

Minireview

Beyond the Hox complex

Nipam H Patel* and Victoria E Prince†

Address: *Department of Organismal Biology and Anatomy and Howard Hughes Medical Institute and †Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL 60637, USA.

Correspondence: Nipam H Patel. E-mail: npatel@midway.uchicago.edu

Published: 8 November 2000

Genome Biology 2000, 1(5):reviews1027.1–1027.4

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2000/1/5/reviews/1027>

© GenomeBiology.com (Print ISSN 1465-6906; Online ISSN 1465-6914)

Abstract

The Hox complex is an example of a gene cluster created by tandem duplications. Recent findings suggest the Hox complex may be just part of a larger chromosomal assemblage of homeobox-containing genes that existed in the ancestor to all vertebrates.

All the genes in all the organisms that exist today presumably derive in one way or another from the contents of the relatively small genomes that were present in the microbes that inhabited the earth some 3.5 billion years ago. Through mutation, gene rearrangement, duplication, and so on, the process of evolution has generated the remarkable array of genes that we now study. As complete genome sequences become available, we can begin to track the evolutionary history of many gene families, and in so doing we can begin to understand some of the details of the evolution of life.

Duplication of genes clearly plays an important role in generating molecular diversity. Sometimes, these duplications arise through the duplication of entire chromosomes or large chromosomal regions. Other times, duplications appear as tandem copies of genes - both *Drosophila* and *Caenorhabditis elegans* contain hundreds of gene pairs that appear to be the result of such tandem duplications [1] - and a number of these tandem duplications are thought to arise through unequal crossing over at meiosis. On occasion, this process of localized duplication can lead to the creation of a large cluster of genes. For example, the human genome contains several clusters of odorant receptor genes, with some clusters containing more than a dozen genes [2].

Our current thinking is that, once duplicated, tandem gene pairs often take on separable genetic functions. This can happen in any of a number of ways. Through changes in the coding region, the protein products of the two genes may

take on biochemically distinct functions. Alternatively, depending on the relative boundaries of the unequal crossing-over event and the position of the original gene's enhancer elements, the two copies may 'inherit' different components of the original gene's expression pattern, thus subdividing the functions of the original gene; they may also be forced to share common regulatory elements through this process. Alternatively, changes in expression pattern between two copies might arise through mutations in one or more enhancers some time after the duplication. Obviously, this process of duplication and divergence can continue over and over again, and many other occurrences are also possible (for example, exon shuffling, generation of alternative transcripts, and the evolution of novel enhancer elements). When we look at the genomes of extant organisms, we can identify a number of gene families that have been generated through these sorts of evolutionary processes. One gene family, the homeobox-containing family of genes, has been particularly well studied in this regard.

Piecing together ancestral genomes

In a recent paper, Pollard and Holland [3] have made good use of the ever-expanding pool of information freely available on the databases to reinvestigate the evolutionary relationships between the large number of so-called *Antennapedia*-superclass homeobox genes. Homeobox genes encode transcription factors, many of which are involved in regulating developmental processes in animals

and plants. Primarily from sequence alignment data, homeobox genes can be subdivided into evolutionarily related families, including the Hox family. In addition to having closely related homeobox sequences, the Hox family of genes also appears to have retained a clustered organization in the genome over hundreds of millions of years of evolution and through several rounds of gene duplication [4]. For example, Hox genes have maintained their clustered organization in all vertebrate lineages examined so far. For the most part, however, we have generally thought of the remaining homeobox-containing genes as being dispersed through the genome, with a few examples of tandem duplicates. What Pollard and Holland argue, however, is that there once existed a much larger array of homeobox-containing genes - all relatively close to one another on a single chromosome. They term this group the *Antennapedia*-superclass homeobox genes. Although *Antennapedia* itself is a Hox gene, the superclass contains many more distantly related homeobox-containing genes that are not part of the Hox family, including several other recognized families such as the engrailed and NK families.

How is it that Pollard and Holland can make an argument for such a large cluster of genes in a common ancestor that lived hundreds of millions of years ago? They do so not by relying on information from any one organism, but by combining data from multiple species. One small subset of the linkage, that of a couple of the NK class genes, is relatively straightforward: multiple NK genes are localized to one chromosomal band in humans and orthologs of these genes reside in one restricted chromosomal location in *Drosophila*. But the majority of their argument, however, is derived in a more ingenious manner. They are able to line up linked arrays of genes in a 'jigsaw' manner between humans and mice, taking advantage of a wide array of additional genes that happened to be linked to these homeobox-containing genes (such as genes for keratins, hedgehog proteins and FGF receptors).

Why does this work? It works because there is a reasonable probability that if two genes were near to one another in the common ancestor to all vertebrates, they will still show detectable linkage in any given extant vertebrate. Such patterns of evolutionarily conserved linkage give rise to what we term synteny between genomes. For example, the available human and mouse data suggest that approximately 180 chromosomal inversions and translocations could 'convert' one genome into the other (in terms of gene order) [5]. Taking an even bigger evolutionary step, it is estimated that when two genes are adjacent (on the same cosmid) in the pufferfish, *Fugu rubripes*, there is roughly a 40-50% chance that the orthologous genes will still show linkage with each other in humans [6]. Given that fish and humans last shared a common ancestor 400 million years ago [7], that means that 800 million years of combined independent evolutionary history has roughly a 40-50% chance of separating two adjacent

genes. Given that the last common ancestor of all vertebrates lived about 600 million years ago [7], this means that any two genes that were adjacent to each other in this ancestral animal would have roughly a 50% chance of still being linked to each other today.

There are obviously a number of assumptions implicit in Pollard and Holland's argument. The rate of chromosomal aberrations may be higher in some lineages than others and the empirical data do not distinguish between what happens when there is a functional consequence to the separation of two genes (such as when they share enhancer regions) versus when there is no consequence to the rearrangement. Despite these uncertainties, however, it still follows that if genes A, B, E, G and H are linked in that order in one vertebrate, and genes A, C, D, F and H are linked in that order in a second vertebrate, there is a chance that genes A, B, C, D, E, F, G, and H were all linked together in the common ancestor of both vertebrates. The more linkage data that are available, the greater the ability to predict accurately what was present in the common ancestor. The implication is that the arrangement of genes in the ancestor to all vertebrates might be deduced once we know the order of genes in a number of diverse extant vertebrates.

Ancient homeobox gene clusters

From their data, Pollard and Holland [3] suggest the following evolutionary scenario (see Figure 1). In some common ancestor of all animals, there existed an ArcheHox complex, possibly composed of three genes. Each of these three genes underwent tandem duplications to generate the progenitor clusters for the genes that Pollard and Holland term EHGBox genes, ProtoHox genes, and Proto-NKL genes. Some time before the divergence in the lineages leading to humans and to jellyfish, the ProtoHox gene complex was duplicated, with one lineage leading to the ParaHox group, which includes Cdx, Xlox and Gsx (leaving it unlinked from the rest of the *Antennapedia* superfamily genes), and the other to the Hox family (called Extended Hox by Pollard and Holland because it contains two additional genes, Evx and Mox, beyond the traditional Hox set).

Additional tandem duplications are then thought to have occurred, such that the common ancestor of all vertebrates contained a large array of homeobox-containing genes on a single chromosome. Within this array were three evolutionarily distinguishable clusters, the EHGBox cluster (which included Engrailed, HB9 and Gbx), the Extended Hox cluster defined above, and the NKL cluster (which included Msx, NK, Lbx and other genes; see Figure 1). The ParaHox cluster was also present, but was not linked to the other three clusters. Whole chromosome duplications occurred later on within the vertebrate lineage, and some gene copies were lost as well. In addition, random chromosomal rearrangements began to break up this very large array of

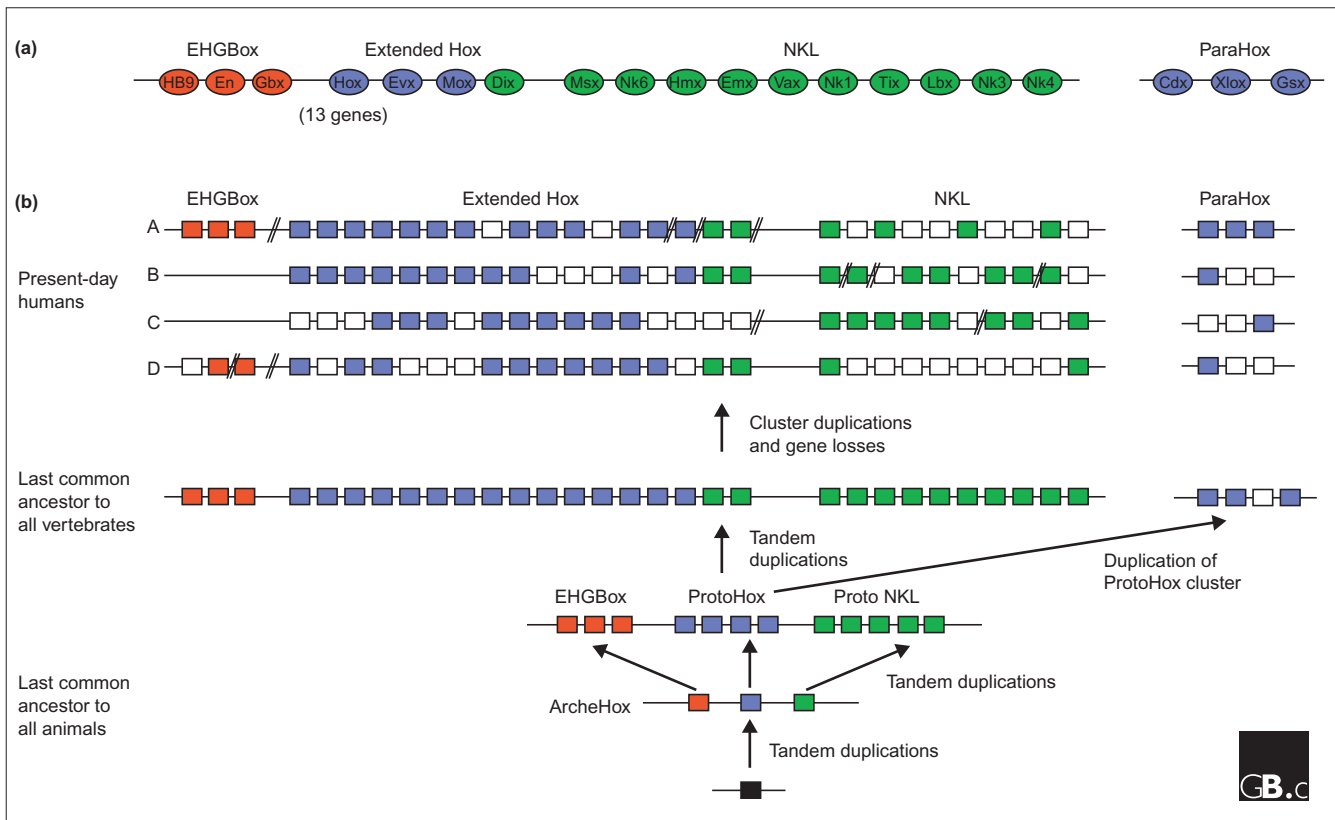


Figure 1

Evolution of homeobox gene organization. Pollard and Holland [3] suggest that the current organization of *Antennapedia*-superclass homeobox genes might have occurred through the illustrated process of gene duplication, chromosome duplication, gene loss, and chromosome rearrangement. **(a)** The genes thought to have been linked into one cluster in the last common ancestor of all vertebrates. **(b)** Scenario of gene duplication (arrows), gene losses (white boxes) and losses of linkage (double slashes) proposed by Pollard and Holland [3]. Adapted from [3].

genes, giving rise to the current pattern of linkage observed today in humans and mice.

This scenario provides us with the opportunity to frame a number of very interesting questions. What led to the formation of such a large cluster of homeobox-containing genes (possibly 30 genes all in one linkage group)? What were the ancestral roles of such an array of transcription factors? How did these functions diversify during evolution? Were there constraints on the breakup of this cluster, and do any constraints still apply? If we narrow these questions down to just the Hox genes, we already have some ideas. We believe that Hox genes played an ancestral role in the axial patterning of the body plan, a role that they still play in animals as diverse as *Drosophila* and mice. Much of the diversification of these genes took place fairly early in animal evolution, before the separation of the protostome lineage (which includes insects) and the deuterostome lineage (which includes vertebrates), but additional duplications via additional tandem duplications and whole chromosome duplications occurred in the vertebrate lineage and may have contributed to the morphological diversification of this group [8]. Within both arthropods and

vertebrates, regulatory changes in these genes also appear to have played a role in morphological diversification. Will the same sorts of things be true for genes of the NKL, ParaHox, and EHGBox genes?

Preserving genome organization

We know that there must be some strong constraints for maintaining clustering for the Hox genes, especially within the vertebrate lineage. This includes a requirement to maintain temporal and spatial colinearity. Colinearity describes how the positions of the genes within the complex correlate with their relative temporal and spatial expression patterns - that is, the genes at one end of the complex are expressed earlier and more anteriorly than those at the other end. Evidence for a requirement of clustering for spatial colinearity has come from extensive analysis of Hox gene regulation in the mouse system, which has revealed that regulatory elements are often shared by more than one gene or even embedded within adjacent genes [9]. The requirement for clustering for temporal colinearity has been addressed in a series of elegant experiments from the lab of Denis Duboule



[10-12]. They have shown that the position of a gene within the cluster affects the time of its expression, and this is at least somewhat independent of the immediately adjacent regulatory sequences that control the spatial aspects of expression. Hence, the most important force maintaining the organization of the cluster may be the selection to maintain temporal colinearity, and this is a property potentially related to the overall organization of the chromatin in this region. The observation that the cluster has been broken up slightly in *Drosophila* and in *C. elegans* [13], where spatial control of expression is tightly regulated but temporal colinearity may no longer be required, lends some support to this view. Given the observations of Pollard and Holland [3], it now becomes important to ask whether there are any selective forces maintaining clustering of the rest of the *Antennapedia*-superfamily homeobox genes in any organisms. At least within the vertebrates, none of the genes outside of the Hox complex appears to have been as tightly constrained as those within the Hox complex, but pairs of genes are possibly being held together by shared regulatory elements. Within *Drosophila*, several of the NKL-class genes sit within a relatively small chromosomal interval, with some clearly immediately adjacent to one another. Functionally, they appear to have distinct roles, but it is interesting to note that a number of them are involved in mesodermal patterning [14-17], potentially reflecting some sort of ancestral function.

The results of Pollard and Holland [3] certainly urge us to look at the homeobox-containing genes in a different way. Continued analysis of genome organization, and the influence this organization has on gene function, will no doubt greatly accelerate our understanding of the evolution of genetic networks and the evolution of all forms of life.

Acknowledgements

We thank Rosie Redfield for discussions on genome rearrangements and Greg Davis for comments on the manuscript.

References

- Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, et al.: **Comparative genomics of the eukaryotes.** *Science* 2000, **287**:2204-2215.
- Glusman G, Sosinsky A, Ben-Asher E, Avidan N, Sonkin D, Bahar A, Rosenthal A, Clifton S, Roe B, Ferraz C, et al.: **Sequence, structure, and evolution of a complete human olfactory receptor gene cluster.** *Genomics* 2000, **63**:227-245.
- Pollard SL, Holland PW: **Evidence for 14 homeobox gene clusters in human genome ancestry.** *Curr Biol* 2000, **10**:1059-1062.
- Garcia-Fernandez J, Holland PW: **Archetypal organization of the amphioxus Hox gene cluster.** *Nature* 1994, **370**:563-566.
- Nadeau JH, Sankoff D: **Counting on comparative maps.** *Trends Genet* 1998, **14**:495-501.
- McLysaght A, Enright AJ, Skrabanek L, Wolfe KH: **Estimation of synteny conservation and genome compaction between pufferfish (*Fugu*) and human.** *Yeast* 2000, **17**:22-36.
- Carroll RL: *Vertebrate Paleontology and Evolution.* New York: WH Freeman, 1998.
- Meyer A, Schartl M: **Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions.** *Curr Opin Cell Biol* 1999, **11**:699-704.
- Gould A, Morrison A, Sproat G, White RA, Krumlauf R: **Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping Hox expression patterns.** *Genes Dev* 1997, **11**:900-913.
- Kondo T, Duboule D: **Breaking colinearity in the mouse HoxD complex.** *Cell* 1999, **97**:407-417.
- Kondo T, Zakany J, Duboule D: **Control of colinearity in AbdB genes of the mouse HoxD complex.** *Mol Cell* 1998, **1**:289-300.
- van der Hoeven F, Zakany J, Duboule D: **Gene transpositions in the HoxD complex reveal a hierarchy of regulatory controls.** *Cell* 1996, **85**:1025-1035.
- McGinnis W, Krumlauf R: **Homeobox genes and axial patterning.** *Cell* 1992, **68**:283-302.
- Nose A, Isshiki T, Takeichi M: **Regional specification of muscle progenitors in *Drosophila*: the role of the msh homeobox gene.** *Development* 1998, **125**:215-223.
- Knirr S, Azpiazu N, Frasch M: **The role of the NK-homeobox gene slouch (S59) in somatic muscle patterning.** *Development* 1999, **126**:4525-4535.
- Jagla K, Frasch M, Jagla T, Dretzen G, Bellard F and Bellard M: **ladybird, a new component of the cardiogenic pathway in *Drosophila* required for diversification of heart precursors.** *Development* 1997, **124**:3471-3479.
- Azpiaz N, Frasch M: **tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of *Drosophila*.** *Genes Dev* 1993, **7**:1325-1340.