Heredity (2005) 94, 145–152

www.nature.com/hdy

REVIEW

Hox, ParaHox, ProtoHox: facts and guesses

J Garcia-Fernàndez
Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal, 645, E-08028, Barcelona, Spain

The Hox gene cluster has captivated the imagination of evolutionary and developmental biologists worldwide. In this review, the origin of the Hox and ParaHox gene clusters by duplication of a ProtoHox gene cluster, and the changes in their gene numbers in major Metazoan Transitions are reviewed critically. Re-evaluation of existing data and recent findings in Cnidarians, Acoels, and critical stages of vertebrate evolution suggest alternative scenarios for the origin, structure, and changes in Hox gene numbers in relevant events of Metazoan evolution. I discuss opposing views and propose that (i) the ProtoHox cluster had only two genes, and not four as commonly believed: a corollary is that the origin of Bilaterians was coincident with the invention of new Hox and ParaHox gene classes, which may have facilitated such a transition; (ii) the ProtoHox cluster duplication was a *cis* duplication event, rather than a *trans* duplication event, as previously suggested, and (iii) the ancestral vertebrate cluster possessed 14 Hox genes, and not the 13 generally assumed. These hypotheses could be verified or refuted in the near future, but they may help critical discussion of the evolution of the Hox/ParaHox family in the metazoan kingdom.
Heredity (2005) **94,** 145–152. doi:10.1038/sj.hdy.6800621
Published online 1 December 2004

Keywords: Hox evolution; Evo-Devo; gene duplication; gene clusters; metazoan transitions; Bilaterians

Introduction

Since Edward Lewis' seminal work on the Bithorax complex of Drosophila (Lewis. 1978), the Hox gene cluster has fascinated evolutionary and developmental biologists alike. Hox genes belong to a class of homeobox genes, a major class of transcription factors regulating many aspects of development (Gellon and McGinnis, 1998). The discovery of the homeobox in 1984 (McGinnis *et al*, 1984) and the finding that similar genes act in similar ways in animals as diverse as flies and mice facilitated the reconciliation of developmental and evolutionary biologists. This came after a century of tormented relationships that followed an initial honeymoon, inspired by the Darwinian 'descent with modification' theory of evolution ('community of embryonic structure reveals community of descent', Darwin, 1859). Since then, experimental embryologists have concentrated on unravelling the mechanisms of embryological processes, while evolutionary biologists have followed the changes of gene frequencies in natural populations. Now, the new field of Evo-Devo, at the frontier of both disciplines, seeks a new 'developmental synthesis of evolution' (Gilbert, 2003). The conservation of Hox genes galvanized the Evo-Devo community and served to define the concept of the metazoan 'zootype', a conserved set of genes patterning the antero-posterior body axis (Slack *et al*, 1993). Soon after that, changes in Hox gene numbers, sequence, and regulation were invoked for body plan evolution and diversification (eg Gellon and McGinnis, 1998; Wagner *et al*, 2003; Amores *et al*, 2004).

Correspondence: J Garcia-Fernàndez, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal, 645, E-08028, Barcelona, Spain. E-mail: jgarcia@bio.ub.es
Received 26 July 2004; accepted 13 October 2004; published online 1 December 2004

What made Hox genes special among developmental regulators is not only their organization in chromosomal complexes, with nine genes in flies and 39 genes in four clusters in mammals (Figure 1), but also the phenomenon of spatial and temporal colinearity. Genes at one end of the cluster are expressed, and pattern the anterior end of the embryo, while genes at the other end of the cluster pattern the posterior end (Duboule and Dollé, 1989). This spatial colinearity (5′ equals posterior, 3′ equals anterior) is a direct consequence in some lineages (mainly vertebrates) of temporal colinearity. Genes at the 3′ end of the cluster are expressed earlier in development than genes at the 5′ end (Duboule, 1994); hence, temporal colinearity in developmental systems that grow from anterior to posterior leads immediately to spatial colinearity. The molecular mechanisms of colinearity are elusive, although chromatin remodelling and the physical topography of chromosomal regions have recently been implicated (Duboule and Deschamps, 2004).

The evolution of Hox genes in metazoans is still not fully understood. Comparison of mammalian and arthropod Hox clusters has led to general agreement that the last common ancestor of Protostomes and Deuterostomes, the two groupings of 'higher' metazoans, had a single Hox cluster composed of seven genes (Figure 1): two from the Anterior Group (paralogous groups (PGs) 1–2 in mammals, Drosophila genes *labial* and *proboscipedia*), one Group 3 gene (PG3, Drosophila *zen* gene), three representatives of the Central Group (PGs4–5, Drosophila *Deformed* and *sex-comb reduced*, and an ancestor of PGs6–8, Drosophila *Antennapedia*, *Ultrabithorax* and *Abdominal-A*), and a single Posterior Group gene (ancestor of PGs9–13, Drosophila *Abdominal-B*). In both Protostome and Deuterostome lineages, this original cluster followed distinct evolutionary pathways, with additional tandem duplications in the central and

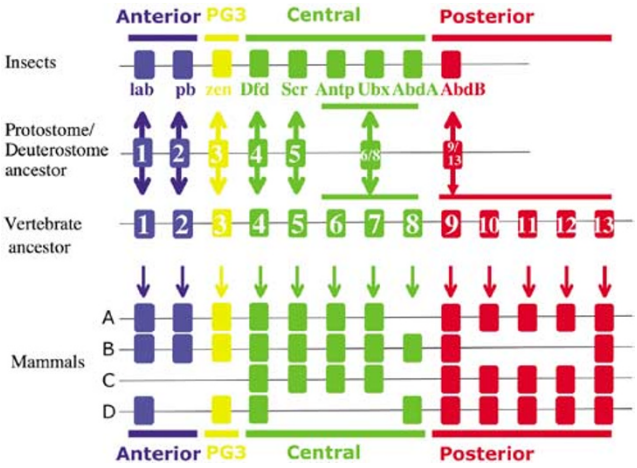


Figure 1 Structure of the insect and mammalian Hox clusters and the cluster structure inferred for the last common ancestor of Protosotomes and Deuterostomes, and for the last vertebrate ancestor, prior to the cluster duplications in the vertebrate lineage (but see Figure 5 for alternative views). Hox genes are grouped in Anterior, Group3, Central, and Posterior classes based in sequence similarities. Numbers and arrows indicate orthology relationships.

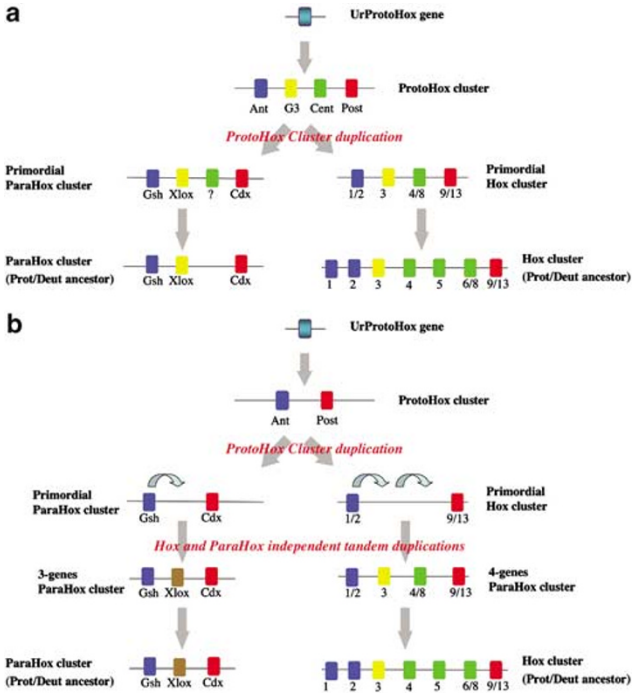


Figure 2 Genesis and evolution of the Hox and ParaHox clusters. (**a**) Four-gene model ProtoHox cluster. A ProtoHox cluster containing four genes, one of each class (Anterior, PG3, Central, and Posterior) duplicated giving raise to the Primordial Hox and ParaHox clusters. The ParaHox cluster lost a Central gene, and the Hox cluster expanded up to seven members in the last common ancestor of Protostomes and Deuterostomes. (**b**) Two-gene model of the ProtoHox cluster. A ProtoHox cluster containing two genes (one Anterior and one Posterior) gave rise, by duplication, to the Primordial, two-gene containing, Hox and ParaHox clusters. The composition of the three- and four-gene containing ParaHox and Hox clusters, respectively, was due to single gene tandem duplications, independently in both clusters. Later in evolution, the four-gene Hox cluster further expanded by tandem duplication, up to seven genes in the last common ancestor of Protostomes and Deuterostomes.

posterior region that account for the present-day composition of the complexes (reviewed in de Rosa *et al*, 1999; Ferrier and Minguillón, 2003).

The Hox cluster was believed to originate by tandem duplication from an ancestral 'central' Hox gene until 1998, with the discovery of the ParaHox cluster (Brooke *et al*, 1998), an evolutionary sister complex of the Hox cluster. This finding indicated that a hypothetical ProtoHox cluster of four genes duplicated early in animal evolution, giving rise to two twin clusters. These would be the primordial Hox cluster, which expanded by *cis* duplication to eight genes in Drosophila, or to 13 paralogous groups in mammals, and the primordial ParaHox cluster, which lost one member and gave rise to the three-gene complex maintained at least in cephalochordates and vertebrates (Figure 2a). Here, I summarize recent data on Hox and ParaHox genes and cluster contents in distinct lineages, and discuss various hypotheses for the evolution of Hox and ParaHox gene clusters. In particular, I discuss: (i) the origin and original structure of the Hox and ParaHox clusters, pointing to the evolutionary changes that accompanied the origin of Bilaterians and the Cambrian Explosion, (ii) the 'accompanying' genes of the Hox and ParaHox clusters, or whether the ProtoHox cluster duplication was a *cis* or a *trans* event, and (iii) the basal content of the Hox gene cluster in the vertebrate lineage.

Guess 1: Two or four? The ProtoHox cluster

The finding of three Hox-related genes closely linked in the amphioxus genome, *AmphiGsh*, *AmphiXlox*, and *AmphiCdx*, plus their phylogenetic relationship to the Hox clusters led to the ProtoHox cluster hypothesis (Brooke *et al*, 1998). Based on sequence similarity, Hox genes can be classified in four groups: the Anterior Group (PGs1–2), Group3, Central Group (PGs4–8) and Posterior Group (PGs9–13). Gsh genes are more closely related to Hox Anterior Group genes than to other ParaHox genes; Xlox is more similar to Hox Group3, and Cdx is more similar to Hox Posterior Group genes. Hence, the model predicts a ProtoHox complex with four genes (Figure 2a): An Anterior ProtoHox gene, ancestor of an Anterior Hox gene (PG1/2) and an Anterior ParaHox (Gsh) gene, a Group 3 ProtoHox gene (ancestor of Hox PG3 and Xlox), a Central ProtoHox gene, (ancestor of a Central Hox gene (PG4/8), a central ParaHox was lost soon after ProtoHox cluster duplication), and a Posterior ProtoHox gene (ancestor of a Posterior Hox gene PG9/13 and Cdx). New discoveries on the ParaHox cluster (eg Finnerty and Martindale, 1999; Yanze *et al*, 2001; Ferrier and Holland, 2003; Cook *et al*, 2004) and reviews on Hox/ParaHox evolution (eg Ferrier and Holland, 2001; Martínez and Amemiya, 2002; Ferrier and Minguillón, 2003) always present an evolutionary framework with a ProtoHox cluster with four genes. Here, I propose an alternative scenario for the structure of the ProtoHox cluster and the primordial Hox and ParaHox clusters, which takes into account a reconsideration of the available data and recent discoveries on lower metazoans.

Brooke *et al* (1998) proposed that the ProtoHox cluster duplicated at the same time as the Cambrian Explosion, when major protostome and deuterostome phyla ap-

peared. However, it was soon found that dipoblast Cnidarians already possessed Hox and ParaHox genes (Finnerty and Martindale, 1999). Hence, ProtoHox cluster duplication must have occurred before the divergence of Cnidarians and Tripoblastic Animals. The search for Hox and ParaHox in Dipoblast animals has been intense. Distinct species of Cnidarians, Anthozoans, Hydrozoans, Cubozoans, and Scyphozoans, have been extensively searched for Hox and ParaHox genes. A major consensus is emerging (summarized in Finnerty, 2003; Finnerty *et al*, 2004): the basal content of the Cnidarian Hox/ParaHox clusters was one Anterior and one Posterior Hox gene, and one Anterior and one Posterior ParaHox gene (Figure 3). Despite intense efforts, no PG3 or central Hox or ParaHox genes have been found in any Cnidarian species. Thus, two conclusions are possible: (i) according to the four-gene ProtoHox cluster model, PG3 Hox and ParaHox were lost in all Cnidarians, and a central ParaHox was lost in all Cnidarians and Bilaterians (Figure 3a), or (ii) The ProtoHox cluster and the primordial Hox and ParaHox clusters had only Anterior and Posterior genes, in other words, the ProtoHox cluster consisted of only two genes (Figure 2b).

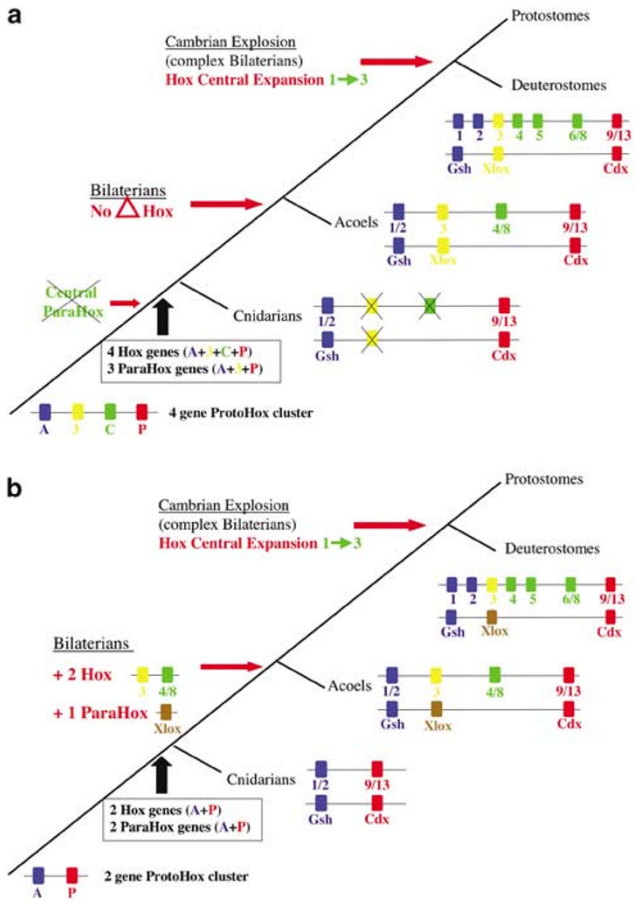


Figure 3 Changes in Hox and ParaHox numbers associated to major Metazoan Transitions. (**a**) Under the four-gene ProtoHox model, the origin of Bilaterians was not accompanied by changes in Hox or ParaHox gene numbers. In addition, Cnidarians lost at least three Hox and ParaHox genes. (**b**) Under the two-gene ProtoHox model, the origin of Bilaterians was coincident by the invention of two Hox (PG3 and Central groups) and one ParaHox (Xlox) classes.

The four-gene model is supported by phylogenetic analyses of the 60 aa homeodomain alone (or the homeodomain plus five flanking aa), of proteins that diverged more than 800 million years ago (Banerjee-Basu and Baxevanis, 2001; Peterson *et al*, 2004). The bootstrap values for any particular grouping of Hox *vs* ParaHox genes range from 40 to 70% (Brooke *et al*, 1998; Finnerty and Martindale, 1999; Banerjee-Basu and Baxevanis, 2001; Minguillón and Garcia-Fernàndez, 2003). These values are far below the confidence rate of phylogenetic analyses with 18S RNA or full protein sequences to establish the relationships among early divergent clades (Hillis and Bull, 1993). In addition, the four-gene cluster model necessarily implies that two classes of Hox genes (Group 3 and Central) and one (Group 3) or two (Group 3 plus Central) ParaHox genes were independently lost in the Cnidarian lineage (Figure 3a). Furthermore, two genes rather than four in Cnidarians may be more consistent with evolutionary considerations (Figure 3b). Hox genes in Bilaterians pattern the antero-posterior body axis. Cnidarian Hox genes display staggered expression along the oral-aboral axis (Finnerty, 2003; Finnerty *et al*, 2004), although data on the polarity of such expression are puzzling (Masuda-Nakagawa *et al*, 2000; Yanze *et al*, 2001; Finnerty, 2003; Finnerty *et al*, 2004). Colinearity of Hox genes in Cnidarians, if any, is consequently difficult to reconcile with only two Hox genes.

Recent data on the Hox complement of early Bilaterians help us to envisage alternative scenarios. Recent progress in molecular phylogeny has shown that Acoelomorphs (*Acoela* + *Nemertodermatida*), former members of the protostomian phylum Platyhelminthes, represent the earliest extant bilaterian clade (Ruiz-Trillo *et al*, 1999, reviewed in Baguñà and Riutort, 2004). Hence, the privileged intermediate position of *Acoelomorpha*, as a simple, unsegmented, acelomated Bilaterian, may help us to understand the evolution of Hox clusters. Do Acoels have 'canonical' higher bilaterian Hox and ParaHox clusters (seven Hox genes + three ParaHox genes), or do they possess an early version of Hox and ParaHox clusters? The Cnidarian/Bilaterian transition may have been accompanied by an increase in the numbers of Hox and ParaHox genes, as they are implicated in the diversification of the antero-posterior body axis. If the four-gene model is correct, and the common ancestor of Cnidarians and Bilaterians already had four Hox genes and three to four ParaHox genes, one would expect to find more complex clusters in Acoels (eg up to the seven-gene Hox cluster of the last common ancestor of Protostomes and Deuterostomes). If the two-gene model is correct and Cnidarians represent this old condition, one would expect to find more than two genes in the Acoel clusters. Very recent data (Cook *et al*, 2004; Baguñà and Riutort, 2004) strongly suggest that Acoels possesses a Hox cluster with four genes (one Anterior, one Group 3, one Central, and one Posterior members) and a ParaHox cluster with three genes, one of each canonical class. Hence, under the four-gene model (Figure 3a), the Cnidarian/Bilaterian transition was not accompanied by an increase in the Hox or ParaHox complement but, if the two-gene model is correct (Figure 3b), the Cnidarian/Bilaterian transition was accompanied by the emergence of two Hox genes (Group 3 and Central), and one ParaHox gene (Xlox). It is

attractive to imagine that clusters with at least three genes were powerful tools to differentially pattern the newly acquired antero-posterior axis of Bilaterians, in contrast to the early two-gene clusters of Cnidarians.

Intermediate hypotheses can also be advanced, for example, a ProtoHox cluster with three genes (Anterior, Group3 and Posterior). In this case, Cnidarians would have lost PG3 and Xlox, and the increase in complexity at the origin of Bilaterians would have been linked only to the origin of a Central Hox gene. This model squares better with the close phylogenetic relationship of PG3 and Xlox. However, the argument still requires that two genes (PG3 and Xlox) were lost independently in the Cnidarian lineage. With a two-gene model, the only assumptions needed are that invention of new genes was linked to an increase in body plan complexity, and that the phylogenetic grouping of Xlox/PG3 is artificial or due to enigmatic functional convergence.

In summary, based on (i) the intensive searches on Cnidarians that depict Hox and ParaHox clusters with two genes, (ii) the finding that the early Bilaterians had Hox and ParaHox clusters with four and three genes, respectively, (iii) the expanded Hox cluster (seven Hox genes) of the basal complex Bilaterians, (iv) the low bootstrap values of Hox and ParaHox grouping, and (v) the suggestion that numbers of Hox and ParaHox genes may well be correlated with increase in body plan complexity, and antero-posterior diversification, I propose the following hypothesis for some of the major Metazoan evolutionary transitions (Figure 3b): First, a ProtoHox cluster with two genes was duplicated predating the Cnidarian/Bilaterian transition. Present Cnidarians have clusters that are direct descendents of the 2-Hox, 2-ParaHox primordial clusters. Second, the Cnidarian/Bilaterian transition was accompanied by the expansion, by tandem duplication, of the Hox and ParaHox clusters independently: in the Hox cluster, a PG3 and a Central Group genes originated by duplication of the Anterior gene, or by shuffling and diversification; in the ParaHox cluster, Xlox originated by tandem duplication of the Anterior ParaHox gene. Third, during early steps of bilaterian evolution, the Hox complex underwent further gene duplications in the central part of the cluster, originating the PG4, PG5, and PGs6–8 founder genes. In Protostome and Deuterostome lineages, the central and posterior Hox genes still generated by duplication the content of their respective Hox cluster. This framework directly invokes changes in Hox/ParaHox gene contents in the major transitions of Metazoans: (i) duplication of the ProtoHox cluster before the Cnidarian/Bilaterian divergence, (ii) expansion of the Hox and ParaHox clusters coincidentally with the origin of Bilaterians, and (iii) further central expansion of the Hox cluster coinciding with the appearance of complex Bilaterians. It is tempting to speculate that these changes in gene numbers were causal to these major transitions.

Whether it contained two or four genes, there is no doubt that the ProtoHox cluster duplicated earlier than the divergence of Bilaterians from Cnidarians. Where the ProtoHox cluster originated or duplicated is unclear. The correct branching of lower metazoans, namely Ctenophores, Placozoans, and Sponges is not clear (eg Martindale *et al*, 2002; Ender and Schierwater, 2003) and searches for a ProtoHox cluster have been unsuccessful. A single Hox-like gene may be present in Ctenophores (Finnerty *et al*, 1996) and no complete Hox-like genes have been reported in sponges. The recent claim that the placozoan gene *Trox2* may be derived from the original Hox/ParaHox gene (the UrProtoHox gene, Figure 2) (Jakob *et al*, 2004), although exciting, warrants further investigation.

Guess 2: *cis* or *trans*? Tandem duplication and the 'coupled' array of Hox and ParaHox clusters

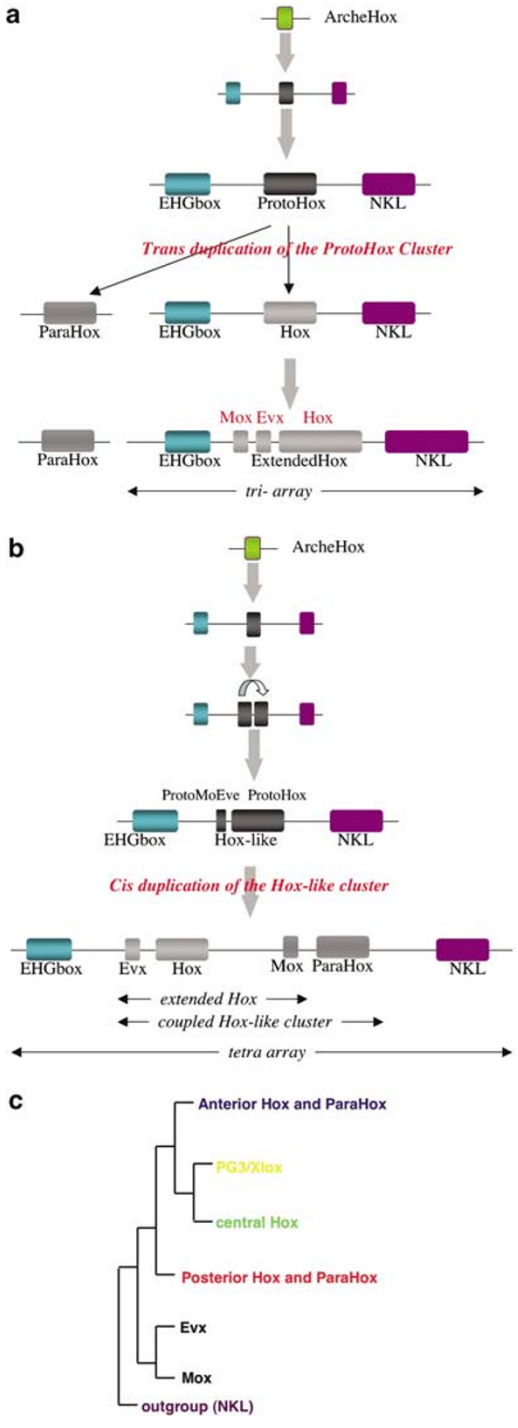
The Hox-like gene Evx is closely linked to the Hox clusters in vertebrates (Dush and Martin, 1992; Amores *et al*, 1998; Powers and Amemiya, 2004), and in a Cnidarian species (Miller and Miles, 1993). The most plausible hypothesis to explain the linkage is that the primordial Hox cluster, before the divergence of Cnidarians and Bilaterians, was already closely linked to Evx. The analyses of the human genome by Pollard and Holland (2000) suggested that four homeobox clusters of the Antennapedia family, the extended Hox cluster (the Hox cluster plus the related homeobox genes Evx and Mox), the NKL cluster (which includes homeobox genes like Nkx, Msx, Dlx, Tlx Emx, or Lbx), and the EHGbox cluster (including En, HB9, Gbx) arose by tandem gene duplication and cluster duplications from an ancestral UrArcheHox gene early in metazoan evolution (Figure 4a). Furthermore, evidence of such clusters is also to be found in the genomes of amphioxus and Drosophila (Castro and Holland, 2003; Luke *et al*, 2003). In such a model, the ProtoHox, EHGbox, and NKL clusters would have arisen by successive tandem duplications from an ancestral cluster of founder genes of each class (ProtoHox, ProtoEHGbox, and ProtoNKL genes). Hence, the ProtoHox cluster would have duplicated nontandemly (by *trans* duplication), with the primordial Hox cluster remaining next to the EHGbox and NKL primordial clusters, whereas the primordial ParaHox cluster would have jumped to other positions of the genome (Figure 4a). This proposal was based on the human genome mapping of 2000, where Evx and Mox were mapped to the *same* end (the 5′end) of the Hox cluster, and assumed that both Evx and Mox arose from tandem duplications of genes of the Hox cluster.

Again, phylogenetic analyses were puzzling: Evx genes fall basal to the Hox/ParaHox clade (Gauchat *et al*, 2000; Kourakis and Martindale, 2000). This position suggests that Evx appeared *before* the duplication event that generated the Hox/ParaHox primordial clusters. The Mox class has rarely been included in phylogenetic analyses, and has vaguely been referred to as the missing ParaHox central gene (Gauchat *et al*, 2000; Hill *et al*, 2003). Two extensive phylogenetic analyses (Banerjee-Basu and Baxevanis, 2001; Minguillón and Garcia-Fernàndez, 2003) suggest that Evx and Mox are closely related, forming a clade basal to the Hox/ParaHox group (Figure 4c). Hence, Evx and Mox arose by duplication of an ancestor (Evx/Mox ancestor or ProtoMoEve) that falls back to the duplication event for the genesis of the Hox/ParaHox clusters (Figure 4b). The pylogenetic data thus suggest that the ProtoMoEve gene was *adjacent* to the ProtoHox cluster.

Several scenarios may be envisaged for tracing the particular duplication events of the ProtoMoEve/ProtoHox cluster (referred to here as the Hox-like cluster) that

fit with current phylogenetic data: (i) an early *cis* duplication of ProtoMoEve into Evx and Mox, and later, the *trans* duplication of the ProtoHox cluster, as suggested by Pollard and Holland (2000). This will result in Evx and Mox *both* next to the *same side* of the Hox cluster. (ii) Evx, Mox, Hox, and ParaHox as result of a single *trans* duplication event of the Hox-like cluster. This will result in Evx next remaining at the 5′ end of the Hox cluster, and Mox at the 5′ end of the ParaHox cluster. To date, Mox has not been found next to the ParaHox cluster in any lineage. And (iii) Evx, Mox, Hox, and ParaHox, as result of a single *cis* duplication event of the extended Hox-like cluster (Figure 4b). This will result in a 'coupled Hox-like cluster' with an array of Evx-Hox-Mox-ParaHox (note Evx is located at the 5′ end of the Hox cluster, and Mox is located at 3′end of the Hox cluster and at 5′ end of the ParaHox cluster). Again, an intact array of such genes has not been found in any lineage. However, the definite mapping of the human and mice genomes (Spring, 2002) indicated that Mox is located about 5 Mb at the 3′ end of the Hox cluster, contrary to what was believed. This location matches with the hypothesis of a tandem duplication event of the coupled Hox-like cluster, by simply including in the model a chromosomal breakage at either side of the ParaHox cluster. Such breakage would have left Evx and Mox at either side of the Hox cluster, and left the ParaHox cluster isolated. A tandem duplication followed by chromosomal breakage nicely squares phylogenetic and current linkage analyses. Interestingly, the characterization of the Hox content in the Urochordate *Oikopleura dioca* shows that Cdx is linked to Hox1 (Seo *et al*, 2004), as predicted by the *cis* duplication model. Whether this linkage is a remnant of the ancestral condition or simply serendipity, and whether the Hox/ParaHox breakage happened only once early in evolution, or many times, remains to be determined.

The tandem duplication followed by chromosomal breakage of the ancestral Hox-like cluster adds a refinement to the model of three homeobox clusters (ProtoHox, EHGbox, and NKL) linked together (Figure 4a). It implies that, early in metazoan evolution (certainly before the Cnidarian/Bilaterian split, see Guess 1), an extended array of *four* homeobox clusters (EGHbox, Hox-with Evx-, ParaHox-with Mox-, and NKL), a tetra-array of Antennapedia-like homeobox primordial clusters had existed (Figure 4b). As none was a *trans*-duplication event, the model also implies that neither duplication generating Antp-like homeobox clusters was the product of a genome polyplodization. By extension, a full genome duplication event predating the Cnidarian/Bilaterian split seems unlikely.



Guess 3: 13 or 14? A vertebrate 14th Hox gene

Mammals have four Hox clusters with representatives of the 13 paralogous groups PG1 to PG13, as the result of gene loss after duplication early in vertebrate evolution

Figure 4 ProtoHox duplication and Antennapedia-like arrays of homeobox clusters. (**a**) Model for the *trans* duplication of the ProtoHox cluster as proposed by Pollard and Holland (2000). The ProtoHox cluster duplicated in *trans*, isolating the ParaHox cluster, and leaving intact an array of three Antp-like clusters: EHGbox, Hox, and NKL. Note that Evx and Mox are both at the *same side* of the Hox cluster. (**b**) Model for the *cis* duplication of the ProtoHox cluster, based on Minguillón and Garcia-Fernàndez (2003). An Evx/Mox ancestor (ProtoMoEve) lied at the 5′ end of the ProtoHox cluster. The *cis* duplication of the Hox-like cluster (ProtoMoEve plus ProtoHox) resulted in an array of four homeobox clusters, with Hox and ParaHox flanked at the 5′end by Evx and Mox, respectively. Subsequently, chromosomal breakage between Mox and the ParaHox cluster account for Evx and Mox lying at *either side* of the Hox cluster. (**c**) Phylogenetic relationship of Evx and Mox with respect to Hox and ParaHox genes. Evx and Mox form a clade basal to all Hox/ParaHox genes. This position implies that an Evx/Mox ancestor (ProtoMoEve) existed before the duplication of the ProtoHox cluster. Adapted from Banerjee-Basu and Baxevanis (2001) and Minguillón and Garcia-Fernàndez (2003).

from a single cluster (Figure 1). Hence, it was believed that such basal vertebrate cluster should have possessed 13 genes, one of each PG. The initial finding of the single Hox cluster of amphioxus, the invertebrate sister group of vertebrates (Garcia-Fernàndez and Holland, 1994), fitted the model nicely. However, further analyses of the cephalochordate cluster surprisingly revealed not a 13-gene cluster, but a 14th gene, *AmphiHox-14* (Ferrier *et al*, 2000). Phylogenetic analyses of the Posterior Hox genes of Chordates (PG9–PG13 or 14) are obscured by their higher evolutionary rate, a phenomenon called Posterior Flexibility (Ferrier *et al*, 2000). Such analyses have not revealed whether *AmphiHox-14* represented a lineage-specific duplication in the amphioxus genome (Figure 5a), or was a remnant of a vertebrate 14th paralogous group, which was subsequently lost early in vertebrate evolution (Figure 5b).

Very recently, a finding by Powers and Amemiya (2004) has strengthened the amphioxus data; the analyses of the HoxA cluster of the coelacanth and the HoxD cluster of the horn shark revealed in each case an additional 14th gene, *A14* and *D14*, respectively. Furthermore, the HoxA cluster of the horn shark also has a pseudogene of a Hoxa14 gene. Sharks are early Gnathostome representatives, and coelachants are early sarcopterygian fishes (lobe-finned fishes and tetrapods, including mammals). These findings necessarily imply that the common ancestor of Gnathostomes already possessed a 14th Hox gene (Figure 5a and b). The phylogenetic analyses of these new 14th genes do not resolve whether *AmphiHox-14* is pro-ortologous of the vertebrate 14th genes, probably as another consequence of Posterior Flexibility. However, position in the cluster and sharing of an intron (Ferrier *et al*, 2000; Powers and Amemiya, 2004) argue in favour of the hypothesis that the single ancestral Vertebrate, Chordate, or Deuterostomian cluster contained 14 PGs.

A 14-gene model may be investigated by searching for such a gene in lower vertebrates (Agnathans), lower Chordates (Urochordates), or other Deuterostomians (Hemichordates and Echinoderms). The available data do not help to resolve the issue. Published and available data on Agnathans, Urochordates, Echinoderms, and Hemichordates suggest multiple posterior genes PGs9–13: at least five in Agnatans (Fried *et al*, 2003), up to six in Urochordates (Spagnuolo *et al*, 2003; Seo *et al*, 2004), at least four in Echinoderms (Martínez *et al*, 1999), and at least three in Hemichordates (Peterson, 2004). Posterior Flexibility, again, hampers the resolution of the particular relationships between deuterostomian posterior genes. In summary, Deuterostomes have multiple posterior genes, up to six members (PGs9–14) in some lineages, and current data suggest that the original cluster that duplicated twice in the vertebrate genome had 14 genes as the most parsimonious hypothesis (Figure 5b). The organization of such a cluster would completely match the single cluster of the cephalochordate amphioxus.

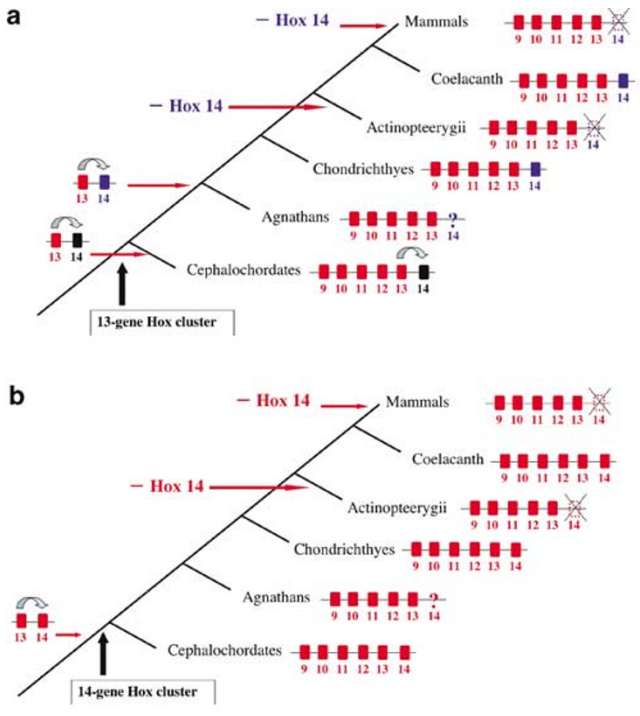


Figure 5 Alternative scenarios for the basal content of the vertebrate cluster, prior to cluster duplications. (**a**) 13-Gene Hox gene model. Independent duplications in the cephalochordate and vertebrate lineages, and independent losses in particular vertebrate lineages account for current Hox gene numbers. (**b**) 14-Gene Hox gene model. Only a single duplication event needs to be assumed prior to the cephalochordate/vertebrate divergence.

Conclusions

The analyses of recent data lead to new hypotheses for the origin and structure of the well-known Hox and ParaHox clusters, at critical crossroads of metazoan evolution. These hypotheses are based on phylogenetic information supplemented by extensive linkage information, and are the most parsimonious models. In summary, I propose:

(i) that the original ProtoHox cluster had only two genes, an Anterior and a Posterior gene, which duplicated before the Cnidarian/Bilaterian divergence. Major Metazoan Transitions, such as the origin of symmetry, the origin of Bilaterians, and the Cambrian Explosion, were accompanied by the increase in the complexity of ProtoHox, Hox, or ParaHox clusters
(ii) that a complex array of four homeobox clusters (EgHbox, extended Hox, ParaHox, and NKL) existed early in metazoan evolution
(iii) that the ancestral vertebrate cluster had expanded to 14 paralogous groups.

These hypotheses remain to be tested which will require extensive analysis of the genomes of animals that illuminate those Metazoan Transitions.

Acknowledgements

I thank Jaume Baguñà, Gemma Marfany, Carolina Minguillón, and Peter Holland for encouraging and passionate discussions, and to Robin Rycroft for checking the english. The author's research is funded by Grants BMC2002-03316 (Ministerio de Ciencia y Tecnología, Spain), by the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya (Distinció per la Promoció de la Recerca Universitaria),

and by European Community's Human Potential Programme HPRN-CT-2002-00263 'Neurogenome'.

References

Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A *et al* (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science* **282**: 1711–1714.

Amores A, Suzuki T, Yan YL, Pomeroy J, Singer A, Amemiya C *et al* (2004). Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. *Genome Res* **14**: 1–10.

Baguñà J, Riutort M (2004). The dawn of bilaterian animals. The case of acoelomorph flatworms. *BioEssays* **26**: 1046–1057.

Banerjee-Basu S, Baxevanis AD (2001). Molecular evolution of the homeodomain family of transcription factors. *Nucleic Acids Res* **29**: 3258–3269.

Brooke NM, Garcia-Fernàndez J, Holland PWH (1998). The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* **392**: 920–922.

Castro LF, Holland PWH (2003). Chromosomal mapping of ANTP class homeobox genes in amphioxus: piecing together ancestral genomes. *Evol Dev* **5**: 459–465.

Cook CE, Jiménez E, Akam M, Saló E (2004). The Hox gene complement of acoel flatworms, a basal bilaterian clade. *Evol Dev* **6**: 154–163.

Darwin C (1859). *On the Origin of Species*. John Murray: London.

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

Duboule D (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development* (Suppl): 135–142.

Duboule D, Deschamps J (2004). Colinearity loops out. *Dev Cell* **6**: 738–740.

Duboule D, Dollé P (1989). The structural and functional organization of the murine HOX gene family resembles that of Drosophila homeotic genes. *EMBO J* **8**: 1497–1505.

Dush MK, Martin GR (1992). Analysis of mouse Evx genes: Evx-1 displays graded expression in the primitive streak. *Dev Biol* **151**: 273–287.

Ender A, Schierwater B (2003). Placozoa are not derived cnidarians: evidence from molecular morphology. *Mol Biol Evol* **20**: 130–134.

Ferrier DEK, Holland PWH (2001). Ancient origin of the Hox gene cluster. *Nat Rev Genet* **2**: 33–38.

Ferrier DEK, Holland PWH (2003). Ciona intestinalis ParaHox genes: evolution of Hox/ParaHox cluster integrity, developmental mode, and temporal colinearity. *Mol Phylogenet Evol* **24**: 412–417.

Ferrier DEK, Minguillón C (2003). Evolution of the Hox/ParaHox gene clusters. *Int J Dev Biol* **47**: 605–611.

Ferrier DEK, Minguillón C, Holland PWH, Garcia-Fernàndez J (2000). The amphioxus Hox cluster: deuterostome posterior flexibility and Hox14. *Evol Dev* **2**: 284–293.

Finnerty JR (2003). The origins of axial patterning in the metazoa: how old is bilateral symmetry? *Int J Dev Biol* **47**: 523–529.

Finnerty JR, Martindale MQ (1999). Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria. *Evol Dev* **1**: 16–23.

Finnerty JR, Master VA, Irvine S, Kourakis MJ, Warriner S, Martindale MQ (1996). Homeobox genes in the Ctenophora: identification of paired-type and Hox homologues in the atentaculate ctenophore, Beroe ovata. *Mol Mar Biol Biotechnol* **5**: 249–258.

Finnerty JR, Pang K, Burton P, Paulson D, Martindale MQ (2004). Origins of bilateral symmetry: Hox and dpp expression in a sea anemone. *Science* **304**: 1335–1337.

Fried C, Prohaska SJ, Stadler PF (2003). Independent Hox-cluster duplications in lampreys. *J Exp Zoolog Part B Mol Dev Evol* **299**: 18–25.

Garcia-Fernàndez J, Holland PWH (1994). Archetypal organization of the amphioxus Hox gene cluster. *Nature* **370**: 563–566.

Gauchat D, Mazet F, Berney C, Schummer M, Kreger S, Pawlowski J *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.

Gellon G, McGinnis W (1998). Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *BioEssays* **20**: 116–125.

Gilbert SF (2003). Opening Darwin's black box: teaching evolution through developmental genetics. *Nat Rev Genet* **4**: 735–741.

Hill A, Wagner A, Hill M (2003). Hox and paraHox genes from the anthozoan Parazoanthus parasiticus. *Mol Phylogenet Evol* **28**: 529–555.

Hillis DM, Bull JJ (1993). An empirical test of bootstraping as a method for assessing confidence in phylogenetic analysis. *Sys Biol* **42**: 182–192.

Jakob W, Sagasser S, Dellaporta S, Holland P, Kuhn K, Schierwater B (2004). The Trox-2 Hox/ParaHox gene of Trichoplax (Placozoa) marks an epithelial boundary. *Dev Genes Evol* **214**: 170–175.

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

Lewis EB (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**: 565–570.

Luke GN, Castro LF, McLay K, Bird C, Coulson A, Holland PWH (2003). Dispersal of NK homeobox gene clusters in amphioxus and humans. *Proc Natl Acad Sci USA* **100**: 5292–5295.

Martindale MQ, Finnerty JR, Henry JQ (2002). The Radiata and the evolutionary origins of the bilaterian body plan. *Mol Phylogenet Evol* **24**: 358–365.

Martínez P, Amemiya CT (2002). Genomics of the HOX gene cluster. *Comp Biochem Physiol B Biochem Mol Biol* **133**: 571–580.

Martínez P, Rast JP, Arenas-Mena C, Davidson EH (1999). Organization of an echinoderm Hox gene cluster. *Proc Natl Acad Sci USA* **96**: 1469–1474.

Masuda-Nakagawa LM, Groer H, Aerne BL, Schmid V (2000). The HOX-like gene Cnox2-Pc is expressed at the anterior region in all life cycle stages of the jellyfish Podocoryne carnea. *Dev Genes Evol* **210**: 151–156.

McGinnis W, Hart CP, Gehring WJ, Ruddle FH (1984). Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of Drosophila. *Cell* **38**: 675–680.

Miller DJ, Miles A (1993). Homeobox genes and the zootype. *Nature* **365**: 215–216.

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**: R12.1–R12.8.

Peterson KJ (2004). Isolation of Hox and ParaHox genes in the hemichordate *Ptychodera flava* and the evolution of deuterostome Hox genes. *Mol Phylogenet Evol* **31**: 1208–1215.

Peterson KJ, Lyons JB, Nowak KS, Takacs CM, Wargo MJ, McPeek MA (2004). Estimating metazoan divergence times with a molecular clock. *Proc Natl Acad Sci USA* **101**: 6536–6541.

Pollard SL, Holland PWH (2000). Evidence for 14 homeobox gene clusters in human genome ancestry. *Curr Biol* **10**: 1059–1062.

Powers TP, Amemiya CT (2004). Evidence for a Hox14 paralog group in vertebrates. *Curr Biol* **14**: R183–R184.

Ruiz-Trillo I, Riutort M, Littlewood DT, Herniou EA, Baguñà J (1999). Acoel flatworms: earliest extant bilaterian Metazoans, not members of Platyhelminthes. *Science* **283**: 1919–1923.

Seo HC, Edavardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A *et al* (2004). *Hox* cluster disintegration with

persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* **431**: 67–71.

Slack JM, Holland PWH, Graham CF (1993). The zootype and the phylotypic stage. *Nature* **361**: 490–492.

Spagnuolo A, Ristoratore F, Di Gregorio A, Aniello F, Branno M, Di Lauro R (2003). Unusual number and genomic organization of Hox genes in the tunicate *Ciona intestinalis*. *Gene* **309**: 71–79.

Spring J (2002). Genome duplication strikes back. *Nat Genet* **31**: 128–129.

Wagner GP, Amemiya C, Ruddle F (2003). Hox cluster duplications and the opportunity for evolutionary novelties. *Proc Natl Acad Sci USA* **100**: 14603–14606.

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