Extensive Chordate and Annelid MacrosyntenY Reveals Ancestral Homeobox Gene Organization

Jerome H. L. Hui,†,1 Carmel McDougall,‡,1 Ana S. Monteiro,§,1 Peter W. H. Holland,1 Detlev Arendt,2 Guillaume Balavoine,3 and David E. K. Ferrier*1

1Department of Zoology, University of Oxford, Oxford, United Kingdom
2EMBL Heidelberg, Heidelberg, Germany
3Institut Jacques Monod, CNRS/University Paris Diderot, France
†Present address: Faculty of Life Sciences, University of Manchester, Manchester, United Kingdom.
‡Present address: School of Biological Sciences, University of Queensland, Brisbane, Australia.
§Present address: Centro Andaluz de Biología del Desarrollo, Universidad Pablo de Olavide, Seville, Spain.

*Corresponding author: E-mail: dekf@st-andrews.ac.uk.

Abstract

Genes with the homeobox motif are crucial in developmental biology and widely implicated in the evolution of development. The Antennapedia (ANTP)-class is one of the two major classes of animal homeobox genes, and includes the Hox genes, renowned for their role in patterning the anterior–posterior axis of animals. The origin and evolution of the ANTP-class genes is a matter of some debate. A principal guiding hypothesis has been the existence of an ancient gene Mega-cluster deep in animal ancestry. This hypothesis was largely established from linkage data from chordates, and the Mega-cluster hypothesis remains to be seriously tested in protostomes. We have thus mapped ANTP-class homeobox genes to the chromosome level in a lophotrochozoan protostome. Our comparison of gene organization in Platynereis dumerilii and chordates indicates that the Mega-cluster, if it did exist, had already been broken up onto four chromosomes by the time of the protostome–deuterostome ancestor (PDA). These results not only elucidate an aspect of the genome organization of the PDA but also reveal high levels of macrosyntenY between P. dumerilii and chordates. This implies a very low rate of interchromosomal genome rearrangement in the lineages leading to P. dumerilii and the chordate ancestor since the time of the PDA.

Key words: Platynereis dumerilii, Antennapedia-class, homeobox, protostome–deuterostome ancestor.

Introduction

The Mega-cluster hypothesis provides the principal current framework for understanding the origin and early evolution of the homeobox-containing genes of the Antennapedia-class (ANTP-class) in animals. The ANTP-class contains such important developmental control genes as the Hox, ParaHox, and NK genes, widely involved in multiple developmental processes across the animal kingdom. The Mega-cluster hypothesis proposes that deep in animal ancestry, a precursor gene of the ANTP-class (the Proto-ANTP gene) underwent several rounds of tandem duplications to form a cluster of genes that were the evolutionary precursors for the genes of the Hox cluster, the ParaHox cluster, the NK cluster, and several closely related gene families (Pollard and Holland 2000; Garcia-Fernandez 2005) (fig. 1). This Mega-cluster then broke apart in several locations as lineages diverged from this ancestral state, at locations hypothesized to have been effectively random across different animal lineages, apart from in functionally constrained clusters such as the Hox cluster. Different extents of Mega-cluster remains are therefore predicted to reside in different lineages, such that comparison of the different evolutionary relics of the Mega-cluster should allow its composition to be deduced.

Alternatives to the Mega-cluster hypothesis can be envisaged that correspond to either of two general forms. First, at least some nontandem duplications could have been involved, either of genes or of groups of genes, before all the component gene families of the hypothesized Mega-cluster had arisen, or a precursor cluster prior to the Mega-cluster may have split apart, such that the complete Mega-cluster never in fact existed. This type of scenario can be considered as a “weak” form of the Mega-cluster hypothesis since even though the complete Mega-cluster did not exist under such a scenario, significant portions of the ANTP-class still evolved within clustered arrangements. The second alternative is that ANTP-class genes secondarily came to be clustered in distinct animal lineages after originating in locations scattered around ancestral genomes. These different scenarios would be expected to lead to different ANTP-class linkages in distinct lineages. The “strong” and weak forms of the Mega-cluster...
**FIG. 1.** Summary of the ANTP-class Mega-cluster hypothesis. Deep in animal ancestry, a single ProtoANTP gene tandemly duplicated to generate the ANTP-class homeobox gene families of animals, a selection of which are shown here. The resultant Mega-cluster is hypothesized to have then broken in several places, presumably independently on different animal lineages. The diagonal lines show the breaks deduced to be present in the chordate ancestor, which separated these genes into three distinct chromosomes. Precise gene order is not known in the ancestral Mega-cluster except within the ParaHox, Hox, and NK clusters. Conventionally, these ANTP-class genes are classified as ParaHox (pale gray boxes), Hox-like (dark gray boxes), and NK-like (black boxes), with Dlx being an NK-like gene (black) linked with Hox-like genes. However, this nomenclature may be misleading and unjustified (see text for details) and we instead propose the use of "Hox-linked" and "NK-linked", in which case Dlx is Hox-linked (dark gray) in chordates. Note the relative timing of the duplications to generate the full complement of ProtoHox, HoxL, NK, and NKL genes is still debated and some components may well have arisen before others (e.g., NK expansion before Hox; Larroux et al. 2007).

The hypothesis would be expected to lead to similar, but not identical, sets of linked ANTP-class genes in distinct lineages if the cluster breaks in effectively random locations. Distinguishing between the strong and weak forms of the Mega-cluster hypothesis requires determination of ANTP-class linkage patterns across a diversity of animal lineages. In contrast, if the ANTP-class linkages evolved from a secondary "coming together" from ancestral scattered locations, then we would expect a largely random assortment of ANTP-class linkages across distinct animal lineages. To date, a comprehensive analysis of the evolution of the Mega-cluster has been hampered by two major problems.

The first problem is the paucity of chromosome-level mapping data. Such data are available from chordates, which ancestrally, and in the prototypical chordate genome of amphioxus (*Branchiostoma floridana*), had/have the Hox cluster and the ANTP-class families Evx, Mox, Gbx, En, Mnx, and Dlx on one chromosome, the ParaHox cluster on a second chromosome, and the remains of the NK cluster on a third chromosome (Castro and Holland 2003; Luke et al. 2003; Castro et al. 2006). To extend the depth of animal ancestry to which the extent of the hypothesized Mega-cluster can be traced, comparisons to protostomes are required. Chromosome-scale resolution in protostomes is currently restricted to the ecdysozoan superphylum (Coulier et al. 2000; Castro and Holland 2003). Unfortunately, the low chromosome number in such ecdysozoan models as *Drosophila* and *Caenorhabditis elegans* means that these systems provide little resolution, due to the difficulty in distinguishing ancestral linkage patterns from chance secondary associations. Also, a high level of genome rearrangement has been described for insects and nematodes (Zdobnov et al. 2005; Zdobnov and Bork 2007), further confounding elucidation of ancestral linkages. Here, we use an alternative protostome for comparison, the polychaete *Platynereis dumerilii*, which resides in the lophotrochozoan superphylum rather than in the Ecdysozoa. *P. dumerilii* has a haploid number of 14 (Jha et al. 1995), a chromosome number potentially similar to that of the protostome–deuterostome ancestor (PDA), if the eumetazoan ancestor had about 13 linkage groups (Putnam et al. 2007) and the chordate ancestor about 17 (Putnam et al. 2008). Combined with a genome size of approximately $1 \times 10^9$ bp, this reduces the likelihood of chance secondary associations of gene linkage that have confounded the insect and nematode comparisons.

The second problem is the difficulty in obtaining robust phylogenetic information from the ANTP-class genes. Dlx is a key gene in the Mega-cluster hypothesis because Dlx is conventionally deemed to have greater sequence similarity to the genes of the NK cluster than to those of the Hox cluster, such that Dlx is often referred to as an "NK-like" gene rather than a "Hox-like" gene. Since this supposed NK-like gene is linked to the Hox cluster in chordates, it has been held to provide evidence for the ancient linkage of NK-like and Hox-like genes in the Mega-cluster, with an interchromosomal translocation occurring that separated all of the NK-like genes, except for Dlx, from the Hox-like genes (Pollard and Holland 2000; Garcia-Fernandez 2005) (fig. 1). The lack of robust resolution of many interfamilial relationships within the ANTP-class in molecular phylogenies based on homeodomain sequences, particularly the relationship of the Dlx family to the Hox and NK genes, has important implications for the Mega-cluster hypothesis. Notably, Dlx is not robustly classified as being more similar to Hox or NK genes via molecular phylogenies; in different analyses, Dlx has been placed closer to the NK cluster genes, the Hox cluster genes, or even outside both NK and Hox.
genes together, and never with strong support values (Howard-Ashby et al. 2006; Monteiro et al. 2006; Ryan et al. 2006; Larroux et al. 2007; Takatori et al. 2008; reviewed in Ferrier 2008). Furthermore, in an attempt to discover diagnostic residues for both the so-called Hox-like and the NK-like groups, the Dlx family could not be accommodated (Fonseca et al. 2008). This emphasizes the weakness of the Hox-like and NK-like superfamilies. Chromosome location and linkage patterns are perhaps a more robust guide to ancestral associations between ANTP-class families; hence, our proposal that the Hox-like and NK-like be abandoned and “Hox-linked” and “NK-linked” adopted as more informative alternatives (Ferrier 2008).

The key question is then, which genes are linked to the Hox and NK clusters and do bilaterian lineages exhibit overlapping patterns of Hox- and NK-linked genes? This would imply an intact Mega-cluster in the PDA and independent breakups along diverging lineages. The alternative scenario, of matching patterns of Hox-linked and NK-linked gene families in diverse bilaterian lineages, would instead imply that the Mega-cluster had already broken up to this extent by the time of the PDA or that an extensive Mega-cluster did not in fact exist. Here, we test this by determining the locations of the ANTP-class genes to discover whether the patterns of Hox-linked and NK-linked genes in chordates and a protostome like Platynereis overlap or instead show the same divisions.

We have mapped ANTP-class homeobox genes to the chromosome level in a lophotrochozoan protostome, the annelid P. dumerilii, for the first time and compared this gene organization in P. dumerilii with available whole-genome sequences. Our data reveal that the Mega-cluster, if it did exist, had already broken up onto four chromosomes by the time of the PDA. These results not only elucidate an aspect of the genome organization of the PDA but also reveal high levels of macro-synteny between P. dumerilii and chordates. This implies a very low rate of interchromosomal genome rearrangement in-between the PDA but also reveal high levels of macrosynteny being not only elucidate an aspect of the genome organization of P. dumerilii, and the parameters were 94 °C for 5 min, 35 times (94 °C for 1 min, 55 °C for 1 min, and 72 °C for 1.5 min), and 72 °C for 10 min. Cloning and sequencing were performed as described above. 3′ RACE was carried out following the manufacturer’s instructions (SMART RACE cDNA Amplification Kit, Clontech). Gene-specific primer PdMox3’ RACE-F1, 5′-CCATACACACTCCGTAGGACAGCATG-3′, was designed based on the sequence information obtained from PCR with degenerate primers. PCR was performed on 3′-RACE-Ready cDNA using the gene-specific primer and the universal primer, UPM (5′-CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT-3′). Five cycles (94 °C for 30 s and 72 °C for 3 min), five cycles (94 °C for 30 s, 70 °C for 30 s, and 70 °C for 3 min), and 25 cycles (94 °C for 30 s, 68 °C for 30 s, and 70 °C for 3 min) of amplification were carried out. The PCR products were analyzed on a 1% agarose gel. Purification of DNA, ligation, transformation, and sequencing were performed as described above. Sequences were aligned with other Mox genes as shown in supplementary fig. S1, Supplementary Material online. The PdMox cDNA is deposited in Genbank with accession number HQ263150.

Materials and Methods
Gene Cloning
Products from polymerase chain reactions (PCRs) were resolved by 1% agarose gel electrophoresis and the bands of appropriate sizes purified with the GFX PCR DNA and Gel Band Purification Kit (Amersham Biosciences). The purified DNA was ligated with the pGEM-T Easy Vector (Promega) and transformed into XL1-Blue chemically competent cells. The QIAprep Spin Miniprep Kit (Qiagen) was used to purify the plasmids, and sequencing of plasmids was done with BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems).

Degenerate primers NewPdMoxF1 5′-AAACCAAGGAAG-GA(A/G)AG(A/G)AC(A/G)GC-3′, NewPdMoxF2 5′-ACCA-AACA(C/T)CAGATCC(A/G)(T/G)GAA, and NewPdMoxRP 5′-TTCCATTTGATTCC(T/G)GTTTG-3′ were designed based on sequences from the Mox genes of other bilaterians (supplementary fig. S1, Supplementary Material online). NewPdMoxF2 was used as the nested primer after the primary PCR of NewPdMoxF1 and NewPdMoxRP. PCR was performed on the genomic DNA of P. dumerilii, and the parameters were 94 °C for 5 min, 35 times (94 °C for 1 min, 55 °C for 1 min, and 72 °C for 1.5 min), and 72 °C for 10 min. Cloning and sequencing were performed as described above. 3′ RACE was carried out following the manufacturer’s instructions (SMART RACE cDNA Amplification Kit, Clontech). Gene-specific primer PdMox3’ RACE-F1, 5′-CCATACACACTCCGTAGGACAGCATG-3′, was designed based on the sequence information obtained from PCR with degenerate primers.

BAC Library Screening
Genomic bacterial artificial chromosome (BAC) and phage clones were isolated as described in Hui et al. (2009) using cDNA fragments as probes (details in supplementary material, Supplementary Material online). Probes containing the ANTP-class homeobox genes were generated with the PCR DIG labelling mix kit (Roche, Germany). DNA was purified with the GFX PCR DNA and Gel Band Purification Kit (Amersham Biosciences). Filters containing the BAC library constructed for P. dumerilii (BACPAC Resources, Children’s Hospital Oakland Research Institute, http://bacpac.chori.org/library.php?id=209) were screened with DIG-labeled probes according to the DIG nucleic acid detection kit (Roche, Germany). Hybridization was carried out at 42 °C overnight with gene-specific probe (25 ng/ml DIG Easy Hyb buffer) and washed (2× saline sodium citrate [SSC]/0.1% sodium dodecyl sulfate [SDS], room temperature, and 2× 15 min; 0.5× SSC/0.1% SDS, 60 °C, and 2× 15 min). Hybridization signals were detected by CDP-star, ready-to-use (Roche, Germany), and exposed to BioMax Light film (Kodak).

BAC Clone Preparation and Sequencing
Each BAC clone was purified with the QIAfilter Plasmid Midi kit (Qiagen) and prepared for chromosomal fluorescent in-situ hybridization (FISH) after confirming the gene content by restriction enzyme digestion (EcoRI and HindIII as frequent cutters, NotI to check the insert size by cutting the insert with restriction sites flanking the polylinker.
promoter regions of the vector), gene-specific primer PCR, and sequencing. To discover whether some BAC clones were overlapping and to design primers for genome walking at the ends of BAC clones, some BAC clones were subjected to end sequencing. Steps were carried out as described in the protocol of the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems) with primers pTARBAC2.1T7, 5′-GCCGCTAATACGACTCACTATAGGGAG-3′, and pTARBAC2.1Sp6, 5′-ACTTGGCTGTTTTTACAAATT-3′, with the following modifications. Each reaction consisted of 10 pmol sequencing primer, 2.5 µl BigDye sequencing buffer, 2 µl ready reaction premix, and 0.5–1 µg of DNA template and was prepared and run through thermal cycles of 96 °C for 1 min and 50 cycles (96 °C for 30 s, 50 °C for 10 s, and 60 °C for 4 min).

Genomic Phage Library Screening
Since some of the screenings of the BAC library did not provide any positive signals with several probes, including PduEmx, PduGbx, PduEngrailed, PduVax, PduNat, PduLbx, and PduNKS probes, a genomic phage library of P. dumerilii, synthesized by Lofstrand Labs Ltd., was screened. Preparation of the host strains, phage infection, plating cultures, and plaque lifts were all carried out as described in the manufacturer’s instructions (Lambda FIX II Library, Stratagen). The primary library contained 5 × 10^{10} pfu/0.4 ml of SM buffer (100 mM NaCl, 8 mM MgSO_{4}·7H_{2}O, 50 mM Tris–HCl, 0.002% gelatin) with 7% dimethyl sulfoxide (DMSO); whereas the amplified library contained 2.6 × 10^{10} pfu/ml of SM buffer with 7% DMSO. The host bacteria strain was XL-1 Blue MRA (P2). The average insert size contained within the vector is about 17 kb. Two amplified phage library plates were screened at ~200,000 pfu on each plate. Plaques were transferred onto a charged nylon membrane (Hybond N+, Amersham Biosciences) with hybridization, washing, and detection steps carried out as for the BAC library screening described above. Secondary and tertiary screenings were further performed before a single plaque was isolated for large-scale DNA preparation with Wizard DNA Clean-up Systems (Promega). All phages were subjected to gene-specific primer sequencing before performing chromosomal FISH. A phage containing PduLbx and a phage containing PduNKS were confirmed.

Chromosomal FISH
Two-color FISH was performed as described in Hui et al. (2009). Briefly, genomic clones were labeled with either digoxigenin or fluorescein-12-dUTP and nick translation and detected with DIG-rodamine Fab fragments plus Texas-Red anti-sheep antibody or Alexa Fluor 488 signal-amplification kit, respectively. Pairs of ANTP-class probes were hybridized to metaphase chromosome spreads, which were counterstained with 4,6-diamidino-2-phenylindole. Images were captured with a Zeiss Axioskop microscope equipped with an Axiocam camera. All images were processed with whole-layer color adjustment on the complete image with Adobe Photoshop 7.0.

Results

Dlx Is Hox-Linked Rather than NK-Like
In P. dumerilii, we find that PduDlx is linked to the Hox cluster, resembling the situation in chordates. A BAC clone containing PduDlx localizes to the telomeric end of a chromosome to which BAC clones containing Hox cluster genes also localize, with the Hox probes labeling a site closer to the centromere (e.g., PduLox5, fig. 2e), except for the Hox gene PduPost1 (see below). Given the ambiguity of the interfamily relationships between the ANTP-class homeobox genes, particularly that of Dlx relative to the Hox and NK cluster gene families in molecular phylogenies (see Introduction), we here refer to Dlx as a Hox-linked gene in both Platynereis and chordates rather than use the possibly misleading term NK-like.

Absence of Hox and ParaHox “Cluster” Linkage
The Hox and ParaHox clusters were postulated to have originated from the duplication of a single ancestral cluster that existed early in metazoan evolution, the ProtoHox gene cluster (Brooke et al. 1998). It is unclear whether this separation into Hox and ParaHox clusters was a tandem duplication (Minguillon and Garcia-Fernandez 2003), a trans-duplication (Brooke et al. 1998), or a break of a single long cluster into two (Ryan et al. 2007). In P. dumerilii, the Hox locus (intact apart from PduPost1, see below; Monteiro AS and Ferrier DEK, unpublished) and the ParaHox locus (split in two but on a single chromosome; Hui et al. 2009) are located on two different chromosomes. PduDvax, which is representative of the ParaHox chromosome (fig. 2a), labels a chromosome that is distinct from the Hox cluster-bearing chromosome labeled by PduLox5 (fig. 2c). This is the first example of Hox and ParaHox clusters being mapped to two different chromosomes in a lophotrochozoan protostome.

Hox-Linked Genes and PduPost1
In the formulation of the Mega-cluster hypothesis, the chordate Hox chromosome contains several other ANTP-class genes in addition to the Hox genes and Dlx. These include three genes originally called the EHGbox genes (Engrailed [en], HB9/Mnx, and Gbx), plus Mox/Mex and Evx; this last gene having been a member of the ancestral Hox cluster and retained in this location in chordates and cnidarians (Miller and Miles 1993; Pollard and Holland 2000; Garcia-Fernandez 2005; Ryan et al. 2007; Holland et al. 2008). Here, we show that P. dumerilii orthologues of Mnx, Mox, and Evx are still Hox-linked (fig. 2a, b, and d). PduEvx, despite still being Hox-linked in Platynereis, is however no longer a member of the Platynereis Hox cluster, with PduPost2 (and the rest of the Hox cluster) being relatively close to the centromere, whereas PduEvx is located closer to a telomere (fig. 2h). This disruption of the “posterior” end of the Hox cluster is further demonstrated by loss of the posterior Hox gene PduPost1 from the cluster (fig. 2). Here, we show that not only has PduPost1 been relocated out of the Hox cluster (fig. 2c) but also that it has been translocated to the “NK cluster” chromosome (discussed below) (fig. 2h and k).
**FIG. 2.** Chromosomal FISH mapping the positions of pairs of ANTP-class homeobox genes in the protostome *Platynereis dumerilii*. All the localized *P. dumerilii* ANTP-class homeobox genes are located on only four chromosomes, represented by colored boxes: red, Hox chromosome (a–e); green, NK chromosome (f–l); orange, NK2 chromosome (m–n); purple, ParaHox chromosome (o). The pale blue box (p–v) represents the gene pairs that distinguish the four *Platynereis* ANTP-class chromosomes from each other. Scale bars = 20 µm. For further details, see supplementary material, Supplementary Material online.
NK Cluster Gene Organization

In addition to Hox and ParaHox clusters, there is a further ANTP-class gene cluster in some animals, the NK cluster of arthropods. In *Drosophila*, the NK cluster contains *tinman/NK4*, *bagpipe/NK3*, *ladybird-early* and *ladybird-late* (together being the fly representatives of Lbx), *C15/Tlx*, and *slouch/NK1* (Jagla et al. 2001). Ancestrally, the insect NK cluster also contained *Msx/ Drop and Hmx/NK5* (Garcia-Fernandez 2005; Richards et al. 2008). NK gene clustering in insects is in marked contrast to the situation in chordates, in which the ancestral NK cluster has been broken up. Despite this disruption, in amphioxus (a proxy for the ancestral chordate condition), NK4, NK3, NK1, NK6, NK7, Lbx, Tlx, and Msx are all on one chromosome (Luke et al. 2003; Butts et al. 2008; Putnam et al. 2008; Wotton et al. 2009). The NK cluster of the PDA thus probably consisted of nine NK genes (Msx, NK4, NK3, Lbx, Tlx, NK7, NK6, NK1, and Hmx/NK5), with a five to six gene core tending to remain clustered in insects (Butts et al. 2008). We mapped seven NK cluster genes in *P. dumerilii* (*PduMsx*, *PduNK4*, *PduNK3*, *PduLbx*, *PduTlx*, *PduNK6*, and *PduNK1*) and provide the first evidence that the NK cluster genes are located within a single chromosome in a lophotrochozoan (fig. 2f–l). In *P. dumerilii*, *PduNK4*, *PduNK3*, and *PduMsx* are tightly linked together close to the telomere-end of the long arm of the NK chromosome (figs. 2i and j). *PduLbx* and *PduTlx* are also close to each other but are located close to the centromere of the long arm of the NK chromosome (figs. 2f and g). *PduNK1* is then further separated from these other NK components and is located at the opposite telomere from *PduMsx* (fig. 2i).

Separation of the NK2 Genes

NK2 genes are another NK gene family in animals, but these have not been accommodated in the Mega-cluster hypothesis until now. In *P. dumerilii*, the NK2 family (*PduNK2.1, PduNK2.2a, and PduNK2.2b*) is on a distinct chromosome from those bearing the NK cluster genes, the Hox-linked genes, and the ParaHox genes (fig. 2m, n, p, q and u). In all animals examined to date, NK2 genes are on a different chromosome from the NK cluster genes, except for the *scarecrow* (*scro*) gene of the mosquito *Anopheles gambiae*, which we propose is a secondary association in this insect (see Discussion).

Discussion

Hox and ParaHox Clusters Were not Linked in the PDA

The location of Hox and ParaHox cluster loci on distinct chromosomes in *Platynereis* and amphioxus, and by extension the PDA, implies that if the ParaHox cluster originated via a large tandem duplication, then this must have been followed by translocation before divergence of the protostome and deuterostome lineages. In contrast to the *Platynereis*/amphioxus situation, some instances of Hox and ParaHox gene linkages are seen in other invertebrates, such as linkage of *OdiCdx* to *OdiHox1* (See et al. 2004) in the urochordate *Oikopleura dioica*, and the linkage of Cdx to the Hox cluster in the recently sequenced genome of the crustacean *Daphnia pulex*. In these cases, however, various losses or duplications of ParaHox genes in both *O. dioica* and arthropods (*Daphnia* and insects), as well as scattering of the Hox cluster in *O. dioica* through the small compacted genome or the extensive genome rearrangements across the sequenced Ecdyssozoa (see Introduction), prevent confident resolution of ancestral organization, as opposed to lineage-specific secondary associations. Consequently, we favor the scenario in which the Hox and ParaHox clusters were arranged on two separate chromosomes in the PDA, an arrangement that has been conserved in both *Platynereis* and amphioxus (fig. 3).

NK Genes: Comparable Breaks in a Lophotrochozoan and Chordates and Ancient Separation of NK2

The *Platynereis* NK cluster genes have been separated into three ‘segments’ along a single chromosome (fig. 2f–l). Intriguingly, the positions of these breaks in the *Platynereis* cluster are the same as those found in amphioxus, except for a further separation of NK6 away from Lbx and Tlx in the *Platynereis* lineage (fig. 3). Perhaps shared regulatory elements or long-range regulatory mechanisms acting over these distinct subsets of the NK cluster genes might be the basis by which these subsets have been conserved in the chordate and lophotrochozoan lineages.

NK2 genes are not linked to NK cluster genes in any animal examined to date, except for the mosquito *scro* gene. However, since *A. gambiae* has only three chromosomes, the linkage of mosquito *scro* with the NK cluster genes is most likely a secondary rather than ancient association. Therefore, we suggest that separation of the NK2 family from the ancestral ProtoNK cluster-bearing chromosome pre-dates the PDA, either via a rare trans-duplication such that the NK2 family was never in the Mega-cluster or the NK2 family originated in the cluster and then translocated away before the origin of the PDA (fig. 3). Chromosome-scale mapping data from nonbilateralian lineages may resolve these two possibilities.

Genome Architecture and ANTP-Class Macrosynteny

Linkage of the ANTP-class homeobox genes on a select few chromosomes is likely to be due to a combination of evolutionary history (e.g., descended from tandem arrays) combined with functional constraints, such as those known for the long-range regulation of subsets of these genes such as the Hox cluster. Additionally, these genes may well be targets of common regulatory factors which may have led to their retention in particular regions of the genome (Janga et al. 2008) or are subject to some other form of nonrandom distribution constraint that is poorly understood at present (Hurst et al. 2004; Batada and Hurst 2007; Koonin 2009). This in turn may account for the unusual occurrence of the posterior Hox gene (*PduPost1*) re-locating to the NK chromosome in *P. dumerilii*. The
translocation of this gene may have been constrained by a need to exist in a genomic region conducive to its appropriate and viable regulation. Other possibilities for this arrangement include inheritance from a Mega-cluster arrangement with the Hox and NK clusters linked (unlikely, as another lophotrochozoan Post1, in *Lottia gigantea*, is still within the Hox cluster) or a chance relocation of PduPost1 to the NK cluster chromosome.

The possibility that translocating ANTP-class homeobox genes may move to a nonrandom location has major implications for the Mega-cluster hypothesis. This hypothesis is based on a supposed NK-like gene (Dlx) being linked to the Hox cluster. Whether or not Dlx really is an NK-like gene or not is one problem for the Mega-cluster hypothesis (see above). A second problem is that if Dlx did begin life with the NK cluster and not with the Hox cluster, then there may be a very limited range of genomic locations that it could translocate to. Thus, Dlx linkage to the Hox cluster may not be indicative of (or the result of descent from) a linkage of Hox- and NK-related genes but instead may reflect Dlx translocating from the NK array and being restricted to a small number of locations in which it could be viably accommodated, one of which is the Hox chromosome.

The Status of the Mega-Cluster Hypothesis

Given the above possibility of constrained genome architecture restricting the viable destinations of translocating genes, is there then any evidence that a Mega-homeobox cluster existed at all? The recently sequenced genome of a noninsect arthropod, *Daphnia pulex* (which has a haploid number of 12), is intriguing, as a proportion of the NK cluster genes (NK4, NK3, Lbx, and Tlx) are on the same scaffold as the Hox cluster of *D. pulex*, as is the only remaining *Daphnia* ParaHox gene, Cdx. Perhaps, this linkage is evidence for the ancient linkage of the Hox, ParaHox, and NK clusters in a Mega-cluster. If this were the case, however, it is odd that none of the other Hox-linked or NK cluster and NK-linked genes are on this *Daphnia* scaffold, including Evx, Dlx, Mox, En, Mnx, Gbx, NK1, or NK5, at least some of which would be expected in the intervening distance. Also, the extensive macrosynteny that we show between *Platynereis* and chordates implies that this *Daphnia* organization is secondarily derived from a highly rearranged ancestor somewhere in ecdysozoan evolution. This would also be consistent with the unusual organization of the *Daphnia* genome, with one chromosome constituting 25% of the total nuclear DNA and being much longer than all of the remaining chromosomes (Colbourne et al. 2011). Of the remaining chromosomes, another 30% of the nuclear DNA is found on chromosomes 2−4, so that only four of the 12 *Daphnia* chromosomes contain approximately 55% of the nuclear DNA. This implies that there are a number of gene linkages that result from secondary associations evolving on the *Daphnia* lineage, which were not present in an ancestor with genes distributed more uniformly across the karyotype. A confident understanding of the linkages seen in *D. pulex* requires the availability of more ecdysozoan genomes, particularly those with haploid numbers that have not been greatly reduced by chromosome fusions.

Further evidence may come from diploblast genomes. Diploblast and chordate genomes show surprisingly high levels of synteny (Putnam et al. 2007; Srivastava et al. 2008), which contrasts with an almost complete lack of synteny between insects, nematodes, and chordates (Zdobnov et al. 2005). Comparisons performed between genomes of the cnidarian *Nematostella vectensis* and vertebrates allowed the identification of Putative Ancestral Linkage groups (PALs) of the ancestral eumetazoan (Putnam et al. 2007). Unfortunately, these PALs do not accommodate many of the *Nematostella* ANTP-class gene-containing scaffolds and do not reveal any Hox-linked and NK-linked associations that...
unambiguously resolve the veracity of the Mega-cluster hypothesis, except for one possible exception. *Nematostella* scaffold 3 is part of the Hox PAL, and as well as containing Hox genes, also contains Msx, a component of the ancestral bilaterian NK cluster (see above). This then may be a remnant from, and evidence for, an ancestral Mega-cluster. Chromosome-level mapping in *Nematostella* and further diploblast genomes assembled to large contig sizes will be needed to test this speculation. Notwithstanding the limits of the *Nematostella* data in the present context, Putnam et al. (2007) speculated that only a few breaks or fusions per chromosome had happened since the eumetazoan ancestor in the *Nematostella* and chordate lineages. Our *Platynereis* data reinforce and extend this view. *Platynereis* also has a very conservative genome and the ecdysozoan models of arthropods and nematodes have undergone extensive genome rearrangements that are not typical for protostomes as a whole.

The genome-level organization of the ANTP-class homeobox genes is thus most likely to have evolved via a pre-PDA Mega-cluster stage, or, a state close to a Mega-cluster but with only very few breaks before the evolution of all the constituent homeobox gene families (the weak Mega-cluster scenario, see Introduction). If this weak Mega-cluster scenario is true, then there were only a maximum of three breaks (either non-tandem duplications or pre-Mega-cluster translocations) prior to the PDA, to produce the four separate components; the ParaHox, Hox-linked, NK-linked, and NK2 still seen in *Platynereis* and amphioxus. What is clear is that there is a surprisingly high degree of similarity of the organization of ANTP-class genes between *Platynereis* and the chordate ancestor, which is inconsistent with a secondary coming together or clustering of these genes from widely scattered locations in an ancestral genome.

**Conclusions**

Fossil data show that the major diversification of animal phyla, including polychaetes, occurred by the end of the Cambrian period (Conway Morris and Peel 2008). The polychaete *P. dumerilii* is hypothesized to have retained many primitive morphological characteristics of the PDA, including ancestral cell types and developmental mode (Arendt et al. 2004; Tessmar-Raible et al. 2007; Dray et al. 2010). At the genomic level, this animal has also proven to be less derived than several other invertebrate systems, retaining a high proportion of ancestral introns (Raible et al. 2005) and elements of microsynteny (Hui et al. 2009). We now reveal a strikingly low level of homeobox gene rearrangement and conserved macrosynteny with chordates. These data imply that the *P. dumerilii* lineage evolved significantly slower than many other protostome lineages since the Cambrian. Although it remains a distinct possibility that the ancestral Mega-cluster existed and the PDA contained the vast majority of the hypothetical Mega-cluster components on a single chromosome, we err on the side of caution here until more genome sequences are available along with further chromosome-level mapping work similar to that described here. Our conservative interpretation of the 550-million-year-old PDA arrangement of the ANTP-class homeobox genes is that they were split into four distinct chromosomes: the Hox, ParaHox, NK, and NK2, an arrangement that persisted to the present day in the *Platynereis* and amphioxus lineages.

**Supplementary Material**

Supplementary text, figures S1 and S2, and table S1 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

Work in the authors’ laboratories is supported by the Royal Society and the Biotechnology and Biological Sciences Research Council. J.H.L.H. was supported by the Hong Kong and China Oxford Scholarship Funds and the Merton College Dorus scholarship and C.M. was supported by a Commonwealth scholarship. Author Contributions: J.H.L.H., C.M., and D.E.K.F. wrote the manuscript. D.E.K.F. conceived, designed, and supervised the project. J.H.L.H. prepared chromosome material, performed RACE and FISH, and isolated most of the BAC clones; C.M. isolated the Dlx BAC clones and performed the Dlx FISH; A.S.M. isolated Hox BAC clones; D.A. and G.B. provided cDNA clones; G.B. provided animals; P.W.H.H. co-supervised the study. All authors approved the final version of the manuscript.

**References**


