Research

Genesis and evolution of the *Evx* and *Mox* genes and the extended Hox and ParaHox gene clusters

Carolina Minguillón*† and Jordi Garcia-Fernàndez*

Address: *Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 645, E-08028 Barcelona, Spain. †Current address: Division of Developmental Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.

Correspondence: Jordi Garcia-Fernàndez. E-mail: jgarcia@bio.ub.es

Published: 23 January 2003

Genome ***Biology*** 2003, **4**:R12

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2003/4/2/R12

Received: 24 September 2002
Revised: 31 October 2002
Accepted: 9 December 2002

Abstract

Background: Hox and ParaHox gene clusters are thought to have resulted from the duplication of a ProtoHox gene cluster early in metazoan evolution. However, the origin and evolution of the other genes belonging to the extended Hox group of homeobox-containing genes, that is, *Mox* and *Evx*, remains obscure. We constructed phylogenetic trees with mouse, amphioxus and *Drosophila* extended Hox and other related *Antennapedia*-type homeobox gene sequences and analyzed the linkage data available for such genes.

Results: We claim that neither *Mox* nor *Evx* is a Hox or ParaHox gene. We propose a scenario that reconciles phylogeny with linkage data, in which an *Evx*/*Mox* ancestor gene linked to a ProtoHox cluster was involved in a segmental tandem duplication event that generated an array of all Hox-like genes, referred to as the 'coupled' cluster. A chromosomal breakage within this cluster explains the current composition of the extended Hox cluster (with *Evx*, Hox and *Mox* genes) and the ParaHox cluster.

Conclusions: Most studies dealing with the origin and evolution of Hox and ParaHox clusters have not included the Hox-related genes *Mox* and *Evx*. Our phylogenetic analyses and the available linkage data in mammalian genomes support an evolutionary scenario in which an ancestor of *Evx* and *Mox* was linked to the ProtoHox cluster, and that a tandem duplication of a large genomic region early in metazoan evolution generated the Hox and ParaHox clusters, plus the cluster-neighbors *Evx* and *Mox*. The large 'coupled' Hox-like cluster *Evx*Hox/*Mox*ParaHox was subsequently broken, thus grouping the *Mox* and *Evx* genes to the Hox clusters, and isolating the ParaHox cluster.

Background

Homeobox genes have crucial roles during embryogenesis and have been deeply studied from the point of view of the evolution of development. Changes in their number and regulation may have been instrumental in body-plan evolution and diversification [1]. Whether the physical linkage of many homeobox genes is maintained by regulatory constraints or is simply a reflection of their evolutionary origin by tandem gene duplication has not yet been fully elucidated. The clustering of the Antennapedia superclass of homeobox genes in contemporary genomes is proposed to be the outcome of tandem gene duplication and cluster duplications from an ancestral *UrArcheHox* gene during metazoan evolution [2,3]. However, genome rearrangements, clade-specific

duplications and gene losses obscure the complete evolutionary chronicle.

The analysis of the human genome led Pollard and Holland to suggest that four such clusters, namely the extended Hox, the ParaHox, the NKL and the EHGbox clusters, arose by successive tandem gene duplications and cluster duplications from an ancestral *UrArcheHox* gene early in metazoan evolution [3]. The extended Hox array includes the Hox cluster genes plus the former orphan classes *Evx* and *Mox*. The evolutionary sister of the Hox cluster, the ParaHox cluster, is believed to have resulted from the non-tandem duplication of a four-gene ProtoHox cluster that gave rise to the primordial Hox and ParaHox clusters [4]. Hence, Hox and ParaHox genes have the same evolutionary age. Although extensive studies have been performed to trace the origin and evolution of the Hox genes [5-7] and more recently the ParaHox plus Hox genes [4,8,9], *Evx* and *Mox* have rarely been considered in these analyses. They have been unified into the extended Hox group, owing to their linked disposition in the genome of certain organisms; for example, *Evx* genes are closely linked to the 5′ end of the Hox gene cluster in most vertebrates and in a cnidarian species [10-12]. Likewise, *Mox* genes map near the opposite extreme of the HoxA and HoxB clusters in the human genome. These linkage data prompted Pollard and Holland to propose that *Evx* and *Mox* genes originated during the tandem duplication events that produced the ancestral Hox cluster genes [3]. In a phylogenetic tree, Hox genes alone do not form a monophyletic clade, but a clade containing both Hox and ParaHox genes. *Evx* genes fall basal to the Hox/ParaHox clade [8,13,14], while the *Mox* gene has vaguely been referred to as a ParaHox gene and suggested to represent the missing ParaHox gene related to the central group (PG4 to PG8) of Hox genes. [14]. Unfortunately, most studies on Hox/ParaHox relationships do not include the *Mox* class [2,8,9]. Nonetheless, the two views of the evolutionary relationship between the *Mox* and the Hox and ParaHox genes (Hox-related or ParaHox-related) are contradictory. If *Mox* genes are derived from the tandem duplication of a particular Hox gene (and thus linked to the Hox gene cluster), they are not ParaHox genes. If *Mox* is a descendant of the missing central ParaHox gene, it is not a Hox gene, although it is linked to the Hox cluster. Following the same reasoning, if *Evx* is the sister of Hox plus ParaHox genes, it cannot have originated from the tandem duplication of a Hox gene.

All these discordant points of view led us to construct phylogenetic trees and search for data backing up the proposed evolutionary relationships between the extended Hox group (including *Evx* and *Mox*) and ParaHox genes. We discuss outlines that may not have been considered yet, and draw an evolutionary scenario, which attests that both *Evx* and *Mox* were generated in the same duplication event that gave rise to the Hox and ParaHox clusters.

Results and discussion

Mox and *Evx* are neither Hox nor ParaHox genes

Phylogenetic trees constructed with the homeodomain and the homeodomain plus flanking residues showed similar topologies. Figure 1 shows a neighbor-joining (NJ) unrooted tree with the homeodomain plus flanking residues of amphioxus, mouse and *Drosophila* sequences. Maximum parsimony (MP) trees showed the same relationships (data not shown). The resulting quartet puzzling (QP) tree was a comb-like tree without any clear internal relationship. QP is based on a maximum likelihood analysis of quartets and is believed to be too conservative. Furthermore, none of the clades below 50% support is retrieved at the final tree, which may be due to the few amino-acid positions of the data, the current lack of any reliable amino-acid model for the evolution of homeodomain-containing proteins and the stringency of the QP method. The trees obtained had three outstanding features (Figure 1). First is the consistent grouping of the already proposed relationship for the Hox and ParaHox genes: that is, *Cdx* is the posterior ParaHox gene more closely related to the posterior group of Hox genes; *Xlox/Pdx1* is the ParaHox gene more closely related to group 3 of Hox genes; and *Gsx* is the ParaHox gene more closely related to the anterior group of Hox genes. Second is the lack of a ParaHox central gene, as only Hox genes are grouped within the central group. The third feature is the grouping of *Mox* and *Evx* class genes. The bootstrap value that supports this relationship is 60%, higher than values reported elsewhere for Hox/ParaHox relationships [4,8,14]. Two major conclusions can be drawn from the analyses: *Mox* is not the central ParaHox gene, and not only *Evx* but also *Mox* genes are equally related to both Hox and ParaHox genes, suggesting an early origin for both classes.

To investigate these relationships further, we constructed various phylogenetic trees to which we added the sequences of other closely related *Antennapedia*-type homeobox genes, which have been shown to be linked to the extended Hox cluster in certain mammalian genomes [3], that is, the Dlx and the Msx classes of NKL homeobox genes and the Engrailed, the Gbx and HB-9 classes of EHGbox homeobox genes. As before, similar topologies were obtained when trees were constructed with the homeodomain or with the homeodomain plus 10 flanking residues each side, and by NJ or MP analyses. Figure 2 shows an unrooted NJ tree (Figure 2a) or the same tree rooted with selected EHG class genes (En) as outgroup sequences (Figure 2b). Again, none of the trees revealed a close relationship between *Mox* and the central Hox genes. Besides, the resulting trees groups together *Evx* and *Mox* classes, in a basal position with respect to the monophyletic Hox and ParaHox group.

Scenarios for the origin and evolution of the extended Hox and ParaHox clusters

Kourakis and Martindale [8] have pointed out that if a sister of the UrProtoHox gene (which gave rise to the ProtoHox

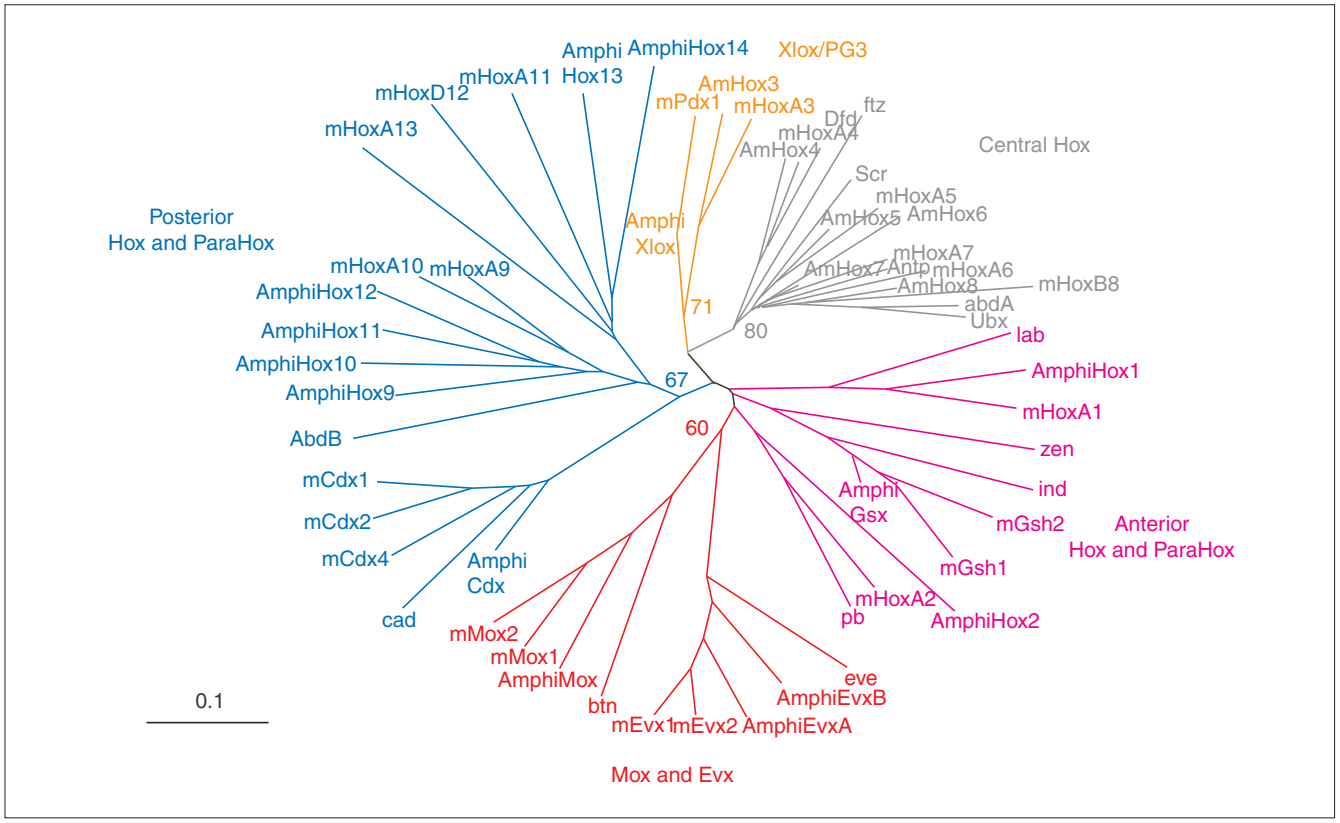

Figure 1
Unrooted neighbor-joining phylogenetic tree. The tree relates the amino-acid sequences of the homeodomain plus 10 flanking residues on both sides in the Hox, ParaHox, *Evx* and *Mox* protein sequences from mouse, amphioxus and *Drosophila*. The numbers refer to bootstrap values. Major groupings are indicated by color codes. Note that *Mox* and *Evx* group together in a monophyletic group.

cluster by tandem duplication) was linked to it, the association of *Evx* with the Hox cluster in certain phyla might be the remnant of such linkage. If this is so, a ParaHox *Evx*-type gene is expected to be adjacent to and 5′ of the *Cdx* gene, provided the Hox/ParaHox split involved genes adjacent to the ProtoHox cluster. This is supported by the presence of genes for tyrosine kinase receptors and collagens, among others, in the vicinity of both Hox and ParaHox clusters ([15] and Figure 3). Our phylogenetic data attractively suggest that *Mox* may well be this gene. Furthermore, careful checking of the mouse and human genomes revealed that, with the exception of mouse *Mox2*, *Mox* and *Evx* genes are linked to the Hox clusters, but at either side of it: whereas *Evx* is tightly linked to the 5′ end of the Hox cluster (under 50 kb), *Mox* is loosely linked to its 3′ end (more than 5 Mb) (Figure 3).

Linkage data and phylogenetic trees allow us to envisage a feasible scenario for the extended Hox/ParaHox cluster origin and evolution (Figure 4). We propose that an ancestral precursor of *Mox* and *Evx* genes (here referred to as the *Evx/Mox* ancestor) was linked to the *UrProtoHox* gene (step 1). The ProtoHox cluster was then generated by tandem duplication of the *UrProtoHox* gene, thus forming, with the *Evx/Mox* ancestor gene, an ancestral Hox-like cluster (step 2). Tandem duplication of the whole cluster and adjacent regions gave rise to the 'coupled' Hox-like cluster (*Evx* plus primordial Hox cluster and *Mox* plus primordial ParaHox cluster, step 3). Thereafter, chromosomal breakage between *Mox* and the primordial ParaHox cluster caused the loose linkage of *Mox* at the anterior end of the Hox cluster (step 4). Finally, the further independent evolution of the primordial Hox and ParaHox clusters (expansion by internal tandem duplications in the former and loss of the central gene in the latter) accounts for the current composition of the extended Hox and the ParaHox arrays in chordates (step 5). Note that steps 4 and 5 are interchangeable, and that Hox cluster expansion and ParaHox reduction may have preceded chromosomal breakage.

Alternative scenarios that include the non-tandem duplication of the ancestral Hox-like cluster would require further steps, including the jumping of *Mox* across clusters. An ancient duplication of the *Evx/Mox* ancestor gene, followed by inversion of *Evx*/ProtoHox plus a local (non-tandem) duplication restricted to the ProtoHox cluster, would account as well for the present situation. Although they

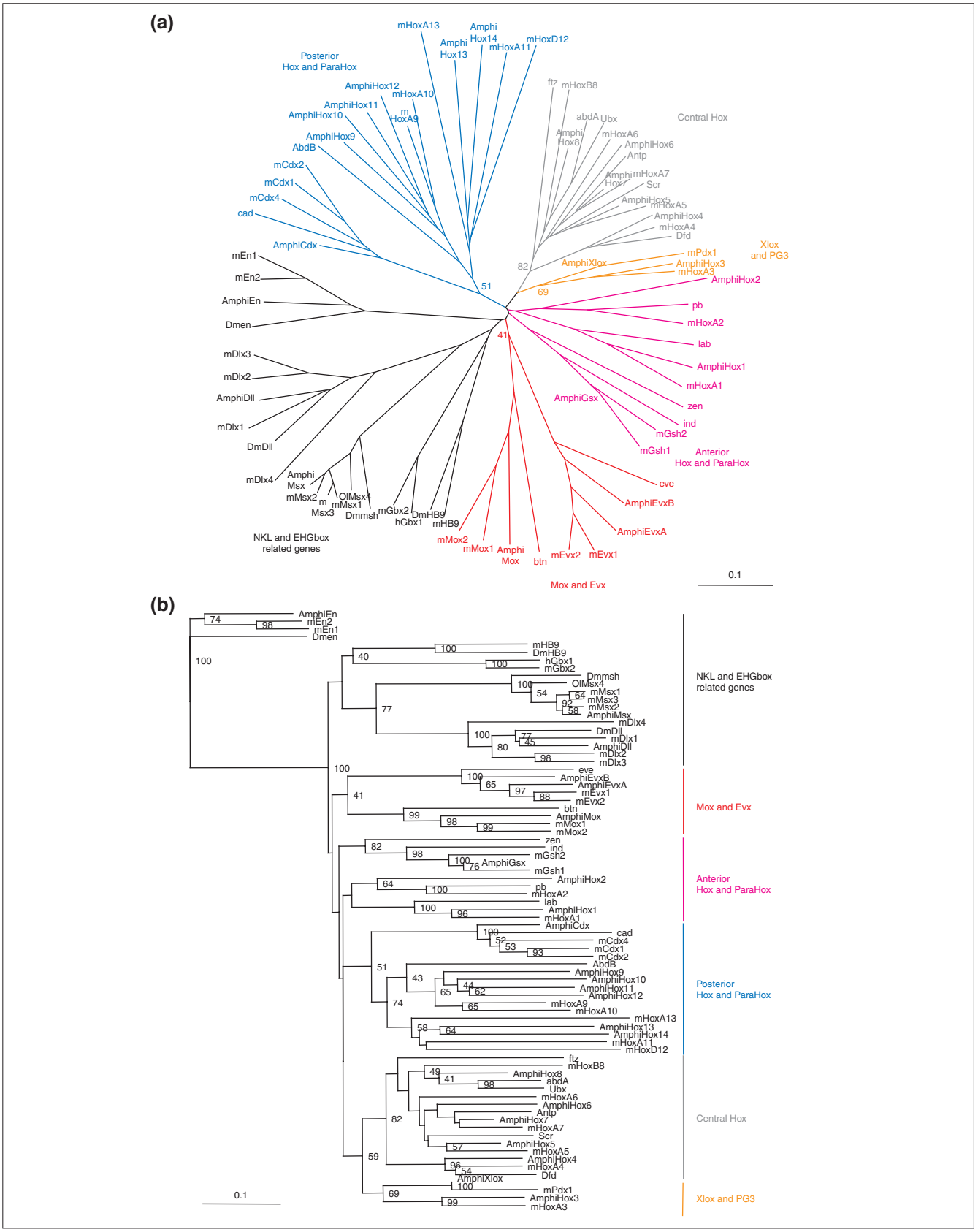

Figure 2 *(see legend on the following page)*

cannot be formally discarded, these scenarios seem unlikely, as they demand more events of gene duplication and local rearrangements than the model proposed here. Furthermore, current linkage data for non-homeobox genes in the vicinity of the Hox and ParaHox clusters (see below) suggest that a larger region was implicated in these duplication events.

The evolutionary scenario proposed here stresses not only the ancient origin of both *Mox* and *Evx* classes but also the necessity of a tandem duplication event to originate the extended Hox and ParaHox clusters. Moreover, not only the ProtoHox cluster, but also neighboring regions (including the *Evx/Mox* ancestor gene), were tandemly duplicated.

Current linkage data strongly favor the proposed outline. It has been proposed that a segmental (non-tandem) duplication restricted to the ProtoHox cluster was involved in the genesis of the extended Hox and ParaHox gene clusters [3,4]. This seems unlikely, as in the neighborhood of the mammalian Hox and ParaHox clusters, there are members of other gene families (for example, tyrosine kinase receptors and collagens ([15] and Figure 3), implying that a larger syntenic region can be traced back to the time of ProtoHox cluster duplication.

This evolutionary scenario nicely squares linkage data on Hox and ParaHox syntenic regions with phylogenetic evidence. It involves regional tandem duplication and chromosomal breakage but no polyploidization events or gene losses at either side of the ParaHox cluster. Such breakage can be dated before the duplication of the Hox and ParaHox clusters at the origins of vertebrates [4,16], since *Mox1* and *Mox2* are linked to the HoxB and HoxA clusters in humans, respectively (Figure 3). However, current linkage data in protostomes do not allow us to trace such breakage further back or determine whether such breakage took place independently in specific lineages. The *Drosophila Evx* homolog, *even-skipped*, is not linked to the Hox cluster and the *Mox* homolog, *buttonless*, is not in the proximity of the Hox cluster nor close to the *cad* and *ind* ParaHox genes [17]. The fly genome is probably highly derived from the protostome ancestor, as is the *Caenorhabditis elegans* genome, which lacks two ParaHox genes and the *Mox* gene [18]. Unfortunately, no linkage data from other invertebrates are available. Moreover, cnidarians probably have Hox and ParaHox clusters derived from the primordial clusters (step 4 in Figure 4). Interestingly *Evx* is linked to Hox genes in anthozoans [9,12], but nothing is known about the chromosomal position of the cnidarian *Mox* homolog with respect to Hox or ParaHox genes [19]. Thus, the existence of cnidarian *Mox* and *Evx* genes, plus Hox and ParaHox, places the tandem duplication of the ancestral Hox-like cluster in early metazoan evolution, before cnidarian divergence.

Conclusions

Most studies dealing with the origin and evolution of Hox and ParaHox clusters have not included the Hox-related genes *Mox* and *Evx*. We have constructed phylogenetic trees with Hox, ParaHox, *Mox* and *Evx* genes and analyzed the available linkage data in mammalian genomes. We support an evolutionary scenario in which an ancestor of *Evx* and *Mox* was linked to the ProtoHox cluster, and that a tandem duplication of a large genomic region early in metazoan evolution generated the Hox and ParaHox clusters, plus the cluster-neighbors *Evx* and *Mox*. The large 'coupled' Hox-like cluster *Evx*Hox/*Mox*ParaHox was subsequently broken, thus grouping the *Mox* and *Evx* and the Hox clusters, and isolating the ParaHox cluster. Whether this breakage happened only once early in evolution, or multiple times in several places is unknown. It is tempting to speculate that a particular extant lineage retains an unbroken version of the 'coupled' cluster.

Materials and methods

Hox, ParaHox, *Evx*, *Mox*, *Msx*, *Gbx* and *Dlx* sequences were obtained from public databases [20]. Trees were constructed with mouse (when available), amphioxus and *Drosophila* sequences. Gene names and accession numbers are as follows: mouse *Mox2* (m*Mox*2, P32443); mouse *Mox1* (m*Mox*1, P32442); amphioxus *Mox* (Amphi*Mox*, AAM09689); *Drosophila buttonless* (btn, AAF56025); mouse *Evx1* (m*Evx*1, P23683); mouse *Evx2* (m*Evx*2, P49749); amphioxus *EvxA* (Amphi*Evx*A, AAK58953); amphioxus *EvxB* (Amphi*Evx*B, AAK58954); *Drosophila even-skipped* (eve, P06602); mouse *Gsh1* (mGsh1, P31315); mouse *Gsh2* (mGsh2, P31316); amphioxus *Gsx* (AmphiGsx, AAC39015); *Drosophila ind* (ind, AAK77133); mouse *Hoxa1* (mHoxa1, P09022); mouse *Hoxa2* (mHoxa2, P31245); amphioxus *Hox1* (AmphiHox1, BAA78620); amphioxus *Hox2* (AmphiHox2, BAA78621); *Drosophila labial* (lab, P10105); *Drosophila proboscipedia* (pb, P31264); *Drosophila zerknüllt* (zen, AAF54087); mouse *Pdx1* (mPdx1, P52946); amphioxus *Xlox* (AmphiXlox, AAC 39016); mouse *Hoxa3* (mHoxa3, P02831); amphioxus *Hox3* (AmphiHox3, CAA48180); mouse *Hoxa4* (mHoxa4, P06798); mouse *Hoxa5* (mHoxa5, P20719); mouse *Hoxa6* (mHoxa6, P09092); mouse *Hoxa7* (mHoxa7, P02830); mouse *Hoxb8* (mHoxb8, P09078); amphioxus *Hox4* (AmphiHox4, BAA78622); amphioxus

Figure 2 *(see figure on the previous page)*
Phylogenetic trees. **(a)** Unrooted neighbor-joining tree; **(b)** rooted trees. The trees' topology suggests that *Evx* and *Mox* group together as a sister group of the Hox/ParaHox clade. See text for discussion.

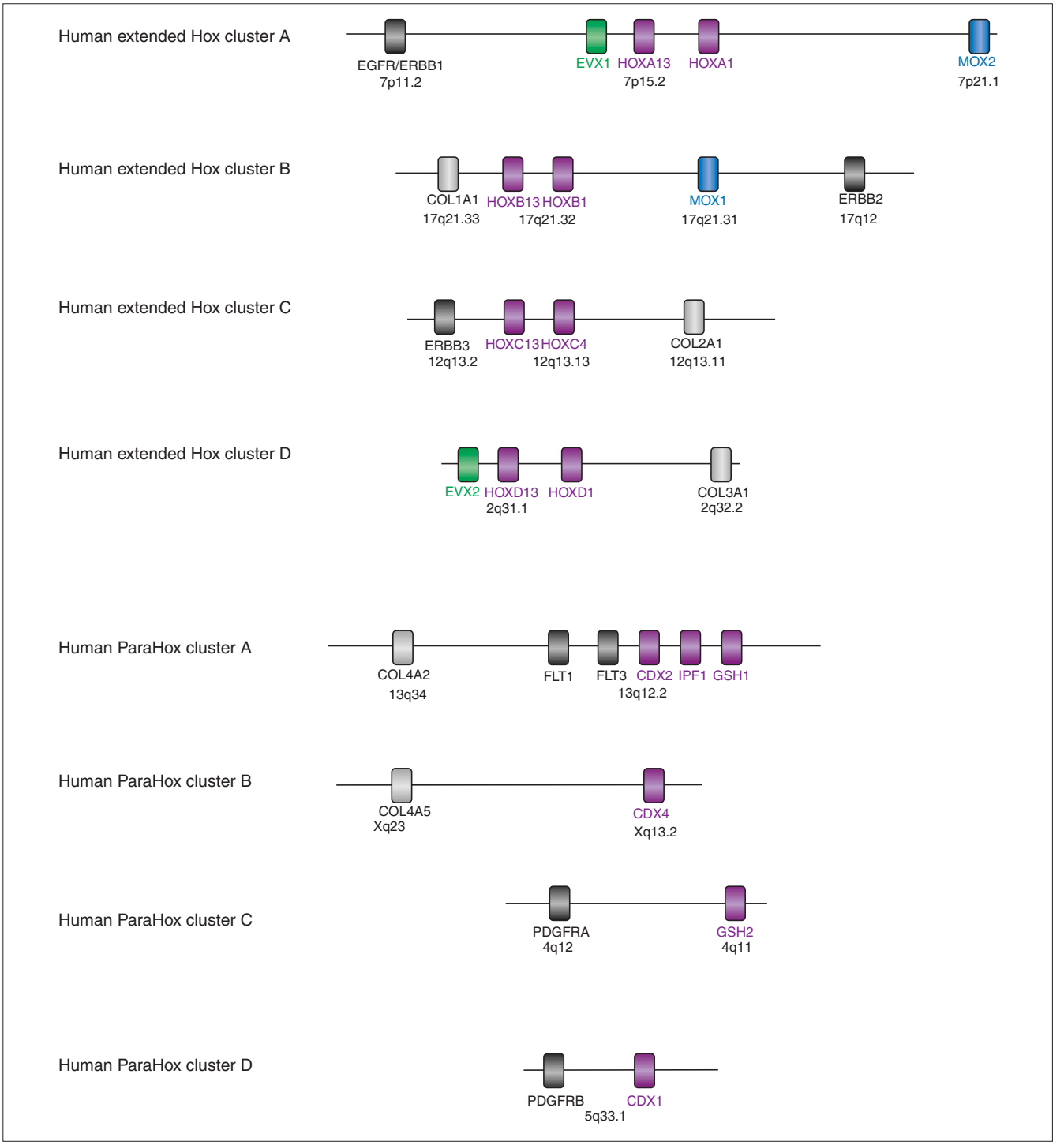

Figure 3
Synteny in the Hox/ParaHox regions in the human genome. The boxes represent the genes (only the 3′- and the 5′-most Hox genes have been depicted for clarity). Chromosomal positions are indicated below. Hox and ParaHox genes are depicted in purple, *Mox* in blue, *Evx* in green and non-homeobox genes such as those for collagens and tyrosine kinase receptors in light and dark gray, respectively.

Hox5 (AmphiHox4, BAA78622); amphioxus *Hox6* (AmphiHox4, BAA78622); amphioxus *Hox7* (AmphiHox4, BAA78622); amphioxus *Hox8* (AmphiHox4, BAA78622); *Drosophila Deformed* (Dfd, P07548); *Drosophila Sex combs reduced* (Scr, P09077); *Drosophila fushi tarazu* (ftz, P02835), *Drosophila Antennapedia* (Antp; P02833); *Drosophila*

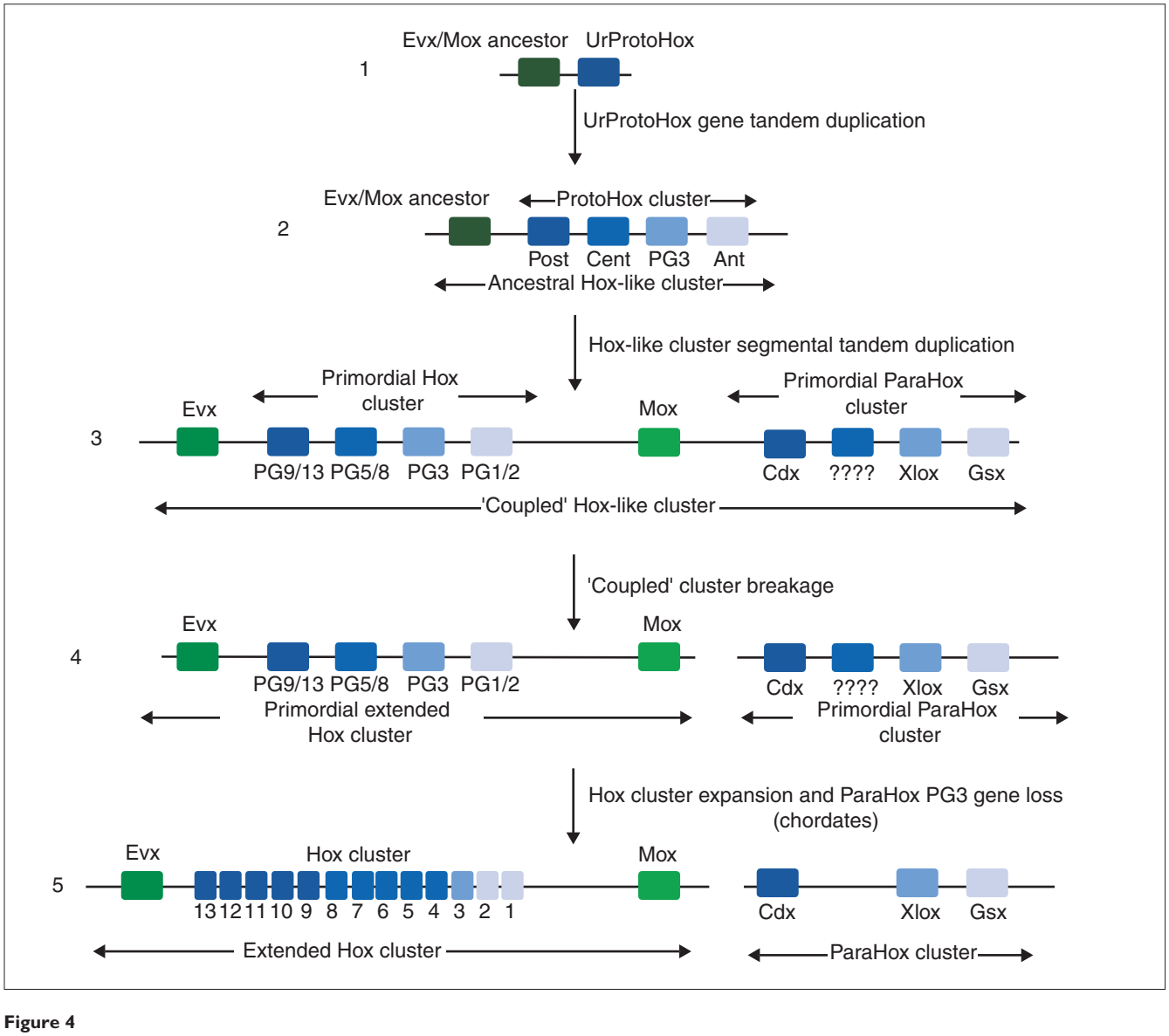

Figure 4
Evolutionary scenario proposed for the origin and evolution of the extended Hox and ParaHox gene clusters. It implies an *Evx*/*Mox* ancestor gene initially linked to the *UrProtoHox* gene (step 1) that gives rise after duplication to *Evx* and *Mox* genes, which is paralleled by the generation of primordial Hox and ParaHox clusters, forming a continuous array containing all the Hox-like genes (the 'coupled' Hox-like cluster; step 3). See text for details.

Ultrabithorax (Ubx, P02834); *Drosophila abdominal-A* (AbdA, P29555); mouse *Cdx1* (mCdx1, P18111); mouse *Cdx2* (mCdx1, P43241); mouse *Cdx4* (mCdx4, Q07424); amphioxus *Cdx* (AmphiCdx, AAC39017); *Drosophila caudal* (cad, P09085); mouse *Hoxa9* (mHoxa9, P09631); mouse *Hoxa10* (mHoxa10, P31310); mouse *Hoxa11* (mHoxa11, P31311); mouse *Hoxd12* (mHoxd12, P23812); mouse *Hoxa13* (mHoxa13, Q62424); amphioxus *Hox9* (AmphiHox9, S47607); amphioxus *Hox10* (AmphiHox10, CAA84522); amphioxus *Hox11* (AmphiHox11, AAF81909); amphioxus *Hox12* (AmphiHox12, AAF81903); amphioxus *Hox13* (AmphiHox13, AAF81904); amphioxus *Hox14* (AmphiHox14, AAF81905); and *Drosophila Abdominal-B* (AbdB, P09087). Selected Antennapedia-type homeobox genes (because of their linkage disposition to the Hox gene cluster in certain genomes), that also were used are: amphioxus *distal-less* (AmphiDll, P53772); amphioxus *Msx* (AmphiMsx, CAA10201); amphioxus *engrailed* (AmphiEn, AAB40144); *Drosophila msh* (Dmmsh, CAA59680); *Drosophila distal-less* (DmDll, AAB24059); *Drosophila engrailed* (DmEn, P02836); *Drosophila HB9* (DmHB9, NP648164); mouse *Dlx1* (mDlx1, Q64317); mouse *Dlx2* (mDlx2, P40764); mouse *Dlx3* (mDlx3, Q64205); mouse *Dlx4* (mDlx4, P70436); mouse *Msx1* (mMsx1, P13297); mouse *Msx2* (mMsx2, Q03358); mouse *Msx3* (mMsx3, P70354); *Oryzias latipes Msx4* (OlMsx4, BAA88311); human *Gbx1* (hGbx1, Q14549)

and mouse *Gbx2* (mGbx2; P48031); mouse *engrailed1* (mEn1, P09065); mouse *engrailed2* (mEn2, P09066); mouse *HB9* (mHB9, NP064328). Sequences from other organisms were omitted as the full set of genes is not available or the homeobox is not fully sequenced.

Trees were constructed using the homeodomain sequence alone or the homeodomain plus ten flanking residues on both sides. The phylogenetic methods used were maximum parsimony (MP), neighbor joining (NJ) and quartet puzzling (QP). First, an alignment was constructed using the ClustalX program [21] and was then edited by eye. NJ trees were inferred by either ClustalX or MEGA 2.0 [22] using a Poisson model for amino-acid evolution. Nodal support was assessed by 1,000 bootstrap replicates. MP trees were inferred using the MEGA 2.0 program, by applying the close-neighbor-interchange method with 1,000 bootstrap replicates. A QP tree was inferred by TREE-PUZZLE 5.0 [23], using the JTT model [24] with a Gamma distribution (eight categories inferred from the data) and 10,000 replicates.

Linkage information was obtained from the human and mouse genome working draft web page [25].

Additional data files

The alignments used to construct the trees in Figure 1 and Figure 2 are available with the online version of this article.

Acknowledgements

We are indebted to Iñaki Ruiz, Gemma Marfany and Ricard Albalat for many discussions, Robin Rycroft and Ivana Miño for checking the English version of the manuscript, and Josep Gardenyes for help with figures. We are particularly indebted to two anonymous referees for extremely fruitful suggestions. This study was supported by grants PB98-1261-C02-02 and BMC2002-03316 (Ministerio de Ciencia y Tecnología, Spain). and by the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya. C.M. held a CIRIT (Generalitat de Catalunya) predoctoral fellowship.

References

1. Gellon G, McGinnis W. **Shaping animal body plans in development and evolution by modulation of Hox expression patterns.** *BioEssays* 1998, **20:**116-125.
2. Finnerty JR, Martindale MQ. **Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria.** *Evol Dev* 1999, **1:**16-23.
3. Pollard SL, Holland PWH. **Evidence for 14 homeobox gene clusters in human genome ancestry.** *Curr Biol* 2000, **10:**1059-1062.
4. Brooke NM, Garcia-Fernàndez J, Holland PWH. **The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster.** *Nature* 1998, **392:**920-922.
5. Schubert FR, Nieselt-Struwe K, Gruss P: **The Antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution.** *Proc Natl Acad Sci USA* 1993, **90:**143-147.
6. Zhang J, Nei M. **Evolution of Antennapedia-class homeobox genes.** *Genetics* 1996, **142:**295-303.
7. de Rosa R, Grenier JK, Andreeva T, Cook CE, Adoutte A, Akam, M, Carroll SB, Balavoine, G: **Hox genes in brachiopods and priapulids and protostome evolution.** *Nature* 1999, **399:**772-776.
8. Kourakis MJ, Martindale MQ: **Combined-method phylogenetic analysis of Hox and ParaHox genes of the metazoa.** *J Exp Zool* 2000, **288:**175-191.
9. Ferrier DEK, Holland PWH. **Ancient origin of the Hox gene cluster.** *Nat Rev Genet* 2001, **2:**33-38.
10. Dush MK, Martin GR**. Analysis of mouse *Evx* genes: *Evx-1* displays graded expression in the primitive streak.** *Dev Biol* 1992, **151:**273-287.
11. Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL, *et al.*: **Zebrafish hox clusters and vertebrate genome evolution.** *Science* 1998, **282:**1711-1714.
12. Miller DJ, Miles A. **Homeobox genes and the zootype.** *Nature* 1993, **365:**215-216.
13. Bürglin TR. **The evolution of Homeobox genes.** In *Biodiversity and Evolution*. Edited by Arai R, Kato M, Doi Y. The National Science Museum Foundation, Tokyo; 1995: 291-336.
14. Gauchat D, Mazet F, Berney C, Schummer M, Kreger S, Pawlowski J, Galliot B: **Evolution of *Antp*-class genes and differential expression of *Hydra* Hox/paraHox genes in anterior patterning.** *Proc Natl Acad Sci USA* 2000, **97:**4493-4498.
15. Spring J: **Genome duplication strikes back.** *Nat Genet* 2002, **31:**128-129.
16. Garcia-Fernàndez J, Holland PWH. **Archetypal organization of the amphioxus *Hox* gene cluster.** *Nature* 1994, **370:**563-566.
17. **FlyBase** [http://flybase.bio.indiana.edu]
18. **WormBase** [http://www.wormbase.org]
19. Naito M, Ishiguro H, Fujisawa T, Kurosawa Y: **Presence of eight distinct homeobox-containing genes in cnidarians.** *FEBS Lett.* 1993, **333:**271-274.
20. **NCBI** [http://www.ncbi.nlm.nih.gov]
21. Thompson JD, Higgins DG, Gibson TJ. **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22:**4673-4680.
22. Kumar S, Tamura K, Jakobsen IB, Nei M: **MEGA2: molecular evolutionary genetics analysis software.** *Bioinformatics* 2001, **17:**1244-1245.
23. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18:**502-504.
24. Jones DT, Taylor WR, Thornton JM**: The rapid generation of mutation data matrices from protein sequences.** *Comput Appl Biosci* 1992, **8:**275-282.
25. **Ensembl Genome Browser** [http://www.ensembl.org]