EVOLUTION & DEVELOPMENT 6:3, 154–163 (2004)

The Hox gene complement of acoel flatworms, a basal bilaterian clade

Charles E. Cook,[a,*,1] Eva Jiménez,[b] Michael Akam,[a] and Emili Saló[b]

[a]University Museum of Zoology, Department of Zoology, Downing Street, Cambridge CB2 3EJ, UK
[b]Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 645, 08028-Barcelona, Spain
*Author for correspondence (email: ccook@phillips.exeter.edu)
[1]Present address: British Antarctic Survery, High Cross, Madingley Road, Cambridge CB3 0ET, UK.

SUMMARY Several molecular data sets suggest that acoelomorph flatworms are not members of the phylum Platyhelminthes but form a separate branch of the Metazoa that diverged from all other bilaterian animals before the separation of protostomes and deuterostomes. Here we examine the Hox gene complement of the acoel flatworms. In two distantly related acoel taxa, we identify only three distinct classes of Hox gene: an anterior gene, a posterior gene, and a central class gene most similar to genes of Hox classes 4 and 5 in other Bilateria. Phylogenetic analysis of these genes, together with the acoel *caudal* homologue, supports the basal position of the acoels. The similar gene sets found in two distantly related acoels suggest that this reduced gene complement may be ancestral in the acoels and that the acoels may have diverged from other bilaterians before elaboration of the 8- to 10-gene Hox cluster that characterizes most bilaterians.

INTRODUCTION

The Hox genes encode a family of transcription factors that regulate anterior/posterior patterning in many bilaterian animal phyla. Most bilaterians possess a Hox cluster comprising genes of at least eight distinct types, or paralogy groups (PGs), that are shared between the most distantly related lineages, implying that the last common ancestor of all these phyla possessed a cluster of at least similar complexity (deRosa et al. 1999). Typically, these genes are clustered, and genes at one end of this cluster (PGs 1–3) specify anterior structures, whereas those at the other (PGs 9–13) specify posterior structures. The principal exceptions to this pattern are the nematodes, some of which possess a much reduced and scrambled Hox gene set. However, it is now clear that this is in large part the result of gene loss (Aboobaker and Blaxter 2003).

Because genes of the Hox cluster play such an important role in patterning the bilaterian body plan, the origins of the Hox gene cluster have attracted considerable attention. Homeobox genes of the Hox-related superclass (the "Antp superclass" of Holland; Holland 2001) probably arose during the diversification of the protistan eukaryotes and are certainly present in sponges (Seimya et al. 1994), which constitute the basal branches of the Metazoa. They may have arisen as a single supercluster, including the NK class, *evx*, and other non-Hox homeobox genes (Holland 2001; Minguillón and Garcia-Fernàndez 2003). However, Hox genes proper (the "Hox class" of Holland [Holland 2001], including both Hox and ParaHox genes) appear to be restricted to bilaterians and diploblasts (cnidarians and ctenophores). They have not been identified in sponges, despite extensive searches (and claims to the contrary that have proven erroneous; Ferrier and Holland 2001).

Extensive sampling of Hox genes from cnidarians has revealed a restricted Hox gene set compared with most bilaterian phyla. Clear orthologues are present for anterior (PG1) and posterior (PGs 9–13) Hox genes (Finnerty and Martindale 1999; Gauchat et al. 2000; Ferrier and Holland 2001). In at least one case these Hox genes are linked to an even-skipped class homeobox gene, as they are in some bilaterians. Sequences less clearly related to central class Hox genes are also present. Thus, a ProtoHox cluster containing at least two and possibly more Hox genes must predate the split of cnidarians and bilaterians.

This ProtoHox cluster underwent an early duplication to give rise to two paralogous clusters that have survived in the genomes of living animals: the definitive Hox cluster and a sister cluster, the ParaHox cluster, which contains genes with sequences related to the anterior (*Gsx*) and posterior (caudal/*Cdx*) Hox genes and, in bilaterians, PG3 (*Xlox*) (Brooke et al. 1998; Finnerty et al. 2003). Kourakis and Martindale (2000) and Ferrier and Holland (2001) proposed that the duplication that generated the Hox and ParaHox clusters took place before the Cnidaria/Bilateria split. This implies that Hox genes of class 3 and ParaHox genes of the Xlox class also arose before this split, though these genes have not yet been identified in any cnidarian.

It is not yet clear when the Hox genes acquired a role in axial patterning. Because aspects of both expression and function are conserved between deuterostomes and protostomes, this role clearly predates the radiation of the major bilaterian lineages (Akam 1989). Expression data from Cnidaria are open to a variety of interpretations, suggesting either a role in cell type specification or a role in body patterning (Gauchat et al. 2000; Yanze et al. 2001; Finnerty et al. 2003; Hill et al. 2003). Parallels in the role of corresponding Hox and ParaHox genes suggest an involvement in axial patterning that predates the duplication of these two clusters (Brooke et al. 1998; Finnerty et al. 2003). This is clearest for the posterior Hox and caudal genes, which both play widely conserved roles in posterior patterning in bilaterians.

Until recently, there appeared to be no living animal group that might allow us to examine the Hox cluster of a bilaterian lineage that arose after the divergence of bilaterians from diploblasts but before the origination of the three great bilaterian clades defined by molecular phylogenetic analysis: the ecdysozoans and the lophotrochozoans, which together comprise the protostomes, and the deuterostomes. 18S ribosomal RNA data, supported by a range of other analyses, suggested that all bilaterians traditionally placed basally within the tree (i.e., the acoelomate and pseudocoelomate phyla) are included within one of these three clades (Adoutte et al. 2000; Peterson and Eernisse 2001).

However, several molecular data sets have now identified the acoel and nemertodermatid flatworms (Acoelomorpha) as basal bilaterians, either as a single clade or two deeply branching lineages. This was first suggested for acoels on the basis of 18S ribosomal gene sequences (Ruiz-Trillo et al. 1999) and has now been confirmed with data for the myosin heavy chain II gene (Ruiz-Trillo et al. 2002), 28S RNA sequences (Telford et al. 2003), and with analyses of myosin, histone H3, elongation factor 1-alpha, and 18S RNA data sets (Giribet 2003). Support for the separation of the Acoela from the other flatworms was also provided by neuroanatomy (Reuter et al. 2001; Gustafsson et al. 2002) and mitochondrial codon usage (Telford et al. 2000). Contradictory data for nemertodermatids have been recognized as resulting from the incorrect attribution of specimens from which sequences were obtained (Jondelius et al. 2002).

The acoels have long been recognized as a particularly simple group among the platyhelminths. For Libby Hyman they epitomized basal bilaterians (Hyman 1951) and served in this respect in zoology textbooks for decades. Although it is now thought that they are in some respects secondarily simplified (e.g., the absence of an epithelial gut), they share with nemertodermatids a number of synapomorphies that may well be primitive, including the lack of a centralized nervous system and lack of nephridiocytes. Recent detailed morphological studies of central nervous system and muscle organization have failed to identify synapomorphies that link them to the platyhelminths (Reuter et al. 2001; Gschwentner et al. 2003). They exhibit a unique form of duet spiral cleavage (Henry et al. 2000) that sets them apart from all other bilaterians and share uniquely with nemertodermatids details of epidermal ciliary structure (Ehlers 1985, 1986; Smith et al. 1986).

Here we show that the Hox gene set of acoel flatworms is distinct from that of other bilaterians. Extensive screens using two distantly related acoel species have recovered representatives of only three distinct PGs. We suggest that this represents a primitively simple Hox cluster and that acoel flatworms derived from the common protostome/deuterostome ancestor before the elaboration of an eight-gene Hox cluster.

METHODS

Symsagittifera roscoffensis (Graf 1891) (formerly *Convoluta roscoffensis*) is a small (5 mm) acoel flatworm found intertidally on beaches in Europe. Adult individuals do not feed but rely on metabolites produced photosynthetically by a symbiotic alga, usually *Platymonas convolutae* or *Tetraselmis tethratele* (Keeble 1912; Douglas 1983). We collected live *S. roscoffensis* from a disjunct population on a cobble beach near East Aberthaw, Wales (grid reference, ST045685) (Mettam 1979) and maintained them in a laboratory incubator at 8–12°C with 16 h of darkness and 8 h of light. We also collected *S. roscoffensis* from the type location at Roscoff, France that were frozen or processed for DNA or RNA extraction immediately.

DNA and RNA were extracted from *S. roscoffensis* using commercial kits (Qiagen-tip 20 cat. no. 10023, Qiagen, Crawley, UK). Short homeobox fragments were amplified from genomic DNA and by reverse transcriptase polymerase chain reaction (PCR) on RNA (extracted from whole organisms) using 10 different degenerate primer sets as previously described and then cloned (deRosa et al. 1999; Cook et al. 2001). Sequences of over 150 homeobox-containing clones from the Welsh population yielded only two different homeobox-containing gene fragments. Fifty clones from the Roscoff population yielded a slightly more diverse population of fragments but no additional clearly distinct genes. We used BLASTP searches (http://www.ncbi.nlm.nih.gov/blast/) and then manual alignments with putative homologues to identify these genes. We assigned one gene, *SrHox1*, to paralogue group 1. One gene, *SrHox4/5*, showed similarities to PG4 and PG5 but could not be assigned unambiguously to either. Short fragments of a posterior and a caudal gene were amplified using gene-specific degenerate primers and designated *SrPost* and *SrCdx*. Sequences were extended using inverse PCR (deRosa et al. 1999) on DNA extracted from the Welsh population, and for all four genes we recovered at least the 180 nucleotides of the homeodomain. Sequences were submitted to GenBank with accession numbers AY117546–AY117550.

Paratomella rubra Rieger and Ott 1971 is a very small (0.5–1.5 mm) acoel flatworm found on nontidal sandy beaches along the Mediterranean coast. This species does not have a photosynthetic symbiont. We collected live *P. rubra* from a sandy beach at Sitges,

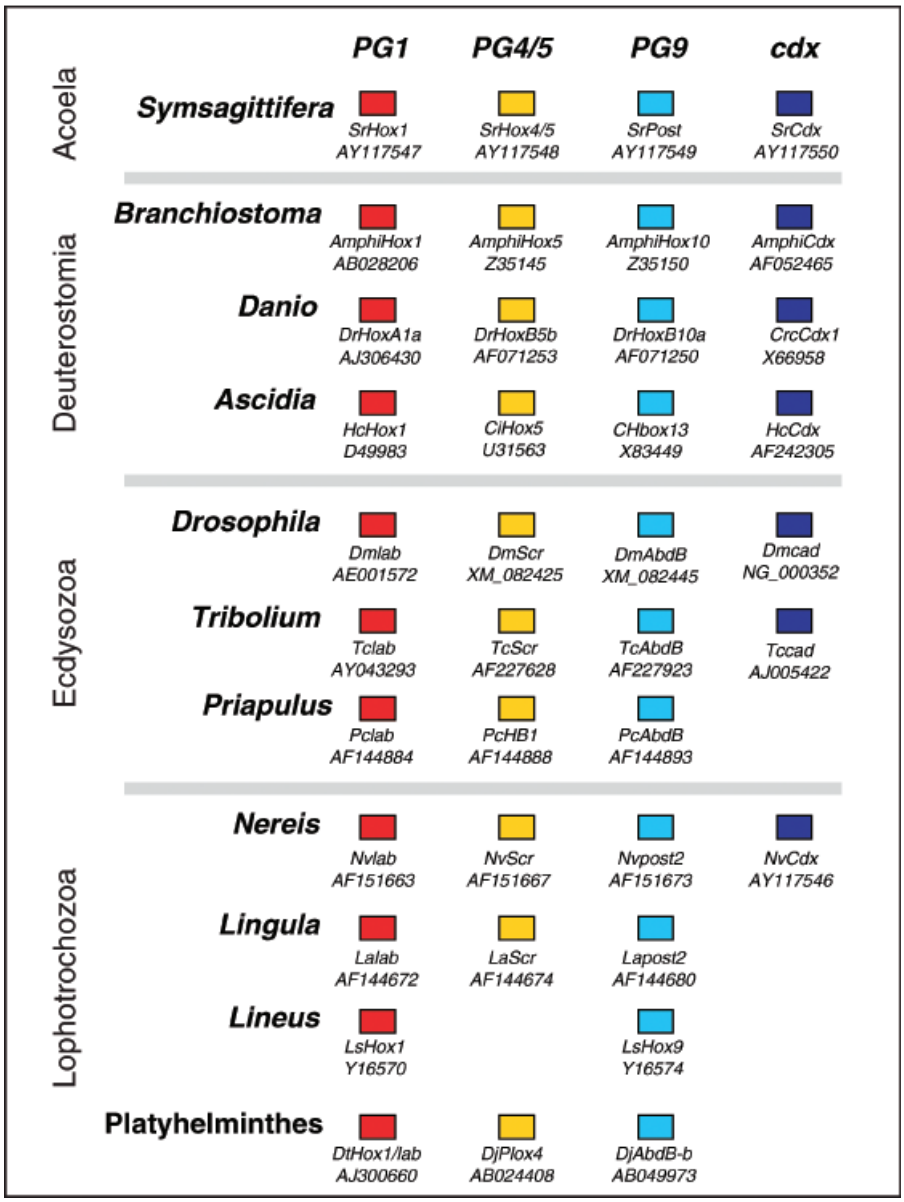

Fig. 1. Cartoon showing genes used for phylogenetic analysis. Gaps indicate genes not available for that taxon and encoded as missing for the phylogenetic analysis. Taxon names refer to genera except Ascidia and Platyhelminthes, for which genes from two species were used. Gene names and GenBank accession numbers are given below cartoon of each gene. The first few letters of each gene name abbreviates each taxon as follows: Amphi, *Branchiostoma floridae*; Dr, *Danio rerio*; Hr, *Herdmania curvata*; Ci, *Ciona intestinalis*; Dm, *Drosophila melanogaster*; Tc, *Tribolium casteneum*; Pc, *Priapulus caudatus*; Nv, *Nereis virens*; La, *Lingula anatina*; Ls, *Lineus sanguineus*; Dt, *Discocelis tigrinia*; Dj, *Dugesia japonica*; Sr, *Symsagittifera roscoffensis*.

Catalunya, Spain. The animals were frozen at -70°C or processed (within 48 h) for RNA extraction upon return to the laboratory. RNA was extracted using Trizol reagent (Life Technologies, Gaithersburg, MD, USA) and retrotranscribed with SuperscriptIII (Invitrogen, Paisley, UK). Short homeobox fragments were amplified from complementary DNA using nine different degenerate primer combinations (Cook et al. 2001) and then cloned. Sequences of 70 homeobox-containing clones yielded fragments of only four genes. These were identified as above and by comparison to the *S. roscoffensis* homeobox sequences as a paralogue group 1 gene (*PrHox1*), two posterior genes (*PrPostA* and *PrPostB*), and a caudal gene (*PrCdx*). Sequences were been submitted to GenBank with accession numbers AY282605–AY282608.

For phylogenetic analysis, we compared the sequences of the three *S. roscoffensis* Hox genes and caudal with genes of the corresponding PGs in diverse representatives of other bilaterian clades. Because the precise correspondence between *SrHox4/5* and other bilaterian genes is uncertain, we assembled alternative data sets that included either PG4 or PG5 genes from other taxa. Similarly, we assembled multiple data sets that used either *Post1* or *Post2* genes as the posterior Hox representative from lophotrochozoan taxa and only one of the PG9–PG14 homologues as the posterior Hox representative from chordates. Complete homeodomain sequences for all four genes were not available for some lophotrochozoan taxa or for the ecdysozoan *Priapulus caudatus*, so some residues were encoded as missing. The second acoel, *P. rubra*, was not included in the phylogenetic analysis because complete homeodomain sequences are not available for this species.

We used PAML and MrBayes to identify the best available Markov process model of amino acid sequence evolution (Yang

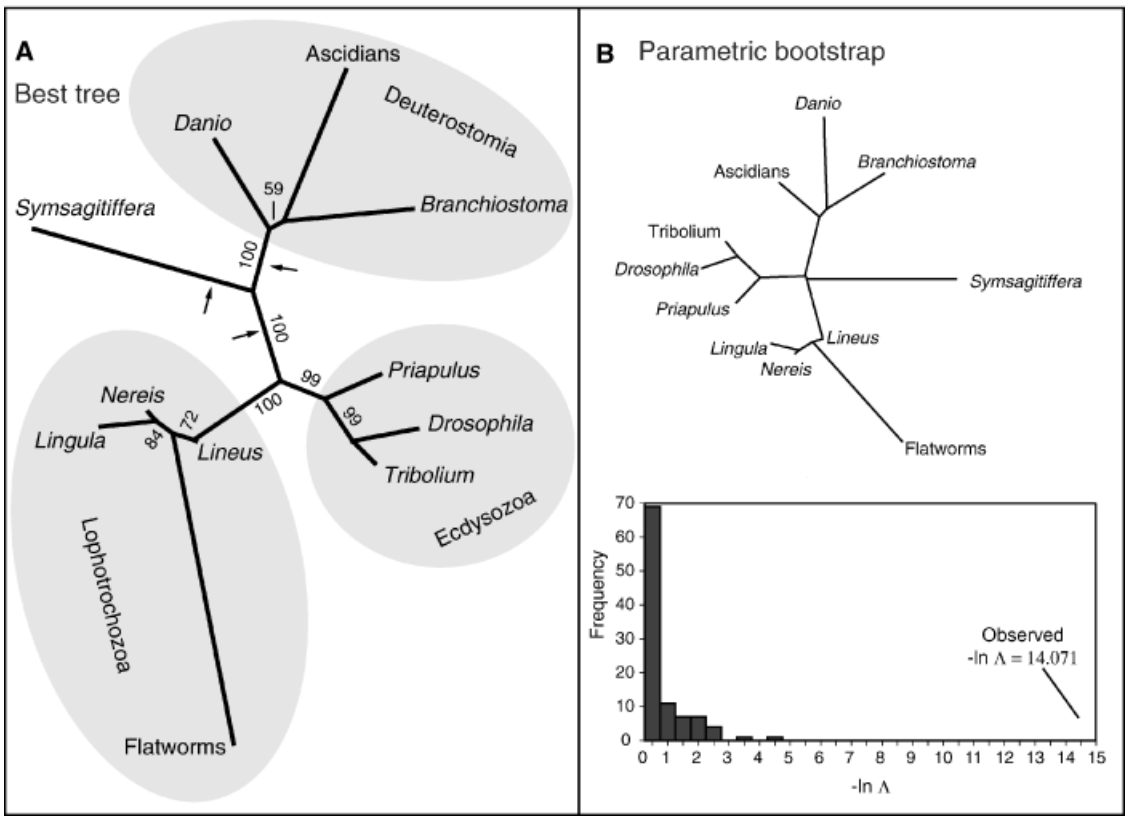

Fig. 2. Phylogenetic analysis of concatenated homeobox amino acid sequences from four genes and eleven taxa. Taxa are identified by genus. For full species names and sequence sources see Figure 1. (A) Unrooted maximum likelihood tree generated using PAML. The best tree identified by MrBayes was topologically identical but with slightly different branch lengths. Numbers next to branches indicate percentage of post-burnin trees in the MrBayes analysis that contained that branch. This tree strongly supports placement of the acoel flatworms outside of the protostomes. Three plausible positions for a root are marked with arrows. (B) Parametric bootstrap analysis: We identified the best tree conforming to the null hypothesis that *Symsagittifera roscoffensis* is within the Lophotrochozoa (for details see supporting material). We then used this tree to generate artificial data sets for parametric bootstrapping. For each data set the ratio between the likelihoods of the best tree and the best tree conforming to the null hypothesis was calculated. These values were binned and plotted (inset graph in B). All the values of this test statistic were less than the value of this same statistic for the original data set, so the null hypothesis is rejected. *Symsagittifera roscoffensis* is therefore not included within the Lophotrochozoa, and joins the bilaterian lineage outside of the Ecdysozoa+Lophotrochozoa clade.

1997; Huelsenbeck and Ronquist 2001) for use with each data set as previously described (Cook et al. 2001). We compared different models using a likelihood ratio test (Huelsenbeck and Crandall 1997; Cook et al. 2001) and chose for all data sets a model that estimated a gamma shape parameter with four categories and used the empirical+F amino acid substitution matrix where F was the matrix of Whelan and Goldman (2001). More complex models with more than four rate categories, or that treated individual genes separately, were not significantly better than this model.

We then used this model to estimate phylogenies in all data sets by quartet puzzling using Tree-Puzzle (Strimmer and von Haeseler 1997; Huelsenbeck and Ronquist 2001). Results for all data sets were topologically identical, and we chose for more extensive analysis the data set represented in Figure 1 because it had the fewest amino acid residues encoded as missing.

We then used PAML and MrBayes to analyze this data set more extensively. We first ran the data set in MrBayes for 100,000 generations using four chains. A graph of the likelihood values of every 100th tree showed that the likelihood values had peaked after 10,000 generations. We doubled this and used 20,000 generations as the "burnin." Every 50th tree for the 80,000 post-burnin generations was stored and these 1600 trees were used to make a consensus tree in which the frequency of occurrence of each clade gives some indication of support for that clade.

We then used PAUP (Swofford 1998) to generate a set of 2835 trees in which the three ecdysozoan and three deuterostome taxa were constrained to monophyly. These constraints prevented testing of unlikely relationships but did test all possible positions of *S. roscoffensis* within the Lophotrochozoa. The tree identified by PAML as having the highest likelihood was topologically identical to the tree with the highest likelihood from among all the trees generated by MrBayes (Fig. 2A).

We tested the possibility that *S. roscoffensis* is within the Lophotrochozoa using a parametric bootstrapping test. We

estimated, using PAML, the maximum likelihood value for each of a set of 315 trees that conformed to the null hypothesis constraint that *S. roscoffensis* is within the lophotrochozoan clade (Huelsenbeck and Crandall 1997; Cook et al. 2001). The ratio between the maximum likelihood value for the best overall tree and for the best tree under the null hypothesis was calculated and used as a test statistic. We used PAML and the parameters of the best tree under the null hypothesis (alpha with four rates and amino acid frequencies) to generate 100 artificial data sets. For each artificial data set the best overall tree and the best tree

```
                          |                          homeobox                          |
PG       DmAntp      QE  RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN  KTKGEP

1        SrHox1      VQ  TRG..TNF.NK.LT..............A......TS.T.N.T.V..........Q..LL  .EGKLS  AY117547
   Loph  ChHox1      GQ  PNL..TNF.NK.LT.........K....A......A..G.N.T.V..........Q..RM  .EG.VV  AF163856
   Loph  DtHox1lab   AV  KGN..TNF.NK.LT......Y..K....A......N..G.N.T.V..........Q..RM  EKPSST  AJ300660
   Loph  Pnox3       NN  NIT..TNF.NK.LT.........Q....A......KSMT.S.T............Q.RRQ  .ARKIE  L41848
   Loph  HtLox7      SS  NNL..TNF.NK.LT.........K....A......ST.G.N.T.V..........H..RL  .ESNNV  Y10888
   Loph  Nvlab       GQ  PNM..TNF.NK.LT.........K....A......A..G.N.T.V..........Q..RM  .ETNVS  AF151663
   Ecd   Tclab       NC  LNT..TNF.NK.LT.........K....A......S..Q.N.T.V..........Q..RM  .EGLI.  AY043293
   Ecd   Dmlab       NT  NNS..TNF.NK.LT..............A......NT.Q.N.T.V..........Q..RV  .EGLI.  AE001572
   Ecd   Cslab       GS  NGS..TNF.TK.LT.......Y.K....A......T..Q.N.T.V..........Q..RM  .EGLLV  AJ00743
   Ecd   LsHox1      GQ  PNT..TNF.NK.LT.........K....A......A..G.N.T.V..........Q..RM  .EGLVQ  Y16570
   Deut  Amhox1      SQ  QNSI.TNF.TK.LS.........K....A..V...AT.E.N.T.V..........Q..RE  .EGLA.  X15575
   Deut  BfHox1      SG  PNN..TNF.TK.LT.......Y.K....A..V...A..N.N.T.V..........Q..RE  .EN.FS  Z35142
   Deut  PmHox1w     GG  IATH.TNFSTK.LT.........K....A..V...A..Q.N.T.V..........Q..RE  .EGHSL  AF434665
   Deut  HfHoxA1     GQ  PNTV.TNF.TK.LT.........K....A..V...A..Q.N.T.V..........Q..RE  .EGLTS  AF224262
   Deut  DrHoxA1a    GQ  PNTV.TNFSTK.LT.........K....A..V...AS.Q.N.T.V..........Q..RE  .EGLL.  AJ306430
   Deut  DrHoxB1a    GP  QNTI.TNF.TK.LT........SK....A..V...AT.E.N.T.V..........Q..RE  .EGLA.  AF071251
   Deut  DrHoxC1a    TS  SGGS.TNF.TK.LT.........K....A......NP.Q.S.T.V..........Q..ML  REGLAQ  AJ306432
   Deut  HsA1        GQ  PNAV.TNF.TK.LT.........K....A..V...AS.Q.N.T.V..........Q..RE  .EGLL.  XM_088185
   Deut  HsB1        GS  PSGL.TNF.TR.LT.........K..S.A..V...AT.E.N.T.V..........Q..RE  REG.RV  NM_002144
   Deut  HsD1        SP  SSAI.TNFSTK.LT.........K....A......NC.H.NDT.V..........Q..RE  REGLLA  NM_024501

4        SrHox4/5        C..T.TA...F........................NL.A...................D.  NNLK    AY117548
   Loph  NvDfd           S..T.TA...H.V...............................................  KLPN    AF151666
   Loph  HrLox6          D..A.TS.S.................NG........S.G...................D.  RLPN    AF004386
   Loph  HlLox20         N..T.TS...H...............S.........V.N.S.................DH  PNHP    AF006638
   Loph  HmLox1          N..L.TS.S.Q................K...V...QM.K.S...................  SQEN    L07298
   Loph  DjPlox4         F..S..A.S.Q.........YY.Q...........NSV..S..............Y..DV  LSKN    AB024408
   Loph  GtHoxD          Y..C..A.S.Q.........YY.Q...........NSV..S..............Y..DV  VLKT    X95417
   Ecd   CsDfd           P..Q.TA...H.I...........................S.................D.  KLPN    AJ007432
   Ecd   FcDfd           T..Q.TA...H.I........Y..............S.V.S..............Y..D.  KLPN    AF361329
   Ecd   TcDfd           P..Q.TA...H.I........Y..............T.V.S.................D.  K       AH004892
   Deut  BfHox4          T..S.TA...Q.V.......................S.G...................D.  RLPN    AB028208
   Deut  MsHoxA4         P..S.TA...Q.A..................V....TM..S...V..............H  KLPN    AF089743
   Deut  DrHoxD4         P..S.TA...Q.V....................S..T.S.S.................DH  KLPN    Y14548
   Deut  HfHoxA4         P..S.TA...Q.V.......................T...S...V.............DH  KKLP    AF224262
   Deut  HsHoxD4         P..S.TA...Q.V.......................T...S.................DH  KLPN    XM_042818
5  Loph  EscScr          S..S.TS...H..........Y.K..............N....................H  KLSH    AY052756
   Loph  NvScr           S..T.TS...H...........................N....................H  KLGH    AF151667
   Ecd   AfScr           T..Q.TS.............................T...S..................H  KKIA    X70080
   Ecd   AkScr           A..Q.TS....................................................H  KMSS    AF011281
   Ecd   DmScr           T..Q.TS....................................................H  KMAS    XM_082425
   Ecd   FcScr           T..Q.TS....................................................H  KMAT    AF361330
   Deut  BfHox5          N..T.TA.....................................................  KLKS    Z35145
   Deut  CiHox5          S..T.TA..............Y...........V..T.......................  KLSS    AJ002028
   Deut  HeHbox9         S..S.TA...............................G....................H  NVKS    U31563
   Deut  HfHoxA5         G..A.TA...................................................D.  KLKS    AF224262
   Deut  DrHoxb5b        G..A.TA.................................S.................D.  KLKS    AF071253
   Deut  PmHoxw5         G..S.TA..........................V.N....S.................D.  KLKS    AF071234

Cdx      SrCdx       TRT KDKY.VV..DR.RA...N..RSAQ.I.I..KS.L.MQVG.S...V........A.ER.VS  RKV     AY117550
   Loph  NvCdx       TRT KDKY.VV..DH.R........YS..I.I..KA.L.QN.N.S...V........A.ER.Q.  .KR     AY117546
   Loph  DtCdx       IRT CDKY..V..EQ.K.......LTQK.VNA..KS.M.R..Q....V........A.ER.LF  .RS     AJ300663
   Ecd   Agcad       TRT KDKY.VV..DQ.R........YT..I.I..KA.L.QN.Q.S...V........A.DR.QK  .KA     AF119382
   Ecd   Bmcad       TRT KDKY.VV.SDH.R........YS..I.I..KA.L.VS.G.S...V........A.ER.QV  .KR     D16683
   Ecd   Pccad       TRT KDKY.VV.SDH.R........YS..I.I..KS.L.SM.G.S...V...                      AF272155
   Ecd   Tccad       TRT KDKY.VV..DL.RI......T.SK.I.IK.KS.L.EN.G.S............A.ER.Q.  .KR     AJ005422
   Deut  GgCdxA      TRT KDKY.VV..DH.R........YS..I.I..KA.L.A..G....V........A.ER.V.  .K.     AB046532
   Deut  GgCdxB      TRT KEKY.VV..DH.R........C...I.I..KS.L.AN.G.S...V........A.ER.II  .K.     AF353624
   Deut  GgCdxC      TRT KDKY.VV..DH.R........YS..I.I..KA.L.SS.G.S...V........A.ER.I.  .K.     U80614
   Deut  Stcad1      TRT KDKY.VV..DQ.R........YS..I.I..KA.L.AN.G.S...V........A.ER.IT  .KR     AF417197
   Deut  Stcad2      TRT KDKY.VV..DH.R........YS..I.I..KA.L.AS.A.....V........A.ER.V.  .K.     AF417198
   Deut  Stcad3      TRT KEKY.VV..DH.R........YS..I.I..KT.L.AS.R.S...V........A.ER.LF  .K.     AF417199
```

conforming to the null hypothesis were identified. Because it was not possible to exhaustively search the tree space for each data set, we performed a MrBayes analysis on 10 of the data sets using four chains with 40,000 generations and a burnin of 10,000 generations and then calculated a consensus tree using every 10th tree from the final 30,000 generations. This consensus tree was identical for all 10 data sets. We then generated the set of 135 trees that conformed to the constraint set suggested by this consensus tree. We also identified the 40 best trees from the original data set. We then used PAML to identify, for each of the 100 artificial data sets, which of these 175 trees was the best tree and which was the best that conformed to the null hypothesis tree and compared the maximum likelihood values of both trees to generate the test statistics graphed in Figure 2B.

RESULTS

We isolated Hox genes from two species of acoel flatworms, *S. roscoffensis* and *P. rubra*. These represent two widely divergent branches among the acoels (Hooge et al. 2002). For one of these species, *S. roscoffensis*, we used a wide range of degenerate Hox primers on animals collected at two different locations, as well as primers specific for posterior Hox and caudal ParaHox genes. Short homeobox fragments were extended by inverse PCR to obtain the complete homeobox and flanking sequence for each gene, allowing unambiguous identification. In past studies using the same methodology, we recovered all or most of the predicted complement of Hox genes from arthropods and an annelid (deRosa et al. 1999; Cook et al. 2001). However, we only recovered three Hox genes from *S. roscoffensis*, corresponding to PGs 1 (*SrHox1*), 4 or 5 (*SrHox4/5*), and a posterior gene (*SrPost*) (Fig. 3).

Using a different array of degenerate primer sets, short homeobox fragments were also isolated from a sample of the much larger French population of *S. roscoffensis*. Hox sequences recovered were identical to sequences for *SrHox1*, *SrHox4/5*, and *SrPost*. However, an additional central class gene sequence differed from the sequence of the Welsh central class gene at 5 of 84 nucleotides. Only one change was not silent, a leucine to proline at homeobox position 42 in the turn region between helix II and helix III. Both genes share an unusual leucine at homeodomain position 39 and an unusual leucine at position 41. Given this similarity, we believe that this central class gene from the French population is either an allelic variant of the *SrHox4/5* found in the Welsh population or the product of a relatively recent gene duplication. Specific searches for this variant in the Welsh population have been unsuccessful.

The less extensive survey of *P. rubra* recovered short fragments of only the same Hox PGs—a single gene of PG1 (*PrHox1*) and fragments of two distinct posterior genes (*PrPostA* and *PrPostB*). Specific searches with primers designed to amplify Hox and ParaHox group 3 genes were unsuccessful. We also recovered clear orthologues of the caudal ParaHox gene from both species (*SrCdx* and *PrCdx*). These results suggest that acoels have a reduced set of Hox genes, either due to gene loss, such as has been shown in *Caenorhabditis elegans* and other nematodes (Aboobaker and Blaxter 2003), or because acoels descend from an ancestral organism that did not yet have the "full" bilaterian complement of Hox genes.

Of the Hox groups represented by the acoel sequences, only the posterior genes have lineage-specific amino acid signatures. We compared *SrPost*, *PrPostA*, and *PrPostB* to posterior genes from a wide range of other Bilateria (Fig. 4) and found that this gene does not have the signature residues that are associated with ecdysozoan or lophotrochozoan sequences. (There are no signature residues common to all deuterostome posterior Hox genes.)

It is clear from the alignment in Figure 4 that *PrPostA* and *PrPostB* are not orthologues of the lophotrochozoan *Post1* and *Post2* genes. A phylogenetic analysis of posterior genes (not shown) placed the two *P. rubra* posterior genes together in a clade with the *S. roscoffensis* posterior gene. Other duplications of posterior genes have also occurred in the ecdysozoan, lophotrochozoan, and deuterostome lineages, so we are not surprised to observe such a duplication in the acoel lineage (Ferrier et al. 2000).

We used data from the *S. roscoffensis* Hox and *Cdx* genes to further explore the position of the acoels within the Bilateria. Concatenated amino acid sequences of the four genes were compared with those of their homologues in other bilaterian taxa as described above and in the supporting information. We found, with strong statistical support

Fig. 3. Alignment of bilaterian paralogue group 1, paralogue groups 4/5, and caudal-related sequences. *Drosophila melanogaster Antp* used as a reference sequence. All available lophotrochozoan sequences for which complete homeodomain sequences are known are included. Many deuterostome and ecdysozoan sequences are very similar or identical to each other so only representative sequences are shown to save space. *Symsagittifera* sequences (SrHoxl, SrHox 4/5 and CrCdx) are shown at the top of each block of sequences, followed by Lophotrochozoan (Loph), Ecdysozoan (Ecd) and Deuterosome (Deut) sequences of the same paralogy groups. In the case of SrHox 4/5, sequences of both PG4 and PG5 genes are listed. PG1 and caudal homologues are identifiable by signature residues that are present in all bilaterian sequences. PG4 and PG5 sequences are quite similar. Within each of the Lophotrochozoa, Ecdysozoa, and Deuterostomia PG4 and PG5 genes can be distinguished using signature amino acid residues within the homeodomain and other residues in flanking regions (deRosa et al. 1999; Telford 2000), but there are no signature residues that can unambiguously distinguish between PG4 and PG5 genes for all bilaterians. We do not have the flanking sequences for SrHox4/5, and the available sequence is insufficient for positive assignment to either one or the other paralogue group.

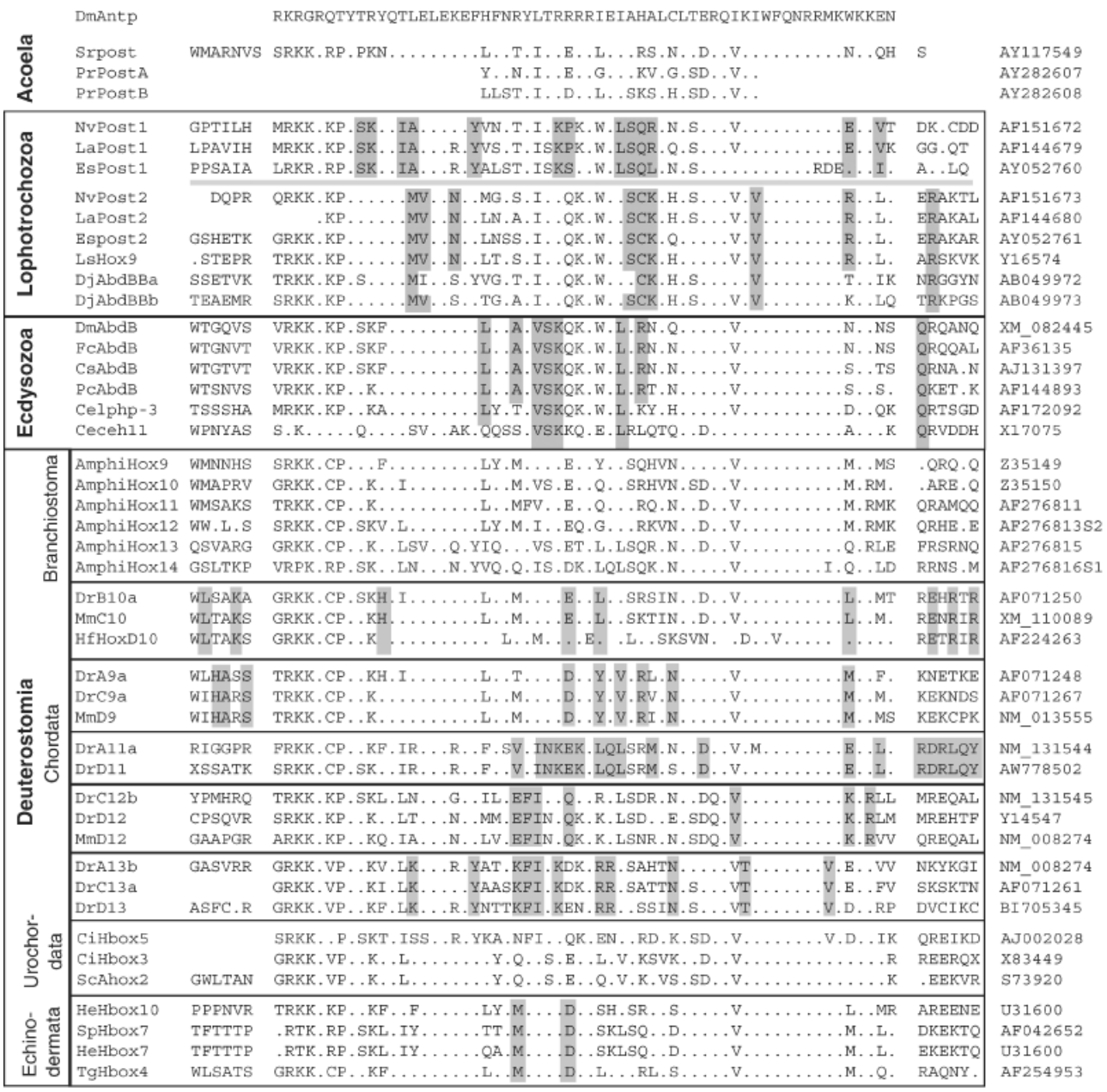

Fig. 4. Comparison of the acoel posterior class genes *SrPost*, *PrPostA*, and *PrPostB* with other bilaterian posterior class Hox gene sequences. The homeodomain (60 residues) and some flanking residues are shown. The *Drosophila melanogaster Antp* sequence is used as a reference. The lophotrochozoan *Post1* and *Post2* sequences and the ecdysozoan *AbdB* sequences include signature residues (highlighted) that are diagnostic for each of those clades. Deuterostome posterior genes are more diverse in sequence, and signature residues are restricted to individual vertebrate paralogue groups (deRosa et al. 1999; Ferrier et al. 2000). The acoel genes contain none of these signatures. Platyhelminthes are represented by the triclad flatworm *Dugesia japonica*, which has two *Post2*-like genes (*DjAbd*) that appear to represent an independent duplication of only that gene. A phylogenetic analysis of bilaterian posterior genes (tree not shown) resolves separate clades for lophotrochozoan, ecdysozoan, acoel, and some vertebrate paralogue groups, but the relationships between these clades are not resolved. Right column shows GenBank accession numbers.

(Fig. 2B), that *S. roscoffensis* is not included within the main branches of any of the three great bilaterian lineages.

We attempted to root the tree with either putative Cnidarian homologues for PG1, posterior, and *Cdx* genes or with ParaHox genes as outgroups for the Hox cluster genes and a Hox posterior gene as the outgroup for *Cdx*. In both cases the branch leading to the outgroup was much longer than any of the ingroup branches. We used MrBayes, as

above, to explore these data sets in more detail. In both cases the position at which the outgroups joined the tree was supported in fewer than 50% of the trees used to make the consensus. We also performed a maximum likelihood distance matrix/neighbor-joining bootstrap analysis, with 500 replicates, on both data sets. In both cases the branch leading to the outgroup received less than 50% support. In addition, the position at which the outgroup joined the tree was different in all four analyses. We therefore conclude that we are unable to root the trees from our analyses reliably using the cnidarian or ParaHox sequences as outgroups. However, if we are to retain the deuterostomes, protostomes, ecdysozoans, and lophotrochozoans as monophyletic groups, the tree can only be rooted at one of the branches indicated by an arrow in Figure 2A. Rooting the tree at one of these points would put the acoels basal in the deuterostomes, basal in the protostomes, or basal in all Bilateria and sister to all other bilaterian taxa. This last hypothesis is supported by molecular phylogenetic analyses of three other genes (Ruiz-Trillo et al. 1999, 2002; Jondelius et al. 2002; Telford et al. 2003).

DISCUSSION

Our data for two acoel species suggest that the Hox cluster of this whole taxon contains fewer than the 8–10 Hox genes characteristic of most bilaterian phyla. The genes we isolated fall into only three PGs recognizable by comparison with other bilaterians—anterior genes closely similar to PG1/*labial*; a central class gene most resembling Hox 4/5 (possibly duplicated); and posterior class genes, most similar to the PG9/10 genes of deuterostomes. These posterior genes are clearly distinct from the *Abd-B* genes of Ecdysozoa and the *Post 1* and *Post 2* genes of Lophotrochozoa. The acoel Hox genes are all more similar to their bilaterian counterparts than they are to cnidarian genes. Phylogenetic analyses of their concatenated amino acid sequences refute the hypothesis that they cluster with other lophotrochozoan sequences, suggesting that the acoels are not platyhelminths but a basal clade of Bilateria. This result is consistent with previous studies that use molecular data but in contradiction to the conclusions of Peterson and Eernisse (2001), who combined molecular and morphological data sets.

Hox cluster evolution

Current models propose that three or four distinct classes of Hox gene arose before the divergence of cnidarians and bilaterians (Ferrier and Holland 2001). For two of these there is direct evidence—clear orthologues of anterior (labial-like/PG 1) and posterior (PGs 9–14) Hox genes are present in cnidarians. The presence of a gene specifically related to PG3 and a single central class gene (related to PGs 4–8) has also been proposed, though no cnidarian sequence can be firmly attributed to these lineages.

We found three of these four classes represented in acoels. However, we found no gene in acoels that can be specifically related to the Hox3 lineage, despite specific searches using primers targeted to amplify such genes. It remains possible that a Hox group 3 gene exists but has diverged to the point where it is no longer amplified efficiently. Alternatively, such a gene may have been lost secondarily in these acoel lineages—gene losses are frequent events in evolution. However, the fact that neither these acoels nor any cnidarian has been shown to contain group 3-related Hox or ParaHox genes suggests that the origin of these genes may in fact have been later than Ferrier and Holland proposed. However, that would then require reconsideration of models for the origin of Hox and ParaHox gene clusters.

Whatever the case for Hox 3, we suggest (Fig. 5) that several of the gene duplications that generated the 8–10 genes of the canonical bilaterian cluster may have occurred after the split between acoels and other bilaterians but before the divergence of protostomes and deuterostomes. Such a scenario would be consistent with the idea that the "true" basal bilaterian had a relatively simple body plan, whereas the protostome/deuterostome ancestor (the "PDA" of Erwin and Davidson [2002]), with its more elaborated Hox cluster, might have been a more complex creature.

Papillon et al. (2003) proposed that the Hox genes of the chaetognath *Spadella cephaloptera* retain a basal bilaterian characteristic—specifically a "hybrid" gene that retains characteristics of both posterior and central Hox genes. No such gene has been recovered in our studies, though the primer sets used should have recovered such a gene if present. Given that the phylogenetic position of the chaetognaths is highly uncertain, it seems premature to assume that such a hybrid gene is a primitive characteristic of bilaterian Hox clusters rather than a derived feature of Chaetognatha, perhaps arising by gene conversion.

Role of Hox genes in basal bilaterians

Acoel flatworms are direct developers—none has been shown to produce a planktonic larva, even though they are a marine group, some of which might have been expected to retain such larvae if they once possessed them. Their basal position among the bilaterian animals therefore provides some support for the view that ancestral metazoans were also direct developers, rather than indirect developers with planktonic larvae (Peterson et al. 1997). In this context, and given that the anterior/posterior patterning role of the Hox genes appears to predate the radiation of the three major bilaterian clades, we might anticipate that the three distinct classes of Hox gene we identified in acoel worms will be involved in defining distinct domains along the body axis

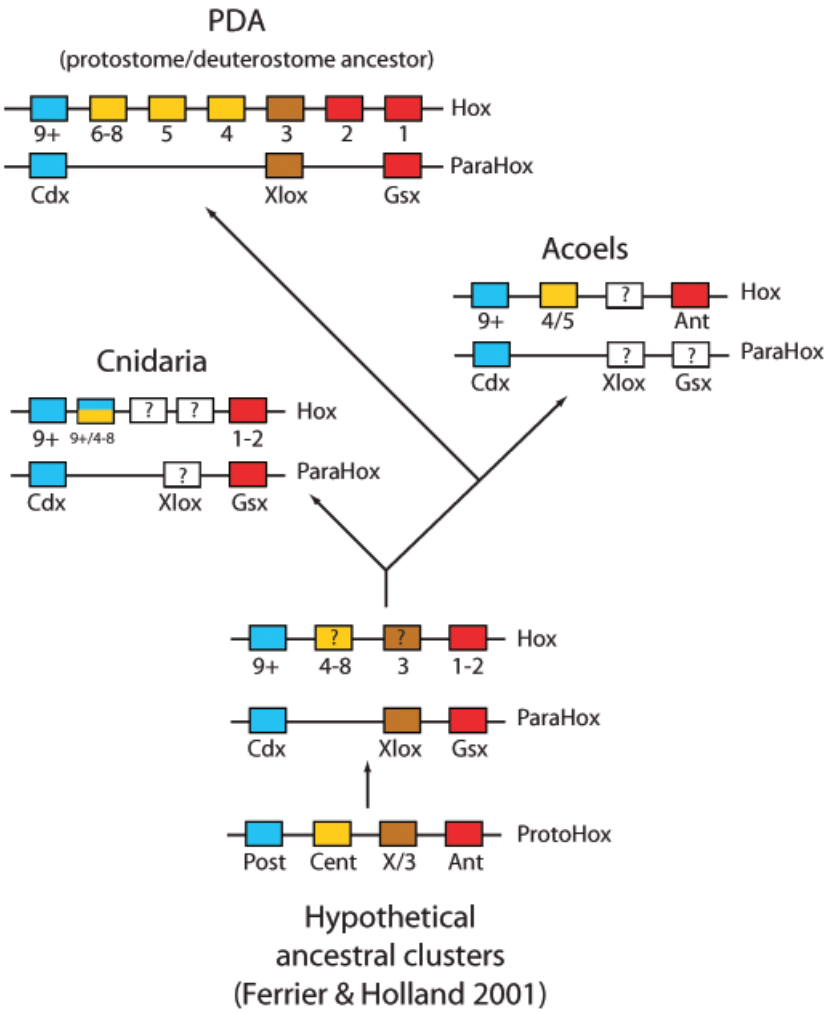

Fig. 5. Summary diagram of Hox cluster complements for basal animal groups. The lower three diagrams show hypothetical structures proposed by Ferrier and Holland (2001) for the ProtoHox cluster before duplication and the Hox and ParaHox clusters that resulted from an ancient cluster duplication. The upper half of the diagram shows the Hox gene complement documented for cnidarians and acoels and the minimal complexity of the Hox cluster in the last common ancestor of protostomes and deuterostomes (PDA), as inferred by de Rosa et al. (1999). Uncolored boxes indicate genes inferred to have been present ancestrally by the Ferrier/Holland model but not yet documented in extant species.

during embryogenesis, when the adult body plan is specified. This remains to be tested.

Acknowledgments

We thank C. Mettam and M. Havard for making us aware of the Cardiff population of *S. roscoffensis* and S. Bourlat and M. Telford for assistance in collecting them. We also thank N. Goldman for helpful discussions regarding phylogenetic analysis. E. S. thanks J. Baguñà for his unfailing encouragement along this work. This work was supported by the BBSRC in Cambridge and by the funds from the Ministerio de Educación y Ciencia, Spain, PB98-1261-C02-01 and BMC2002-03992. E. J. was funded by a BRD from the Universitat de Barcelona.

REFERENCES

Aboobaker, A. A., and Blaxter, M. L. 2003. Hox gene loss during dynamic evolution of the nematode cluster. *Curr Biol.* 13: 37–40.

Adoutte, A., Balavoine, G., Lartillot, N., Lespinet, O., Prud'homme, B., and de Rosa, R. 2000. The new animal phylogeny: reliability and implications. *Proc. Natl. Acad. Sci. USA* 97: 4453–4456.

Akam, M. 1989. Hox and HOM: homologous gene clusters in insects and vertebrates. *Cell* 57: 347–349.

Brooke, N. M., Garcia-Fernàndez, J., and Holland, P. W. 1998. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* 392: 920–922.

Cook, C. E., Smith, M. L., Telford, M. J., Bastianello, A., and Akam, M. 2001. Hox genes and the phylogeny of the arthropods. *Curr. Biol.* 11: 759–763.

deRosa, R., Grenier, J. K., Andreeva, T., Cook, C. E., Adoutte, A., and Akam, M. 1999. Hox genes in brachiopods and priapulids and protostome evolution. *Nature* 399: 772–776.

Douglas, A. E. 1983. Establishment of symbiosis in *Convoluta roscoffensis*. *J. Mar. Biol. Assoc.* 63: 409–418.

Ehlers, U. 1985. *Das Phylogenetische System der Plathelminthes*. Gustav Fischer Verlag, Stuttgart.

Ehlers, U. 1986. Comments on a phylogenetic system of the Platyhelminthes. *Hydrobiologia* 132: 1–12.

Erwin, D. H., and Davidson, E. H. 2002. The last common bilaterian ancestor. *Development* 129: 3021–3032.

Ferrier, D. E., and Holland, P. W. 2001. Ancient origin of the Hox gene cluster. *Nat. Rev. Genet.* 2: 33–38.

Ferrier, D. E., Minguillón, C., Holland, P. W., and Garcia-Fernàndez, J. 2000. The amphioxus Hox cluster: deuterostome posterior flexibility and Hox14. *Evol. Dev.* 2: 284–293.

Finnerty, J. R., and Martindale, M. Q. 1999. Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria. *Evol. Dev.* 1: 16–23.

Finnerty, J. R., Paulson, D., Burton, P., Pang, K., and Martindale, M. Q. 2003. Early evolution of a homeobox gene: the parahox gene Gsx in the Cnidaria and the Bilateria. *Evol. Dev.* 5: 331–345.

Gauchat, D., et al. 2000. Evolution of Antp-class genes and differential expression of *Hydra* Hox/paraHox genes in anterior patterning. *Proc. Natl. Acad. Sci. USA* 97: 4493–4498.

Giribet, G. 2003. Molecules, development and fossils in the study of metazoan evolution: Articulata versus Ecdysozoa revisited. *Zoology* 106: 303–326.

Gschwentner, R., Mueller, J., Ladurner, P., Rieger, R., and Tyler, S. 2003. Unique patterns of longitudinal body-wall musculature in the Acoela (Plathelminthes): the ventral musculature of *Convolutriloba longifissura*. *Zoomorphology* 122: 87–94.

Gustafsson, M. K., et al. 2002. Neuropeptides in flatworms. *Peptides* 23: 2053–2061.

Henry, J. Q., Martindale, M. Q., and Boyer, B. C. 2000. The unique developmental program of the acoel flatworm, *Neochildia fusca*. *Dev. Biol.* 220: 285–295.

Hill, A., Wagner, A., and Hill, M. 2003. Hox and paraHox genes from the anthozoan *Parazoanthus parasiticus*. *Mol. Phylogenet. Evol.* 28: 529–535.

Holland, P. W. 2001. Beyond the Hox: how widespread is homeobox gene clustering? *J. Anat.* 199: 13–23.

Hooge, M. D., Haye, P. A., Tyler, S., Litvaitis, M. K., and Kornfield, I. 2002. Molecular systematics of the Acoela (Acoelomorpha, Platyhelminthes) and its concordance with morphology. *Mol. Phylogenet. Evol.* 24: 333–342.

Huelsenbeck, J. P., and Crandall, K. A. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annu. Rev. Ecol. Syst.* 28: 437–466.

Huelsenbeck, J. P., and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

Hyman, L. H. 1951. *The Invertebrates: Platyhelminthes and Rhynchocoela. The Acoelomate Bilateria.* McGraw-Hill, London.

Jondelius, U., Ruiz-Trillo, I., Baguñà, J., and Riutort, M. 2002. The Nemertodermatida are basal bilaterians and not members of the Platyhelminthes. *Zool. Script.* 31: 201–215.

Keeble, F. 1912. *Plant-Animals. A Study in Symbiosis.* Cambridge University Press, Cambridge, UK.

Kourakis, M. J., and Martindale, M. Q. 2000. Combined-method phylogenetic analysis of Hox and ParaHox genes of the metazoa. *J. Exp. Zool.* 288: 175–191.

Mettam, C. 1979. A northern outpost of *Convoluta roscoffensis* in South Wales. *J. Mar. Biol. Assoc.* 59: 251–252.

Minguillón, C., and Garcia-Fernàndez, J. 2003. Genesis and evolution of the Evx and Mox genes and the extended Hox and ParaHox gene clusters. *Genome Biol.* 4: art. no.-R12.

Papillon, D., Perez, Y., Fasano, L., Le Parco, Y., and Caubit, X. 2003. Hox gene survey in the chaetognath *Spadella cephaloptera*: evolutionary implications. *Dev. Genes Evol.* 213: 142–148.

Peterson, K. J., and Eernisse, D. J. 2001. Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. *Evol. Dev.* 3: 170–205.

Peterson, K. J., Cameron, R. A., and Davidson, E. H. 1997. Set-aside cells in maximal indirect development: evolutionary and developmental significance. *Bioessays* 19: 623–632.

Reuter, M., Raikova, O. I., Jondelius, U., Gustafsson, M. K. S., Maule, A. G., and Halton, D. W. 2001. Organisation of the nervous system in the Acoela: an immunocytochemical study. *Tissue Cell* 33: 119–128.

Ruiz-Trillo, I., Riutort, M., Littlewood, D. T. J., Herniou, E. A., and Baguñà, J. 1999. Acoel flatworms: earliest extant bilaterian metazoans, not members of Platyhelminthes. *Science* 283: 1919–1923.

Ruiz-Trillo, I., et al. 2002. A phylogenetic analysis of myosin heavy chain type II sequences corroborates that Acoela and Nemertodermatida are basal bilaterians. *Proc. Natl. Acad. Sci. USA.* 99: 11246–11251.

Seimya, M., Ishiguro, H., Miura, K., Watanabe, Y., and Kurosawa, Y. 1994. Homeobox-containing genes in the most primitive metazoa, the sponges. *Eur. J. Biochem.* 221: 219–225.

Smith, J. P. S. III, Tyler, S., and Rieger, R. M. 1986. Is the Turbellaria polyphyletic? *Hydrobiologia* 132: 13–21.

Strimmer, K., and von Haeseler, A. 1997. Likelihood-mapping: a simple method to visualize phylogenetic content. *Proc. Natl. Acad. Sci. USA* 94: 6815–6819.

Swofford, D. L. 1998. *PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4.0b10.* Sinauer Associates, Sunderland, MA.

Telford, M. J. 2000. Turning Hox "signatures" into synapomorphies. *Evol. Dev.* 2: 360–364.

Telford, M. J., Herniou, E. A., Russell, R. B., and Littlewood, D. T. J. 2000. Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. *Proc. Natl. Acad. Sci. USA* 97: 11359–11364.

Telford, M. J., Lockyer, A. E., Cartwright-Finch, C., and Littlewood, D. T. J. 2003. Combined large and small subunit ribosomal RNA phylogenies support a basal position of the acoelomorph flatworms. *Proc. R. Soc. Lond. B Biol. Sci.* 270: 1077–1083.

Whelan, S., and Goldman, N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18: 691–699.

Yang, Z. 1997. PAML: a program package phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 15: 555–556.

Yanze, N., Spring, J., Schmidli, C., and Schmid, V. 2001. Conservation of Hox/ParaHox-related genes in the early development of a cnidarian. *Dev. Biol.* 236: 89–98.