Hox Gene Loss during Dynamic Evolution of the Nematode Cluster

A. Aziz Aboobaker* and Mark L. Blaxter
Institute of Cell, Animal and Population Biology
University of Edinburgh
Edinburgh EH9 3JT
Scotland
United Kingdom

Summary

Hox genes are important: their central role in anterior-posterior patterning provides a framework for molecular comparison of animal body plan evolution [1]. The nematode *Caenorhabditis elegans* stands out as having a greatly reduced Hox gene complement [2]. To address this, orthologs of *C. elegans* Hox genes were identified in six species from across the Nematoda, and they show that rapid homeodomain sequence evolution is a general feature of nematode Hox genes. Some nematodes express additional Hox genes belonging to orthology groups that are absent from *C. elegans* but present in other bilaterian animals. Analysis of the genomic environment of a newly identified *Brugia malayi* Hox6-8 ortholog (*Bm-ant-1*) revealed that it lay downstream of the *Bm-egl-5* Hox gene and that their homeodomain exons are alternately *cis* spliced to the same 5′ exon. This organization may represent an intermediate state in Hox gene loss via redundancy. The Hox clusters of nematodes are the product of a dynamic mix of gene loss and rapid sequence evolution, with the most derived state observed in the model *C. elegans*.

Results and Discussion

Hox genes are homeodomain transcription factors found closely clustered in the genomes of animals [3]. Comparison of Hox gene complements between phyla suggests that the ancestral protostome cluster contained at least nine members (ortholog groups) and that most extant taxa have representatives of all of them [4]. One species that does not follow the conservative nature of Hox gene evolution is the nematode *C. elegans* [2, 5]. Only six Hox genes are present, and they are organized as three gene pairs across 5 Mb of chromosome III. They can be assigned to only four ortholog groups [5] (the anterior gene *Ce-ceh-13*; two distinct central genes, *Ce-lin-39* and *Ce-mab-5*; and three posterior group genes, *Ce-egl-5*, *Ce-php-3*, and *Ce-nob-1* [6]). Firm assignment of the *Ce-lin-39* and *Ce-mab-5* genes to canonical ortholog groups of other protostomes has been contentious, due to their extreme sequence divergence [3]. While this may be a plesiomorphic retention of a simple, ancestral cluster [5], current hypotheses of bilaterian phylogeny suggest that nematodes are a crown member of the protostome superphylum Ecdyso-

*Correspondence: aziz.aboobaker@ed.ac.uk

zoa [7, 8]; this finding implies that at least five Hox genes have been lost during the evolution of *C. elegans* (Figure 1). In order to understand the evolutionary processes leading to the state observed in *C. elegans*, we have exploited the robust nematode molecular phylogeny [9] to examine Hox gene evolution across the phylum.

C. elegans is a rhabditid, a member of nematode Clade V described by Blaxter et al. [9]. We have surveyed a range of species that embraces the diversity of the phylum: *Pristionchus pacificus* (Clade V), *Strongyloides stercoralis* (Clade IV), *Meloidogyne javanica* (Clade IV), *Ascaris suum* (Clade III), *Brugia malayi* (Clade III), and *Trichinella spiralis* (Clade I). We were able to identify orthologs of all the *C. elegans* Hox genes, except the divergent posterior gene *nob-1*, across the phylum. These new sequences clarify orthology assignment of nematode Hox genes (Figure 1), and, in particular, they affirm that *mab-5* is a *ftz/Lox5* ortholog [10]. The absence of the highly conserved *Deformed/Hox4* LPNTK C-terminal peptide [4] suggests that *lin-39* is a *Sex combs reduced/Hox5* ortholog. The posterior gene *nob-1* is a recent duplication event in the *Caenorhabditis* lineage [2, 6].

Additional Hox genes, orthologs of bilaterian genes that are definitively absent from the complete genome of *C. elegans* and from the 7x coverage of the *C. briggsae* genome [11], were identified in nematodes from Clades IV, III, and I. A *Hox3* ortholog (*hox-3*) was found in *M. javanica*, *B. malayi*, *A. suum*, and *T. spiralis*. A central *antennapaedia-like* gene (*ant-1*) was found in *B. malayi*, *A. suum*, and *T. spiralis* (see the Supplementary Material available with this article online for aligned sequences). These genes were not recovered by using identical procedures from *P. pacificus*. Hox cluster reduction in nematodes is therefore not due to a single event, but has been occurring through the evolution of the phylum. Loss of *hox-3* and *ant-1* can be inferred to have taken place since *C. elegans* last shared a common ancestor with nematodes in Clades IV and III, respectively (Figure 1). The remaining Hox gene losses either occurred before the radiation of the nematode crown group, or they occurred within the nematode lineage.

As Hox genes play core roles in development, it is difficult to conceive how loss of multiple genes could arise [12]. The exquisite specificity of Hox protein regulation of transcription is in part mediated via interactions between their N-terminal, nonhomeodomain portions with DNA binding partners [13]. Cloning of full-length cDNAs from the *B. malayi egl-5* and *ant-1* genes revealed that they have an identical 5′ sequence (see the Supplementary Material). Genomic analysis showed that their homeodomain exons are alternatively spliced to a shared 5′ exon that encodes an N-terminal domain with similarity to the *Ce*-EGL-5 N terminus (Figure 2A and the Supplementary Material). This confirms that *ant-1* is clustered with other Hox genes, and it strongly suggests that an *Ubx/AbdA* gene is absent in this species. These two genes have overlapping but distinct temporal expression patterns, with that of of *Bm-ant-1* contained
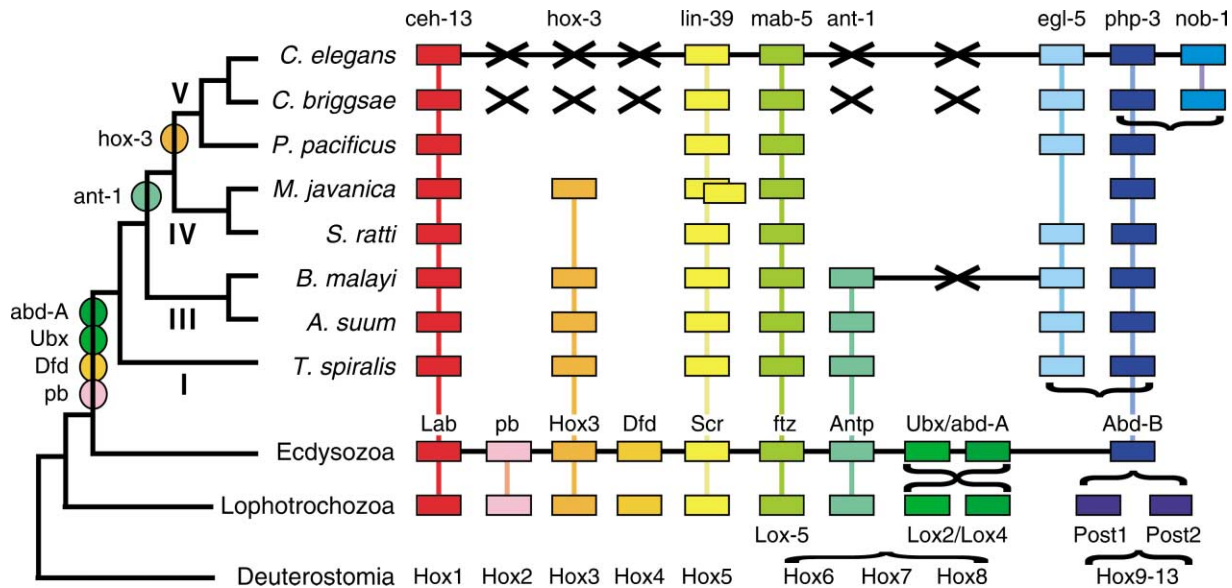
Figure 1. Hox Genes in Nematodes

The isolated Hox genes are indicated by boxes that are colored by orthology group. Horizontal lines indicate linkage, and crosses indicate Hox genes that are missing from the *C. elegans* genome and 7x coverage of the *C. briggsae* genome [11] and the absence of a central gene between *B. malayi ant-1* and *egl-5*. Colored vertical lines through boxes indicate confident orthology assignment based on characteristic residue and phylogenetic analyses (see the Supplementary Material). Curly brackets indicate possible lineage-specific gene duplications. The tree on the left summarizes the robust SSU rDNA phylogeny for the nematodes [9]. Circles colored by orthology groups on the tree are positioned to indicate the earliest possible time of Hox gene loss from the *C. elegans* lineage based on the data presented here. The *Pp-mab-5* and *Pp-lin-39* genes have been described previously [30, 31].

entirely within that of *Bm-egl-5* (Figure 2B). We hypothesize that this organization may represent an intermediate state in the evolutionary processes that have lead to the absence of *ant-1* in *C. elegans*. Sharing an N terminus may result in partial overlap between binding partners and target sites and, subsequently, full redundancy and gene loss.

Current models of the modes of evolution of Hox gene function involve gene duplications [14], micro [15] and macro [16] changes in expression pattern, and changes in sequences outside the 60 amino acid homeodomain [17, 18]. In general, the homeodomains evolve slowly, but, when Hox genes are divorced from homeotic function, as has happened with *Hox3* and *ftz* genes in the diptera [10, 19–21], their homeodomains are observed to evolve more rapidly. The independently duplicated posterior-group Hox genes of deuterostomes also have elevated rates of substitution [22]. We find that all of the nematode ortholog groups show elevated substitution rates across the phylum when compared to genes from arthropods [21] (Figure 3 and the Supplementary Material for details of analysis) and other bilaterians (data not shown, see the Supplementary Material). By analogy to other systems, the functions of all the nematode Hox genes may have changed rapidly across the phylum, as constraint on all the Hox homeodomains has been lost.
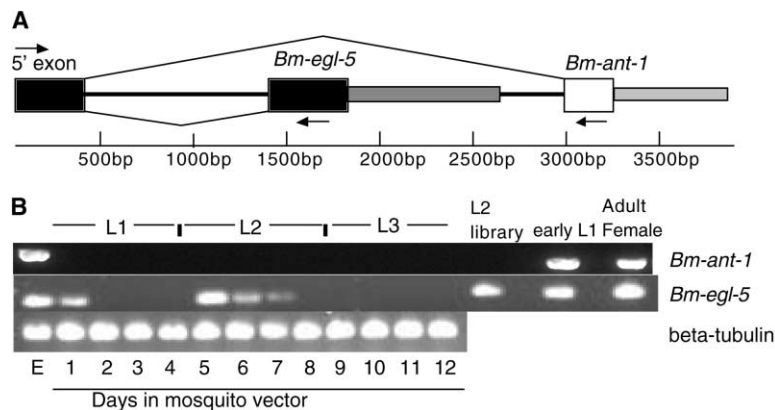

Figure 2. *Bm-egl-5* and *Bm-ant-1* Hox Genes Share a 5′ Exon

(A) *Bm-ant-1* and *Bm-egl-5* share the same 5′ exon and are in close proximity in the genome (see the Supplementary Material for alignments). The black boxes are the *egl-5* exons; the white box is the *ant-1* 3′ exon. Gray boxes indicate 3′ UTRs, and arrows represent primers used for life cycle expression analysis.

(B) *Bm-ant-1* and *Bm-egl-5* have overlapping but different expression patterns (see the Supplementary Material for Experimental Procedures). *Bm-ant-1* mRNA is detected by RT-PCR in developing embryos (E) and in early L1 larvae before entry into the mosquito vector. *Bm-egl-5* mRNA is detected in embryos, in early L1, and during the first day of the life cycle, and then again between the larval molts (which occur at days 4–5 and days 8–9) during days 5, 6, and 7, correlating with the presence of the transcript in an L2 cDNA library. Both mRNAs are detected in adult female cDNA, and this is probably reflective of embryonic expression in utero. β-tubulin mRNA levels were measured from the same samples to confirm the evenness of the amount of starting material.
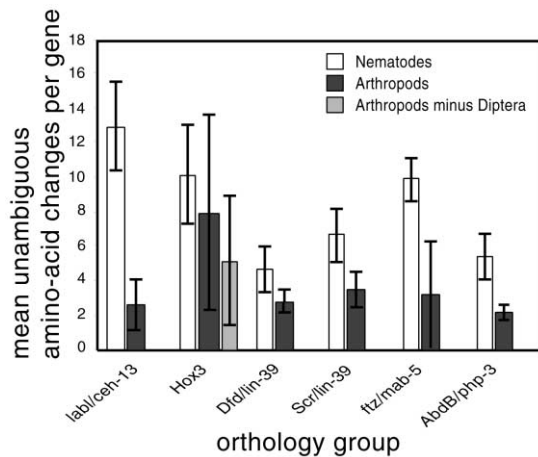
Figure 3. Elevated Homeodomain Evolutionary Rates in the Nematoda

Orthologous homeodomains were aligned across the bilaterians or protostomes depending on whether orthology could be confidently assigned. Maximum parsimony was used to infer the number of unambiguous amino acid changes in each arthropod and nematode sequence since the last common ancestor of the two phyla, based on a specified species phylogeny [9, 28] (using Maclade 4.02 [29], see the Experimental Procedures and the Supplementary Material for details and alignments used). These data were used to calculate the mean number of changes for each orthology group in each phylum (the error bars represent one standard deviation of the mean). For the Hox3 orthology group, two separate analyses, including and excluding the rapidly evolving dipteran fly sequences [20], were performed. The mean evolutionary rate of the homeodomain is higher for all nematode genes assessed than that for corresponding arthropod genes, including the rapidly evolving *Hox3/zen/bicoid* [20] and *ftz* genes [10]. The *lin-39* gene was compared to both *Dfd* and *Scr* genes. The *egl-5* genes have not been included, as it remains unclear whether they derive from a nematode-specific duplication or whether they are orthologous to a posterior gene lost from arthropods. The very divergent *Sr-php-3* gene was excluded from the analysis.

Nematodes are renowned for sharing a conservative body plan. The model *C. elegans* has a strongly determinative, lineage-driven mode of development [23], resulting in an invariant cell lineage and eutely. Hox gene functions in *C. elegans* have been evolving within this deterministic developmental mode, and their expression is now cell lineage, and not cell position, dependent [24]. Is this linked to Hox gene loss? Not all nematodes have an invariant lineage [25], but the loss of Hox genes is not clearly related to the evolution of this mode of development. Both "new" nematode Hox genes are found in Clade III and IV species that have a cell lineage during early embryogenesis very similar to that of *C. elegans* [26]. We suggest a three-step process in which a lineage-dependent mechanism of development was first adopted, ultimately releasing some Hox genes from a core role in positional identity pathways, followed by recruitment of these genes to new function in the context of lineage. Once a gene is released from its essential role, it is free to be lost, possibly through the exon-sharing mechanism observed for *B. malayi ant-1* and *egl-5*, or to move rapidly through sequence space to assume novel functions.

Experimental Procedures

Hox genes were identified and clones were extended by using techniques described elsewhere [4, 21], with additional degenerate primer sets used to allow for divergent nematode homeodomains and the inclusion of a subtractive PCR screening approach of a large number of cloned inserts used to allow the identification of rare inserts (see the Supplementary Material for details). PCR from *B. malayi* genomic DNA was used to investigate the organization of the *Bm-ant-1* and *Bm-egl-5* genes. Characteristic residue and phylogenetic analyses (using PAUP*4.0b10 [27]) were used to assign sequences to orthology groups (see the Supplementary Material for details). Rates of substitution in homeodomains were measured by constraining a tree of the homeodomains to a phylogeny of the species derived from other genes [9, 28] (see the Supplementary Material for these trees). We compared the mean number of unambiguous amino acid changes in each orthology group across Nematoda and Arthropoda since their last common ancestor by using MacClade 4.03 [29].

Supplementary Material

Supplementary Material including additional Experimental Procedures, sequence alignments of the *Bm-eg15/ant-1* genes with their *C. elegans* orthologs, alignments of the nematode Hox domain sequences identified in this study and a set of bilaterian orthologs, and the phylogenetic trees summarized in Figure 3 is available at http://images.cellpress.com/supmat/supmatin.htm. Nexus alignments of Hox genes from different phyla, along with two supplementary tables, are available on the authors' web site at http://www.nematodes.org/evo_devo.html.

References

1. Carroll, S.B. (1995). Homeotic genes and the evolution of arthropods and chordates. Nature *376*, 479–485.
2. Ruvkun, G., and Hobert, O. (1998). The taxonomy of developmental control in *Caenorhabditis elegans*. Science *282*, 2033–2041.
3. Burglin, T.R. (1994). A comprehensive classification of homeobox genes. In Guidebook to the Homeobox Genes, D. Duboule, ed. (Oxford, United Kingdom: Oxford University Press), pp. 25–71.
4. de Rosa, R., Grenier, J.K., Andreeva, T., Cook, C.E., Adoutte, A., Akam, M., Carroll, S.B., and Balavoine, G. (1999). Hox genes in brachiopods and priapulids and protostome evolution. Nature *399*, 772–776.
5. Burglin, T.R., and Ruvkun, G. (1993). The *Caenorhabditis elegans* homeobox gene cluster. Curr. Opin. Genet. Dev. *3*, 615–620.
6. Van Auken, K., Weaver, D.C., Edgar, L.G., and Wood, W.B. (2000). *Caenorhabditis elegans* embryonic axial patterning requires two recently discovered posterior-group Hox genes. Proc. Natl. Acad. Sci. USA *97*, 4499–4503.
7. Aguinaldo, A.M., Turbeville, J.M., Linford, L.S., Rivera, M.C., Garey, J.R., Raff, R.A., and Lake, J.A. (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. Nature *387*, 489–493.
8. 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.
9. Blaxter, M.L., De Ley, P., Garey, J.R., Liu, L.X., Scheldeman, P.,

Vierstraete, A., Vanfleteren, J.R., Mackey, L.Y., Dorris, M., Frisse, L.M., et al. (1998). A molecular evolutionary framework for the phylum Nematoda. Nature *392*, 71–75.

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

11. The *Caenorhabditis briggsae* Genome Sequencing Consortium. (2002). The *Caenorhabditis briggsae* Genome Project: (http://genome.wustl.edu/projects/cbriggsae/index.php).

12. Holland, P.W. (1999). The future of evolutionary developmental biology. Nature *402*, C41–C44.

13. Passner, J.M., Ryoo, H.D., Shen, L., Mann, R.S., and Aggarwal, A.K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. Nature *397*, 714–719.

14. Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y.L., et al. (1998). Zebrafish hox clusters and vertebrate genome evolution. Science *282*, 1711–1714.

15. Stern, D.L. (1998). A role of Ultrabithorax in morphological differences between *Drosophila* species. Nature *396*, 463–466.

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

17. Ronshaugen, M., McGinnis, N., and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. Nature *415*, 914–917.

18. Galant, R., and Carroll, S.B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. Nature *415*, 910–913.

19. Telford, M.J., and Thomas, R.H. (1998). Of mites and zen: expression studies in a chelicerate arthropod confirm *zen* is a divergent Hox gene. Dev. Genes Evol. *208*, 591–594.

20. Stauber, M., Prell, A., and Schmidt-Ott, U. (2002). A single Hox3 gene with composite *bicoid* and *zerknullt* expression characteristics in non-Cyclorrhaphan flies. Proc. Natl. Acad. Sci. USA *99*, 274–279.

21. 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.

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

23. Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev. Biol. *100*, 64–119.

24. Cowing, D., and Kenyon, C. (1996). Correct Hox gene expression established independently of position in *Caenorhabditis elegans*. Nature *382*, 353–356.

25. Voronov, D.A., and Panchin, Y.V. (1998). Cell lineage in marine nematode *Enoplus brevis*. Development *125*, 143–150.

26. zurStrassen, O. (1906). Die Geschicte der T-Reisen von *Ascaris megalocephala* als Grundlage zu einer Entwickelungsmechanik dieser Spizies. Zoologica, Stuttgart *17*, 39–342.

27. Swofford, D.L. (1998). PAUP* Phylogenetic Analysis Using Parsimony (* and Other Methods), Version 4 (Sunderland, MA: Sinauer).

28. Giribet, G., Edgecombe, G.D., and Wheeler, W.C. (2001). Arthropod phylogeny based on eight molecular loci and morphology. Nature *413*, 157–161.

29. Maddison, W.P., and Maddison, D.R. (1989). Interactive analysis of phylogeny and character evolution using the computer program MacClade. Folia Primatol *53*, 190–202.

30. Jungblut, B., and Sommer, R.J. (1998). The *Pristionchus pacificus mab-5* gene is involved in the regulation of ventral epidermal cell fates. Curr. Biol. *8*, 775–778.

31. Eizinger, A., and Sommer, R.J. (1997). The homeotic gene *lin-39* and the evolution of nematode epidermal cell fates. Science *278*, 452–455.