Biochem Genet (2009) 47:100–116
DOI 10.1007/s10528-008-9210-6

Hox Genes in the Parasitic Platyhelminthes *Mesocestoides corti*, *Echinococcus multilocularis*, and *Schistosoma mansoni:* Evidence for a Reduced *Hox* Complement

Uriel Koziol · Ana I. Lalanne · Estela Castillo

Received: 12 January 2008 / Accepted: 15 August 2008 / Published online: 22 January 2009

Abstract Little is known about the *Hox* gene complement in parasitic platyhelminthes (Neodermata). With the aim of identifying *Hox* genes in this group we performed two independent strategies: we performed a PCR survey with degenerate primers directed to the *Hox* homeobox in the cestode *Mesocestoides corti*, and we searched genomic assemblies of *Echinococcus multilocularis* and *Schistosoma mansoni*. We identified two *Hox* genes in *M. corti*, seven in *E. multilocularis*, and nine in *S. mansoni* (including five previously reported). The affinities of these sequences, and other previously reported *Hox* sequences from flatworms, were determined according to phylogenetic analysis, presence of characteristic parapeptide sequences, and unusual intron positions. Our results suggest that the last common ancestor of triclads and neodermatans had a *Hox* gene complement of at least seven genes, and that this was probably derived by gene loss from a larger ancestral *Hox* complement in lophotrochozoans.

Keywords *Hox* · Parahox · Platyhelminthes · Cestode · *Mesocestoides corti* · Neodermata

Introduction

Hox genes are found in all metazoan phyla except poriferans (Larroux et al. 2007). They are active in distinct domains along the main body axes and direct the

U. Koziol · A. I. Lalanne · E. Castillo (✉)
Sección Bioquímica, Facultad de Ciencias, Universidad de la República, CP 11400, Montevideo, Iguá 4225, Uruguay
e-mail: castillo@fcien.edu.uy

Present Address:
A. I. Lalanne
Unité du Développement des Lymphocytes, Institut Pasteur, Paris, France

morphogenesis of segment-specific structures via the activation of downstream target genes (reviewed by Gehring 2007). These genes are organized into a so-called "*Hox* cluster" (Bürglin 1994). A *Hox* cluster consists of anterior, central, and posterior genes; the central genes have been identified only in bilaterians, but not in cnidarians, which are a basal metazoan phylum that arose before the divergence of bilaterians (Chourrout et al. 2006). The central *Hox* genes are diverse, and their diversity is considered the basis of morphological complexity in bilateral development (Ogishima and Tanaka 2007). Understanding the evolution of the central *Hox* genes should thus lead to understanding the evolution of bilaterian body plans. The reconstruction of *Hox* cluster evolution in the Metazoa should provide valuable data for understanding the evolution of bilaterian body plans and the relationship between genetic and morphological complexity (reviewed by Ferrier 2007).

It has been proposed that lophotrochozans have an ancestral *Hox* complement composing 10–11 genes (de Rosa et al. 1999; Balavoine et al. 2002; Kulakova et al. 2007): *Hox1*, *Hox2*, *Hox3*, *Hox4*, *Hox5*, *Lox5* (which is probably orthologous to *ftz* from ecdysozoans; Telford 2000b), *Hox7* (named after the *Lineus sanguineus LsHox7*, and similar to ecdysozoan *Antp*; we have followed this nomenclature after Kulakova et al. 2007), *Lox2*, *Lox4*, *Post-1*, and *Post-2*. Genes clearly orthologous to *Hox1-5* are found in both protostomes and deuterostomes. The genes *Lox5*, *Lox2*, *Lox4*, *Post-1*, and *Post-2* are, on the other hand, characteristic of lophotrochozoans (Balavoine et al. 2002). *Lox5* genes are characterized by the presence of the conserved "Lox5" parapeptide, which is N-terminal to the homeodomain. *Lox2* and *Lox4* genes share with ecdysozoan *Ubx* and *Abd-A* the "Ubd-A" parapeptide. It is not clear whether *Lox2* and *Lox4* originated as an independent duplication of an ancestral "Ubd-A" gene in lophotrochozoans, or whether the last common ancestor of protostomes already had two "Ubd-A" genes.

Metazoans also possess an ancestral Parahox cluster (Garcia-Fernàndez 2005). *Hox* and Parahox clusters probably originated from the duplication of an ancestral ProtoHox cluster; therefore, *Hox* genes are not monophyletic, with anterior *Hox* genes being most related to the *Gsx* Parahox genes, *Hox3* genes with *Xlox* Parahox genes, and posterior *Hox* genes to *Cdx* Parahox genes. All three Parahox genes have been found in several lophotrochozoans (Ferrier and Holland 2001; Kulakova et al. 2008).

Platyhelminthes constitute a group of organisms that display a range of diverse life histories from free-living to parasitic. As they lack a coelom, segmentation, elaborated organs, and anus, they were considered to be among the basal groups of bilaterians. However, analyses based on 18S ribosomal and homeobox sequences placed Platyhelminthes as a derived bilaterian phylum, among the lophotrochozoans (Aguinaldo et al. 1997; Balavoine 1997). Also, many recent morphological analyses coincide in placing the platyhelminthes not as basal to all bilaterians but as derived protostomes, together with other spiralian phyla (Ax 1996; Nielsen 2001). The exact position of platyhelminthes among lophotrochozoans is not clear. Indeed, they have been placed at a wide range of positions, from being basal to all other lophotrochozans (Helmkampf et al. 2008) to a sister relationship with Annelida (Lartillot and Philippe 2008).

Information about *Hox* genes in plathyhelminthes is mostly about planarians (Tricladida) (Bartels et al. 1993; Tarabykin et al. 1995; Bayascas et al. 1997; Orii et al. 1999; Saló et al. 2001; Nogi and Watanabe 2001). Orthology relationships of *Hox* genes of triclads have been well established (Saló et al. 2001), in groups PlHox1–PlHox9. These groups have been proposed to be orthologous to *Hox1-5*, *Lox5*, *Lox4*, and *Post-2*. Few *Hox* genes have been isolated from parasitic flatworms (Neodermata). Studies included *Schistosoma mansoni* (Pierce et al. 2005; Webster and Mansour 1992), in which *Hox1*, *Hox4*, *Lox5*, and *Lox4* orthologs were identified, and the cestode *Taenia asiatica* (Kim et al. 2007). Recently, Olson (2008) extensively reviewed the *Hox* genes from the platyhelminthes, performing a phylogenetic analysis together with other selected lophotrochozoan sequences. This analysis included, also, unpublished sequences from *Hymenolepis microstoma*. Orthology assignment between *Hox* genes of triclads and neodermatans, or of platyhelminthes and other lophotrochozoans, was based exclusively on phylogenetic relationships of the homeodomain, and only some orthology groups were clearly recovered. Information about Parahox genes in platyhelminthes is scarcer; although they have been searched for extensively (Saló et al. 2001; Olson 2008), only *Xlox* and *Cdx* orthologs have been found in the polyclad *Discocelis tigrina* (Saló et al. 2001).

Our study was directed at obtaining *Hox* gene sequences from neodermatans, using two different strategies. We used the *Schistosoma mansoni* and *Echinococcus multilocularis* genomic assembled contigs from the Sanger Institute to search for *Hox* and Parahox genes; this approach should be insensitive to the amplification bias of degenerate PCR strategies. Furthermore, the 7× coverage of the *S. mansoni* genome makes unlikely the possibility that existing *Hox* genes might be missed. We also searched for *Hox* genes by degenerate PCR in the cyclophyllidean cestode *Mesocestoides corti*. We have inferred orthology relationships among *Hox* genes of neodermatans, triclads, and other lophotrochozoans by means of phylogenetic analysis, presence of characteristic parapeptides, and unusual intron positions. Our results suggest that the last common ancestor of triclads and neodermatans had a reduced *Hox* complement compared to other lophotrochozoans, probably due to gene loss.

Materials and Methods

Mice infected with *M. corti* tetrathyridia were kindly donated by Laura Dominguez and Jenny Saldaña of Facultad de Quimica, Universidad de la Republica, Uruguay. Parasite removal and culture were made following Britos et al. (2000). Tetrathyridia were cultured in vitro to obtain segmented worms, showing an elongate body and numerous proglottids, and separated manually. DNA was extracted from tetrathyridia following McManus et al. (1985). RNA was extracted from tetrathyridia and segmented worms using Trizol (Gibco). cDNA was synthesized from 5 μg tetrathyridia or adult worm RNA with poly dT primer and Superscript II reverse transcriptase (Invitrogen).

Isolation of *Hox* cDNA Fragments

The strategy employed (Tarabykin et al. 1995) was to amplify *Hox* homeobox genes from tetrathyridia cDNA with degenerated primers directed to the coding region of the first and third helices of the homeodomain. Primer sequences are S01, GARYTNGARAARGARTT, and S02, CKNCKRTTYTGRAACAA. Cycling conditions were as described by Tarabykin et al. (1995). PCR bands were excised from agarose gels and cloned into pGEM-T-Easy vector (Promega). The 200 recombinant plasmids were sequenced by CTAG Service of Facultad de Ciencias (Uruguay) using a Perkin-Elmer ABI Prism 377 automated DNA sequencer.

Southern Blot Assay

Southern blot assay was performed following Sambrook et al. (1989). A 0.8% agarose gel was loaded with 10 μg *M. corti* or 20 μg mouse DNA digested with *Eco*RI and *Sal*I. The radioactive probes were synthesized by *Eco*RI digestion of the recombinant plasmids containing *MvHox1* or *MvHox7* fragments. The excised inserts were labeled with 32P-αATP using the Prime-a-Gene kit (Promega).

Search for *Hox* and Parahox genes in *S. mansoni* and *E. multilocularis* genomic assemblies

We searched for *Hox* and Parahox genes in genomic contigs from *S. mansoni* genome version 3.1 and *E. multilocularis*. These sequence data were produced by the *Schistosoma* and *Echinococcus* Sequencing Groups at the Sanger Institute and can be obtained from ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/ and ftp://ftp.sanger.ac.uk/pub/pathogens/Echinococcus/. Searching for *Hox* genes in genomic contigs was done by Blastn in the Sanger Blast Server (http://www.sanger.ac.uk/cgi-bin/blast/submitblast/s_mansoni and http://www.sanger.ac.uk/cgi-bin/blast/submitblast/Echinococcus), using *Mus musculus* HoxA1; *Euprymna scolopes* EsAntp, EsPost-1, and EsPost2; *Nereis virens* Gsx and Cdx; and *Capitella* sp. Xlox protein sequences as queries (accession numbers are provided in Fig. 2). The expected cut-off value was 10 in order to avoid missing homeoboxes with several introns. A list containing all the contigs that were hit was analyzed both manually and by Blastp (after joining the conceptual translation of exons) against the GenBank nr database. Approximately 100 Blast hits were analyzed for each species, resulting in the retrieval of more than 50 homeodomains, not only from the ANTP class but also from others such as POU, paired, LIM, SINE, and ZF (this suggests the search for *Hox* and *Hox*-like genes was exhaustive). Homeoboxes and flanking regions from genes identified as *Hox* or *Hox*-like were then recovered, and their exons joined manually when introns were present, based on amino acid similarity to other *Hox* genes and the presence of canonical splice sites. In some cases (i.e., *EmHox1*), similarity did not allow the confident recovery of the complete homeobox; in these cases the sequence was retrieved only to the nearest possible canonical splice site to the Blast hit. We also searched for *Hox* genes in unassembled reads of these organisms, in the *S. mansoni* Genome Database (GeneDB; http://www.genedb.org/genedb/smansoni/), and in ESTs from all platyhelminthes in GenBank.

Phylogenetic Analyses

Conceptually translated amino acidic sequences were aligned using ClustalW software (Thompson et al. 1994). Alignments included numerous sequences from all *Hox* and Parahox orthologous groups belonging to other lophotrochozoans, together with the deuterostome *Mus musculus* and the ecdysozoan *Drosophila melanogaster*. Only some sequences from each planarian orthology group, PlHox1-PlHox9, were included, because genes within these groups are very similar and have been extensively characterized before (Saló et al. 2001; Olson 2008). Unrooted phylogenetic analyses were performed, using only homeodomain sequences. Similar results were obtained when performing rooted analyses using *Evx* and *Mox* genes from *Mus musculus* and *Drosophila melanogaster* as outgroups, except that bootstrap support values for some clades were lower (data not shown). Maximum parsimony and neighbor-joining (NJ) phylogenetic trees were constructed using Mega (Kumar et al. 2004). Poisson correction was used as the substitution model for NJ; similar results were obtained using other models. Bootstrap support values were estimated using 1000 replicates. Maximum likelihood analysis was performed using the ProML application from BioEdit (Hall 1999), using the Jones-Taylor-Thornton model. All analyses gave very similar results, except in nodes with very low support; therefore, only the NJ analysis is reported.

Results

We found two *Hox* genes in *M. corti* by degenerate PCR, seven in the *E. multilocularis* genome contigs, and nine in the *S. mansoni* genome contigs and unassembled reads (including the five genes previously reported by Pierce et al. 2005). No Parahox genes were found, although *Evx* and *Mox* orthologs, which are basal to Hox/Parahox genes (Minguillón and Garcia-Fernàndez 2003), were clearly identified in both *S. mansoni* and *E. multilocularis* (data not shown). We have classified these sequences, together with those of triclads, according to homeodomain sequence similarity and parapeptides (Fig. 1), phylogenetic analysis (Fig. 2), and intron positions. Below, we describe these sequences and classify them according to the orthology groups to which they probably belong.

Hox1 Orthologs

Hox1 orthologs from the platyhelminthes have been recovered before in the polyclad *Discocelis tigrina* (*Distox*-A, Saló et al. 2001), in several triclads (PlHox1 group; Saló et al. 2001), in *S. mansoni* (*SmHox1*; Pierce et al. 2005), in *Echinostoma trivolvis* (*ETOX-A*, L19170; see Olson 2008), and in *Taenia asiatica* (in which two paralogous genes were found, *TasHox1a* and *TasHox1b*; Kim et al. 2007).

We found a *Hox1* ortholog in *E. multilocularis* (*EmHox1*, Contig_0006768) and in *M. corti* (*MvHox1*, AY187806). In the case of *MvHox1*, one of the primers annealed 5′ of the expected site. This circumstance allowed us to have a more

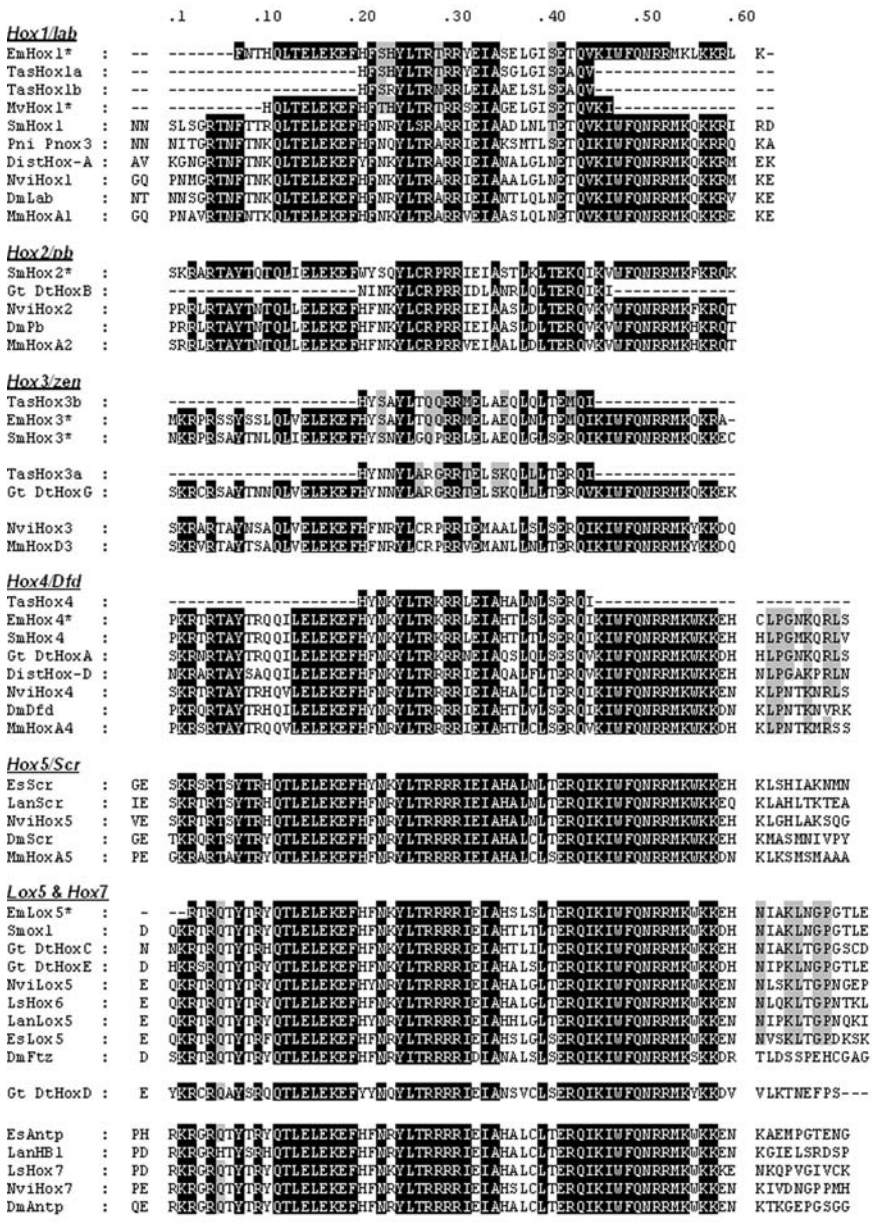

Fig. 1 Alignments of homeodomain and flanking sequences from *M. corti*, *E. multilocularis*, and *S. mansoni* with related sequences from other lophotrochozoans, the ecdysozoan *D. melanogaster* and the deuterostome *Mus musculus*. Numbers above correspond to standard positions for the homeodomain. Alignments of *Hox5* and *Hox7* sequences are included for comparison. Within each alignment, positions that are absolutely conserved are in white letters with black shading. Residues marked with gray shading within the homeodomain are referred to in the main text. Residues marked with gray shading in the *Hox4/Dfd*, *Lox5,* and *Lox4* alignments correspond to the Dfd, Lox5, and Ubd-A parapeptides, respectively. Sequences obtained in this study are marked with an asterisk. Abbreviations and accession numbers are given in Fig. 2

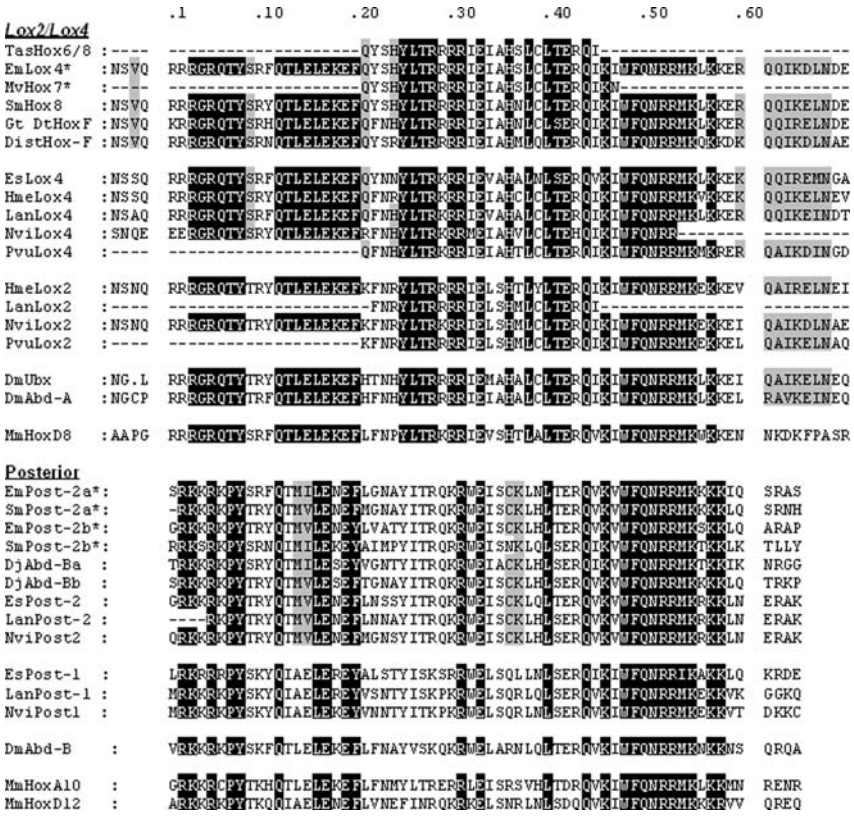

Fig. 1 continued

complete homeodomain sequence for this gene. To confirm that the isolated sequence is indeed from *M. corti* and to estimate the gene copy number, we performed Southern blot analysis. When working with parasites the contamination with host material is a concern. For this reason, we included mouse genomic DNA lanes in the Southern blots. The membranes hybridized with *MvHox1* probe exhibit one single band for *M. corti* DNA digested with *Sal*I and *Eco*RI (Fig. 3). This suggests that *MvHox1* is a single copy gene. Mouse DNA did not reveal any band. We have also confirmed by RT-PCR with gene-specific primers that *MvHox1* is expressed in both the tetrathyridium stage and in the adult segmented worms (data not shown).

In phylogenetic analysis, all putative *Hox1* genes from the platyhelminthes cluster with moderate support with *Hox1* orthologs from other bilaterian phyla (Fig. 2); furthermore, all *Hox1* genes from cestodes are clustered too. Alignment of these sequences reveals that although some characteristic residues of *Hox1* genes are conserved in some cyclophyllidean cestodes (such as threonine in position 43), these sequences are rather divergent. In position number 23, most homeodomains have an asparagine residue, but *Hox1* representatives of the cestodes *M. corti*,

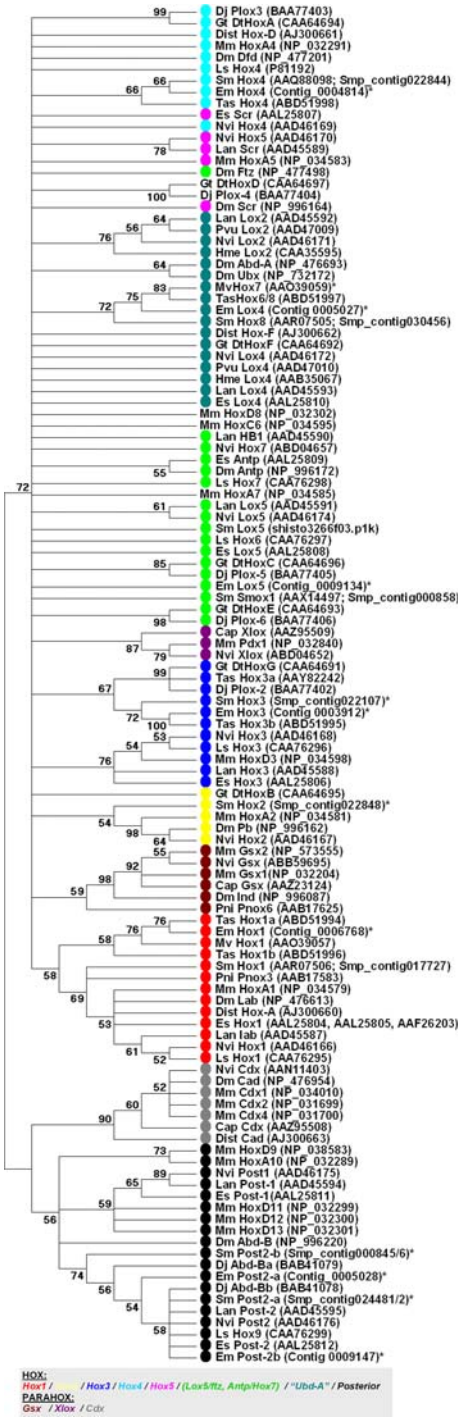

Fig. 2 Neighbor-joining tree of *Hox* and Parahox sequences from platyhelminthes and other bilaterians. Bootstrap support values are given next to nodes in percentages. Nodes with less than 50% support have been collapsed. Genbank accession numbers (and/or contig numbers for *S. mansoni* and *E. multilocularis*) are provided to the right. Sequences obtained in this study are marked with an asterisk. Orthology groupings are indicated by the color codes defined below the tree. Members of the PlHox5 group (*Gt DtHoxD* and *Dj Plox-4*) are not color-coded. Species abbreviations: Cap, *Capitella* sp. (Annelida); Dist, *Discocelis tigrina* (Polyclada); Dm, *Drosophila melanogaster* (Arthropda); Dj, *Dugesia japonia* (Tricladida); Em, *Echinococcus multilocularis* (Cestoda); Es, *Euprymna scolopes* (Mollusca); Gt, *Girardia tigrina* (Tricladida); Hme, *Hirudo medicinalis* (Annelida); Lan, *Lingula anatina* (Brachiopoda); Ls, *Lineus sanguineus* (Nemertea); Mm, *Mus musculus* (Vertebrata); Nvi, *Nereis virens* (Annelida); Pni, *Polycelys nigra* (Tricladida); Pvu, *Patella vulgata* (Mollusca); Sm, *Schistosoma mansoni* (Trematoda); Tas, *Taenia asiatica* (Cestoda)

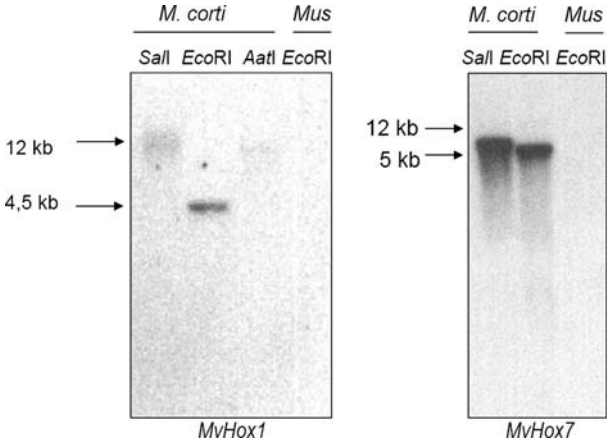

Fig. 3 Southern blot analysis of *MvHox1 and MvHox7*. *Mesocestoides corti* DNA (10 μg) or *Mus musculus* (20 μg) was digested with the restriction enzyme indicated above and hybridized with a radioactive probe generated with the sequence available for *MvHox1* and *MvHox7*. The size of the molecular weight marker fragments is indicated on the left

E. multilocularis, and *T. asiatica* show serine/threonine. In addition, homeodomains of *Hox1* orthologs have arginine or lysine in position 24, but *MvHox1*, *EmHox1*, and *TasHox1a* display the amino acid histidine. In position 29 all *Hox1* proteins have an alanine, except cestode ones.

Olson (2008) labeled several other genes from *Echinococcus* spp. (Hbx1 and Hbx2) as *Hox-1*-like genes; however, close inspection clearly demonstrates that these genes are actually from the NKL subclass.

Hox2 Orthologs

In *S. mansoni*, we have found a putative *Hox2* ortholog, *SmHox2* (Smp_contig022848). It is located in scaffold Smp_scaff000314, in which *SmHox4* is also located. It has a phase-0 intron between codons 46 and 47 of the homeobox. This homeodomain is present in Smp_166150 in the GeneDB; it seems that this CDS is incorrectly assembled, fusing the homeodomain to a serine/threonine kinase.

Phylogenetic analysis groups this sequence with *Hox2* genes from other bilaterians with moderate support. The amino acid sequence shares several conserved residues characteristic of *Hox2* (such as T11) or both *Hox2* and *Hox3* genes (C27, P29). Before this work, the only putative ortholog of *Hox2* genes recovered from any platyhelminth was a small fragment from *Girardia tigrina*, *DtHoxB* (Bayascas et al. 1997; Saló et al. 2001). In our analysis, *DtHoxB* is clustered with low support with *Hox3* genes (this node is collapsed in Fig. 2), but given its similarity to *SmHox2* and the limited information in this short sequence, this might be an artifact.

Hox3 Orthologs

In triclads, the PlHox3 group has been linked to *Hox3* genes due to overall sequence similarity and low phylogenetic support (Balavoine and Telford 1995; Bayascas et al. 1997; Saló et al. 2001), although it is highly divergent. In neodermatans, putative PlHox3 orthologs have been found in *T. asiatica* (two paralogs, *TasHox3a* and *TasHox3b*) and a short sequence similar to *TasHox3b* was found in *Echinostoma trivolvis* (*ETOX-E*, L19216). We have found similar sequences in both *S. mansoni* (*SmHox3*, Smp_contig022107; Smp_164610 in the GeneDB) and *E. multilocularis* (*EmHox3*, Contig_0003912). All these sequences are clustered together with good bootstrap support, constituting a clade that is basal to both *Hox2* and *Hox3* genes (this node is collapsed in Fig. 2). Therefore, the assignment of these sequences to the *Hox3* group remains tentative.

This clade is clearly resolved in phylogenetic analyses into two groups, with very good support; the differences between them are apparent in the alignment (Fig. 1). One branch (*Hox3a* group) contains genes from triclads and *TasHox3a*; the other (*Hox3b* group) has genes from the neodermatans *E. multilocularis*, *S. mansoni*, and the *TasHox3b* gene. This latter group would also include *ETOX-E*, which is very similar to these sequences and grouped with them in the analysis by Olson (2008).

One explanation for this topology could be that the last common ancestor of triclads and neodermatans possessed two *Hox3* genes, and that these were either selectively lost or artifactually not recovered in searches for *Hox* genes in different lineages (*Hox3b* in triclads, *Hox3a* in most neodermatans). Alternatively, *TasHox3a* could be a *T. asiatica*-specific paralog that converged with *Hox3* sequences of triclads.

Central Genes

Among central genes, resolution is usually very low in phylogenetic analysis using the homeodomain sequence (Balavoine et al. 2002), and this is the case in our analysis. We were able, however, to establish orthology relationships using parapeptide sequences.

Hox4 Orthologs

Hox4 genes have been described from several platyhelminthes (*Distox-D* from *D. tigrina*, PlHox4 group from triclads, *SmHox4* from *S. mansoni,* and the short sequence *TasHox4* from *T. asiatica*). Olson (2008), based on phylogenetic analyses, labeled members of the PlHox4 group as belonging to the *Hox1*-like genes; however, the presence of the Dfd parapetide clearly demonstrates the affinity of these sequences to *Hox4* genes, as suggested by Saló et al. (2001).

One putative *Hox4* ortholog from *E. multilocularis* (*EmHox4*; Contig_0004814) was found in our search. In the phylogenetic analysis, the three neodermatan sequences clustered together with good support, but *Hox4* genes formed part of a large polytomy of central genes. We were able to unite them, however, thanks to the presence of the Dfd parapeptide (Balavoine et al. 2002) in all of these sequences (Fig. 1). This parapeptide is very well conserved in all sequences from the

platyhelminthes. This orthology group would also include *Fhhbx2* (X66824) from *Fasciola hepatica* and *ETOX-B* (L19171), which are very similar to *SmHox4*.

Lox5 Orthologs

Lophotrochozoan *Lox5* genes can be clearly distinguished by the presence of the conserved Lox5 parapeptide (Balavoine et al. 2002). In triclads, the *Lox5* gene has duplicated, and therefore the PlHox6 group has two genes per species. In *S. mansoni*, Pierce et al. (2005) also found two *Lox5* paralogs: *Smox1* and the more divergent *SmLox5*, in which the Lox5 parapeptide is less well conserved. *Smox1* is placed in scaffold Smp_scaff000004, together with the *SmPost-2b* gene. *SmLox5*, on the other hand, is located in an unplaced read in *S. mansoni* assembly 3.1, (schisto_3266f03.p1k).

We have found a *Lox5* ortholog in the *E. multilocularis* genome, *EmLox5* (Contig_0009134), with a well-conserved Lox5 parapeptide (Fig. 1). A very similar sequence from *Taenia solium* was identified in several ESTs (Genbank acc. nos. EL756693, EL759423, EL762691). This is the first identification of *Lox5* orthologs in cestodes.

The PlHox6 genes from *Girardia tigrina* (*DtHoxC*, *DtHoxE*; Bayascas et al. 1997), *Smox1*, and *EmLox5* all possess two very unusual introns within the homeodomain; the first is a phase-2 intron in codon 24, and the second is a phase-0 intron between codons 51 and 52. The more divergent *SmLox5* shares the first intron position but not the second one. This gives further support to the hypothesis that all these genes are orthologous.

Planarian PlHox5 genes have been tentatively proposed to be orthologous to *Hox5*, based on overall similarity (Bayascas et al. 1997; Saló et al. 2001). It is interesting that the PlHox5 genes from the tricladida also share these intron positions. The *Girardia tigrina DtHoxD* gene (Bayascas et al. 1997) is known to have a phase-2 intron in codon 24. We also compared by Blastn the recently described cDNA sequence of the *HoxD*-like gene from *Schmidtea mediterranea* (EU082824; Iglesias et al. 2008) with the *S. mediterranea* draft genome assembly (AAWT00000000) and confirmed that it shares both intron positions. Furthermore, PlHox5 genes have Q in position 6 of the homeodomain, which is characteristic of *Lox5*, *Hox7*, *Lox2*, and *Lox4* genes, instead of T6, as found in *Hox5* genes. All these results suggest that PlHox5 genes are probably *Lox5* orthologs (and therefore paralogs of PlHox6) that lost the Lox5 parapeptide after the duplication, and that *Hox5* genes have not been found in the triclads. Similarly, we have been unable to find *Hox5* genes in the *S. mansoni* and *E. multilocularis* genomes.

Lox2/*Lox4* Orthologs

Several genes most similar to *Lox2* and *Lox4* have been recovered from the platyhelminthes before, although in many cases these sequences are too small to allow observation of the Ubd-A parapeptide. As pointed out by Olson (2008), these sequences are very similar to each other and are therefore easily identifiable. These include *Distox-F* from *D. tigrina*, the triclad PlHox7/8 group, the gene *ETOX-C*

from *Echinostoma trivolvis* (L19172; see Olson 2008), *SmHox8* from *S. mansoni*, and *TasHox6/8* from *T. asiatica*.

In *M. corti*, we obtained a short sequence, *MvHox7* (AY187808), that is 100% identical at the amino acid level to *TasHox6/8*. Southern blot analysis demonstrates that there is a single copy of this gene in *M. corti* (Fig. 3). In *E. multilocularis*, we identified the gene *EmLox4* (Contig_0005027), which has a well-conserved Ubd-A parapeptide and several characteristic residues within the homeodomain (Fig. 1).

All "Ubd-A" genes from the platyhelminthes are most similar to the *Lox4* genes of other lophotrochozoans, and they cluster together in our phylogenetic analysis (albeit with very low support; this node is collapsed in Fig. 2). They present two synapomorphic residues (Q21, [R/K]60) and two plesiomorphic residues (H24 and A35) described for *Lox4* (Telford 2000a). No genes with *Lox2*-specific residues have been found in the platyhelminthes.

Post-2 Orthologs

Because of their divergent nature, few posterior genes have been recovered from flatworms by degenerate-PCR surveys. The only published posterior gene sequences belong to the triclad PlHox9 group, very simlar to *Post-2* genes, which would include an unpublished *G. tigrina* gene (GtAbdB-b, Saló et al. 2001), *Dugesia japonica Abd-Ba* and *Abd-Bb* (Nogi and Watanabe 2001), *S. mediterranea Abd-Ba* (EST EG409633, and others) and *AbdB-b* (EST EG404975, and others), and a *Dugesia ryukyuensis AbdB-b* EST (BW635170). We have found two posterior genes in *S. mansoni*: *SmPost-2a* (two exons, in Smp_contig024481 and Smp_contig024482, which are contiguous in scaffold Smp_scaff000397; the first exon is present in Smp_087070 in GDBase, although it is incorrectly assembled) and *SmPost-2b* (two exons, in Smp_contig000845 and Smp_contig000846, which are contiguous in scaffold Smp_scaff000004; it is not present in GeneDB). An EST similar to part of the *SmPost-2b* gene is present in *Schistosoma japonicum* (AA143933). Two posterior genes were also identified in *E. multilocularis*: *EmPost-2a* (Contig_0005028) and *EmPost2-b* (Contig_0009147). All these genes contain a phase-0 intron between codons 44 and 45 of the homeodomain.

All identified posterior genes from flatworms are clustered with good support with *Post-2* genes from other lophotrochozoans. Indeed, the alignment clearly shows that these genes have several *Post-2*-specific residues, such as M14, [I/V]15, C36, and K37, in addition to P7, which is characteristic of all posterior genes. The phylogenetic relationship among them is ambiguous, however, and very sensitive to parameter changes (data not shown). Thus, it is not possible to determine whether two *Post-2* genes were present in the common ancestor of triclads and neodermatans, or whether independent duplications have occurred.

Three unpublished posterior genes were reported by Olson (2008) in the cestode *Hymenolepis microstoma*. One of these clearly clustered with *Post-2* genes; the other two (*Post-1a* and *Post-1b*) were referred to as divergent *Post-1*-like genes, with uncertain phylogenetic affinities. We have not been able to identify any *Post-1*-like genes in *S. mansoni* and *E. multilocularis.*

Problematic Sequences

In previous degenerate-PCR surveys of *Echinococcus granulosus* and *M. corti*, we have found several short homeobox fragments (Genbank acc. nos. AF095860-62, AY187809-13). These sequences have been included in analyses by other groups (Kim et al. 2007; Olson 2008). Strikingly, there are no sequences similar to these in *E. multilocularis*. Furthermore, these sequences are most similar to vertebrate *Hox* genes, although they are not identical to any sequence in the GenBank nr database. Specifically, the *EgHox10* and *MvHox10* sequences are most similar to vertebrate *Hox*10 paralogs, but the amino acid identity with the best Blastp match (NP_032289) is only 63%.

Southern blot analyses under high stringency conditions failed to detect these sequences in either *M. corti* or *E. granulosus*, except in the case of *EgHox1* where multiple bands were observed (data not shown). We have provisionally ruled out the possibility that they originated by contamination from the hosts *Bos taurus* or *Mus musculus*, because they are not present in the sequenced genomes, but they could have originated from some other contamination source.

It is possible that some of these sequences are true *Hox* or *Hox*-like genes from *E. granulosus* and *M. corti*. Therefore, we have chosen not to discontinue these sequences from the GenBank database yet, although we question their origin, and they should be used with caution.

Similarly, three Hox/Parahox sequences were found in *E. multilocularis* reads with very high similarity to vertebrate Hox/Parahox genes, even at the nucleotide level in introns and UTRs, when compared to *Mus musculus* and *Rattus norvegicus* (reads emu-907i19.p1k, emu-923f09.p1k, emu-956k24.p1k; data not shown). These reads are not present in the assembled contigs. We have interpreted them as possibly originating from host contamination.

In our analysis, we have been unable to find *Hox7* orthologs in platyhelminthes. The only published sequence that shows similarity to this gene group is the very short sequence L19173, from *Echinostoma trivolvis*. This sequence, however, is also very similar to other orthology groups, and strangely, it is 100% identical at the nucleotide level to the *CTs-Dfd* homeobox (S76416) from the annelid *Ctenodrilus serratus*.

Finally, we have reanalyzed some sequences from the planaria *Polycelis nigra* and *Phagocata woodworthi* that have been considered orphans (Orii et al. 1999; Olson 2008). *Pnox6* (Balavoine and Telford 1995; AAB17625) from *Polycelis nigra* shows the highest similarity in Blast searches to *ind*, the *D. melanogaster* ortholog of the *Gsx* Parahox gene. In all our phylogenetic analyses, it clusters with moderate support with *Gsx* genes (Fig. 2). Its sequence is, however, rather divergent. The short sequences *PwoxF* and *PwoxG* (L19174 and L19179) from *Phagocata woodworthi* are very similar to *Pnox6*. The *Pnox5* homeodomain fragment (Balavoine and Telford 1995) shows the highest similarity in Blast searches to *Mnx/exex* genes, especially to the *Mnx* gene from the placozoan *Trichoplax adherens* (DQ355807).

Discussion

In this study we have found Hox genes in three neodermatan species and attempted to determine relationships of orthology among them and with other lophotrochozoans. According to our results and our reinterpretation of the PlHox5 group, all Hox groups previously identified in triclads (PlHox1-9) are present in the neodermatans: *Hox1* (PlHox1), *Hox2* (PlHox2), *Hox3* (PlHox3), *Hox4* (PlHox4), *Lox5* (PlHox5, PlHox6), *Lox4* (PlHox7/8), and *Post-2* (PlHox9). This implies that four genes that have been proposed to be present in the ancestral lophotrochozoan are missing in neoophoran platyhelminthes: *Hox5*, *Hox7*, *Lox2,* and *Post-1*. (The *Post-1*-like genes from *H. microstoma* reported by Olson in 2008 have not been published, preventing us from including them in our analysis. They did not group, however, with other *Post-1* genes in that work.) It is possible that one of these groups has been missed in our search and those of others; however, we consider this unlikely, especially in the case of *S. mansoni*, in which there is good genomic coverage and an independent degenerate-PCR strategy, and in triclads, where extensive research has been done. Furthermore, analysis of the *S. mediterranea* draft genome sequences did not produce *Hox* genes from any of the orthology groups missing in this work (data not shown).

We have also been unable to find Parahox genes in neodermatans, which parallels the results of PCR surveys. Parahox-like genes were found in *Discocelis tigrina* (Saló et al. 2001), and *Pnox6* from *Polycelis nigra* might be a *Gsx* ortholog, as described before. Therefore, Parahox genes could be missing specifically in neodermatans.

Therefore, we propose that the last common ancestor of triclads and neodermatans had a *Hox* gene complement of at least seven or eight genes (depending on whether *Post-2* duplication occurred only once or independently in several lineages). If Platyhelminthes is the most basal lophotrochozoan phylum studied yet, the absence of some of these genes (*Lox2*, and perhaps *Post-1* and *Hox7*) could be interpreted as plesiomorphous (Fig. 4). Indeed, the presence of a single *Lox2/Lox4* gene in triclads was interpreted as evidence of a basal position of platyhelminthes among lophotrochozoans (Saló et al. 2001). If the absence of *Hox7*, *Post-1*, and *Lox2* is plesiomorphous, then the last common lophotrochozoan ancestor would have had a *Hox* gene complement of only eight genes [*Hox1-5*, a *Lox5/Hox7* ancestor, a single Ubd-A gene (*Lox2/Lox4* ancestor), and a single *Post-1/Post-2* ancestor].

If, however, platyhelminthes are not basal to all other studied lophotrochozoan species (including members of Annelida, Mollusca, Brachiopoda, and Bryozoa; Balavoine et al. 2002; Kulakova et al. 2007), the clear implication is that *Hox5*, *Hox7*, *Lox2*, and *Post-1* have been lost in the lineage leading to neoophoran platyhelminthes (the second scenario in Fig. 4). If Platyhelminthes is the most basal lophotrochozoan phylum, this scenario would still be equally valid.

On the basis of several grounds, we favor the gene-loss scenario. First, there is no consensus on the position of platyhelminthes among lophotrochozoans, and they cannot thus be assumed to be basal. Second, all *Hox* genes from the platyhelminthes show characteristics that link them to the corresponding specific genes of other

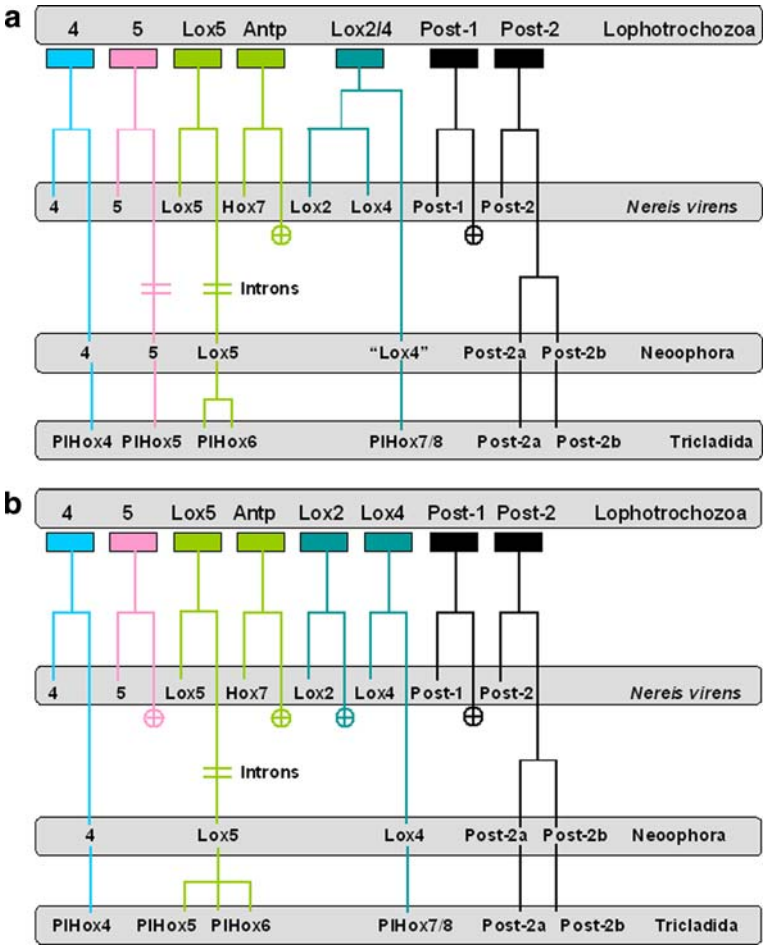

Fig. 4 Two possible evolutionary scenarios for central and posterior *Hox* genes in neoophoran platyhelminthes. **a** Scenario inferred by Saló et al. 2001. Supposing playthelminthes were basal lophotrochozoans, some of the gene absences could be plesiomorphous (as the absence of *Lox2* is interpreted here); also, the PlHox5 group is interpreted as orthologous to *Hox5*, making the intron positions in PlHox5 and PlHox6 convergent. **b** In this scenario, which maximizes the number of possible gene losses, the last common ancestor of all lophotrochozoans is hypothesized to have had 11 genes, four of which were lost in the lineage leading to neoophoran platyhelminthes. Color coding as in Fig. 2

lophotrochozoans (they do not show primitive characteristics). For example, *Post-2* genes from the platyhelminthes share several synapomorphic residues with those of other lophotrochozoans. Assuming that they are basal to both *Post-1* and *Post-2* genes from other lophotrochozoans, these synapomorphies would have been subsequently lost in *Post-1* genes. Similar arguments can be made for the position of *Lox4* genes from flatworms. Finally, *Lox5* genes from the platyhelminthes possess a well-conserved Lox5 parapeptide; if these genes are basal to *Lox5* and *Hox7* from other lophotrochozoans, this parapeptide would have had to be secondarily lost in

Hox7 genes. Furthemore, Telford (2000b) has proposed that *Lox5* is orthologous to ecdysozoan *ftz* genes, and that *Hox7* could be orthologous to *Antp*. If this is correct, then both genes would have been present in the last common protostome ancestor and, therefore, in the last common lophotrochozoan ancestor.

Acknowledgments The authors thank Laura Dominguez and Jenny Saldaña for providing infected mice, and Madelón Portela for technical assistance and the Automatic Sequencing of the Faculty of Sciences. We are very grateful to the *Echinococcus multilocularis* and *Schistosoma mansoni* Sequencing Groups at the Sanger Institute for permitting us to work with the genomic assemblies, and to Najib M. El-Sayed (University of Maryland), and Klaus Brehm (Universität Würzburg). This work was supported by IFS, DINACYT, and a M.Sc. fellowship from PEDECIBA to A.I.L.

References

Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387:489–493

Ax P (1996) Multicellular animals—a new approach to the phylogenetic order in nature—vol I. Springer Verlag, Berlin

Balavoine G (1997) The early emergence of platyhelminths is contradicted by the agreement between 18S rRNA and *Hox* genes data. C R Acad Sci III 320:83–94

Balavoine G, Telford MJ (1995) Identification of planarian homeobox sequences indicates the antiquity of most Hox/homeotic gene subclasses. Proc Natl Acad Sci USA 92:7227–7231

Balavoine G, de Rosa R, Adoutte A (2002) *Hox* clusters and bilaterian phylogeny. Mol Phylogenet Evol 24:366–373

Bartels JL, Murtha MT, Ruddle FH (1993) Multiple *Hox*/HOM-class homeoboxes in Platyhelminthes. Mol Phylogenet Evol 2:143–151

Bayascas JR, Castillo E, Munoz-Marmol AM, Salo E (1997) Planarian *Hox* genes: novel patterns of expression during regeneration. Development 124:141–148

Britos L, Dominguez L, Ehrlich R, Marin M (2000) Effect of praziquantel on the strobilar development of Mesocestoides corti in vitro. J Helminthol 74:295–299

Bürglin T (1994) A comprehensive classification of homeobox genes. In: Duboule D (ed) Guidebook to the homeobox genes. Oxford University Press, Oxford

Chourrout D, Delsuc F, Chourrout P, Edvardsen RB, Rentzsch F, Renfer E, Jensen MF, Zhu B, de Jong P, Steele RE, Technau U (2006) Minimal ProtoHox cluster inferred from bilaterian and cnidarian *Hox* complements. Nature 442:684–687

de Rosa R, Grenier JK, Andreeva T, Cook CE, Adoutte A, Akam M, Carroll SB, Balavoine G (1999) *Hox* genes in brachiopods and priapulids and protostome evolution. Nature 399:772–776

Ferrier D (2007) Evolution of *Hox* gene cluster. In: Papageorgiou S (ed) HOX gene expression. Springer Science + Business Media, New York

Ferrier DE, Holland PW (2001) Sipunculan Parahox genes. Evol Dev 3:263–270

Garcia-Fernàndez J (2005) *Hox*, Parahox, ProtoHox: facts and guesses. Heredity 94:145–152

Gehring W (2007) The homeobox as a key for understanding the principles of the genetic control of development. In: Papageorgiou S (ed) HOX gene expression. Springer Science + Bussiness Media, New York

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98

Helmkampf M, Bruchhaus I, Hausdorf B (2008) Multigene analysis of lophophorate and chaetognath phylogenetic relationships. Mol Phylogenet Evol 46:206–214

Iglesias M, Gomez-Skarmeta JL, Saló E, Adell T (2008) Silencing of Smed-betacatenin1 generates radial-like hypercephalized planarians. Development 135:1215–1221

Kim KH, Lee YS, Jeon HK, Park JK, Kim CB, Eom KS (2007) *Hox* genes from the tapeworm *Taenia asiatica* (Platyhelminthes: Cestoda). Biochem Genet 45:335–343

Kulakova M, Bakalenko N, Novikova E, Cook CE, Eliseeva E, Steinmetz PR, Kostyuchenko RP, Dondua A, Arendt D, Akam M, Andreeva T (2007) *Hox* gene expression in larval development of the

polychaetes *Nereis virens* and *Platynereis dumerilii* (Annelida, Lophotrochozoa). Dev Genes Evol 217:39–54

Kulakova MA, Cook CE, Andreeva TF (2008) Parahox gene expression in larval and postlarval development of the polychaete *Nereis virens* (Annelida, Lophotrochozoa). BMC Dev Biol 8:61

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Larroux C, Fahey B, Degnan SM, Adamski M, Rokhsar DS, Degnan BM (2007) The NK homeobox gene cluster predates the origin of *Hox* genes. Curr Biol 17:706–710

Lartillot N, Philippe H (2008) Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. Philos Trans R Soc Lond B Biol Sci 363:1463–1472

McManus DP, Knight M, Simpson AJ (1985) Isolation and characterisation of nucleic acids from the hydatid organisms, Echinococcus spp. (Cestoda). Mol Biochem Parasitol 16:251–266

Minguillón C, Garcia-Fernàndez J (2003) Genesis and evolution of the Evx and Mox genes and the extended *Hox* and Parahox gene clusters. Genome Biol 4:R12

Nielsen C (2001) Animal evolution: interrelationships of the living phyla. Oxford University Press, Oxford

Nogi T, Watanabe K (2001) Position-specific and non-colinear expression of the planarian posterior (Abdominal-B-like) gene. Dev Growth Differ 43:177–184

Ogishima S, Tanaka H (2007) Missing link in the evolution of *Hox* clusters. Gene 387:21–30

Olson PD (2008) *Hox* genes and the parasitic flatworms: new opportunities, challenges and lessons from the free-living. Parasitol Int 57:8–17

Orii H, Kato K, Umesono Y, Sakurai T, Agata K, Watanabe K (1999) The planarian HOM/HOX homeobox genes (Plox) expressed along the anteroposterior axis. Dev Biol 210:456–468

Pierce RJ, Wu W, Hirai H, Ivens A, Murphy LD, Noel C, Johnston DA, Artiguenave F, Adams M, Cornette J, Viscogliosi E, Capron M, Balavoine G (2005) Evidence for a dispersed *Hox* gene cluster in the platyhelminth parasite Schistosoma mansoni. Mol Biol Evol 22:2491–2503

Saló E, Tauler J, Jimenez E, Bayascas JR, Gonzalez J, Garcia Fernandez J, Baguña J (2001) *Hox* and parahox genes in flatworms: characterization and expression. Amer Zool 41:652–663

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Tarabykin VS, Lukyanov KA, Potapov VK, Lukyanov SA (1995) Detection of planarian Antennapedia-like homeobox genes expressed during regeneration. Gene 158:197–202

Telford MJ (2000a) Turning *Hox* "signatures" into synapomorphies. Evol Dev 2:360–364

Telford MJ (2000b) Evidence for the derivation of the *Drosophila fushi tarazu* gene from a *Hox* gene orthologous to lophotrochozoan Lox5. Curr Biol 10:349–352

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Webster PJ, Mansour TE (1992) Conserved classes of homeodomains in Schistosoma mansoni, an early bilateral metazoan. Mech Dev 38:25–32