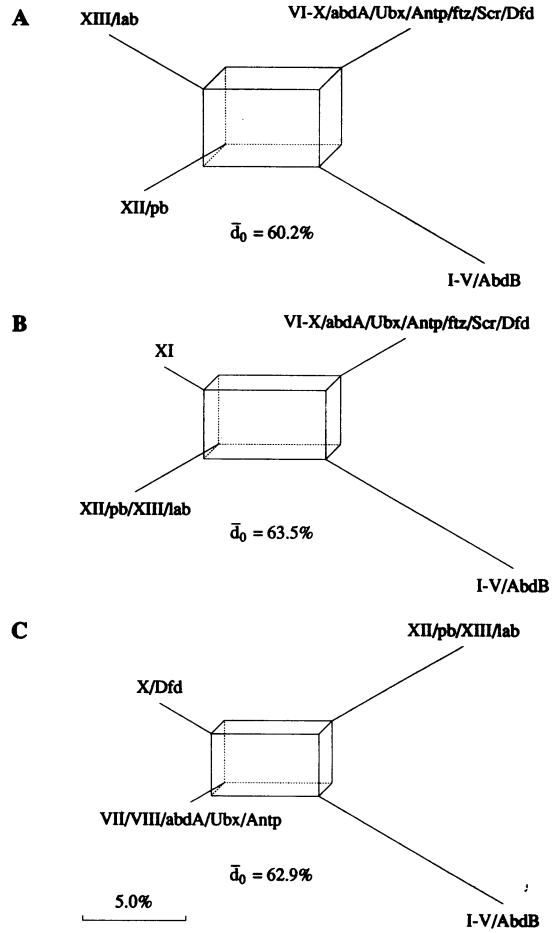

FIG. 6. Statistical geometry analysis for the 3' class. The geometries were calculated as in Fig. 4. The mean geometry for the nodes between paralog group XII and *pb*, paralog group XIII and *lab*, the genes of the 5' class, and those of the central class (*A*), paralog group XI, paralog groups XII and XIII, and Drosophila *pb* and *lab*, the 5' class, and the central class (*B*) and the *Antp*-like genes, the *Dfd*-like genes, the 3' class and the 5' class (*C*) are shown.

3'-class gene (*ceh-13*) (36) and two members of the central class (*mab5*, *ceh-15*) (37, 38) have been identified. A possible, but diverged, member of the 5' class might be the fourth gene in the cluster, *ceh-11* (36). Genes showing such remote similarity to the *AbdB*-like genes have also been cloned in cnidarians (*SAox3*, *Cnox-1*) (13, 14). Genes more related to *AbdB* may have simply been missed due to the use of *Antp*-like sequences in most screenings. The other cnidarian genes are definite members of the 3' class, probably belonging to the *pb*-like group (*SAox2*, *Cnox-2*, and *cnox2*) (39) and the *lab*-like group (*SAox1*, *cnox1*). Although the physical linkage of the cnidarian genes has not been shown, the observation of the *pb*- and *lab*-like genes raises the possibility of a four-gene homeobox cluster in metazoans without a coelom.

Conclusions. Most of the homeobox genes examined are part of the regulatory network controlling embryonic development in metazoans. It is tempting to link the radiation of the major types of these homeobox genes to the origin of metazoans. The Antennapedia-type genes play a key role in specifying segment identity along the anterior–posterior axis in insects and vertebrates. From our analysis of the *Drosophila* and human homeobox gene clusters we propose the following model for the evolution of these genes (Fig. 7). Starting from a single ancestral gene, the three precursors of the 3', central, and 5' classes of the Antennapedia-type genes are derived. The insect and vertebrate derivatives of these three classes correspond to the anterior (3' class), central (central class), and posterior (5'

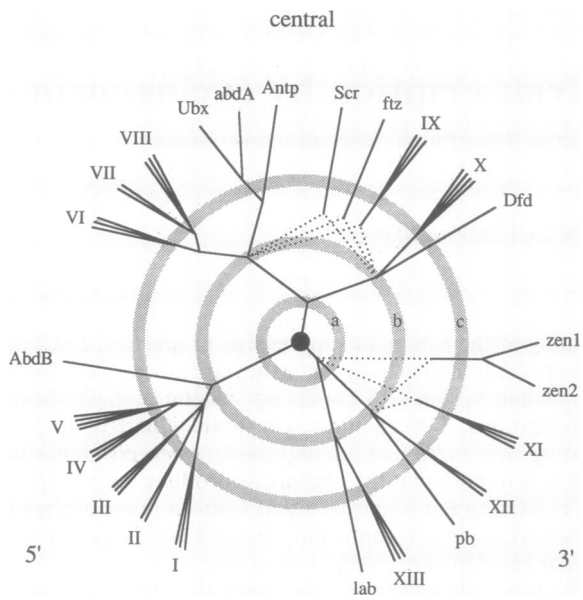

FIG. 7. Schematic diagram showing the evolutionary relationships of the *Drosophila* and human Antennapedia-type homeobox genes, starting from the proposed common precursor (filled circle). Dashed lines indicate alternative connections in uncertain cases. The shaded rings represent the approximate time points for the diploblast/triploblast divergence (a), the insect/vertebrate divergence (b), and the vertebrate cluster duplication (c).

class) regions of the embryo. The division into three classes fits nicely with the proposed function of the Antennapedia-type genes to determine the head, trunk, and tail in the ancestral arthropod (11, 41). The next duplication in the ancestral cluster, dividing the 3′ class into the *pb*- and *lab*-like genes, preceded the separation of diploblasts and triploblasts, possibly more than 1 billion years ago (42). Before the divergence of vertebrates and arthropods, another duplication separated the *Antp* and *Dfd* precursors in the central class. Thus, together with the *AbdB* precursor, the common ancestor of higher metazoans in the Precambrian already contained a cluster of at least five genes. Independently during the evolution of insects and vertebrates, further duplications generated the 11 genes found in *Drosophila* and the 13 genes proposed for the vertebrate ancestral cluster. The number of genes was increased in vertebrates by duplications of the whole cluster. *Drosophila*, on the other hand, used different strategies, such as multiple promoters and alternative splicing, to increase the complexity and the coding potential of its anterior–posterior differentiation control system.

Studies on the molecular evolution of the Antennapedia-type homeobox genes not only provide us with more information about the evolution of gene families but may well contribute to our understanding of the most basic mechanisms underlying the genetic control of embryonic development. Moreover, the striking conservation of the homeobox sequences back to diploblastic metazoans, together with the probable role of these genes in regional specification processes, allows us to examine phylogenetic relationships in the evolution of the metazoans. Further studies on the Antennapedia-type homeobox genes in sponges, *Trichoplax*, planarians, or protozoans might even shed some light on the still mysterious origin of metazoans, a question recently brought up again on the basis of analyses of 28S rRNA sequences (42).

We thank Thomas Kaufman for sharing sequence information. We are grateful to Susanne Dietrich and Michael Gross for fruitful discussions and to Guillermo Oliver and Patrick Trembley for comments during the preparation of the manuscript. F.R.S. was supported by the Max-Planck-Gesellschaft and the Deutsche Forschungsgemeinschaft.

1. Scott, M. P., Tamkun, J. W. & Hartzell, G. W. I. (1989) *Biochim. Biophys. Acta* **989**, 25–48.
2. Laughon, A. (1991) *Biochemistry* **30**, 11357–11367.
3. McGinnis, W. & Krumlauf, R. (1992) *Cell* **68**, 283–302.
4. Beemann, R. W. (1987) *Nature (London)* **327**, 247–249.
5. Stuart, J. J., Brown, S. J., Beeman, R. W. & Denell, R. E. (1991) *Nature (London)* **350**, 72–74.
6. Kappen, C., Schughart, K. & Ruddle, F. H. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5459–5463.
7. Duboule, D. & Dollé, P. (1989) *EMBO J.* **8**, 1497–1505.
8. Graham, A., Papalopulu, N. & Krumlauf, R. (1989) *Cell* **57**, 367–378.
9. Holland, P. W. H. & Hogan, B. L. M. (1988) *Genes Dev.* **2**, 773–782.
10. Kessel, M. & Gruss, P. (1990) *Science* **249**, 374–379.
11. Akam, M., Dawson, I. & Tear, G. (1988) *Development Suppl.* **104**, 123–133.
12. Bürglin, T. R., Ruvkun, G., Coulson, A., Hawkins, N. C., McGhee, J. D., Schaller, D., Wittmann, C., Müller, F. & Waterston, R. H. (1991) *Nature (London)* **351**, 703.
13. Murtha, M. T., Leckman, J. F. & Ruddle, F. H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10711–10715.
14. Schierwater, B., Murtha, M., Dick, M., Ruddle, F. H. & Buss, L. W. (1991) *J. Exp. Zool.* **260**, 413–416.
15. Hart, C. P., Fainsod, A. & Ruddle, F. H. (1987) *Genomics* **1**, 182–195.
16. Barad, M., Jack, T., Chadwick, R. & McGinnis, W. (1988) *EMBO J.* **7**, 2151–2161.
17. Schughart, K., Kappen, C. & Ruddle, F. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7067–7071.
18. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
19. Felsenstein, J. (1985) *Evolution* **39**, 783–791.
20. Eigen, M., Winkler-Oswatitsch, R. & Dress, A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5913–5917.
21. Acampora, D., D'Esposito, M., Faiella, A., Pannese, M., Migliaccio, E., Morelli, F., Stornaiuolo, A., Nigro, V., Simeone, A. & Boncinelli, E. (1989) *Nucleic Acids Res.* **17**, 10385–10402.
22. Stornaiuolo, A., Acampora, D., Pannese, M., D'Esposito, M., Morelli, F., Migliaccio, E., Rambaldi, M., Faiella, A., Nigro, V., Simeone, A. & Boncinelli, E. (1990) *Cell Differ. Dev.* **31**, 119–127.
23. D'Esposito, M., Morelli, F., Acampora, D., Migliaccio, E., Simone, A. & Boncinelli, E. (1991) *Genomics* **10**, 43–50.
24. Simeone, A., Acampora, D., Nigro, V., Faiella, A., D'Esposito, M., Stornaiuolo, A., Mavilio, F. & Boncinelli, E. (1991) *Mech. Dev.* **33**, 215–228.
25. McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. & Gehring, W. J. (1984) *Cell* **37**, 403–408.
26. Scott, M. P. & Weiner, A. J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4115–4119.
27. Regulski, M., Harding, K., Kostriken, R., Karch, F., Levine, M. & McGinnis, W. (1985) *Cell* **43**, 71–80.
28. Kuroiwa, A., Kloter, U., Baumgartner, P. & Gehring, W. J. (1985) *EMBO J.* **4**, 3757–3764.
29. Hoey, T., Doyle, H. J., Harding, K., Wedeen, C. & Levine, M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 4809–4813.
30. Rushlow, C., Doyle, H., Hoey, T. & Levine, M. (1987) *Genes Dev.* **1**, 1268–1279.
31. Karch, F., Bender, W. & Weiffenbach, B. (1990) *Genes Dev.* **4**, 1573–1587.
32. Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992) *Comput. Appl. Biosci.* **8**, 189–191.
33. McAlpine, P. J. & Shows, T. B. (1990) *Genomics* **7**, 460.
34. Dollé, P., Izpisúa-Belmonte, J.-C., Falkenstein, H., Renucci, A. & Duboule, D. (1989) *Nature (London)* **342**, 767–772.
35. Dolecki, G. J., Wang, G. & Humphreys, T. (1988) *Nucleic Acids Res.* **16**, 11543–11558.
36. Schaller, D., Wittmann, C., Spicher, A., Müller, F. & Tobler, H. (1990) *Nucleic Acids Res.* **18**, 2033–2036.
37. Costa, M., Weir, M., Coulson, A., Sulston, J. & Kenyon, C. (1988) *Cell* **55**, 747–756.
38. Kamb, A., Weir, M., Rudy, B., Varmus, H. & Kenyon, C. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4372–4376.
39. Schummer, M., Scheurlen, I., Schaller, C. & Galliot, B. (1992) *EMBO J.* **11**, 1815–1823.
40. Cribbs, D. L., Pultz, M. A., Johnson, D., Mazzulla, M. & Kaufman, J. C. (1992) *EMBO J.* **11**, 1437–1445.
41. Akam, M. (1989) *Cell* **57**, 347–349.
42. Christen, R., Ratto, A., Baroin, A., Perasso, R., Grell, K. G. & Adoutte, A. (1991) *EMBO J.* **10**, 499–503.