EVOLUTION & DEVELOPMENT 2:5, 284–293 (2000)

The amphioxus Hox cluster: deuterostome posterior flexibility and *Hox14*

David E. K. Ferrier,[a,b] Carolina Minguillón,[a] Peter W. H. Holland,[b] and Jordi Garcia-Fernàndez[a,*]

[a]Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 645, 08028 Barcelona, Spain; [b]Division of Zoology, School of Animal and Microbial Sciences, University of Reading, Whiteknights, Reading, RG6 6AJ, England

*Author for correspondence (email: jgarcia@bio.ub.es)

SUMMARY The amphioxus (*Branchiostoma floridae*) Hox cluster is a model for the ancestral vertebrate cluster, prior to the hypothesized genome-wide duplications that may have facilitated the evolution of the vertebrate body plan. Here we describe the posterior (5′) genes of the amphioxus cluster, and report the isolation of four new homeobox genes. Vertebrates possess 13 types of Hox gene (paralogy groups), but we show that amphioxus possesses more than 13 Hox genes. Amphioxus is now the first animal in which a Hox14 gene has been found. Our mapping and phylogenetic analysis of amphioxus "Posterior Class" Hox genes reveals that these genes are evolving at a faster rate in deuterostomes than in protostomes, a phenomenon we term Posterior Flexibility.

INTRODUCTION

Hox genes are transcription factors that pattern the anterior–posterior axis of almost all animal embryos (the Zootype) (Slack et al. 1993). A striking feature is the clustered organization of the genes, with a chromosomal position reflecting their site of action in the embryo. Genes at the 3′ end of the cluster pattern the anterior, genes at the 5′ end pattern the posterior, and the central genes pattern the middle. Changes in Hox expression can correlate with morphological evolution (e.g., crustacean limbs [Averof and Patel 1997], butterfly prolegs [Warren et al. 1994], vertebrae [Gaunt 1994, Burke et al. 1995], fin to limb transition [Sordino et al. 1995], and snake trunk homogeneity [Cohn and Tickle 1999]), and altered organization may also be linked with evolution (e,g., absence of *Fugu* group 7 [Aparicio et al. 1997], extra teleost clusters and bony fish diversity [Amores et al. 1998]). In particular, evolution of the vertebrates involved the origin of new cell types and organ systems, and increased body organization complexity. This coincided with genome-wide duplications followed by maintenance of multiple members of many gene families, including Hox genes (Holland et al. 1994).

Mammals have four Hox clusters, teleosts at least four (*Fugu* has four, and zebrafish seven), and lampreys probably have three (Aparicio et al. 1997, Amores et al. 1998, Sharman and Holland 1998). Each higher vertebrate cluster contains a selection of the 13 different types (paralogy groups) of vertebrate Hox genes, with no single cluster containing all 13 genes. Of a possible 52 genes, the mouse only has 39. Amphioxus (*Branchiostoma*, Cephalochordata) is the invertebrate sister group to the vertebrates, with a simpler vertebrate-like body plan and a single Hox gene cluster. The previous characterization of the amphioxus cluster indicated its archetypal nature relative to vertebrates, at least up to the start of the posterior end of the cluster (Garcia-Fernàndez and Holland 1994). The amphioxus cluster has no missing genes, unlike each individual vertebrate cluster.

Hox genes are divisible into classes, based on their sequence, cluster position, and embryonic expression domains. The Posterior genes form one of these classes, containing the deuterostome genes in groups 9 and above, and the protostome *AbdB, Post1,* and *Post2* genes (Izpisúa-Belmonte et al. 1991, Schubert et al. 1993, de Rosa et al. 1999). Until recently, it was thought that only deuterostomes contained multiple Posterior genes, whereas protostomes had only one. It is now clear that multiple Posterior genes exist in several protostomes, and that their sequence relationships provide strong characters in support of a new classification of the protostomes into two major groups (Ecdysozoa and Lophotrochozoa) (Aguinaldo et al. 1997, de Rosa et al. 1999). Ecdysozoa have a Posterior gene clearly related to *Drosophila AbdB* (the first member of the Posterior gene class to be described), whereas Lophotrochozoa possess two Posterior genes (*Post1* and *Post2*), neither of which is more closely related to the *AbdB* class, or to any particular vertebrate paralogy group.

We continued the genomic walk into the posterior (5′) region of the amphioxus Hox cluster, to refine our understanding of chordate Hox gene evolution. A surprising finding

was that amphioxus has more than 13 Hox genes. Amphioxus thus possesses the largest number of Hox genes in a single cluster of any animal so far characterized. Molecular phylogenetic analysis shows that the Posterior AmphiHox genes do not separately group with the individual vertebrate paralogy groups, or with any other deuterostome Posterior gene. This lack of resolution between deuterostome subphyla contrasts with clear interphyletic groupings of protostome Posterior genes, implying that deuterostome Posterior genes are evolving at a faster rate than those of protostomes. We call this distinctive behaviour of the 5′ deuterostome genes Posterior Flexibility.

MATERIALS AND METHODS

Genomic walking

Two amphioxus genomic libraries were used, each made from single animals collected from Tampa Bay (Florida, USA). A lambda library of approximately 6×10^5 independent pfu was constructed in a partially filled-in FIX II vector (Stratagene, La Jolla, CA, USA), following methods described in Garcia-Fernàndez et al. (1993) and the manufacturer's instructions. Cosmid library MPMGc117 (Max-Planck Institut für Molekulare Genetik, Berlin, Germany) was also used (Lehrach et al. 1990). Walking started with a probe situated 10kb 5′ of *AmphiHox10*, and was performed as described previously (Garcia-Fernàndez and Holland 1994).

Homeobox isolation

The whole of the genomic walk was screened with a degenerate oligonucleotide SO2(CKNCKRTTYTGRAACCA), as described in Garcia-Fernàndez and Holland (1994), which recognizes the highly conserved region of the third helix of Antp class homeoboxes. Fragments to which SO2 hybridized were subcloned into pBluescript and sequenced, either manually or on automated sequencers (ABI Prism [Perkin-Elmer, Foster City, CA, USA] and ALF [Pharmacia, Uppsala, Sweden]). For the genes that contained a homeobox intron, the remainder of the homeobox was isolated by hybridization with a degenerate oligo to the first helix SO1(GARYTNGARAARGARTT), or for *AmphiHox14*, by PCR with SO2 and the oligonucleotide DFH14 (AARAARCGSTGTCCSTACAC) on a phage clone from the genomic walk.

Phylogenetic analysis

The sequences used in the phylogenetic comparisons with the AmphiHox Posterior genes reported here (Accession numbers AF276811 to AF276817) were obtained from public databases, and manually aligned around their homeodomains. The homeodomain sequences with six additional amino acids on either end were used to construct trees by parsimony (PAUP beta 4.0 version) and Neighbour-joining (Phylip) (Swofford 1998; Felsenstein 1993), with the parameters given in Fig. Legend 4, and by maximum likelihood (Puzzle, from http://www.zi.biologie.uni-muenchen.de/~strimmer/puzzle.html) with the parameters given in Fig. Legend 5. The smaller selection of vertebrate sequences used in the maximum likelihood analysis permitted faster computer analysis. Three members of each vertebrate paralogy group were selected so as to encompass the bulk of the diversity seen within each group.

RESULTS AND DISCUSSION

Distinct organization of the posterior region of the AmphiHox cluster

We have undertaken a genomic walk in the 5′ direction from the *AmphiHox10* gene (Garcia-Fernàndez and Holland 1994), and identified four new homeobox genes. We designate these genes *AmphiHox11*, *AmphiHox12*, *AmphiHox13*, and *AmphiHox14*. Note that the numbers in the gene names do not imply orthology to vertebrate paralogy groups, but solely reflect genomic order (Fig.1A). These four newly discovered homeobox genes, together with the previously described *AmphiHox9* and *10* genes, are most closely related to the "Posterior class" of vertebrate and invertebrate Hox genes (Fig. 2), as judged from BLAST comparisons of homeodomains. We therefore call this the "Posterior" region of the amphioxus Hox cluster. Like the other genes of the Hox cluster, all of the Posterior genes in amphioxus have the same transcriptional orientation.

In some respects the organization of the Posterior region of the amphioxus Hox cluster is strikingly different from a consensus mammalian cluster (Fig. 1B). First, the amphioxus cluster is at least three times longer than the average human cluster, yet most of the relative intergenic distances are the same (the dotted lines of Fig.1B tend to be parallel). Second, there are some deviations from this rule, notably the intergenic distance for 12–13 where amphioxus is more compact, and the relative distance between 9 and 10, which is much larger in amphioxus. Third, three of the Posterior AmphiHox genes have introns in their homeoboxes (*AmphiHox11*, *12*, and *14*). Homeobox introns are not present in any vertebrate Posterior genes so far isolated. *AmphiHox12* has an intron between codons 37 and 38 of the homeobox, a rare site for a homeobox intron. The position of the homeobox intron in *AmphiHox11* and *14* is the same as in *Drosophila AbdB* (Bürglin 1994), between codons 44 and 45 of the homeobox. This is a common intron site for *Antp* class genes (Bürglin 1994), and may be a "hotspot" for intron insertion, or an ancient intron site.

A chordate Hox 14

Amphioxus has a fourteenth Hox gene. All vertebrates so far examined have no more than 13 Hox paralogy groups, and it has been assumed that the vertebrate ancestor had only 13 Hox genes, prior to the cluster duplications during vertebrate origins (Amores et al. 1998, Holland et al. 1994). Previous analysis of the Amphioxus Hox cluster showed its prototypical nature relative to vertebrates (Garcia-Fernàndez and Holland 1994); hence, we expected to find only 13 AmphiHox genes. After cloning *AmphiHox13* we continued the genomic walk to

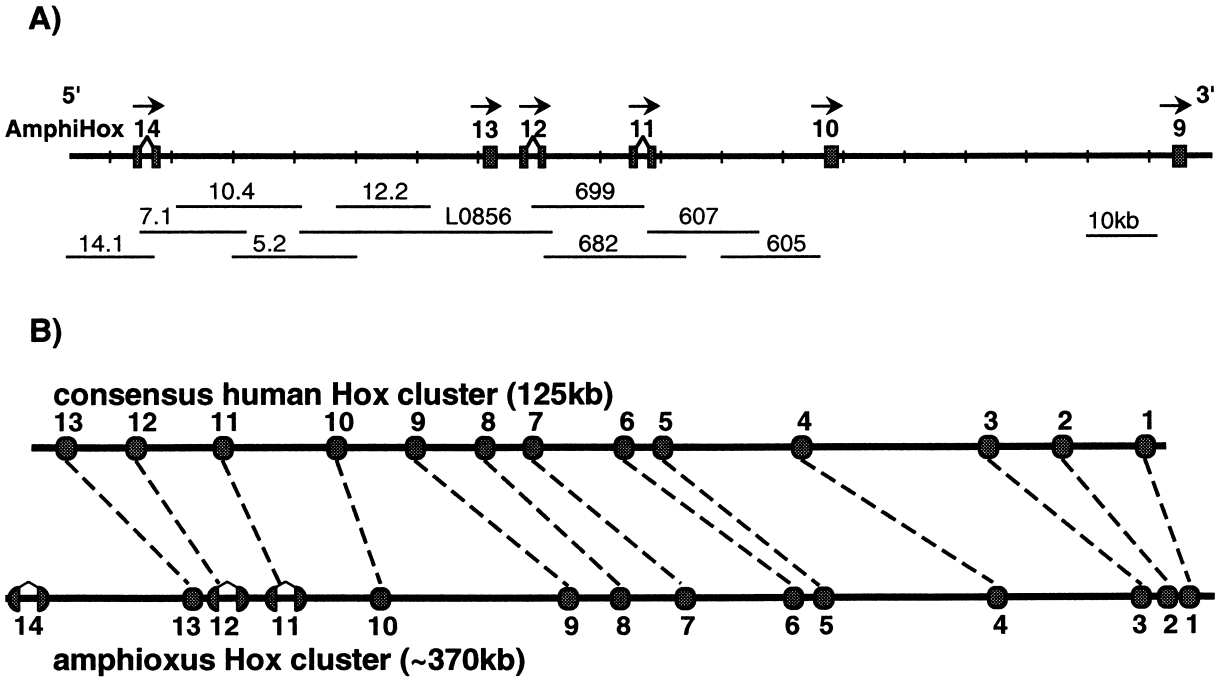


Fig. 1. (A) The genomic walk from *AmphiHox10*. The homeoboxes of the six Posterior class amphioxus Hox genes are shown as gray boxes on a continuous line, which represents the chromosome. *AmphiHox11, 12,* and *14* have homeobox introns (not drawn to scale). All of the genes have the same transcriptional orientation. Beneath the chromosome are a selection of the clones from the genomic walk, with the clone number shown. The region between *AmphiHox9–10* is described in Garcia-Fernàndez and Holland (1994). (B) A comparison of the amphioxus Hox cluster to a consensus human cluster. The intergenic distances of the consensus human cluster are averages of the distances from the four Hox clusters (A, B, C, and D) (Duboule 1994; Acampora et al. 1989). Gray ovals mark the positions of the homeoboxes. The amphioxus and human clusters are drawn to different scales, but show the relative intergenic distances. The dotted lines connect putative orthologues, and reveal the conservation of relative intergenic distances except for between groups 1–2, 9–10, and 12–13. Group 1 genes are at the 3′ end of the cluster, and group 13/14 at the 5′ end.

try to establish linkage to the *AmphiEvx* genes, as a sign that we had reached the 5′ end of the Hox cluster. Vertebrate *Evx* genes are located on the 5′ ends of some Hox clusters, which probably represents the ancestral condition (Amores et al. 1998). Although we have cloned the *AmphiEvx* genes, we have not been able to determine whether they are linked to the Hox cluster (data not shown). However, the genomic walk beyond *AmphiHox13* led to the discovery of *AmphiHox14*.

AmphiHox14 is clearly a Posterior class Hox gene. BLAST searches with its homeodomain preferentially identify deuterostome Posterior Hox proteins, and it clearly groups with the Posterior genes in phylogenetic trees. Its precise evolutionary relationship with the other AmphiHox genes and with the vertebrate Posterior genes is obscure, however. Two scenarios are possible. First, *AmphiHox14* could be the result of a tandem gene duplication specific to the cephalochordate lineage, so that no other chordate possesses an orthologue or a semi-orthologue (a semi-orthologue is a gene resultant from duplication of an orthologue [Sharman 1999]). Second, *AmphiHox14* may represent the ancestral vertebrate condition, with vertebrate group 14 genes having subsequently been lost. There is also the possibility that some vertebrates still possess a group 14 gene, but that it has simply not been isolated yet. In this regard it is noteworthy that despite all of the concentrated effort on describing mammalian Hox clusters, *Hoxb-13* was only isolated as recently as 1996 (Zeltser et al. 1996), four years after it had been thought all mammalian Hox genes were known (Scott 1992). Furthermore, the increasingly divergent nature of the homeoboxes further 5′ in the cluster would make it harder to isolate such genes by the conventional methods of degenerate PCR and homeobox hybridizations.

The relationship of *AmphiHox14* to the other AmphiHox genes, and to the vertebrate Hox genes, is ambiguous from phylogenetic analyses. By parsimony analysis, *AmphiHox14* falls onto a Posterior class polytomy (see below), resolving with no particular Posterior gene. Neighbour-joining analysis, however, does group it weakly with *AmphiHox13* (bootstrap 77%), as does maximum likelihood analysis (Puzzle value 84). This is consistent with *AmphiHox14* having originated by a tandem duplication from *AmphiHox13*, specifically in the cephalochordate lineage. Despite this tentative

```
DmeAntp          FGKCQE  RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN  KTKGEP

AmphiHox
AmphiHox9        WMNNHS  SRKK.CP...F.........LY.M....E..Y..SQHVN.....V..........M..MS  .QRQ.Q
AmphiHox10       WMAPRV  GRKK.CP..K..I.......L..M.VS.E..Q..SRHVN.SD..V..........M.RM.  .ARE.Q
AmphiHox11       WMSAKS  TRKK.CP..K..........L..MFV..E..Q...RQ.N..D..V..........M.RMK  QRAMQQ
AmphiHox12       WW.L.S  SRKK.CP.SKV.L.......LY.M.I..EQ.G...RKVN..D..V..........M.RMK  QRHE.E
AmphiHox13       QSVARG  GRKK.CP..K..LSV..Q.YIQ...VS.ET.L.LSQR.N..D..V..........Q.RLE  FRSRNQ
AmphiHox14       GSLTKP  VRPK.RP.SK..LN...N.YVQ.Q.IS.DK.LQLSQK.N.....V.........I.Q..LD  RRNS.M

group 9
BreA9a           WLHASS  TRKK.CP..KH.I.......L..T....D..Y.V.RL.N.....V..........M..F.  .NETKE
HsaB9            WLHARS  SRKK.CP..K..........L..M....D..H.V.RL.N.S...V..........M..M.  .EQ.KE
BreC9a           WIHARS  TRKK.CP..K..........L..M....D..Y.V.RV.N.....V..........M..M.  .E.NDS
HsaD9            WIHARS  TRKK.CP..K..........L..M....D..Y.V.RI.N.....V..........M..MS  .E.CPK

group 10
HsaA10           WLTAKS  GRKK.CP..KH.........L..M....E..L..SRSVH..D..V..........L..M.  RENRIR
BreB10a          WLSAKA  GRKK.CP.SKH.I.......L..M....E..L..SRSIN..D..V..........L..MT  REHRTR
MmuC10           WLTAKS  GRKK.CP..KH.........L..M....E..L..SKTIN..D..V..........L..M.  RENRIR
BreD10           WLTAKS  GRKK.CP..KH.........L..M....E..L..SKSVS..D..V..........L..MS  RENRIR

group 11
BreA11a          RIGGPR  FRKK.CP..KF.IR...R..F.SV.INKEK.LQLSRM.N..D..V.M........E..L.  RDRLQY
HsaC11           APNAPR  TRKK.CP.SKF.IR...R..F..V.INKEK.LQLSRM.N..D..V..........E..LS  RDRLQY
BreD11           XSSATK  SRKK.CP.SK..IR...R..F..V.INKEK.LQLSRM.S..D..V..........E..L.  RDRLQY

group 12
BreC12b          YPMHRQ  TRKK.KP.SKL.LN...G..IL.EFI..Q..R.LSDR.N..DQ.V..........K.RLL  MREQAL
HsaC12           ??????  SRKK.KP.SKL.LA...G..LV.EFI..Q..R.LSDR.N.SDQ.V..........K.RLL  LREQAL
BreD12           CPSQVR  SRKK.KP..K..LT...N..MM.EFIN.QK.K.LSD..E.SDQ.V..........K.RLM  MREHTF
MmuD12           GAAPGR  ARKK.KP..KQ.IA...N..LV.EFIN.QK.K.LSNR.N.SDQ.V..........K.RVV  QREQAL

group 13
BreA13b          GASVRR  GRKK.VP..KV.LK...R.YAT.KFI.KDK.RR.SAHTN.....VT.......V.E..VV  NKYKGI
HsaB13           ACAFRR  GRKK.IP.SKG.LR...R.YAA.KFI.KDK.RK.SA.TS.S....T.......V.E..VL  AKVKNS
BreC13a          ??????  GRKK.VP..KI.LK.....YAASKFI.KDK.RR.SATTN.S...VT.......V.E..FV  SKSKTN
BreD13           ASFC.R  GRKK.VP..KF.LK...R.YNTTKFI.KEN.RR..SSIN.S...VT.......V.D..RP  DVCIKC
HsaD13           MCVYRR  GRKK.VP..KL.LK...N.YAI.KFINKDK.RR.SA.TN.S...VT.......V.D..IV  SKLKDT

other deuterostomes
CiHbox5          ??????  SRKK..P.SKT.ISS..R.YKA.NFI..QK.EN..RD.K.SD..V........V.D..IK  QREIKD
CiHbox4          ??????  QR.R.RP..K..LS...R..GA.EFIS.EM.EQ..VRVG.ND..V.............RMQ  HRGEQS
CiHbox3          ??????  GRKK.VP..K..L........Y.Q..S.E..L.V.KSVK..D..V..............R  REERQX
StyelaAhox2      GWLTAN  GRKK.VP..K..L........Y.Q..S.E..Q.V.K.VS.SD..V..............K  .EEKVR
HeHbox10         PPPNVR  TRKK.KP..KF..F......LY.M....D..SH.SR..S.....V..........L..MR  AREENE
SpHbox7          TFTTTP  .RTK.RP.SKL.IY......TT.M....D..SKLSQ..D.....V..........M..L.  DKEKTQ
HeHbox7          TFTTTP  .RTK.RP.SKL.IY......QA.M....D..SKLSQ..D.....V..........M..L.  EKEKTQ
TgHbox4          WLSATS  GRKK.CP..KF.........L..M....D..L...RL.S.....V..........M..Q.  RAQNY.

protostomes
DmeAbdB          WTGQVS  VRKK.KP.SKF.........L..A.VSKQK.W.L.RN.Q.....V..........N..NS  QRQANQ
CsaAbdB          WTGTVT  VRKK.KP.SKF.........L..A.VSKQK.W.L.RN.N.....V..........S..TS  QRNA.N
PcaAbdB          WTSNVS  VRKK.KP..K..........L..A.VSKQK.W.L.RT.N.....V..........S..S.  QKET.K
Y75B8A.1/php-3   TSSSHA  MRKK.KP..KA.........LY.T.VSKQK.W.L.KY.H.....V..........D..QK  QRTSGD
egl5/ceh11       WPNYAS  S.K.....Q....SV..AK.QQSS.VSKKQ.E.LRLQTQ..D.............A...K  QRVDDH
Y75B8A.2/nob-1b  WAISHD  G.KK..P.KKD.ISR..Y.YSV.Q...NK..S.LSAQ.M.D.K.V.V.........D..LR  QRHSG.
Nvi-Post1        GPTILH  MRKK.KP.SK..IA.....YVN.T.I.KPK.W.LSQR.N.S...V..........E..VT  DK.CDD
Lan-Post1        LPAVIH  MRKK.KP.SK..IA...R.YVS.T.ISKPK.W.LSQR.Q.S...V..........E..VK  GG.QT?
Nvi-Post2        ??DQPR  QRKK.KP......MV..N..MG.S.I..QK.W..SCK.H.S...V.V........R..L.  ERAKTL
Lan-Post2        ??????  ????.KP......MV..N..LN.A.I..QK.W..SCK.H.S...V.V........R..L.  ERAKAL
LsHox9           .STEPR  TRKK.KP......MV..N..LT.S.I..QK.W..SCK.H.....V.V........R..L.  ARSKVK
```

Fig. 2. Sequence alignment of a representative selection of Posterior Hox proteins. The AmphiHox sequences do not obviously group with individual vertebrate paralogy groups. The alignment includes the homeodomain with six amino acid flanks, and is compared to *Drosophila Antennapedia*. Dots represent identities, and question marks are unknown residues. Abbreviations of species names are given in Fig. legend 4.

result, the possibility remains that the vertebrate ancestor possessed a group 14 Hox gene.

Posterior AmphiHox origins: independent duplications or pro-orthology?

To investigate the evolutionary relationships between the Posterior Hox genes we performed molecular phylogenetic analyses of the encoded sequences. We restricted analysis to the homeodomain plus six amino acid flanks, since many of these flanking residues are characteristic for different vertebrate paralogy groups or for protostome groups, while sequence beyond this region is less phylogenetically informative.

One can envisage two possible phylogenetic tree topologies, a priori (Fig. 3). First, each Posterior AmphiHox gene may be pro-orthologous to one specific vertebrate paralogy group (a pro-orthologue is a gene that is orthologous to the ancestor of the whole set of paralogues of the gene in question [Sharman 1999]). In this scenario, each Amphioxus gene would be expected to fall as a sister to a different vertebrate paralogy group in a phylogenetic tree. Alternatively, the apparent lack of similarity of the AmphiHox genes to the vertebrate counterparts could be due to their origin via a series of tandem duplications independent from those that formed the vertebrate Posterior genes. In this scenario the AmphiHox genes should form a distinct clade from the vertebrate genes. A mixture of these two extreme scenarios is also possible.

The actual outcome of our phylogenetic analysis is shown in Fig. 4. The AmphiHox genes group neither with the individual vertebrate groups, nor strongly with themselves. In

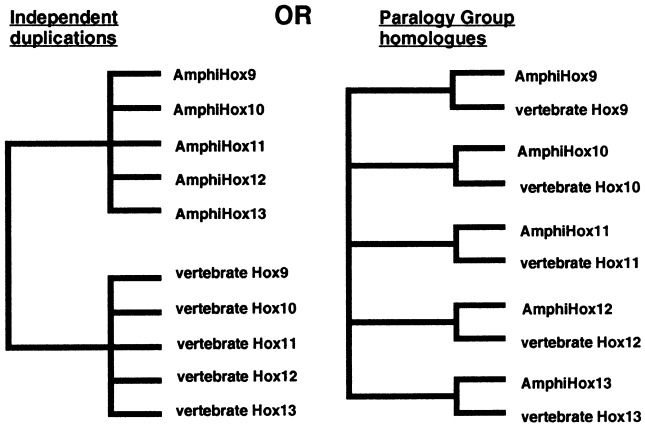


Fig. 3. A schematic representation of the two extreme alternatives for phylogenetic tree topologies, for the behavior of the Posterior AmphiHox genes relative to the vertebrate paralogy groups (as discussed in the text).

the parsimony analysis they remain unresolved on a Posterior gene polytomy (arrow in Fig. 4). In the Neighbour-joining analysis *AmphiHox13* and *14* group together weakly (bootstrap=77%) (tree not shown), but are clearly separated from another very weak grouping of *AmphiHox9–12*. The bootstrap values uniting *AmphiHox9–12* are all less than 36.5%, which is well below the 70% value conventionally taken as the minimum value indicating a valid grouping.

This weak association of some AmphiHox genes was examined more closely with maximum likelihood (Fig. 5). The AmphiHox genes are scattered among the vertebrate paralogy groups (Fig. 5A), forming neither an AmphiHox clade (indicating independent duplication) nor sister groupings with the vertebrate genes (indicating pro-orthology) (Fig. 3). When the scattered tree topology of Fig. 5A is compared to an artificially constructed "independent duplication" model (with the AmphiHox genes grouped together in a single clade), the independent duplication topology is significantly worse than the best maximum likelihood tree (Fig. 5B). Therefore, the extreme version of the independent duplication model, with all Posterior AmphiHox genes arising independently from the vertebrate paralogy groups, can be rejected.

To investigate whether one or more divergent AmphiHox genes were disrupting the tree topolgy (e.g., "pulling" each other away from pro-orthologous, sister-grouping relationships), we constructed maximum likelihood trees of the vertebrate genes with each individual Posterior AmphiHox gene (Figs. 5C–5F, and data not shown). The individual AmphiHox genes still do not resolve as sisters to their putative orthologous vertebrate paralogy groups, with the exception of *AmphiHox9*. Furthermore, the separate AmphiHox genes fall onto different internodes; *AmphiHox13* is between (Hox13–12) and (Hox 9–11), *AmphiHox12* is between (Hox 13–11) and (Hox 9–10), *AmphiHox11* is between Hox9 and (Hox 10–13), and *AmphiHox10* is between (Hox 13–11) and (Hox 9–10). These variable internode positions are strong evidence against the Independent Duplication model. The internode positions of the AmphiHox genes are also inconsistent with pro-orthologous relationships with the vertebrate paralogy groups. However, comparisons of the maximum likelihood trees in Fig. 5 (C–F) with trees in which the AmphiHox genes were moved to sister group positions with their putative orthology groups, revealed that the sister group relationships are not significantly worse than the best maximum likelihood trees (data not shown). The possibility thus remains that the *AmphiHox9–13* genes are pro-orthologous to the vertebrate Hox9–13 paralogy groups, but that these relationships have largely been obscured during the evolution of these sequences.

The Posterior Hox genes of Amphioxus are clearly evolving in a different fashion from the more anterior (3′) Hox genes. *AmphiHox1, 2, 3,* and *4* each group with a different vertebrate paralogy group and the homologous *Drosophila* genes in a phylogenetic tree (Fig. 6), implying descent from a set of four distinct genes in the most recent common ancestor. The middle Hox genes (groups 5–8) are not resolved, due to a sparseness of distinctive residues which are necessary for resolution in the molecular phylogeny programmes. The Posterior genes have comparable numbers of distinctive residues to the anterior (3′) Hox genes, and so the lack of resolution of the Posterior AmphiHox genes with the individual vertebrate groups is not simply due to the cephalochordate–vertebrate divergence being so ancient that orthologies cannot inherently be resolved, as shown by the Anterior gene resolution.

Deuterostome Posterior Flexibility

Posterior Hox genes are not inherently unresolvable between phyla. The Posterior Hox genes of the protostome phyla are clearly resolved into their three types in phylogenetic trees (*AbdB, Post1,* and *Post2* [Fig.4]), in contrast to the deuterostome Posterior Hox genes. The origins of most Bilaterian phyla, including all of those encompassed by these Hox sequences, probably occurred in a short stretch of geological time. The fossil evidence for an explosive radiation of triploblastic phyla is highly suggestive, but also independent molecular evidence points to a concentrated radiation, with short interphyletic internodes in the rDNA trees (Philippe et al. 1994). The behavior of the protostome and deuterostome Hox genes in phylogenetic trees should thus be comparable to each other, unless they are in fact evolving differently in separate phylogenetic lineages.

We suggest that for the deuterostome Posterior Hox genes, the lack of resolution between subphyla is due to the genes evolving at higher rates relative to the protostome Posterior genes, and relative to the more anterior genes. We call this phenomenon Posterior Flexibility.

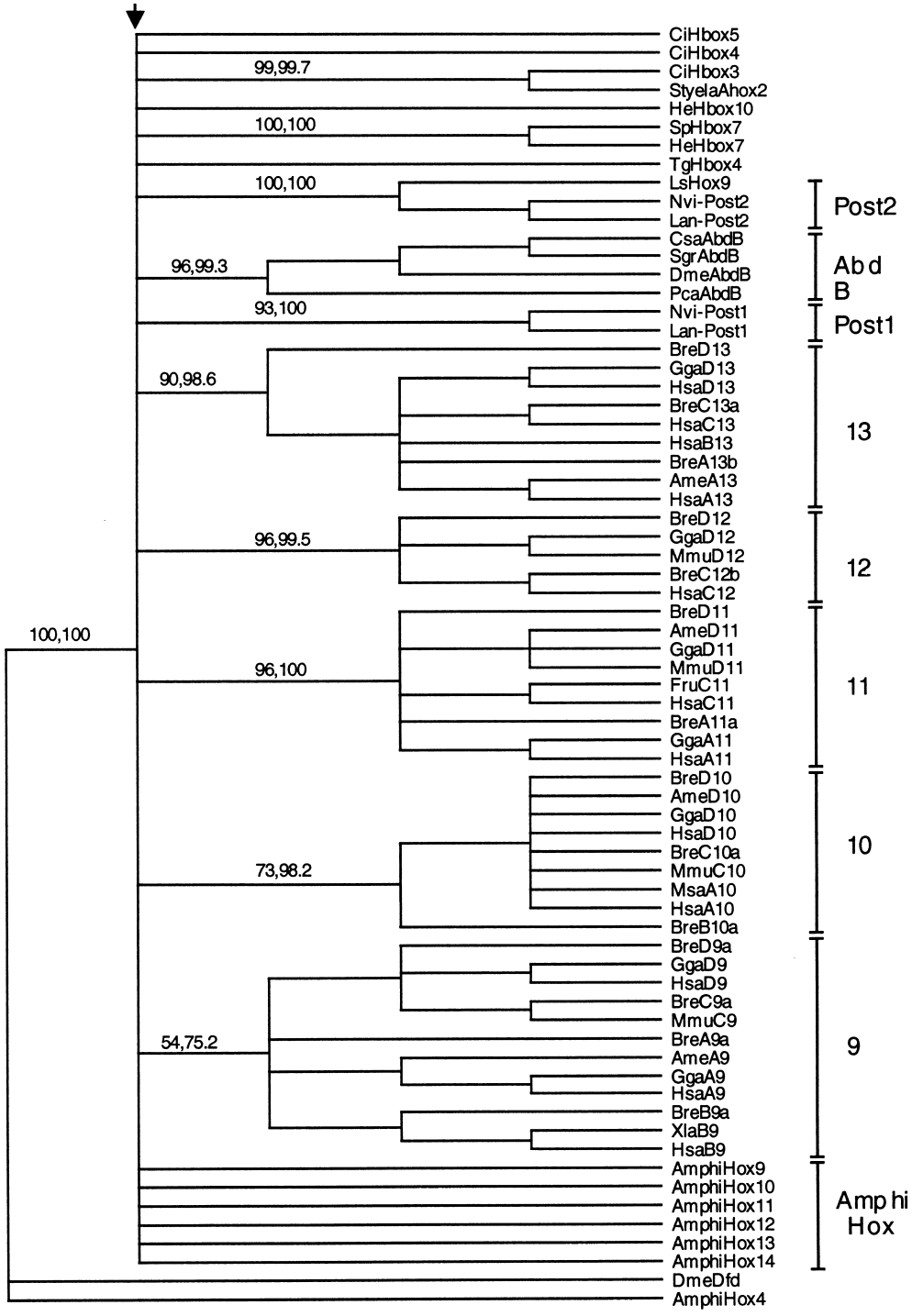
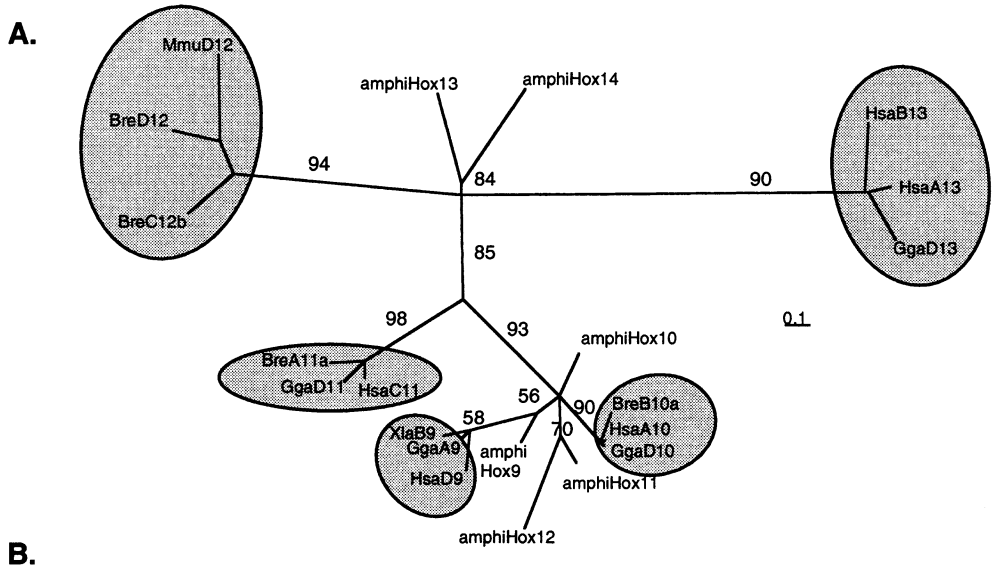
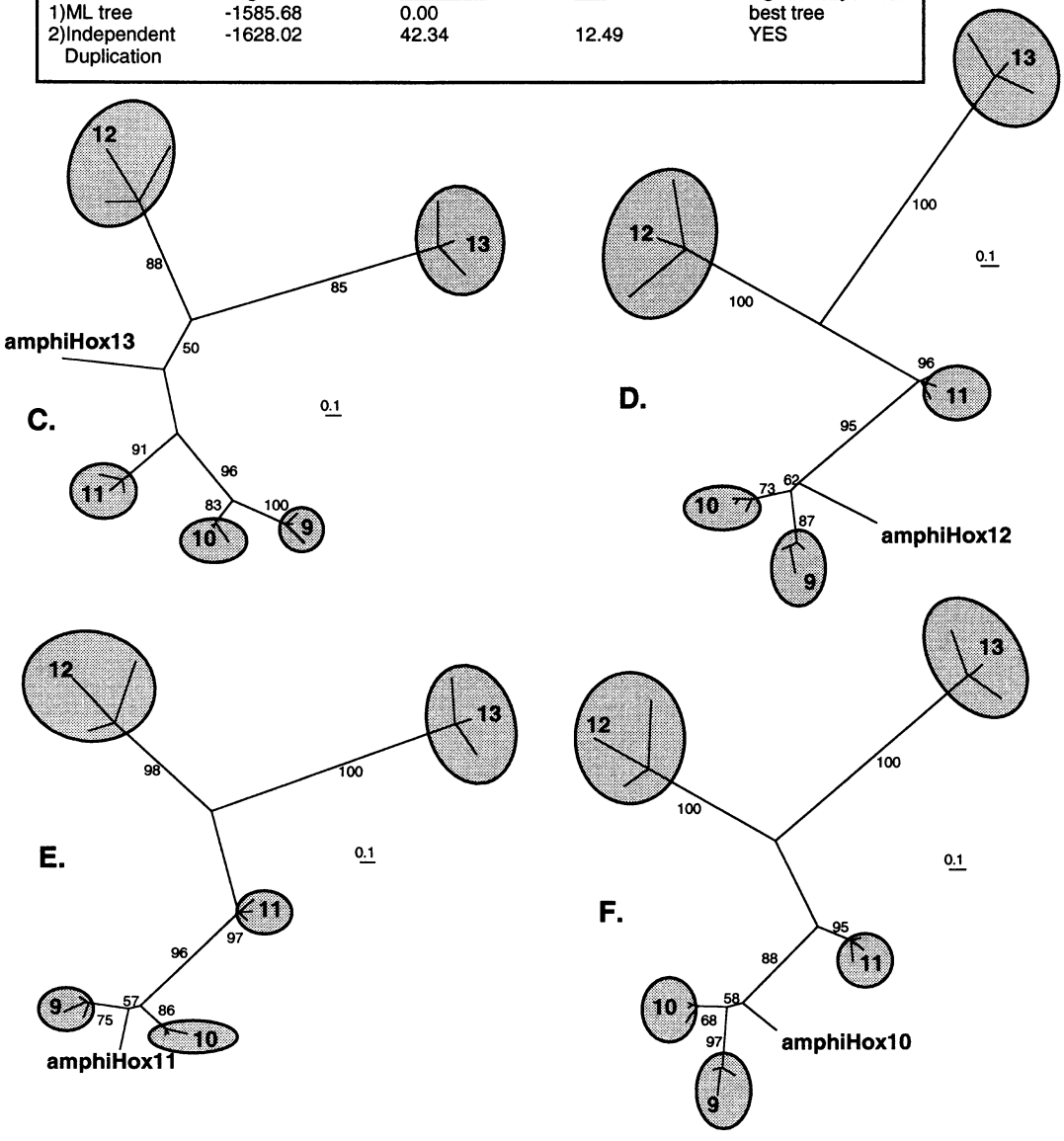


Fig. 4. Phylogenetic tree of Posterior Hox proteins. Distinct clades are formed of each vertebrate paralogy group (numbered 9–13), and each protostome group (*AbdB*, *Post1,* and *Post2*). All of the invertebrate deuterostome sequences fall onto a Posterior gene polytomy (arrow), with the exception of *CiHbox3*/*StyelaAhox2* and *SpHbox7*/*HeHbox7* (see text). The tree was constructed with PAUP, using 100 bootstrap replicates, and collapsing all nodes of less than 50%, with *DmeDfd* and *AmphiHox 4* as the outgroup. The same data set was subjected to Neighbour-joining analysis within the PHYLIP package, using 1000 bootstrap replicates and the Dayhoff substitution model. Virtually the same tree topology was produced, if nodes of less than 80% were collapsed (except for the vertebrate group 9 node). Minor exceptions to the conserved topology were seen in a few of the gene groupings within the vertebrate paralogy groups. Numbers on the branches represent bootstrap percentages from the parsimony and Neighbour-joining analyses, respectively. The numbers on the vertical bars are the vertebrate Hox paralogy group. Species abbreviations are: Ci, *Ciona intestinalis*; Styela, *Styela clava*; He, *Heliocidaris erythrogramma*; Sp, *Strongylocentrotus purpuratus*; Tg, *Tripnuestes gratilla*; Ls, *Lineus sanguineus*; Nvi, *Nereis virens*; Lan, *Lingula anatina*; Csa, *Cupiennius salei*; Sgr, *Schistocerca gregaria*; Dm, *Drosophila melanogaster*; Pca, *Priapulus caudatus*; Bre, *Brachydanio rerio*; Gga, *Gallus gallus*; Has, *Homo sapiens*; Ame, *Ambystoma mexicanum*; Mmu, *Mus musculus*; Fru, *Fugu rubripes*; Msa, *Morone saxitilis*; Xla, *Xenopus laevis*; Amphi, *Branchiostoma floridae.*



B.

Tree	LogL	difference	s.e.	significantly worse
1)ML tree	-1585.68	0.00		best tree
2)Independent Duplication	-1628.02	42.34	12.49	YES



The deuterostome genes are not drifting freely, however, even in the invertebrate deuterostomes. *CiHbox3* and *StyelaAhox2* appear to be orthologous Posterior genes (>99% bootstrap value, Fig. 4) from the two major groups of ascidians. *CiHbox3* is from *Ciona intestinalis*, an Enterogonid, while *Ahox2* is from *Styela clava*, a Pleurogonid (Di Gregorio et al. 1995; Ge et al. 1994). rDNA data suggests that the Enterogonid/Pleurogonid split occurred at a similar time to the lineage split leading to *Xenopus* and mammals (Wada 1998), which occurred about 350 million years ago (Paton et al. 1999). The conservation of a gene over such a length of time indicates that it is evolutionarily constrained. Thus, deuterostome Posterior genes can be resolved within subphyla (Urochordata, Cephalochordata, and Vertebrata), but not between these taxonomic groups. Protostome Posterior Hox genes resolve at a much deeper level, between phyla. The cause of deuterostome Posterior Flexibility is not clear, but it does correlate with deuterostomes having more Posterior class genes than protostomes (provided all, or most, types of protostome Posterior Hox gene have been found). Perhaps the constraining selective pressures on the homeoboxes have been diluted among the deuterostome genes relative to the protostome genes, or the individual deuterostome proteins are interacting with fewer co-factors and target genes than the protostome Posterior proteins.

Another phenomenon concerning a pattern of divergent evolution of Posterior Hox genes has previously been defined: *Laxitas terminalis* (van der Hoeven et al. 1996). This describes the increasing levels of divergence between members of the vertebrate paralogy groups in progressively more 5′/posterior positions; mouse and zebrafish *HoxD10* are almost identical in their homeodomains, while mouse and zebrafish *HoxD13* are only 78.3% identical. The proposed reason for this evolutinary pattern is that many of the functions of the higher vertebrate Posterior Hox genes (e.g., autopod patterning, vertebral sacro-lumbar transition positioning, and penile bone patterning) are not linked to basal vertebrate features. Increased functional variability is also tolerated in the posterior and terminal-patterning 5′ Hox genes, which is reflected in greater sequence divergence. The Posterior Flexibility that we describe here is a distinct phenomenon. Posterior Flexibility describes the behavior of the Posterior Hox genes as a whole, which are evolving differently from the anterior and middle Hox groups. Posterior Flexibility encompasses a much greater level of sequence divergence than *Laxitas Terminalis*, and describes sequence evolution between phyla and subphyla, rather than within the vertebrate subphylum.

CONCLUSION

We performed a genomic walk of 125 kb beyond *AmphiHox10*, at the 5′ end of the Amphioxus Hox gene cluster, and isolated four new homeobox genes. Amphioxus thus has a cluster with a minimum of 14 Hox genes, making it the most Hox gene-rich cluster of any animal so far characterized. Phylogenetic analysis with these new genes reveals deuterostome Posterior Flexibility, whereby deuterostome Posterior class Hox genes are evolving at a faster rate than the anterior (3′) Hox genes or the Posterior Hox genes of protostomes. We favor the scenario in which the vertebrate ancestor had a cluster with at least 13 Hox genes, with *AmphiHox9–13* being pro-orthologues of the vertebrate paralogy groups 9–13, with their relationships having been obscured by Posterior Flexibility. A *Hox14* gene has not been isolated from any other chordate to date. We look forward to the resolution of whether any vertebrate or other deuterostome has a fourteenth Hox gene (Fig. 7).

Acknowledgements

We thank Simon Patton and Carola Burgthof for cosmid library MPMGc117. We thank colleagues from the Departament de Genètica and the Holland lab for discussions, and two anonymous referees for their helpful suggestions. This work was supported by grants from DGESIC to J. G. F. (Ministerio de Educación y Cultura PB95-0579 and PB98-1261-C02-02) and from the BBSRC to P. W. H. H. Collaboration between J. G. F. and P. W. H. H. was facilitated by grants from Acciones Integradas of the British Council/Ministerio de Educación y Cultura. D. E. K. F. held an EC Marie Curie TMR postdoctoral fellowship. C. M. holds a CIRIT (Generalitat de Catalunya) fellowship. We are indebted to the Serveis Científico-Tècnics of the Universitat de Barcelona and to Reading University's AMS sequencing service for automatic sequencing.

Fig. 5. Maximum likelihood analysis of Posterior AmphiHox proteins with the vertebrate paralogy groups. (A) All six Posterior *AmphiHox* proteins compared to the vertebrate groups (circled and shaded). Numbers on the internodes are Puzzle values. The parameters used in the Puzzle maximum likelihood analysis were tree reconstruction by Neighbour-joining and Quartet puzzling, with the Dayhoff substitution model, and 1 invariable and 8 gamma rates calculated from the data, with 1000 replicates. (B) Maximum likelihood user tree comparison of the tree shown in (A) with the independent duplication model, represented by a tree with all of the *AmphiHox* proteins drawn together into a single clade (not shown). The independent duplication model is significantly worse than the best maximum likelihood tree. (C–F) Individual Posterior AmphiHox proteins relative to the vertebrate paralogy groups. (C) AmphiHox13, (D) AmphiHox12, (E) AmphiHox11, (F) AmphiHox10. AmphiHox14 is not shown because it groups with AmphiHox13, and AmphiHox9 is not shown as it resolves as the sister to vertebrate group 9, as expected for a por-orthologue (discussed in text).

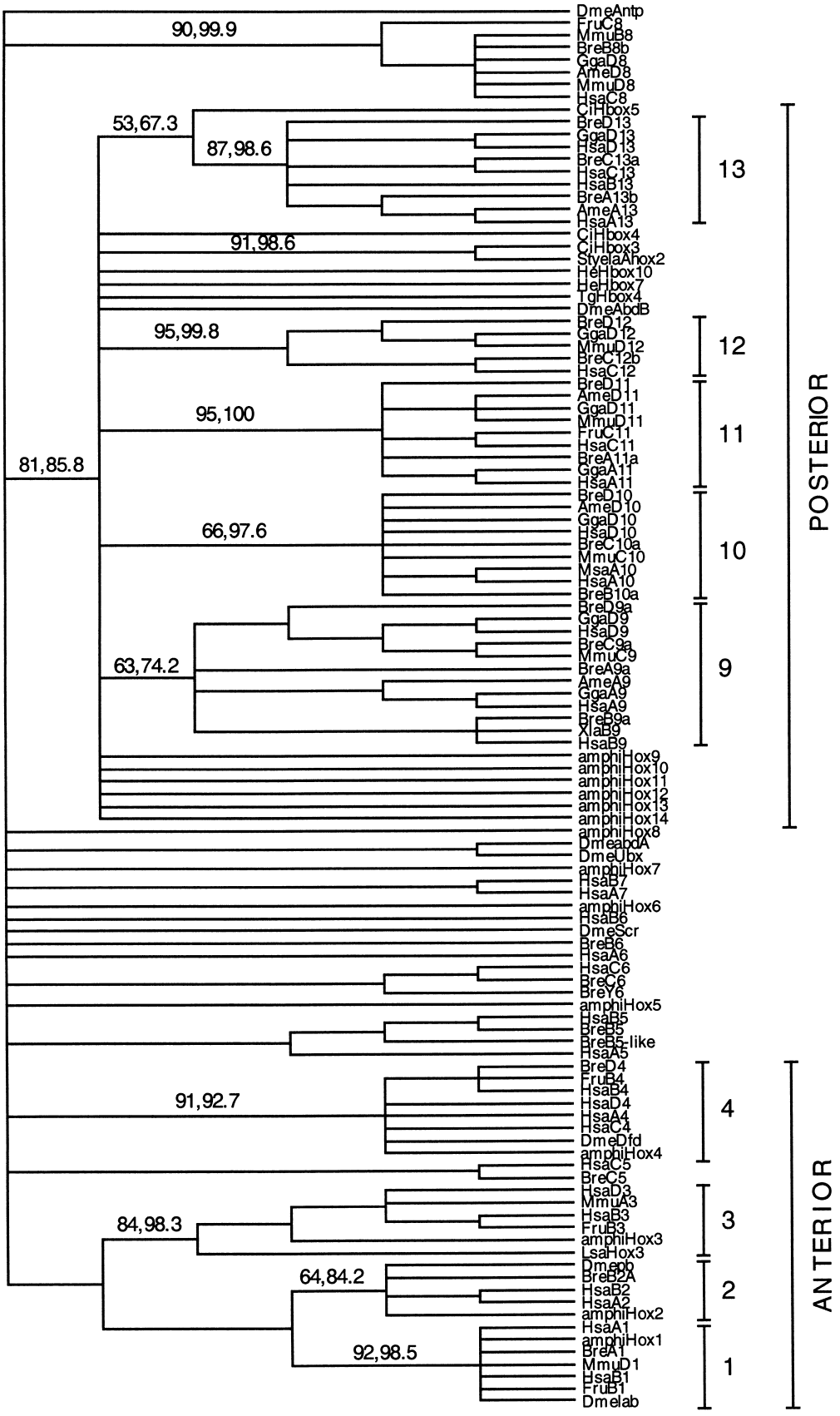


Fig. 6. A phylogenetic tree of selected Hox proteins. Included proteins are from vertebrate paralogy groups 1–13, *AmphiHox1–14*, all *Drosophila* Hox proteins, with the nemertine group 3 protein (*LsaHox3*), and ascidian and sea urchin Posterior class proteins. Abbreviations of species names are given in Fig. legend 4. The numbers on the branches are parsimony and Neighbour-joining bootstrap values, respectively, using the parameters described in Fig. legend 4. Numbers on the vertical bars represent the Hox orthology group.

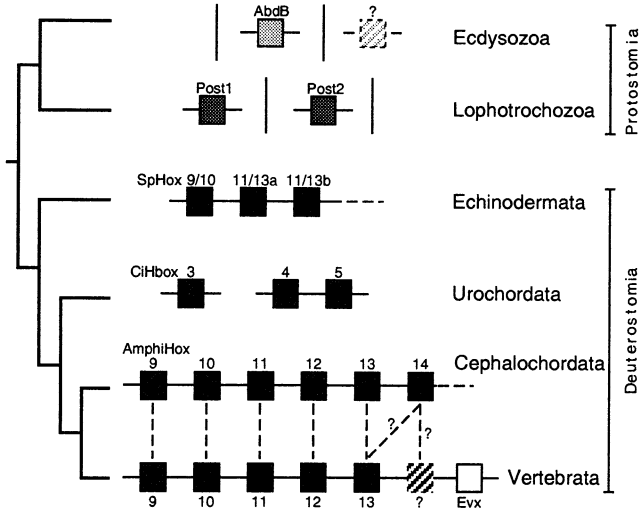


Fig. 7. Phylogenetic schematic of the organization of Bilaterian Posterior Hox genes. The three protostome gene groups (*AbdB*, *Post1,* and *Post2*) are distinct from each other, and from the deuterostome genes. The Ecdysozoa may have a second Posterior gene (de Rosa et al. 1999) (hatched square). Continuous lines between genes denote established physical linkage (see Martinez et al. 1999 for echinoderm linkage). Dotted lines between Cephalochordata and Vertebrata genes represent putative orthology, which has been obscured by relatively rapid rates of sequence evolution (Posterior Flexibility). It is unclear whether *AmphiHox14* is an amphioxus-specific tandem duplication from *AmphiHox13*, or represents an ancestral condition with vertebrate group 14 genes having been lost or not yet found.

REFERENCES

Aguinaldo, A. M. A., et al. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387: 489–493.

Acampora, D., et al. 1989. The human HOX gene family. *Nucleic Acids Res.* 17: 10385–10402.

Amores, A., et al. 1998. Zebrafish *hox* clusters and vertebrate genome evolution. *Science* 282: 1711–1714.

Aparicio, S., et al. 1997. Organization of the *Fugu rubripes Hox* clusters: evidence for continuing evolution of vertebrate *Hox* complexes. *Nature Genetics* 16: 79–83.

Averof, M., and Patel, N. H. 1997. Crustacean appendage evolution associated with changes in Hox gene expression. *Nature* 388: 682–686.

Bürglin, T. R. 1994. A Comprehensive Classification of Homeobox Genes. In D. Duboule (ed.). *Guidebook to the Homeobox Genes*. Oxford University Press, Oxford, pp. 25–71.

Burke, A. C., Nelson, C. E., Morgan, B. A., and Tabin, C. 1995. *Hox* genes and the evolution of vertebrate axial morphology. *Development* 121: 333–346.

Cohn, M. J., and Tickle, C. 1999. Developmental basis of limblessness and axial patterning in snakes. *Nature* 399: 474–479.

de Rosa, R., et al. 1999. Hox genes in brachiopods and priapulids and protostome evolution. *Nature* 399: 772–776.

Di Gregorio, A., et al. 1995. Cloning of ascidian homeobox genes provides evidence for a primordial chordate cluster. *Gene* 156: 253–257.

Duboule, D. (ed.), 1994. *Guidebook to the Homeobox Genes*. Oxford University Press, Oxford.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by author. Department of Genetics, University of Washington, Seattle.

Garcia-Fernàndez, J., Baguñà, J., and Saló, E. 1993. Genomic organization and expression of the planarian homeobox genes *Dth-1* and *Dth-2*. *Development* 118: 241–253.

Garcia-Fernàndez, J., and Holland, P. W. H. 1994. Archetypal organization of the amphioxus *Hox* gene cluster. *Nature* 370: 563–566.

Gaunt, S. J. 1994. Conservation in the Hox code during morphological evolution. *Int. J. Dev. Biol.* 38:549–552.

Ge, T., Lee, H., and Tomlinson, C. R. 1994. Identification of an antennapedia-like homeobox gene in the ascidians *Styela clava* and *S.plicata*. *Gene* 147: 219–222.

Holland, P. W. H., Garcia-Fernàndez, J., Williams, N. A., and Sidow, A. 1994. Gene duplications and the origins of vertebrate development. *Development Suppl.* 125–133.

Izpisúa-Belmonte, J., Falkenstein, H., Dollé, P., Renucci, A. and Duboule, D. 1991. Murine genes related to the *Drosophila AbdB* homeotic gene are sequentially expressed during development of the posterior part of the body. *EMBO J.* 10: 2279–2289.

Lehrach, H., et al. 1990. Hybridisation fingerprinting in genome mapping and sequencing. In K. E. Davis and S. M. Tilghman (eds.). *Genome Analysis 1: Genetic and Physical Mapping*. Cold Spring Harbor Press, New York, pp. 39–81.

Martinez, P., Rast, J. P., Arenas-Mena, C., and Davidson, E. H. 1999. Organization of an echinoderm *Hox* gene cluster. *Proc. Natl. Acad. Sci. USA* 96: 1469–1474.

Paton, R. L., Smithson, T. R., and Clack, J. A. 1999. An amniote-like skeleton from the Early Carboniferous of Scotland. *Nature* 398: 508–513.

Philippe, H., Chenuil, A., and Adoutte, A. 1994. Can the Cambrian explosion be inferred through molecular phylogeny? *Development Suppl.* 15–25.

Schubert, F. R., Nieselt-Struwe, K., and Gruss, P. 1993. The Antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution. *Proc. Natl. Acad. Sci. USA* 90: 143–147.

Scott, M. P. 1992. Vertebrate homeobox gene nomenclature. *Cell* 71: 551–553.

Sharman, A. C., and Holland, P. W. H. 1998. Estimation of *Hox* gene cluster number in lampreys. *Int. J. Dev. Biol.* 42: 617–620.

Sharman, A. C. 1999. Some new terms for duplicated genes. *Semin. Cell. Dev. Biol.* 10: 561–563.

Slack, J. M. W., Holland, P. W. H., and Graham, C. F. 1993. The zootype and the phylotypic stage. *Nature* 361: 490–492.

Sordino, P., van der Hoeven, F., and Duboule, D. 1995. *Hox* gene expression in teleost fins and the origin of vertebrate digits. *Nature* 375: 678–681.

Swofford, D. L. 1998. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. (Sinauer Associates, Sunderland, Massachusetts).

van der Hoeven, F., Sordino, P., Fraudeau N., Izpisúa-Belmonte, J. C., and Duboule, D. 1996. Teleost HoxD and HoxA genes: comparison with tetrapods and functional evolution of the HOXD complex. *Mechanisms of Development* 54: 9–21.

Wada, H. 1998. Evolutionary history of free-swimming and sessile lifestyles in urochordates as deduced from 18S rDNA molecular phylogeny. *Mol. Biol. Evol.* 15: 1189–1194.

Warren, R. W., Nagy, L., Selegue, J., Gates, J., and Carroll, S. 1994. Evolution of homeotic gene regulation and function in flies and butterflies. *Nature* 372: 458–461.

Zeltser, L., Desplan, C., and Heintz, N. 1996. *Hoxb-13*: a new Hox gene in a distant region of the HOXB cluster maintains colinearity. *Development* 122: 2475–2484.