

# The rise and fall of Hox gene clusters

Denis Duboule

Although all bilaterian animals have a related set of Hox genes, the genomic organization of this gene complement comes in different flavors. In some unrelated species, Hox genes are clustered; in others, they are not. This indicates that the bilaterian ancestor had a clustered Hox gene family and that, subsequently, this genomic organization was either maintained or lost. Remarkably, the tightest organization is found in vertebrates, raising the embarrassingly finalistic possibility that vertebrates have maintained best this ancestral configuration. Alternatively, could they have co-evolved with an increased 'organization' of the Hox clusters, possibly linked to their genomic amplification, which would be at odds with our current perception of evolutionary mechanisms? When discussing the why's and how's of Hox gene clustering, we need to account for three points: the mechanisms of cluster evolution; the underlying biological constraints; and the developmental modes of the animals under consideration. By integrating these parameters, general conclusions emerge that can help solve the aforementioned dilemma.

"See my son, here time becomes space"  
Gurnemanz, in Parsifal (R. Wagner)

## Introduction

The discovery and study of Hox gene clusters have been central to the development of many conceptual tools, now widely applied, regarding the structure, function and regulation of animal genomes. For example, the concept that various animals not only share their genes, but also complex genetic systems and that these systems are used at different times and places within the same organism [see references in Kirschner and Gerhart (Kirschner and Gerhart, 2006)]. On an even more basic level, the relatively recent evolution of vertebrate genomes, as well as important revisions of animal phylogenies, have also largely relied upon the composition of these gene clusters (de Rosa et al., 1999). The heuristic value of this genetic system is in itself a remarkable and fascinating topic, which lies outside the scope of this review. However, new paradigms are often associated with undesirable side effects, and the discovery that mice and flies have evolutionary and functionally related Hox 'clusters' (Duboule and Dollé, 1989; Graham et al., 1989) is no exception to this rule. It was indeed quickly assumed that all other animal species would contain a cluster of Hox genes, and many subsequent publications reported the existence, in some species, of a Hox gene 'cluster', when in fact only isolated Hox genes (or fragments thereof) had been obtained (e.g. Duboule, 1994a).

This misleading perception of the prevalence of the Hox cluster has been reinforced by the common, erroneous graphical representation of these loci, in particular in reviews and textbooks,

whenever inter-species comparisons are shown (e.g. de Rosa et al., 1999; Lemons and McGinnis, 2006). The reductionism of the classical scheme, inherited from the first alignments between the mouse and *Drosophila* genes, usually conveys four wrong messages. (1) The horizontal alignment of individual genes, according to which paralogous group they belong to, suggests that they are structurally linked, when this is not always the case. (2) The representation of genes as small boxes suggests that Hox genes are identical to each other, which is rarely the case. (3) The absence of scale suggests that Hox loci from various species are of the same genomic size, which is rarely so. (4) The absence of any other information regarding the DNA content (e.g. the presence or absence of repeats) suggests that intergenic sequences are not important.

An example, albeit not the most striking, of this biased perception is given in Fig. 1, which shows a comparison between the *Drosophila*, the amphioxus and one of the four murine Hox clusters, using two different levels of resolution. A schematic alignment is shown (Fig. 1A) as it usually appears in the literature, with colors to illustrate paralogous groups and with Hox genes represented by boxes. The same comparison is also shown at the correct scale (Fig. 1B), which clearly illustrates the discrepancy between the traditional representation and the physical reality. Although this issue might appear somewhat anecdotal, it is of key importance whenever the functional genomic evolution of the Hox cluster(s) is considered.

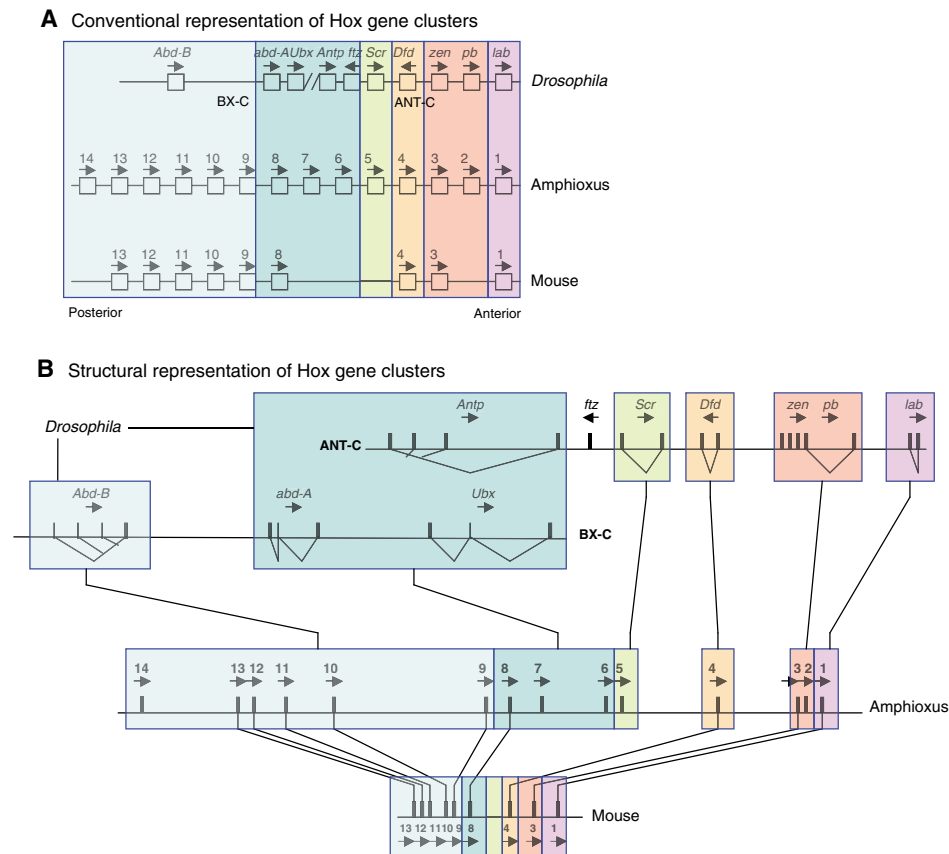
In this review, I discuss why this lack of precision has contributed to the failure to appreciate a crucial problem associated with our understanding of the evolution of vertebrate Hox clusters: that some extant Hox clusters are probably 'better organized' than their ancestral forms. I propose a potential solution to this problem, which relies upon the counter-intuitive view that genome duplications, in some cases, might increase regulatory constraints, thus leading to the consolidation of genetic loci, rather than to their relaxation.

## Collinearity: myth or reality?

Ever since the first alignment between vertebrate Hox and *Drosophila* homeotic (HOM) clusters was proposed (see Akam, 1989), confusion has surrounded the nature of these clusters. As discussed above, this confusion largely stems from the simplified graphical representation of the Hox clusters in these two species, which has subsequently been used to extract conclusions concerning the evolution of this gene family. The fact that the two *Drosophila* gene clusters [the Antennapedia (ANT-C) and Bithorax (BX-C) complementation groups] were artificially juxtaposed to properly align with their vertebrate counterparts, contributed to the perception that a single Hox gene cluster exists in insects. This was somehow then formalized when this collection of HOM genes, distributed at two different loci, became known as a 'HOM complex' (Akam, 1989). Furthermore, not only do *Drosophila* have two separate HOM clusters, but they are different from one another: whereas ANT-C is rather disorganized, with 'foreign' (non-HOM) genes interspersed amongst the HOM genes and with homeotic genes found in both transcriptional orientations,

National Research Centre 'Frontiers in Genetics', Department of Zoology and Animal Biology, University of Geneva, Sciences III, Switzerland. School of Life Sciences, Ecole Polytechnique Fédérale, Lausanne, Switzerland.

e-mails: Denis.Duboule@zoo.unige.ch; Denis.Duboule@epfl.ch



**Fig. 1. Conventional and to-scale representations of Hox clusters.**

The colored boxes correspond to the various paralogy groups – those genes that are most closely related in sequence, and hence are derived from a common ancestral Hox gene. (A) Schematic depicting Hox clusters as they are usually represented in the literature and textbooks. It shows the respective positions of paralogy groups 1 to 14 for the cephalochordate amphioxus and a vertebrate prototypic cluster, and their corresponding *Drosophila* genes located on both ANT-C and BX-C. (B) A more precise representation of the organization within Hox clusters, with the correct relative distances and a clear separation present between the two *Drosophila* sub-clusters. The comparison between A and B highlights the differences that exist between a structural reality and its conceptual interpretation. Only the former, as shown in B, should be considered when discussing the structural and functional evolution of Hox clusters. ANT-C, Antennapedia complex; BX-C, Bithorax complex.

BX-C is somewhat better organized, resembling to some degree vertebrate Hox clusters (Duboule, 1992). Why do such details matter?

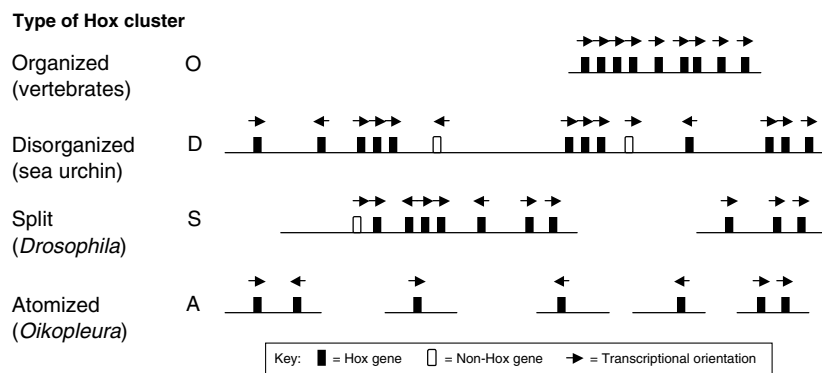
Hox gene clustering is neither a topographic oddity, nor the mere trace of how this gene family originated through local gene duplications. In several cases, it reflects a more profound level of functional organization, which was originally described by Ed Lewis in genetic terms (Lewis, 1978). The collinear correspondence between gene order and the body levels where these genes are expressed during development has, for many years, provided a convenient explanation as to why Hox genes had remained ‘clustered’. Recently, however, this explanation has been challenged in several studies, coinciding with the detailed description of additional animal model systems, such as urochordates (e.g. Seo et al., 2004), where at least some level of coordination in Hox gene expression is observed despite the absence of gene clustering (see Galliot, 2005; Monteiro and Ferrier, 2006). To make sense of these apparently paradoxical datasets (see Lemons and McGinnis, 2006), we need to integrate several parameters, such as the kind of cluster under consideration for a given animal, the precise definition(s) of collinearity(ies), and the relationship between the developmental strategies of different animals and the type of cluster used.

### There are clusters and clusters

In addressing the first issue, a tentative definition of what is meant by ‘clustering’ is required, as the use of ‘clustered’ versus ‘non-clustered’ Hox genes has become arguably limiting. Without abusing an exhaustive number of qualifications [nicely clustered, clustered but separated, clustered in pairs, tightly or loosely clustered (see

Payre and Ternell, 1994)], I propose to define a minimal number of structural organizations to help us to think about the problem. For clarity’s sake, let us consider only four possibilities as a first-line classification of Hox ‘clusters’ (Fig. 2): (1) organized clusters (or ‘type O clusters’, as in the prototypic vertebrate); (2) disorganized clusters (‘type D’, e.g. the sea urchin cluster); (3) split clusters (‘type S’, e.g. the *Drosophila* HOM ‘cluster’); and (4) atomized clusters (‘type A’, e.g. the urochordate *Oikopleura* ‘cluster’), where genes are mostly scattered throughout the genome. Multiple combinations of types can, of course, be found, particularly in ‘type S’ animals, which may have, for example, both type O and type D sub-clusters. Other combinations will undoubtedly be reported as additional genomes are sequenced. Moreover, only a few genomes have been analyzed to the extent that a firm assignment can be given to a cluster type. Also, a bias might exist in the selection of protostome model systems as most of these animals have rather small genomes and develop rapidly. Yet, the final picture might not differ drastically from what we can now contemplate. With this simple analytical tool at hand, we can reconsider the animal phylogeny and superimpose the appropriate types of clusters to hypothesize about their structural evolution (Fig. 3).

Cnidarians have Hox genes (e.g. Gauchat et al., 2000) that are organized in ‘type A clusters’. Although some Hox genes are still found in pairs, the general organization of the few Hox genes is atomized (Chourrout et al., 2006; Kamm et al., 2006). This situation is similar to that found in the flatworm *Schistosoma mansoni* (Pierce et al., 2005) and in nematodes, and can be regarded as type A, although some genes remain in pairs (i.e. they maintain some degree of genomic organization) (Aboobaker and Blaxter, 2003). The prototype of the type S (split) cluster is found in *Drosophila*, where



**Fig. 2. Structural classification of Hox clusters.** Type O (organized) clusters are well organized, with genes tightly arranged and all encoded by the same DNA strand. They are devoid of both ‘foreign’ genes and repeats, yet they may contain non-coding RNAs and miRNAs. Vertebrate clusters provide, so far, the sole example of this organization. Type D (disorganized) clusters are much larger and may contain mixed-up Hox genes (black boxes), or genes in opposite orientations, in addition to non-Hox genes (white boxes) and repeats. Type D examples are found in amphioxus and in sea urchins. Split (type S) clusters can have type O or type D features in each of their sub-clusters, such as in Diptera, whereas the type A (atomized) ‘cluster’ represents the ‘no-cluster’ situation, in which genes are found, at best, in pairs at scattered genomic loci (e.g. in *Oikopleura*).

a chromosomal breakpoint can be either between *Antp* and *Ubx*, or between *Ubx* and *abdominal A* (*abd-A*) (Von Allmen et al., 1996; Negre and Ruis, 2007). Other non-dipteran insects show a type S cluster, as in the moth *Bombix mori*, although a breakpoint lies closer to the ‘anterior extremity’ of the gene series (Yasukochi et al., 2004). By contrast, some insect species have the full complement of Hox genes at a single locus. In these cases, clusters can nevertheless be classified as type D, mostly on account of their large size and apparent high level of ‘disorganization’. For example, in the mosquito *Anopheles gambiae*, a very large cluster that exceeds 700 kb is found (Holt et al., 2002), which contains many interspersed repeats that are mostly absent from type O clusters. Similarly, a large but unique cluster appears to exist in *Tribolium castaneum* (Brown et al., 2002).

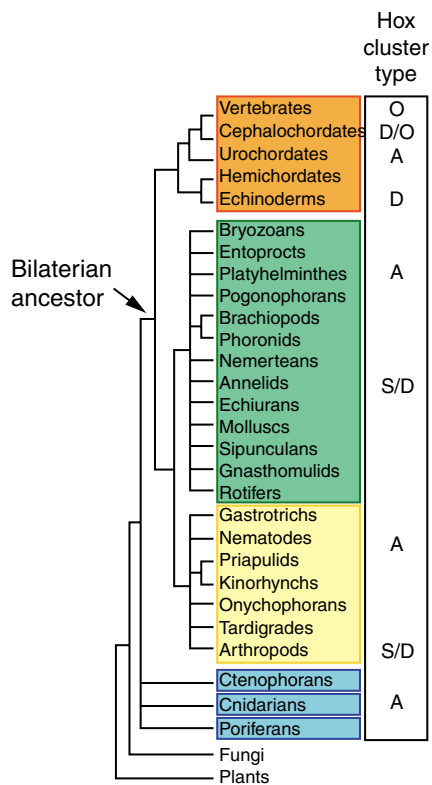
As for ecdysozoans, Hox genes have been described in a variety of lophotrochozoan species, in particular in molluscs (e.g. Callaerts et al., 2002) but also in nemerteans (Kmita-Cunisse et al., 1998) or brachiopods (de Rosa et al., 1999). Yet, so far, the genomic organization of only a flatworm (see above) and the annelid polychaete *Platynereis* is known, and these probably contain a rather intact cluster (D. Ferrier, personal communication). The situation in protostomes is equally heterogeneous. The sea urchin Hox cluster is large and contains genes in opposite transcriptional orientations and at unexpected positions with respect to their paralogous groups, revealing the occurrence of important rearrangements (Cameron et al., 2006) that are typical of type D clusters. By contrast, analyses of *Ciona intestinalis* and *Oikopleura dioica* indicate that urochordates have atomized ‘clusters’ of Hox genes (Ikuta et al., 2004; Seo et al., 2004). In marked contrast, the cephalochordate amphioxus has a well-defined cluster, although some aspects of it make the distinction between type D and type O difficult – for example, its rather large size (at least 450 kb) as compared with vertebrates (Garcia-Fernandez and Holland, 1994) [see references in Minguillon et al. (Minguillon et al., 2005)] and the presence of internal repeats within two intergenic regions (C. Amemiya, personal communication). By and large, vertebrates display the most tightly organized Hox gene clusters, with all genes in the same transcriptional orientation, spanning ~100 kb. These clusters are very rich in conserved non-coding DNA sequences, they are mostly devoid of any repetitive sequences (e.g. Lander et al.,

2001) and the genes typically have very short introns (not visible at the scale used in Fig. 1), thus reinforcing their ‘compacted’ nature, as if only minimal sequence requirements have been conserved along with a fully optimized genetic structure.

### Construction versus destruction

Despite the many animal groups for which genomic data are not available and for which the status of their Hox ‘clusters’ remains unknown, when considering the distribution of cluster types shown in Fig. 3, a surprising conclusion is reached and an embarrassing question raised. Firstly, it becomes clear that most bilateral animals will have, at best, a largely disorganized cluster, most probably a split cluster. Furthermore, a complete fragmentation of the ‘cluster’ is seen in very different groups of animals, and might thus be expected for many species; the textbook Hox cluster might thus be the exception and not the rule. Secondly, whereas various groups display a single Hox gene ‘cluster’ (types O/D), those that can be classified as ‘organized’ are exclusively found in chordates, or even within vertebrates, if one considers the amphioxus cluster to resemble type D, for reasons mentioned above. With this in mind, we can now reconsider the question of the ancestral bilaterian Hox cluster, as well as the potential sequence of structural modifications leading to the situation shown in Fig. 3.

The fact that our bilaterian ancestor had at least one set of clustered Hox genes can be inferred, given that the vertebrate cluster and the *Drosophila* counterpart correspond to each other. Although the exact composition of such an ancestral cluster, in terms of how many genes and which paralogous groups are represented, is still open to debate (see de Rosa et al., 1999; Ryan et al., 2007; Garcia-Fernandez, 2005), *Drosophila* Hox genes, clearly orthologous to vertebrate counterparts, were found in both sub-clusters of the fly, indicating that the cluster had been split at some point in the evolution of Diptera (Duboule and Dollé, 1989; Graham et al., 1989). This explanation is indeed more parsimonious than the opposing scenario, wherein two original clusters would have repeatedly and independently merged into a unique and comparably contiguous series. In this context, and provided the associated constraints (as discussed below) were released, one can imagine how a cluster can progressively become disorganized, or even split into two pieces or more, through recurrent events leading to the atomized situation.



**Fig. 3. Cluster types in various animal species.** A recent animal phylogeny with deuterostomes shown in orange, protostomes in green (lophotrochozoans) and yellow (ecdyzozoans), and ctenophorans, cnidarians and poriferans arbitrarily shown in blue, as outgroups of bilateria. Cluster type is indicated to the right. Amongst the deuterostomes, cephalochordates are still positioned closer to vertebrates than are tunicates (urochordates), contrary to the proposal from Delsuc et al. (Delsuc et al., 2006). This uncertainty does not change the nature of the argument regarding the evolution of chordate Hox clusters (see text). Whereas type A, S and D clusters are found throughout these large groups, genuine type O clusters are only described in vertebrates (see text for a discussion of the type D/O cluster found in cephalochordates).

Yet if we assume that a unidirectional logic prevailed in this process, i.e. from an organized state towards a less organized state, we must naturally conclude that the ‘best organized’ cluster is the closest relative in terms of general structure to the ancestral bilaterian cluster, while all others suffered an evolutionary erosion. Interestingly, whichever criteria are applied to define the ‘best organized Hox cluster’ (see, for example, Fig. 1), vertebrates always score highest, indicating that a direct relationship exists, at least at the level of the structural organization of Hox genes, between the ancestral bilateria and ourselves, from which all other animals are derived. In other words, vertebrates, amazingly, would be the only animals in which the original genomic structure of this crucial gene family has persisted throughout evolution (Fig. 4).

Although this possibility would have pleased early eighteenth-century naturalists (or the actual proponents of intelligent design), we must admit that, for a variety of reasons, this vertebrate-centric view is unlikely to reflect reality, for it would imply a particular phylogenetic link between vertebrates and the bilaterian ancestor, which is otherwise not supported by any theoretical and/or scientific considerations. In addition, although the Hox gene complement of

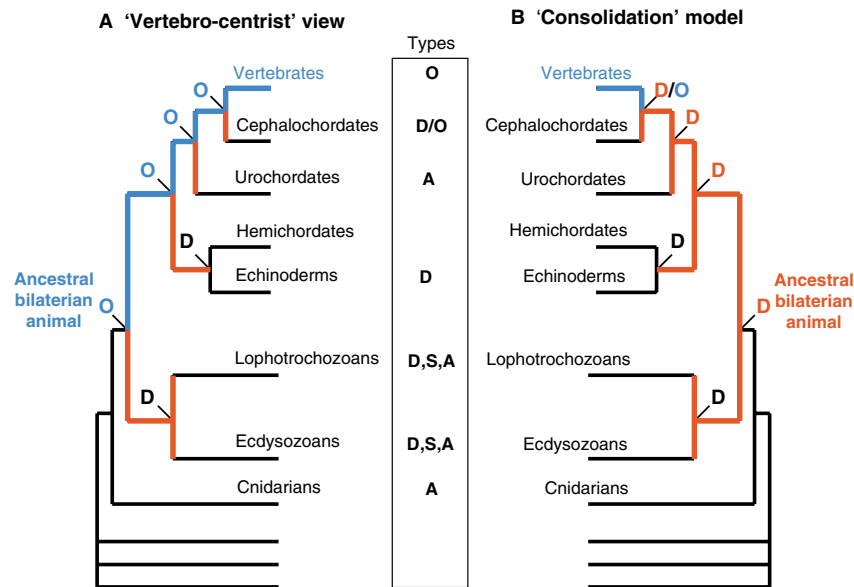
this ancestral animal and of extant vertebrates were certainly related, some paralogy groups found in vertebrates have clearly been produced subsequently (for example, the most-posterior groups), indicating that vertebrate Hox clusters are not the direct and privileged descendants of an ancestral bilaterian Hox gene cluster. But if we agree on this impossibility, we must face the question of how to construct and select for a ‘better-organized’ cluster, rather than how to leave it to become disorganized. This former possibility has received little, if any, attention in the literature, in which the terms ‘expansion’, ‘diversification’ and sometimes ‘simplification’ (e.g. Lemons and McGinnis, 2006) are generally used to qualify the evolution of Hox gene clusters, rather than ‘organization’ or ‘consolidation’. Under this proposed scenario, the ancestral cluster would be type D, and would evolve into a type O along the vertebrate lineage (Figs 4, 5). What kind of mechanism(s) could have underpinned such counter-intuitive processes, assuming that the necessary selective potential existed?

### Global is beautiful

It is difficult to imagine why or how a large gene cluster (say 500 kb) that contains about ten genes in various orientations, together with some repeats, would be progressively transformed into a 100 kb cluster, with the same ten genes encoded now by the same DNA strand and without any repeats. This process of ‘consolidation’ must have represented an ‘added value’, in evolutionary terms, that was selectively favored over either a stabilization or a ‘simplification’ of the cluster (Fig. 5). The various solutions that might account for such consolidation all rely upon the integration of the functions of single genes into a more global mode of operation. Such a communal ability to fulfil a functional task that cannot be fulfilled by any of the genes in isolation was called ‘meta-genic’, and, accordingly, the Hox clusters should be regarded as meta-genes (Duboule, 1994b). For example, a source of novel protein products could result from splicing patterns that become more complex, once neighboring genes are encoded by the same DNA strand. Also, the emergence of global enhancer sequences, located outside the cluster itself, might favor increased gene proximity to facilitate and secure a coordinated transcriptional response (Spitz et al., 2001; Spitz et al., 2003). Optimized, coordinated responses to global regulations might themselves be the favored functional approach of this gene family, in which gene dosage effects and compensatory mechanisms are often observed amongst neighboring genes, helped by largely redundant protein functions (e.g. Zakany et al., 1997; Wellik and Capecchi, 2003).

In this context, the acquisition of global, cluster-wide regulations might have triggered a progressively increasing level of structural organization between neighboring genes, to allow them to respond at the transcriptional level in a more coordinated way (Fig. 6). For example, the reduction of intergenic distances and the elimination of ‘foreign’ sequences (non-Hox transcription units or repeats) can be understood where a group of contiguous Hox genes is recruited to achieve a novel meta-genic function; for example, the recruitment of several genes of the mammalian *Hoxd* cluster by a digit-specific global regulation (Kmita et al., 2002; Spitz et al., 2003). In turn, this process of consolidation is intrinsically directional, as it paves the way for other regulatory co-options to occur, for at least two reasons: first, the functional potential of a coordinated series of regulators is largely greater than that of a single transcription unit, as it may provide more integrated possibilities, including dosage effects and redundancy; second, it is conceivable that strong and remote global enhancer regions might foster the emergence of novel regulatory controls at the same site, owing to the presence of various specific or general





**Fig. 4. The dilemma of bilaterian Hox cluster evolution.** Two possible models can account for the restriction of type O clusters to vertebrates. **(A)** In the 'vertebro-centrist' view, the ancestral bilaterian animal had a type O cluster, which was maintained all the way through to the vertebrate lineage, as illustrated in blue. All other bilaterian animals are derived from this ancestral form, first through a type D cluster (red) and, subsequently, with type S or A clusters. Although this view provides the easiest mechanistic explanation (as it progresses towards disorganization only), it implies that vertebrates are 'direct descendants' of this ancestor, which is at odds with our modern view of animal evolution. **(B)** In the 'consolidation model', the ancestral cluster was of type D, as illustrated in red, which implies that a Hox cluster, at some point during the evolution of deuterostomes, underwent a phase of consolidation (that is, an increased organization), as suggested by the blue color on the top right. Although this view fits better with our current understanding of animal phylogeny, it requires some conceptual tools to explain how such an 'organization' can occur and be selected for.

transcription factors, an increased accessibility or a particular chromosomal architecture, a process referred to as 'regulatory priming' (Gonzalez et al., 2007) (Fig. 6). In other words, cluster consolidation may merely illustrate the evolution of a meta-gene structure and its associated meta-cis regulations (Duboule, 1994b), in a way that is similar to our current views of how a single transcription unit might have appeared and be further stabilized, for example by 'consolidating' exonic sequences or recruiting various regulations.

If we accept the possibility that an ancestral Hox gene cluster was subsequently the target of bi-directional structural evolutions, towards both fragmentation and consolidation, the question of which parameters and/or constraints favored one structure or the other ought to be addressed. But before this, let us reconsider the strange correspondence between type O clusters and the vertebrate lineage, in the light of what I have discussed.

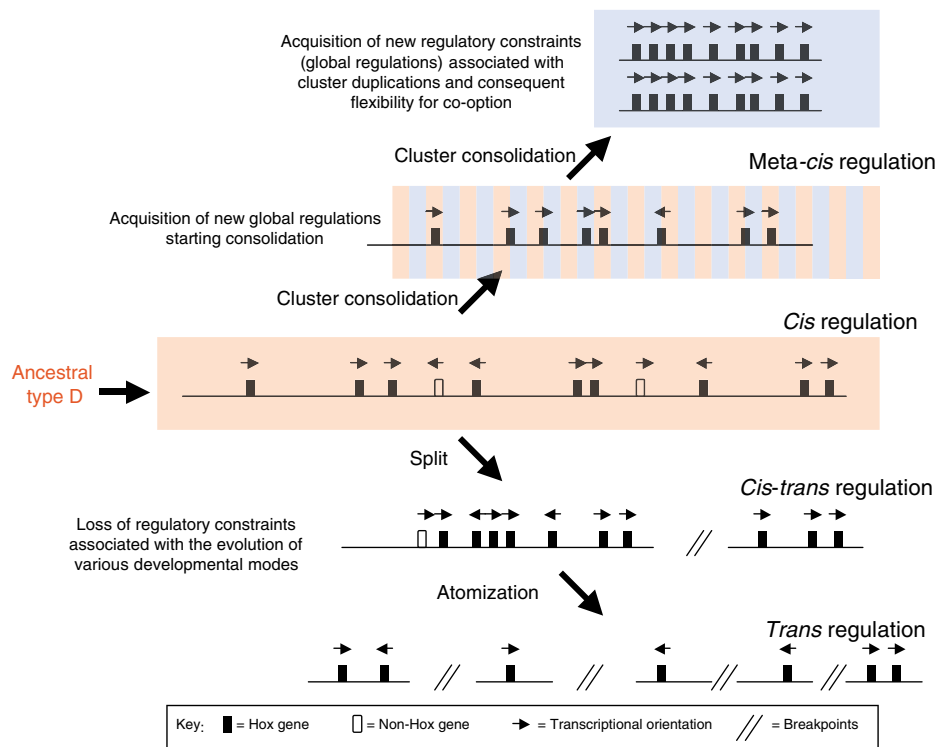
### Why us?

The vertebrate lineage is the only one in which multiple complete Hox clusters have been described, i.e. containing both anterior and posterior types of Hox genes. All vertebrates thus contain a minimum of four Hox clusters, with additional copies present in fish (Hurley et al., 2005) (see below). The mechanisms that operated at the genesis of this cluster amplification are still a matter of debate; in particular, whether or not these four copies were produced by full-genome duplications or by more-restricted DNA segmental amplification. It is likely that the first possibility will be validated, once more genomes are sequenced, in particular those of animals that might provide a link between early chordates and gnathostomes (vertebrates with jaws) such as agnathans species (vertebrates without jaws). Although this mechanistic question is not necessarily

relevant to the problem of Hox cluster functional evolution, it raises another paradox, which, once clarified, might help us to understand the emergence of consolidation: how can one explain that highly consolidated Hox clusters are found in those species that evolved several copies of them? One obvious possibility is that a highly organized (i.e. already fully consolidated) cluster had evolved before the first genome duplication event. If correct, this cluster, under 'consolidation constraints', would no longer be represented in those early chordates such as urochordates or cephalochordates, for which genomic sequence is available. Alternatively, a semi-consolidated cluster could have been duplicated, with some further consolidation occurring thereafter, independently on both duplicated copies.

It is difficult to imagine how convergent cluster organizations may have followed large-scale genome duplication, to reach such a level of similarity amongst, for example, the four human Hox clusters. However, the existence of an already fully consolidated cluster before genome amplification does not make the explanation easier. Our current views regarding the selective advantages of a genome (a gene; a meta-gene) being duplicated arguably predict that the constraints that maintain genes in close proximity would be relaxed in one or other of the two duplicated copies (Ohno, 1970; Holland, 1999). This should favor fragmentation and disorganization, rather than maintenance or further consolidation. Yet, if we consider the emergence of global regulation (and the associated regulatory hubs) as major factors of consolidation, as discussed above, we might be able to explain why cluster duplication did not lead to fragmentation but, instead, to further internal organization (see Figs 5, 6).

In vertebrates, Hox gene functions are essential for the development of various morphological features that are generally considered to be late evolutionary novelties, associated either with



**Fig. 5. Transitions in the structural evolution of Hox clusters.** From the starting type D cluster (red, as in Fig. 4), transitions to type S and, subsequently, to type A 'clusters' (downward) are easy to imagine, and could occur concomitantly with the loss of those constraints that keep genes together. These transitions do not call for any particular mechanism other than the occurrence of chromosomal breakpoints. In some cases, these breaks might be beneficial, along with the evolution of an organism towards a different developmental mode for which clustered Hox genes might be an obstacle rather than a solution (see the text). The arrows towards the top illustrate the mechanism of cluster consolidation, which might occur together with genome duplication. Duplication of the locus allows for more (or for different) global regulations to evolve, which in turn further consolidate the various clusters. In this scheme, a meta-cis structure must exist before genome duplication to provide the basis for further co-options. The color code illustrates the passage from a type D to a type O cluster (see also the 'consolidation model' in Fig. 4).

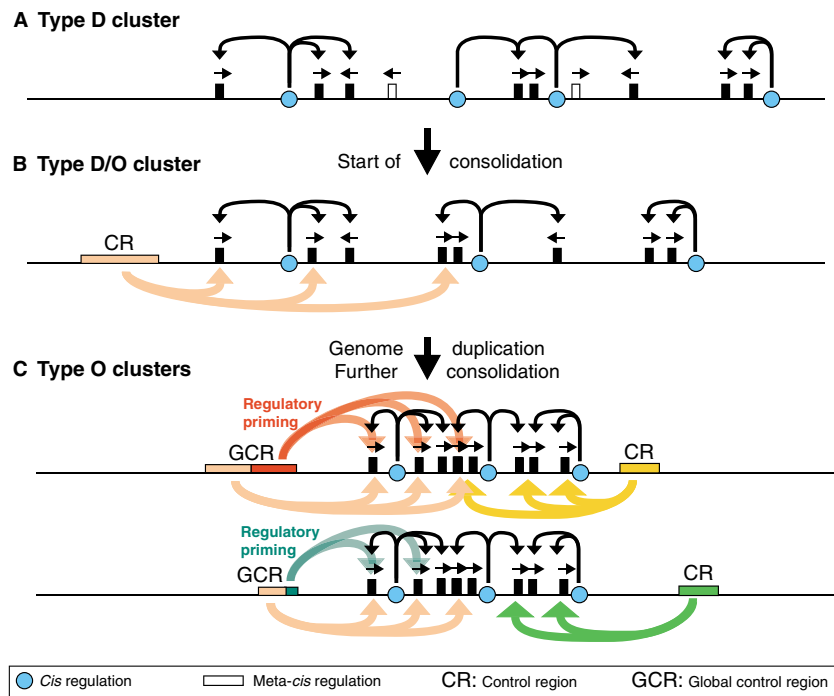
the emergence of this lineage or with important steps of its own evolution, such as the development of skeletal appendages, external genital organs, metanephric kidneys and the branchial apparatus. Much in the same way that duplicated genes can acquire some evolutionary flexibility and be recruited for divergent functions, Hox meta-genes, following their duplication, might have offered novel possibilities for regulations to be co-opted, thus triggering the emergence of these various vertebrate features [see Wagner et al. (Wagner et al., 2003) and references therein].

Such large-scale duplications might have favored the recruitment of novel global regulations, in particular via regulatory priming (Gonzalez et al., 2007) (Fig. 6), together with more-local, gene-specific controls, as in the case of hindbrain patterning (Trainor and Krumlauf, 2001). It is indeed conceivable that a single cluster rapidly experienced serious constraints in its capacity to recruit novel global regulations, without interfering with pre-existing ones. In this view, cluster duplication would favor consolidation, not only by widening the range of regulatory possibilities, but also by actively triggering potential regulatory innovations owing to the presence of an already 'competent' locus architecture. The apparent co-evolution of important regulatory controls for both limbs and genitals in tetrapods may illustrate this effect (Kondo et al., 1997). This increase in the general organization of the clusters, as triggered by the maintenance or re-enforcement of a consolidation process, might have required some internal rearrangements, and there are many examples of paralogous genes being lost in a single, or in several,

clusters. This phenomenon is usually explained by the existence of redundancy, occurring once clusters have been duplicated. Yet the possibility that some Hox genes were lost by positive selection owing to their incompatibility with a newly recruited global regulatory control should not be overlooked.

Although this scenario might account for some convergence in the consolidation of Hox gene clusters after duplication, it is unlikely that all four clusters independently evolved from an ancestral type D to type O organization. Consequently, this process must have started to occur early on, during the early chordate-to-vertebrate transition. In this respect, the actual structure of the cephalochordate amphioxus cluster, although testifying to the existence of a single entire cluster in an early chordate ancestor, is in fact of limited significance, as it cannot be considered to be a direct ancestral form of the vertebrate clusters. First, the phylogenetic position of cephalochordates as the closest relatives of vertebrates has recently been challenged (Delsuc et al., 2006). Second, this cluster itself might have succumbed to either some disorganization or consolidation after cephalochordates separated from the vertebrate (or the vertebrate-urochordate) lineage. Functional analyses will hopefully reveal whether or not global Hox regulation is at work in these animals.

Most importantly, the functional characterization of Hox clusters in agnathans should indicate whether consolidation does indeed go hand-in-hand with duplication, i.e. whether or not the consolidation process in ancestral vertebrate clusters started before the full



**Fig. 6. Meta-cis regulation as a mechanism that triggers vertebrate Hox cluster consolidation.** (A) Schematic of a type D cluster present in the ancestral bilaterian animal. Hox genes (black boxes) are present in various transcriptional orientations (arrows), dispersed over a large fragment, which also includes foreign (non-Hox) genes (white boxes). Genes are kept together owing to the presence of cis-acting regulatory sequences (blue circles), which are shared between several neighboring genes. (B) A type D/O cluster present in an ancestral chordate. Owing to the emergence of a meta-cis regulatory sequence in a control region (CR, brown), the genomic organization of target genes is improved as a consequence of positive selection. Progressively, foreign genes are lost together with some large intergenic fragments, which optimizes the meta-cis control. (C) After genome duplication, consolidation is favored owing to the increasing possibility to recruit novel meta-cis regulation (red, yellow, green boxes and arrows), in particular through the intrinsic capacity of a potent control region (CR in B) to generate regulatory novelties by ‘regulatory priming’ (red and dark green in C; see text). This leads to a concentration of large-scale enhancers to form a ‘global control region’ (GCR), which might continue to recruit and/or evolve novel regulatory elements owing to its intrinsic properties. As a consequence, duplicated clusters are further consolidated.

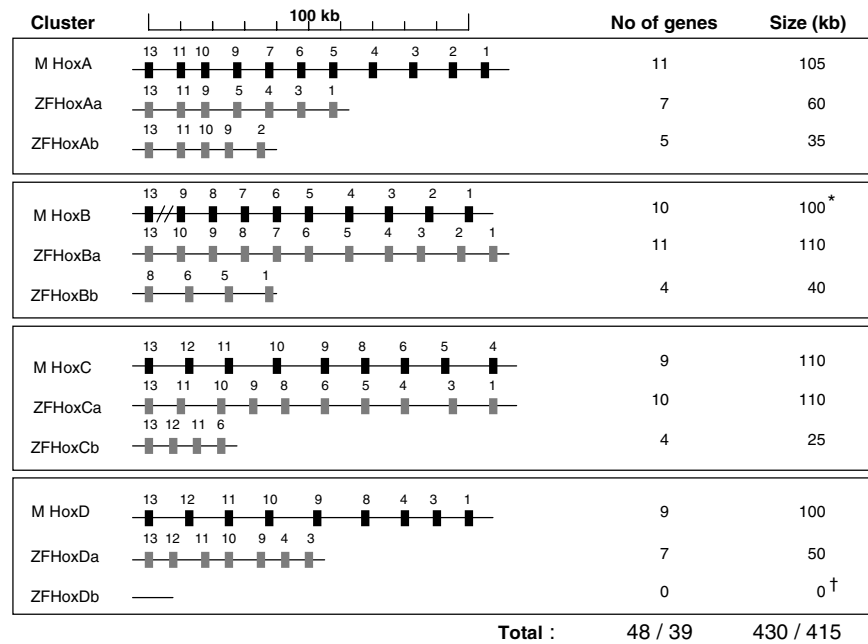
complement was reached, as would be expected from the above discussions. A detailed structure-function analysis of the vertebrate Hox clusters may also help in this endeavor, as it might reveal the remnants of pre-duplication global regulatory mechanisms, as exemplified by a remnant of the mouse *Hoxd* global control region (GCR), which is located at a similar relative position upstream of the *Hoxa* cluster (Lehoczky et al., 2004). The GCR, located upstream of the *Hoxd* cluster (Spitz et al., 2003), contains several globally acting enhancers (Gonzalez et al., 2007), in contrast to its reduced and simplified counterpart on the *Hoxa* cluster (Lehoczky et al., 2004). Analysis of this element(s) in an ancestral cluster, prior to duplication, might indicate whether GCR-associated regulations were lost in the *Hoxa* cluster after duplication, or acquired in the *Hoxd* cluster.

### The teleost alternative

In the above scenario, cluster duplications facilitated the evolution of global regulations, which in turn accompanied the emergence of crucial vertebrate features, leading to increased organism complexity. However, this view seems to be contradicted in the case of teleost fishes, which are of particular interest in this context. It is well accepted that crown teleostei experienced an additional round of genome duplication (Prince et al., 1998; Amores et al., 1998). As a consequence, all teleost fish genomes analyzed so far bear seven to eight Hox gene clusters, depending on the species [e.g. Amores et al. (Amores et al., 2004) and references therein]. These clusters

are often referred to as having been ‘further amplified’, much in the same way as ancestral vertebrates had an ‘amplified’ complement of four clusters derived from two successive genome duplications. Yet, despite the expected loss of several newly duplicated fish Hox genes, as occurred during previous rounds of duplications, cluster consolidation has not been seen. Also, if the passage from one to four clusters was associated with increased complexity in vertebrates (e.g. Holland and Garcia-Fernandez, 1996), what could have been the adaptive value, for teleostei, to have twice as many Hox clusters as us?

Here again, a closer look at the teleost Hox clusters reveals an unexpected picture, as illustrated by the zebrafish, *Fugu* and medaka genomes displaying similar global organizations (Kurosawa et al., 2006), although with some differences that are not relevant to this argument. In contrast to what is commonly believed, a comparison between zebrafish and murine Hox clusters at the same scale (Fig. 7) reveals that teleostei have not experienced a further ‘amplification’ of their complement of Hox clusters, but rather an important reorganization of this gene family, which accompanied (was made possible by) an additional round of genome duplication. In fact, the overall number of Hox genes in fishes (around 48) is close to that in mice (39), despite the existence of seven clusters, not four. This is obviously not due to cluster splitting, as illustrated by paralogous and syntenic relationships, but instead to a massive elimination of Hox genes after the additional duplication. What about consolidation?



**Fig. 7. Comparison between the mouse and zebrafish Hox gene clusters.** Only the numbers of genes and their paralogy groups are shown for comparison, along with the sizes (scale bar on top) of the various clusters. The four murine (M) clusters are shown with Hox genes in black and, below, their zebrafish (ZF) Hox gene counterparts in gray. Mouse *Hoxa*, *Hoxb* and *Hoxc* clusters have two zebrafish counterparts, whereas the mouse *Hoxd* cluster has a single orthologous cluster in zebrafish. On the right, both the number of genes per cluster and the sizes (kb) of the respective clusters are shown. Totals are given at the bottom, which show that despite zebrafish having the higher number of clusters and a slight increase in Hox gene number, the overall lengths obtained when all clusters are artificially considered together are virtually identical in the two animals. \*, The position of the mouse *Hoxb13* gene is not shown at the correct scale, as it is located ~100 kb from *Hoxb9* (see text). The *Hoxb* cluster is ~200 kb, which makes the overall length of the four murine clusters even larger than that of the seven fish counterparts. †, A ZFHoxDb 'cluster' is shown, even in the absence of Hox genes, as this cluster was selectively lost, as indicated by the structure of the remaining genomic locus (Woltering and Durston, 2006).

Only two of the seven fish clusters (*hoxba* and *hoxca*) resemble their mouse counterparts in terms of both gene number and cluster size (Fig. 7). The other five clusters show clear signs of consolidation, as suggested by their compacted sizes. Whereas together the sizes of the four murine Hox clusters come to ~415 kb (not including *Hoxb13*, see Fig. 7 legend), the seven fish clusters together give an overall size of 430 kb. Therefore, a 20% increase in gene number correlates with only a 3% increase in the overall size of the clusters. In this case, again, cluster duplication, if anything, leads to an increased 'organization' rather than to cluster atomization.

This conclusion is based upon the number of genes per DNA length, and also on the potential functional relevance of these small and compacted clusters. It is noteworthy that the most-represented paralogous group in zebrafish is group 13, with six out of the seven clusters containing a gene of this group. *Hox13* gene products are of particular importance in the establishment of the body plan as they somehow terminate the patterning systems, owing to their prevalent function over other Hox products (see below). Remarkably enough, group-13 genes, present in all six zebrafish clusters, are always physically associated with at least two other *Abd-B*-related genes (from groups 12, 11, 10 or 9), thus buffering the deleterious effect that such a product would have, should it be expressed in an anterior gene-like manner. This risk is dealt with differently in the mouse *Hoxb* cluster, which only contains *Hoxb9* as another *Abd-B*-related gene: *Hoxb13* has been maintained at a distance (~100 kb) from the rest of the cluster, such that its expression is late and restricted posteriorly. These different, yet

functionally convergent strategies suggest that these compact zebrafish meta-genes have maintained the same functional organization as that found in other vertebrates.

Even though the regulatory modalities associated with fish Hox clusters have not yet been studied, and hence we do not know to what extent global regulations are present, the redistribution of the Hox informational content into numerous, but small and compact, meta-genes must have given teleostei an increased genetic modularity, allowing for more flexibility in implementing large-scale regulations (for example the activation or extinction of a mini-cluster in one particular structure). The impact of these alternative Hox genomic configurations should perhaps not be considered in an ontogenic context – that of organism complexity – but instead in a phylogenetic context, at the level of the very high number of species found in this animal group owing to their enormous potential for radiation and rapid evolution. Therefore, the evolution of the Hox cluster complement in teleostei might have favored their great evolutionary success, perhaps at the expense of a more robust, but also more constrained, developmental body plan. In this view and in agreement with Wagner and colleagues (Crow et al., 2006), it is not the number of clusters (i.e. the overall number of Hox genes) that correlates with higher organism complexity and/or species diversity but, instead, the functionality and regulatory flexibility of meta-genes.

### Trans-collinearity versus cis-collinearity

If global (meta-) gene regulation accounts for the evolution and stability of type O clusters, what mechanisms maintained type D clusters and prevented them from splitting and further entering into



a phase of fragmentation? Ever since spatial collinearity (the correspondence between the order of Hox genes on the chromosome and their domains of expression) was discovered in vertebrates (Gaunt et al., 1988), it has been considered to be a major constraint that keeps these genes together. Such coordinated gene expression is indeed most easily envisaged as occurring via in-cis mechanisms, such as enhancer sharing (Sharpe et al., 1998), or via large-scale gene regulation (Kmita et al., 2000), rather than via solely trans-acting processes. This view has been recently challenged, following several reports that show that genes belonging to type A clusters (for example, the fully fragmented ‘cluster’ of the larvacean *Oikopleura*) are, to some extent, also expressed with a spatially ‘collinear’ distribution (see Seo et al., 2004; Lemons and McGinnis, 2006; Monteiro and Ferrier, 2006). This apparent paradox is not a surprise, as a similar problem was encountered early on, when single Hox genes were isolated from mammalian type O clusters and studied in transgenic mice in vivo. Such transgenes, when integrated randomly in the genome, could recapitulate part of their spatial expression patterns, indicating that cluster organization is dispensable for establishing some of the expected rostral-to-caudal expression boundaries, at least within a certain spatial window (Krumlauf, 1994).

As in the case of clustering (see above), the paradigmatic value of collinearity has led many of us to describe this process as occurring in settings in which it failed to exist. Statements that mention that ‘collinear expression was maintained in the absence of clustering’ illustrate this problem (see Monteiro and Ferrier, 2006). This confusion understandably reflects the fact that the rostral-to-caudal progressive expression of Hox genes in animals carrying a type A cluster is likely to derive from a genuine collinear mechanism associated with an ancestral type D cluster. Consequently, talking about collinearity in animals that do not have a Hox gene cluster might not be entirely incorrect, when considering this phylogenetic view. To clarify this issue, I suggest that we refer to these distinct situations as either cis- or trans-collinearity, despite the intrinsic paradox that the latter qualification presents. Cis-collinearity defines the correspondence between the physical order of Hox genes and their domains of expression along the body axis, and hence applies to the original phenomenon as described by Ed Lewis in the Bithorax cluster of *Drosophila* (Lewis, 1978), whereas trans-collinearity defines the maintenance of the correct sequence of expression domains along the axis, with respect to paralogous groups, in the partial or complete absence of genomic clustering. Cis-collinearity applies to type O and D clusters and to the internal structure of sub-clusters in animals with type S clusters, whereas trans-collinearity applies to type A ‘clusters’ and to the ‘collinear’ correspondence that might exist between sub-clusters in type S animals (e.g. *Drosophila*). Accordingly, bilateral animals could be classified as being ‘cis-collinear’ (types O and D), ‘cis/trans-collinear’ (type S) or ‘trans-collinear’ (type A; see Fig. 5). This classification might be of use when discussing the relationship between Hox gene (non-) clustering and various developmental modes.

Various, non-mutually exclusive explanations have been proposed to account for trans-collinearity, i.e. the fact that Hox genes maintain their rostral-to-caudal coordinated expression in the absence of a bona fide gene cluster. Clustering might be necessary to refine, coordinate and to stabilize expression domains that are otherwise dictated by regulatory controls lying in the vicinity of the genes themselves. Also, the different readouts of Hox expression that are currently used (e.g. the developing vertebrate spinal cord or sclerotome) might not exert the strongest constraints on the system.

For example, the early collinear expression of Hox genes in epiblast cells during gastrulation (Forlani et al., 2003; Iimura and Pourquie, 2006) might require a strictly clustered organization, whereas subsequent ‘collinear’ domains of expression, such as in the developing spinal cord, might not have such a requirement. Understandably, type O/D clustering might be constrained by a unique site of collinear expression, which would thus require cis-collinearity, other sites having already evolved more gene-specific types of regulation. In the absence of the former constraint, such as in a type A ‘cluster’, fragmentation can thus occur, leading to trans-collinearity.

### To give time to time

Although several examples of trans-collinearity have now been reported, they all are concerned with spatial rather than temporal expression, and the importance of time in keeping Hox genes clustered (Duboule, 1992; Duboule, 1994b) has not yet been challenged (see Garcia-Fernandez, 2005; Monteiro and Ferrier, 2006). Temporal collinearity [the correspondence between Hox gene order and their temporal sequence of activation (Dollé et al., 1989; Izpisua-Belmonte et al., 1991)] might thus have been a major factor in keeping Hox genes together, whenever type O/D clusters are considered. Accordingly, animals carrying type S/A clusters are not expected to implement this regulatory property. Therefore, temporal collinearity can be both crucial and dispensable, depending upon the animal species under consideration. The understanding of these contrasting situations requires consideration of both the biological relevance of this process and the underlying mechanisms.

It is well accepted that all animals with bilateral symmetry use their Hox gene complement to organize their rostral-to-caudal polarity, and that this process relies on similar combinations of Hox gene products. For example, Hox genes related to the *Drosophila* gene *labial* (paralogy group 1; Fig. 1) function in the patterning of the rostral extremities of bilaterian animals, whereas *Abd-B*-related gene(s) (paralogy groups 9 to 13; Fig. 1) pattern caudal structures. Genetic analyses in mice and flies suggest that this genetic circuitry, observed in various structures of the same animal (e.g. the vertebrate limbs or intestine), is unlikely to be strictly based upon a protein combinatorial system [the notion of ‘Hox code’, as proposed by Kessel and Gruss (Kessel and Gruss, 1991), taken in its most orthodox meaning]. Instead, some Hox proteins have intrinsic properties that enable the more-‘posterior’ proteins to counteract, or over-rule, the function of the more-‘anterior’ ones, whenever both products co-exist. For example, the recent combined inactivation of all paralogy group-10 Hox genes in mice generated a strong phenotype, with lumbar segments bearing ribs (Wellik and Capecchi, 2003). However, this complete functional ablation of group-10 function did not elicit any remarkable phenotype in more-caudal regions, in particular where group-11 Hox genes are expressed along with group 10.

This property, called the ‘posterior prevalence’ rule (Duboule, 1991; Duboule and Morata, 1994), requires ‘posterior’ products to be present only at the developing caudal end to avoid problems of mis-specification at the anterior end of the developing embryo via the functional suppression of more-‘anterior’ functions. Various regulatory strategies have evolved that prevent the antagonizing effect of posterior Hox products over anterior functions. For example, in *Drosophila*, a complex interplay of cis-acting sequences controls the expression of the posterior *Abd-B* gene in the most-posterior parasegments only (see Maeda and Karch, 2006). Alternatively, in those animals where the elaboration of more-caudal segments is delayed in time, a mere delay in ‘posterior’ Hox gene

activation might be sufficient to restrict their expression posteriorly, thus preventing their deleterious effects (the ‘Hox clock’) (Duboule, 1994b). The importance of posterior prevalence has been verified in many instances; for example, in the proper determination of pools of motoneurons (Tarchini et al., 2005) and in the early sequential migration of epiblast cells during chick gastrulation (Iimura and Pourquié, 2006).

Even though the mechanisms that underlie temporal collinearity are not yet fully understood, the use of the linear structure of the DNA molecule to support a time device is not difficult to conceptualize (see Deschamps and Van Nes, 2005). For instance, processes involving the directional spreading, or removal, of any kind of molecule or relative distance effects between an enhancer and neighboring target promoters (e.g. Tarchini and Duboule, 2006), could impose a collinear temporal regulation on any series of contiguous transcription units in a way that would be impossible to achieve if a gene’s genomic neighborhood were to disappear. Distinct mechanisms might also co-exist in the same animal in different contexts (tissues, cell-types, structures), because once temporal collinearity constrains genes to stay together, it paves the way for the recruitment of other collinear regulations. For example, the evolution of a global enhancer positioned outside a gene cluster might easily lead to a temporal sequence in the response of target genes, following a relative distance effect or even a stochastic process (Kmita and Duboule, 2003).

Consequently, a correlation must exist between the existence of a Hox gene cluster and the implementation of developmental strategies that determine rostral-to-caudal identities in a temporal sequence. Clustering need not be complete (for example, in those cases in which only a restricted caudal part of the embryo would segment following a temporal progression), but ought to involve those Hox genes that precisely determine these caudal segments. In the absence of this temporal parameter, a major constraint is released and the cluster can freely evolve towards a more disorganized state as described above. An illustration of this process is provided by *Drosophila*, in which both splitting and disorganization has occurred. However, the ANT-C sub-cluster is clearly more disorganized than is the BX-C cluster, which might be because the thoracic and abdominal parts were more exposed to this temporal constraint in the lineage that gave rise to Diptera (i.e. in short-germ insects). Because in *Drosophila* this constraint has also been released, owing to the particular mechanism of abdominal segmentation, it has been proposed that BX-C was permissive and available for rearrangements (Duboule, 1992). Since then, several different breakpoints have been isolated within BX-C, which separate *Ubx* from the two abdominal genes (*abd-A* and *Abd-B*) in *D. melanogaster* (Von Allmen et al., 1996) and in other species such as *Drosophila virilis* (Negre and Ruiz, 2007), that support this view.

### License to split? Inverting the constraints

Diptera provide a good example of the transition from a time sequence-dependent segmentation process (short-germ insects) to a time sequence-independent process, associated with modification of the Hox cluster following the release of the temporal constraint. This slow and progressive disorganization in *Drosophila* was recently explained by the complete release of all the constraints that keep these genes together in other animals (except between the pair of genes *abd-A* and *Abd-B*, which are never found separated). In this view, the existence of Hox clusters in *Drosophila* reflects the ‘phylogenetic inertia’ of the system, i.e. the difficulty of finding an acceptable breakpoint and of maintaining it at the population level (Negre and Ruiz, 2007). Accordingly, one might consider that the

more drastic rearrangements that lead to trans-collinearity, which are observed in other groups of bilaterian animals, also followed the emergence of developmental modes that no longer require a precisely timed sequence of Hox gene activation.

The relationship between cluster organization, the type of collinearity and the existence of larval stages is more difficult to generalize. Hox genes are required for the elaboration of the adult body plan, hence their function might not be crucial in primary larvae, such as in the dipleurula larvae of those sea urchin that show an indirect developmental mode, or in the lophotrochozoans polychaete annelid larvae (Peterson et al., 2000). Whereas these annelid larvae seem to implement temporal collinearity in the late activation of their Hox gene complement, the same is unlikely to be true for sea urchins. In any case, the fact that, in these species, Hox genes may be somewhat ‘set-aside’ much in the same way that set-aside cells may contribute to the definitive body plan, as argued by Davidson and colleagues (Peterson et al., 2000), does not help us to predict whether cis- or trans-collinearity contribute, and in which temporal context. In such indirect developmental modes, the necessity to activate Hox genes in a temporal sequence, and hence to maintain a gene cluster, might depend upon the fate of these set-aside cells, i.e. whether the derived rostral and caudal structures will be generated in a time sequence or concomitantly.

Understandably, the implementation of a completely determinative developmental strategy, i.e. developmental modes in which fates are generally invariable and determined early on, as for example in the ascidian *Oikopleura* (Seo et al., 2004) and in nematodes, must have made temporal collinearity unnecessary (Duboule, 1992). In addition, animals that display either a poorly segmented body plan, or a highly heteromeric series (i.e. ‘segments’ that do not bear any similarity or mechanistic relationship to each other), might no longer require any local enhancer sharing between pairs of neighboring Hox genes, allowing for a complete fragmentation of the cluster. Although these various factors point to a natural tendency towards disorganization, once constraints have been released, the alternative possibility, wherein the mere existence of a cluster itself represents a constraint for the evolving potential of an organism, should not be ignored.

In this somewhat counter-intuitive view, the effects of gene clustering in coordinating gene regulation in time might be detrimental to the implementation of one particular developmental strategy; for example, in the many animals that do develop their rostral and caudal extremities concomitantly. In such cases, the disorganization of the cluster might not reflect a release of some constraints (e.g. the loss of temporal collinearity), but, instead, might have been a necessary step to escape the obligation of activating ‘caudal’ genes only after ‘rostral’ genes, thus favoring, or accompanying, the shift to another developmental mode. In such a scheme, the cluster becomes a constraint, and cluster disorganization, much like the consolidation observed in vertebrates, is seen as an active process that is under positive selection, rather than being the result of tolerated evolutionary erosion.

### In the beginning

Recent comparisons between sets of Hox genes of various animals, including cnidarians, suggest that this original Hox cluster contained a rather substantial number of genes, belonging to all major classes of Hox genes represented in extant species (e.g. de Rosa et al., 1999; Garcia-Fernandez, 2005; Ryan et al., 2007). The fact that some particular groups were amplified subsequently – for example, *Abd-B*-related genes in vertebrates – does not change the basic problem,

which is: what mechanisms kept these genes together and which constraint was applied to this genomic structure, either to prevent fragmentation or to facilitate consolidation.

Although these considerations concerning the evolution of Hox clusters might indeed contribute to our understanding about mechanisms and the relationships between cluster type and developmental modes, they do not help us to understand which type of cluster and, accordingly, which type of collinearity was present, both at the origin of bilaterian animals and just before their radiation. We and others assumed (as discussed above) that the ancestral bilaterian animal had a type D cluster and that spatial cis-collinearity occurred, as virtually all animal classes analyzed to date display this latter property to some extent, either in cis or in trans. Accordingly, it is fair to speculate that such an animal was segmented, or at least displayed some reiteration of modular structures along its rostral-to-caudal axis. It is nevertheless more problematic to infer, either from available experimental data, or from theoretical considerations, whether or not temporal collinearity was implemented and, correspondingly, whether this ancestral animal produced its caudal segments with a time delay with respect to more-rostral structures.

Even though temporal collinearity is considered as the strongest constraint to maintain Hox genes in a single cluster (Duboule, 1992; Monteiro and Ferrier, 2006), it is conceivable that it emerged as a consequence of clustering, i.e. was made possible by the existence of a gene cluster, rather than being the original force that kept the genes together in an ancestral form. To propose an inverted sequence of events, as suggested by Ferrier and Holland (Ferrier and Holland, 2002), would once again, as for clustering itself (see above), support a vertebrate-centrist view that is unlikely to be correct, as it seems that few animal species implement a full temporal collinear mechanism, from the most rostral to the most caudal paralogy group. As discussed previously (Kmita and Duboule, 2003), collinear mechanisms (whether spatial or temporal) can come in different flavors, and hence the quest for a universal collinear process might be futile. In this view, it is perhaps simpler to consider the existence of a gene cluster that might have subsequently triggered the emergence of other collinear strategies (temporal or spatial), and that the various forms of Hox gene clusters that we contemplate today merely testify to particular histories of successive recruitment and abandonment of collinear mechanisms.

The scenario proposed above (temporal collinearity appearing after spatial cis-collinearity) would fit with the intuitive perception that regulatory mechanisms recruited, during evolution, on the top of pre-existing principles (in this case owing to the existence of a cluster, hence of cis-collinearity), should be generally less constrained than the former and thus more prone to being selected against (Duboule and Wilkins, 1998). The fact that temporal collinearity is certainly found in a minority of bilaterians, unlike trans- or cis-spatial collinearity, suggests that the former postdates the emergence of the latter. If true, the question as to when exactly temporal collinearity was recruited on top of a cis-collinear system will have to be answered. This problem will be difficult to address, even with a comprehensive description of many more genomes of animal species with well-described developmental modes. This is owing to the versatility of collinearity; once a cis-collinearity cluster is established and maintained, it can be used as a matrix where various collinear mechanisms can evolve, based on processes as different as DNA replication, chromatin spreading and distance effects, to name but a few. A superficial description of these novel model systems will lead to a simplified conclusion (it is, or it is not, collinear), which cannot faithfully reflect the evolutionary history of the mechanisms involved.

## Conclusion

The epistemic value of the Hox gene family has arguably not yet reached its full potential. The emergence of new model systems, as well as the availability of additional sequenced genomes, will help us to understand the functional evolution of this amazing gene family and will undoubtedly provide novel, widely applicable conceptual tools. In this review, I have discussed two rather counter-intuitive proposals to account for the functional and structural evolution of Hox gene clusters. First, genomic topographies (gene clusters in this case) can evolve towards 'more-organized states' through a process of consolidation. Second, the consolidation of Hox clusters was stimulated, or at least reinforced, after genome duplications, to accompany the emergence of vertebrates, as gene clusters were more readily able to recruit additional global regulatory controls. In the past, several structural, functional or regulatory features of Hox genes were successfully transposed to other genetic systems. Future work will tell us whether similar hypotheses can be proposed for other evolutionarily conserved gene clusters.

I thank M. Akam, C. Amemiya, A. Amores, D. Ferrier, B. Galliot, J. Garcia-Fernandez, S. Kuratani, F. Spitz, L. Wolpert and J. Woltering for communicating data and for discussions; J. Alfred and C. Garvey for careful editing of the manuscript; and G. Perec and the referees for inspiring suggestions. D.D. is supported by funds from the Canton de Genève, the Louis-Jeantet and Claraz Foundations, the Swiss National Research Fund, the National Research Center (NCCR) 'Frontiers in Genetics' and the EU programmes 'Cells into Organs' and 'Crescendo'.

## References

- Aboobaker, A. and Blaxter, M. (2003). Hox gene evolution in nematodes: novelty conserved. *Curr. Opin. Genet. Dev.* **6**, 593-598.
- Akam, M. (1989). Hox and HOM: homologous gene clusters in insects and vertebrates. *Cell* **57**, 347-349.
- 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.
- Amores, A., Suzuki, T., Yan, Y. L., Pomeroy, J., Singer, A., Amemiya, C. and Postlethwait, J. H. (2004). Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. *Genome Res.* **14**, 1-10.
- Brown, S. J., Shippy, T. D., Beeman, R. W. and Denell, R. E. (2002). Tribolium Hox genes repress antennal development in the gnathos and trunk. *Mol. Phylogenet. Evol.* **24**, 384-387.
- Callaerts, P., Lee, P. N., Hartmann, B., Farfan, C., Choy, D. W., Ikeo, K., Fischbach, K. F., Gehring, W. J. and de Couet, H. G. (2002). HOX genes in the septilid squid *Euprymna scolopes*: implications for the evolution of complex body plans. *Proc. Natl. Acad. Sci. USA* **99**, 2088-2093.
- Cameron, R. A., Rowen, L., Nesbitt, R., Bloom, S., Rast, J. P., Berney, K., Arenas-Mena, C., Martinez, P., Lucas, S., Richardson, P. M. et al. (2006). Unusual gene order and organization of the sea urchin Hox cluster. *J. Exp. Zool. B Mol. Dev. Evol.* **306**, 45-58.
- Chourout, D., Delsuc, F., Chourout, P., Edvardson, R. B., Rentzsch, F., Renfer, E., Jensen, M. F., Zhu, B., de Jong, P., Steele, R. E. et al. (2006). Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature* **442**, 684-687.
- Crow, K. D., Stadler, P. F., Lynch, V. J., Amemiya, C. and Wagner, G. P. (2006). The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts. *Mol. Biol. Evol.* **23**, 121-136.
- Delsuc, F., Brinkmann, H., Chourout, D. and Philippe, H. (2006). Tunicates and not cephalochordates are the closest relatives of vertebrates. *Nature* **439**, 965-968.
- 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.
- Deschamps, J. and van Nes, J. (2005). Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development* **132**, 2931-2942.
- Dollé, P., Izpisua-Belmonte, J.-C., Falkentsein, H., Renucci, A. and Duboule, D. (1989). Coordinate expression of the murine HOX-5 homeobox containing gene during limb pattern formation. *Nature* **342**, 767-772.
- Duboule, D. (1991). Patterning in the vertebrate limb. *Curr. Opin. Genet. Dev.* **1**, 211-216.
- Duboule, D. (1992). The vertebrate limb, temporal colinearity and the Hox/HOM gene network. *BioEssays* **14**, 375-384.
- Duboule, D. (ed.) (1994a). *Guidebook to Homeobox Genes*. Oxford: Oxford University Press.



- Duboule, D.** (1994b). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Dev. Suppl.* **1994**, 135-142.
- Duboule, D. and 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.
- Duboule, D. and Morata, G.** (1994). Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet.* **10**, 358-364.
- Duboule, D. and Wilkins, A.** (1998). The evolution of 'Bricolage'. *Trends Genet.* **14**, 54-59.
- Ferrier, D. E. K. and Holland, P. W. H.** (2002). *Ciona intestinalis* ParaHox genes: evolution of Hox/ParaHox cluster integrity, developmental mode, and temporal colinearity. *Mol. Phylogenet. Evol.* **24**, 412-417.
- Forlani, S., Lawson, K. A. and Deschamps, J.** (2003). Acquisition of *Hox* codes during gastrulation and axial elongation in the mouse embryo. *Development* **130**, 3807-3819.
- Galliot, B.** (2005). Relaxed constraints on *Hox* gene clustering during evolution. *Heredity* **94**, 277.
- García-Fernández, J.** (2005). The genesis and evolution of homeobox gene clusters. *Nat. Rev. Genet.* **6**, 881-892.
- García-Fernández, J. and Holland, P. W. H.** (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. and Galliot, B.** (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.
- Gaunt, S. J., Sharpe, P. T. and Duboule, D.** (1988). Spatially restricted domains of homeo-gene transcripts in mouse embryos: relation to a segmented body plan. *Dev. Suppl.* **104**, 169-179.
- Gonzalez, F., Duboule, D. and Spitz, F.** (2007). Transgenic analysis of *Hoxd* gene regulation during digit development. *Dev. Biol.* (in press).
- Graham, A., Papalopulu, N. and Krumlauf, R.** (1989). The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell* **57**, 367-378.
- Holland, P. W. H.** (1999). Gene duplication: past, present and future. *Semin. Cell Dev. Biol.* **10**, 541-547.
- Holland, P. W. H. and García-Fernández, J.** (1996). Hox genes and chordate evolution. *Dev. Biol.* **173**, 382-395.
- Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M., Wides, R. et al.** (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**, 129-149.
- Hurley, I., Hale, M. E. and Prince, V. E.** (2005). Duplication events and the evolution of segmental identity. *Evol. Dev.* **6**, 556-567.
- Iimura, T. and Pourquie, O.** (2006). Collinear activation of *Hoxb* genes during gastrulation is linked to mesoderm cell ingression. *Nature* **442**, 568-571.
- Ikuta, T., Yoshida, N., Satoh, N. and Saiga, H.** (2004). *Ciona intestinalis* Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proc. Natl. Acad. Sci. USA* **101**, 15118-15123.
- Izpisua-Belmonte, J. C., Falkenstein, H., Dollé, P., Renucci, A. and Duboule, D.** (1991). Murine genes related to the *Drosophila* AbdB homeotic gene are sequentially expressed during development of the posterior part of the body. *EMBO J.* **10**, 2279-2289.
- Kamm, K., Schierwater, B., Jakob, W., Dellaporta, S. L. and Miller, D. J.** (2006). Axial patterning and diversification in the cnidaria predate the Hox system. *Curr. Biol.* **16**, 920-926.
- Kessel, M. and Gruss, P.** (1991). Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. *Cell* **67**, 89-104.
- Kirschner, M. and Gerhart, J.** (2006). *The Plausibility of Life: Resolving Darwin's Dilemma*. New Haven, CT: Yale University Press.
- Kmita, M. and Duboule, D.** (2003). Organizing axes in time and space; 25 years of collinear tinkering. *Science* **301**, 331-333.
- Kmita, M., van der Hoeven, F., Zakany, J., Krumlauf, R. and Duboule, D.** (2000). Mechanisms of *Hox* gene colinearity: transposition of the anterior *Hoxb* gene into the posterior *HoxD* complex. *Genes Dev.* **14**, 198-211.
- Kmita, M., Fraudeau, N., Héroult, Y. and Duboule, D.** (2002). A serial deletion/duplication strategy *in vivo* suggests a molecular basis for *Hoxd* genes colinearity in limbs. *Nature* **420**, 145-150.
- Kmita-Cunisse, M., Loosli, F., Bierne, J. and Gehring, W. J.** (1998). Homeobox genes in the ribbonworm *Lineus sanguineus*: evolutionary implications. *Proc. Natl. Acad. Sci. USA* **95**, 3030-3035.
- Kondo, T., Zakany, J., Innis, J. W. and Duboule, D.** (1997). Of fingers, toes and penises. *Nature* **390**, 29.
- Krumlauf, R.** (1994). Hox genes in vertebrate development. *Cell* **78**, 191-201.
- Kurosawa, G., Takamatsu, N., Takahashi, M., Sumitomo, M., Sanaka, E., Yamada, K., Nishii, K., Matsuda, M., Asakawa, S., Ishiguro, H. et al.** (2006). Organization and structure of hox gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes. *Gene* **370**, 75-82.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitxHugh, W. et al.** (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
- Lehoczyk, J. A., Williams, M. E. and Innis, J. W.** (2004). Conserved expression domains for genes upstream and within the *HoxA* and *HoxD* clusters suggests a long-range enhancer existed before cluster duplication. *Evol. Dev.* **6**, 423-430.
- Lemons, D. and McGinnis, W.** (2006). Genomic evolution of *Hox* gene clusters. *Science* **313**, 1918-1922.
- Lewis, E. B.** (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565-570.
- Maeda, R. K. and Karch, F.** (2006). The ABC of the BX-C: the bithorax complex explained. *Development* **133**, 1413-1422.
- Minguillon, C., Gardenyes, J., Serra, E., Castro, L. F., Hill-Force, A., Holland, P. W., Amemiya, C. T. and García-Fernández, J.** (2005). No more than 14, the end of the amphioxus *Hox* cluster. *Int. J. Biol. Sci.* **1**, 19-23.
- Monteiro, A. S. and Ferrier, D. E. K.** (2006). *Hox* genes are not always collinear. *Int. J. Biol. Sci.* **2**, 95-103.
- Negre, B. and Ruiz, A.** (2007). HOM-C evolution in *Drosophila*: is there a need for *Hox* gene clustering? *Trends Genet.* **2**, 55-59.
- Ohno, S.** (1970). *Evolution by Gene Duplication*. Heidelberg: Springer-Verlag.
- Payre, J. and TERNELL, O.** (1994). A cluster of *Hox* genes in *Cantatrix sopranica* L. In *La classification des choses* (ed. G. Percec), pp. 345-351. Geneva University Press.
- Peterson, K. J., Irvine, S. Q., Cameron, R. A. and Davidson, E. H.** (2000). Quantitative assessment of *Hox* complex expression in the indirect development of the polychaete annelid *Chaetopterus* sp. *Proc. Natl. Acad. Sci. USA* **97**, 4487-4492.
- Pierce, R. J., Wu, W., Hirai, H., Ivens, A., Murphy, L. D., Noel, C., Johnston, D. A., Artiguenave, F., Adams, M., Cornette, J. et al.** (2005). Evidence for a dispersed *Hox* gene cluster in the platyhelminth parasite *Schistosoma mansoni*. *Mol. Biol. Evol.* **12**, 2491-2503.
- Prince, V. E., Joly, L., Ekker, M. and Ho, R. K.** (1998). Zebrafish hox genes: genomic organization and modified colinear expression patterns in the trunk. *Development* **125**, 407-420.
- Ryan, J. F., Mazza, M. E., Pang, K., Matus, D. Q., Baxeavanis, A. D., Martindale, M. Q. and Finnerty, J. R.** (2007). Pre-bilaterian origins of the *Hox* cluster and the Hox code: evidence from the sea anemone, *Nematostella vectensis*. *PLoS ONE* **2**, e153.
- Seo, H. C., Edvardsen, R. B., Maeland, A. D., Bjordal, M., Jensen, M. F., Hansen, A., Flaot, M., Weissenbach, J., Lehrach, H., Wincker, P. et al.** (2004). *Hox* cluster disintegration with persistent anterior-posterior order of expression in *Oikopleura dioica*. *Nature* **431**, 67-71.
- Sharpe, J., Nonchev, S., Gould, A., Whiting, J. and Krumlauf, R.** (1998). Selectivity, sharing and competitive interactions in the regulation of *Hoxb* genes. *EMBO J.* **17**, 1788-1798.
- Spitz, F., Gonzalez, F., Peichel, K., Vogt, T., Duboule, D. and Zakany, J.** (2001). Large scale transgenic and cluster deletion analysis of the *HoxD* complex separate an ancestral regulatory module from evolutionary innovations. *Genes Dev.* **15**, 2209-2214.
- Spitz, F., Gonzalez, F. and Duboule, D.** (2003). A Global Control region defines a chromosomal landscape containing the *HoxD* cluster. *Cell* **113**, 405-417.
- Tarchini, B. and Duboule, D.** (2006). Control of *HoxD* genes colinearity during early limb development. *Dev. Cell* **10**, 93-103.
- Tarchini, B., Huynh, T. H., Cox, G. A. and Duboule, D.** (2005). *HoxD* cluster scanning deletions identify multiple defects leading to paralysis in the mouse mutant Ironside. *Genes Dev.* **19**, 2862-2876.
- Trainor, P. and Krumlauf, R.** (2001). *Hox* genes, neural crest cells and branchial arch patterning. *Curr. Opin. Cell Biol.* **13**, 698-705.
- Von Allmen, G., Hogga, I., Spierer, A., Karch, F., Bender, W., Gyurkovics, H. and Lewis, E.** (1996). Splits in fruitfly *Hox* gene complexes. *Nature* **380**, 116.
- Wagner, G. P., Amemiya, C. and Ruddle, F. H.** (2003). *Hox* cluster duplications and the opportunity for evolutionary novelties. *Proc. Natl. Acad. Sci. USA* **100**, 14603-14606.
- Wellik, D. M. and Capecchi, M. R.** (2003). *Hox10* and *Hox11* genes are required to globally pattern the mammalian skeleton. *Science* **301**, 363-367.
- Woltering, J. M. and Durston, A. J.** (2006). The zebrafish *hoxDb* cluster has been reduced to a single microRNA. *Nat. Genet.* **6**, 601-602.
- Yasukochi, Y., Ashakumary, L. A., Wu, C., Yoshida, A., Nohata, J., Mita, K. and Sahara, K.** (2004). Organization of the *Hox* gene cluster of the silkworm, *Bombyx mori*: a split of the *Hox* cluster in a non-*Drosophila* insect. *Dev. Genes Evol.* **214**, 606-614.
- Zakany, J., Fromental-Ramain, C., Warot, X. and Duboule, D.** (1997). Regulation of number and size of digit by posterior *Hox* genes: a dose dependent mechanism with potential evolutionary implications. *Proc. Natl. Acad. Sci. USA* **94**, 13695-13700.