Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements.

Daniel Chourrout, Frédéric Delsuc, Pascal Chourrout, Rolf B. Edvardsen, Fabian Rentzsch, Eduard Renfer, Marit F. Jensen, Baoli Zhu, Pieter de Jong, Robert E. Steele, et al.

MINIMAL PROTOHOX CLUSTER INFERRED FROM BILATERIAN AND CNIDARIAN HOX COMPLEMENTS

D. Chourrout[1*], F. Delsuc[2], P. Chourrout[3], R.B. Edvardsen[1], F. Rentzsch[1], E. Renfer[1], M.F. Jensen[1], B. Zhu[4], P. de Jong[4], R.E. Steele[5] & U. Technau[1*].

[1]Sars International Centre for Marine Molecular Biology, University of Bergen, Thormoehlensgt. 55, 5008 Bergen, Norway

[2]Laboratoire de Paléontologie, Phylogénie et Paléobiologie, Institut des Sciences de l'Evolution, UMR5554-CNRS, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France

[3]American Hospital of Paris, 63, bd Victor Hugo - B.P.109 - 92202 Neuilly-sur-Seine Cedex, France

[4]Children's Hospital and Research Center at Oakland, Oakland, California 94609, USA.

[5]Department of Biology Chemistry; University of California, Irvine; 240D Medical Sciences I; Irvine, CA 92697-1700; USA

*Authors for correspondence:

Daniel.chourrout@sars.uib.no; Tel: +47-55584313

Ulrich.technau@sars.uib.no; Tel: +47-55584340

Bilaterian animals have a Hox gene cluster essential for patterning the main body axis, and a ParaHox gene cluster. Comparison of Hox/ParaHox genes has led to postulate that both clusters originated from the duplication of an ancient cluster named ProtoHox, which contained up to four genes with at least the precursors of anterior and posterior Hox/ParaHox genes (1-3). However, the way genes diversified within the ProtoHox, Hox and ParaHox clusters remains unclear as no systematic study of non-bilaterian animals exists. Our characterisation of the full Hox/ParaHox gene complements and genomic organisation in two cnidarian species (*Nematostella vectensis* and *Hydra magnipapillata*) suggests a ProtoHox cluster simpler than originally thought: (i) both species possess bilaterian-like anterior Hox genes, but their non-anterior genes do not appear as counterparts of either bilaterian central or posterior genes, (ii) two clustered ParaHox genes, *Gsx* and a gene related to *Xlox* and *Cdx*, are found in *Nematostella vectensis*, (iii) we do not find clear phylogenetic support for a common origin of bilaterian *Cdx* and posterior genes, which therefore may have appeared after the ProtoHox cluster duplication. Consequently, the ProtoHox cluster may have consisted of only two anterior genes. Non-anterior genes can have appeared independently in the Hox and ParaHox clusters, possibly after the separation of bilaterians and cnidarians.

Recent molecular phylogenies support that the non-bilaterian Cnidaria is the sister group to bilaterians (**4,5**) and are therefore informative to reconstruct the early history of bilaterian homeobox gene complements. Earlier searches for cnidarian homeobox genes have revealed the presence of anterior-like Hox genes as well as *Gsx*, so that the ProtoHox cluster must have been duplicated prior to the cnidarian-bilaterian split. Earlier reports (**6-13**) have proposed that posterior Hox genes but not central genes or *Hox3* are also present in cnidarians. We used publicly available high coverage genome shotgun sequence to identify the complete set of homeobox genes of two distantly related cnidarians, the freshwater polyp *Hydra magnipapillata* (Hydrozoa) and the sea anemone *Nematostella vectensis* (Anthozoa). *Nematostella* is particularly informative as it is considered to represent the basal group within the Cnidaria (**14,15**).

Eighteen *Nematostella* candidate homeodomains were allocated to the Antennapedia class / Hox subclass and named based upon relationship with vertebrate and fly sequences. Nine of them were clearly not Hox genes, and included the ParaHox gene *Gsx*, four distinct *Mox* genes, *Mnx*, *Gbx*, *Rough*, and *Evx*. Seven were considered as candidate Hox genes, including

five genes already known **(13)** which we renamed to avoid confusions with bilaterian Hox paralogs: *HoxA* (formerly *anthox6*), *HoxB* (novel), *HoxC* (formerly *anthox7*), *HoxDa* and *HoxDb* (two distinct copies with identical homeobox of a gene known as *anthox8*), *HoxE* (formerly *anthox1a*) and *HoxF* (formerly *anthox1*). Finally, two novel genes showed higher divergence from Antp/Hox genes of other species (see below). In contrast, a search in the *Hydra* genome revealed only eight Antp/Hox genes, including one *Mox* gene, one *Gsx* gene and six candidate Hox genes.

To clarify the relationships between bilaterian and cnidarian Hox/ParaHox genes, we carried out phylogenetic analyses (Methods in Supporting Online Material, SOM), including evaluation of.alternative hypotheses that receive some support in the data but would remain hindered with traditional tree building methods. Most reliable trees were obtained with large datasets of homeodomain proteins from putative slow-evolving bilaterian species such as *Nereis virens*, *Cupiennius salei*, *Ptychodera flava* and *Branchiostoma floridae*. First, using only bilaterian sequences, we could recognise the major groups of Hox and ParaHox genes (Fig. 1a and SOM) and confirm that ParaHox genes *Gsx* and *Xlox* cluster with the anterior genes *Hox1/2* and *Hox3*, respectively. Surprisingly, the group of ParaHox *Cdx* genes appeared divergent from all three groups of bilaterian Hox genes, and showed only slight phylogenetic affinity (BP = 38) with posterior Hox genes (Fig. 1a and SOM). The existence of a precursor for Hox posterior genes and *Cdx* in the ProtoHox cluster is therefore questionable.

Neighbour-net analyses clearly showed that within cnidarians *Nematostella* is the slowest evolving species, and the *Nematostella* dataset was consequently used for extensive sequence comparisons. *HoxA, HoxB, HoxC, HoxDa/Db* and one divergent gene *(Hox-related)* clustered with the anterior genes (*Hox1-3*) of bilaterians (Fig. 1b and SOM). The precise relationships within the anterior Hox homeodomains could not be statistically resolved despite weak preferential affinities of *HoxA/HoxB* with *Hox1/2*, and of *HoxC/HoxDa/HoxDb* with *Hox3* in several analyses (not shown). By comparison, *HoxE* and *HoxF* which have been previously tentatively classified as posterior genes (**13**) did not preferentially group with anterior, central or posterior genes (Fig. 1b and SOM). Instead, they formed an independent and strongly supported branch. Finally, one of the two divergent genes showed conflicting affinities in network analyses with *Cdx* and *Xlox* (Fig. 1b) and is therefore most likely a ParaHox gene. This gene named *Xlox/Cdx* is not a second Gsx gene, according to likelihood-based statistical tests (SOM). When Hox genes from other cnidarians were included, the same picture appeared with cnidarian-specific non-anterior Hox genes, although their long branches tended to be attracted towards the fast evolving posterior genes (SOM). Overall, cnidarians possess

(i) *bona fide* anterior Hox genes, (ii) several non-anterior Hox genes not clearly related to central and posterior genes of bilaterians, and (2) *Gsx* and a second candidate ParaHox gene related to *Xlox* and *Cdx*.

Using BAC library screening and sequence reconstruction from shotgun data, we established linkages between distinct homeobox genes in the *Nematostella* genome. The results (Fig. 2a and SOM) showed a main cluster of five genes *(HoxC/HoxDa/HoxDb/Evx/HoxA)* in a tandem array over approximately 50 kilobases. Linkage between the non-Hox gene *Evx* and a Hox gene has been shown earlier in another anthozoan cnidarian (**16**). We extended the cluster sequence and reached two other homeobox genes approximately 200 kilobases downstream: one is *Mnx*, a gene also present in the neighbourhood of chordate Hox clusters (**17**), and the other is the ortholog of the fly gene *Rough* which is on the chromosome arm 3R as the ANT-C and BX-C complexes. Another linkage could be established between the anterior gene *Hox-related* and the non-anterior gene *HoxE*, whereas *HoxB* and *HoxF* appeared to be isolated in two distinct genome locations. In sum, linkages in the *Nematostella* genome provide support for the existence of an ancient cnidarian Hox cluster associating *Evx*, several anterior Hox genes and at least one non-anterior Hox gene. In the *Nematostella* lineage, the cluster was expanded through gene duplications. Our analyses indeed robustly support that two precursor genes have been amplified into *HoxC/HoxDa/HoxDb* and into *HoxE/HoxF* subsets, respectively. A common origin for *HoxA* and *HoxB* is also suggested by the analyses, though with weaker support. The cluster was also affected by a rearrangement placing *Evx* between anterior Hox genes, and by at least two splits. Most interestingly, our linkage studies placed the candidate *Xlox/Cdx* gene immediately downstream of *Gsx*, in tandem orientation. Thus, this association can represent a ParaHox cluster. By comparison, genomic sequence reconstruction of the *Hydra* Hox gene complement provided evidence for strong degeneration of the Hox cluster in *Hydra*: only two non-anterior genes which most likely result from a recent duplication (*HmHoxc1* and *HmHoxc3*) could be linked (Fig. 2c), and *Gsx* was the only ParaHox gene detected.

As lineage-specific duplications in the Hox cluster are rare within bilaterians, we wanted to know whether or not those found in *Nematostella* are limited to the Hox genes. For this we examined potential linkages between all other homeobox genes. Including the nine Hox-like genes mentioned above, we found a total of 139 homeobox genes, a surprisingly high number compared to other invertebrates (Table 1 and SOM). Phylogenetic analyses allowed placing at least 87 of those into 58 known groups of homeobox genes, out of 76 groups known for

bilaterians. They also suggested that 42 unclassified genes have arisen through recent amplifications of maximally ten genes. By comparison, *Hydra* has a considerably smaller number of homeobox genes and gene groups. Since all *Hydra* homeobox gene groups have representatives in the *Nematostella* genome, we assume that the homeobox complement has been dramatically reduced in the *Hydra* lineage (Table 1). Comparisons between the extended homeobox sequences of *Nematostella* allowed the detection of an additional 13 physical clusters (Fig. 2b and SOM). Four of these clusters may have a more ancient origin, either because they have been identified in the genomes of bilaterians or because they associate distantly related genes. The other nine clusters are undoubtedly the result of recent tandem duplications. We also did not detect obvious synteny conservation between the environments of well related but unlinked homeobox genes, as might be expected after whole genome duplication (not shown). Hence, the gene duplications observed among the *Nematostella* Hox genes represent a general phenomenon for the homeobox gene complement.

From our data, the most parsimonious scenario for the evolution of Hox/ParaHox clusters is as follows (Fig. 3): two ProtoHox genes (*P1/2* and *P3*) gave rise to the Hox cluster consisting of two anterior Hox genes *Hox1/2* and *Hox3*, and to the ParaHox cluster with *Gsx* and *Xlox*. Subsequently, internal duplications expanded the Hox cluster, first by adding a precursor of the non-anterior genes. This gene may have appeared before the cnidarian-bilaterian split or independently in both lineages. Similarly, *Cdx* was added in the ParaHox cluster, either in the bilaterian lineage, or before the cnidarian-bilaterian split. After this split, the cnidarian and bilaterian Hox clusters were further expanded through lineage-specific duplications. In this scenario, cnidarians never had *bona fide* central and posterior Hox genes, and perhaps no *Cdx* gene either. Alternative scenarios can be proposed but they appear to be less parsimonious. For example, the precursor of non-anterior genes could have already existed in the ProtoHox cluster but had then to evolve into multiple directions leading to central/posterior Hox genes of bilaterians, *HoxE/F* of cnidarians, and *Cdx*. Or there were even more such precursors in the ProtoHox cluster and some had to be lost in one lineage and/or in one of the two clusters.

Another main angle to interpret the existing data is that cnidarians are in fact simplified bilaterians, with a reduced mesoderm as one of the possible regressions (**13, 18**). A loss of central and posterior genes, or their rapid divergence into current non-anterior Hox genes of cnidarians could have accompanied the process of simplification, as would have the Hox cluster breakdowns too. While this cannot be ruled out completely, extensive EST (**19**) and

genome surveys in *Nematostella* identified a surprising complexity of genes, including many genes lost in bilaterian lineages. This and the considerable degree of gene sequence conservation of the *Nematostella* lineage argues against a highly reduced and derived organism. In addition, more complex Hox gene complements in more basal metazoan phyla such as Placozoa, which would be indicative of gene loss in Cnidaria, could not be found (**20**). We therefore argue, that the common ancestor of cnidarians and bilaterians contained rather simple primordial Hox and ParaHox clusters that had very distinct fates in Cnidaria and Bilateria. The ProtoHox cluster itself may have consisted of only two anterior genes.

METHODS:

Homeodomains were searched into each cnidarian genome data using TBLASTN at low stringency (10^{-8}) with representative query sequences from mouse and the other cnidarian and classified into known classes, subclasses and groups. Alignments of shotgun datasets with expressed sequence tags had indicated that the probability of a given gene to be represented exceeds 98%. Phylogenetic analyses were based on Maximum likelihood, Bayesian and distance-based network analyses of homeodomain proteins. Phylogenetic network analyses allowed to visualise conflicting signals and areas of uncertainty in the data, which appear with homeodomains due to reduced alignment lengths. Genomic sequences were extended using the in-house developed software MàD (for Marche à Droite) which performs walks and jumps using shotgun sequences and their physical links. Linkages between homeobox genes were established through BAC library screening, confirmed using PCR on positive BAC clones, with MàD reconstruction of long sequence contigs. All technical details are provided in Supporting online Information.

Sequences have been deposited in Genbank with accession numbers DQ500742-DQ500879 with a synthetic presentation in Supporting Online Material.

Supporting Online Material accompanies the paper on www.nature.com/nature.

Acknowledgements:

We acknowledge Dan Rokhsar for his efforts in heading the *Nematostella* Genome project at DoE-JGI and the J. Craig Venter Institute and the National Human Genome Research Institute

for the *Hydra* Genome Project. The initial EST screen in *Nematostella* was funded by the DFG (Te311-3/1). The construction of the *Nematostella* BAC library was funded by the National Science Foundation (NSF IBN-0208335).

Authors' contribution

The project was conceived and the manuscript was written by D. Chourrout and U. Technau. The bioinformatic identification and genomic assemblies of the homeobox gene complement in the *Hydra* and *Nematostella* genome datasets was done by D. Chourrout. The software package including MàD was written by P. Chourrout, who conceived and trained it with D. Chourrout. The phylogenetic analyses were carried out by F. Delsuc. R.B.Edvardsen also participated in the homeobox gene classification and phylogenetic analysis. The cloning of cDNAs and genomic clones of Hox-related genes was done by U. Technau, E. Renfer and M.F. Jensen, BAC library screening was done by U. Technau and E. Renfer and BAC colony PCR were performed by M.F. Jensen. The in situ hybridisations were carried out by Fabian Rentzsch. R.E. Steele and U. Technau are members of the *Nematostella* Genome consortium, R.E. Steele is a member of the *Hydra* genome consortium. R.E. Steele, Pieter de Jong and U. Technau were Co-PIs on the NSF grant for the generation of the *Nematostella* BAC library, U. Technau collected the animal material and prepared the DNA and B. Zhu generated the BAC library under supervision of P. de Jong.

Competing interests' statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Daniel.chourrout@sars.uib.no and Ulrich.technau@sars.uib.no

REFERENCES

(1) Ferrier, D.E.K. & P.W. Holland, P.W. Ancient origin of the Hox gene cluster. Nature Reviews Genetics 2, 33-38 (2001).

(2) Garcia-Fernandez, J. The genesis and evolution of homeobox gene clusters. Nature Reviews Genetics 6, 881-892 (2005).

(3) Brooke, N.M., Garcia-Fernandez, J. & Holland, P. The Parahox gene cluster is an evolutionary sister of the Hox gene cluster. Nature 392, 920-922 (1998).

(4) Collins, A.G. Evaluating multiple alternative hypotheses for the origin of Bilateria: an analysis of 18S molecular evidence. PNAS 95, 15468-15463 (1998).

(5) Philippe, H., Lartillot, N. & Brinkmann, H. Multigene analysis of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa and Protostomia. Mol. Biol. Evol. 22, 1246-1253 (2005).

(6) Shenk, M.A., Bode, H.R. & Steele, R.E. Expression of Cnox-2, a HOM/HOX homeobox gene in hydra, is correlated with axial pattern formation. Development 117, 657-667 (1993).

(7) Gauchat, D. et al. Evolution of Antp-class genes and differential expression of Hydra Hox/paraHox genes in anterior patterning. Proc. Natl. Acad. Sci. U S A. 97, 4493-4498 (2000).

(8) Finnerty, J.R. & Martindale, M.Q. Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria. Evol. Dev. 16-23 (1999).

(9) Finnerty et al. Early evolution of a homeobox gene: the parahox gene *Gsx* in the Cnidaria and the Bilateria. Evol Dev. 5(4):331-45 (2003).

(10) Kuhn, K., Streit, B. & Schierwater B. Homeobox genes in the cnidarian *Eleutheria dichotoma*: evolutionary implications for the origin of Antennapedia-class (HOM/Hox) genes. Mol. Phylogenet. Evol. 6, 30-38 (1996).

(11) Kuhn, K., Streit, B. & Schierwater B. Isolation of Hox genes from the scyphozoan *Cassiopeia xamachana*: implications for the early evolution of Hox genes. J. Exp. Zool. 285, 63-75 (1999).

(12) Martinez, D.A. et al. Cnidarian homeoboxes and the zootype. Nature 393, 748-749 (1998).

(13) Finnerty, J.R. et al. Origins of bilateral symmetry: *Hox* and *Dpp* expression in a sea anemone. Science 304, 1335-1337 (2004).

(14) Collins, A.G. Phylogeny of the Medusozoa and the evolution of the cnidarian life cycles. J. Evol. Biol. 15, 418-432 (2002).

(15) Darling, J.A. et al. Rising starlet: the starlet sea anemone, *Nematostella vectensis*. BioEssays 27, 211-221 (2005).

(14) McGinnis, W. & Krumlauf, R. Homeobox genes and axial patterning. Cell 68, 283-302 (1992).

(15) Kmita, M. & Duboule, D. Organizing axes in time and space: 25 years of collinear tinkering. Science 301, 331-333 (2003).

(16) Miller, D.J. & Miles A. Homeobox genes and the zootype. Nature 365, 215-216 (1993).

(17) Castro, L.F.C. & Holland, P.W.H. Chromosomal mapping of ANTP class homeobox genes in Amphioxus: piecing together ancestral genomes. Evol Dev. 5, 459-465 (2003).

(18) Seipel, K. & Schmid, V. Evolution of striated muscle: jellyfish and the origin of triploblasty. Dev. Biol. 282, 14-26 (2005).

(19) Technau, U. et al. Maintenance of ancestral complexity and non-metazoan genes in two basal cnidarians. Trends Genet. 21, 633-639 (2005).

(20) Monteiro, A.S., Schierwater, B., Dellaporta, S.L. & Holland, P.W.H. A low diversity of ANTP class homeobox genes in Placozoa. Evol. Dev. 8, 174-182 (2006).

FIGURE AND TABLE LEGENDS:

Figure 1 | Phylogenetic analyses of bilaterian and *Nematostella* Hox genes based on their homeodomain. More detailed figures including all sequence origins are available in SOM.

(a) Neighbour-Net network of 82 bilaterian Hox and ParaHox genes based on ML distances estimated using a JTT substitution matrix plus a eight-category gamma rate heterogeneity correction (JTT+G_8). Note the large differences in evolutionary rates between the fast evolving Posterior Hox and Cdx genes, and the slow evolving Central Hox genes. Bayesian posterior probabilities (PP) for the major groups have been mapped on the network. A dash (-) indicates that the corresponding split was not observed in the Bayesian analysis. Bootstrap proportions (BP) were inferred using the BioNJ algorithm for the same dataset and JTT+G_8 ML distance estimates. Scale bar represents estimated number of changes per site. Note the position of the Cdx group which is distinct from the posterior Hox group.

(b) Neighbour-Net network of 92 Hox and ParaHox genes from bilaterian and *Nematostella* based on JTT+G_8 ML distance estimates. Note the position of non-anterior *Nematostella* genes (*Nv_HoxE* and *Nv_HoxF*). The major groups of bilaterian Hox and ParaHox proteins remain unaffected by the addition of the *Nematostella* dataset. Bayesian

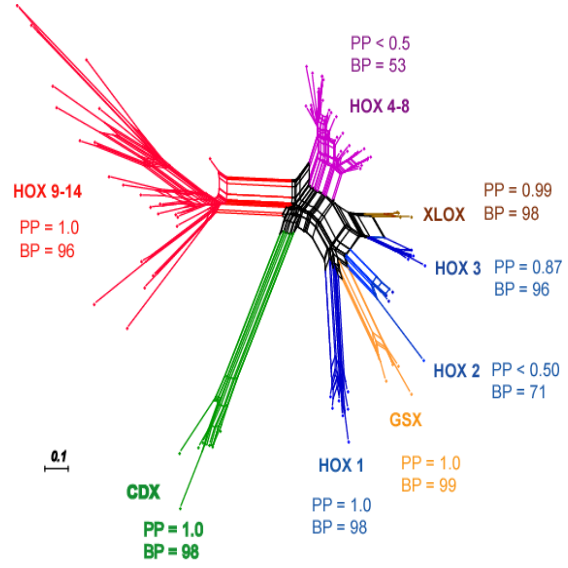
Posterior Probabilities (PP) and Bootstrap proportions (BP) were calculated as for the bilaterian dataset.

Figure 2 | All physical linkages detected between cnidarian homeobox genes. (a). Linkages associating *Nematostella* ANTP/Hox-like genes were validated using BAC library screening with individual Hox-like genes, before an *in silico* genome walk using the local assembler of shotgun traces allowed to reconstruct the cluster DNA sequences. The position of Hox and Hox-related genes including the ParaHox gene *Gsx* and their respective orientations are represented by red arrows, and other linked homeobox genes by blue arrows. Putative intervening non-homeobox genes are represented by circles, filled with green when expression is confirmed by EST alignments; sequence similarity with known genes is detected by BLASTX (SOM). **(b). Linkages associating other *Nematostella* homeobox genes** were validated with in silico genome walk using the local assembler of shotgun traces. **(a). Linkages associating *Hydra* ANTP/Hox-like genes including *Gsx*** were validated with *in silico* genome walk. Genes found in their environments have been annotated using BLASTX and alignments with *Hydra* ESTs (SOM).

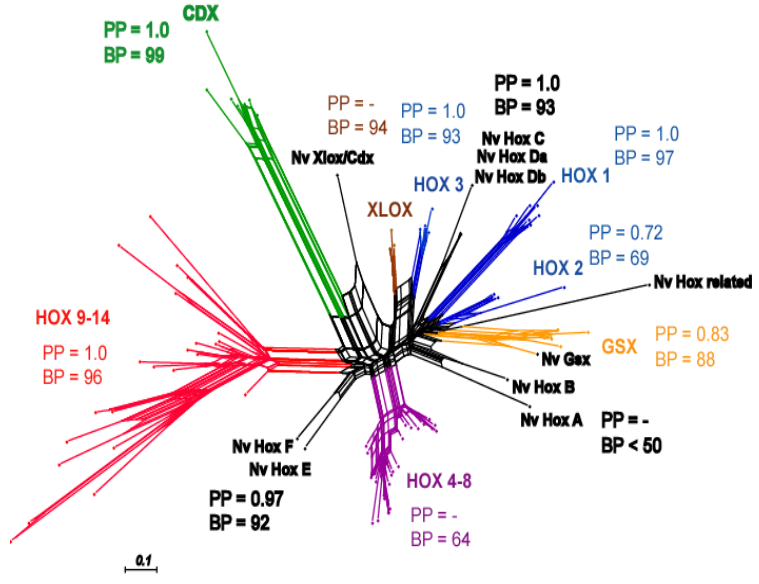
Figure 3 | Parsimonious consensus scenario for the evolution of Hox/ParaHox clusters of bilaterians. The ProtoHox cluster may have contained only two anterior genes (P1/2 and P3) and its duplication generated two equally simple Hox and ParaHox clusters. The prototypal bilaterian Hox cluster may have contained an extra non-anterior Hox gene (BNA), precursor of future central and posterior genes. An independent duplication in the Hox cluster would have generated the precursor of cnidarian non-anterior Hox genes (CNA), which later on became HoxE and HoxF of *Nematostella*. The third ParaHox gene *Cdx* can have appeared through a duplication in the ParaHox cluster, and not in the ProtoHox cluster as generally proposed, either in early bilaterians or before the split of bilaterians and cnidarians. In the latter case, cnidarians had already simplified their ParaHox cluster through the loss of *Xlox*.

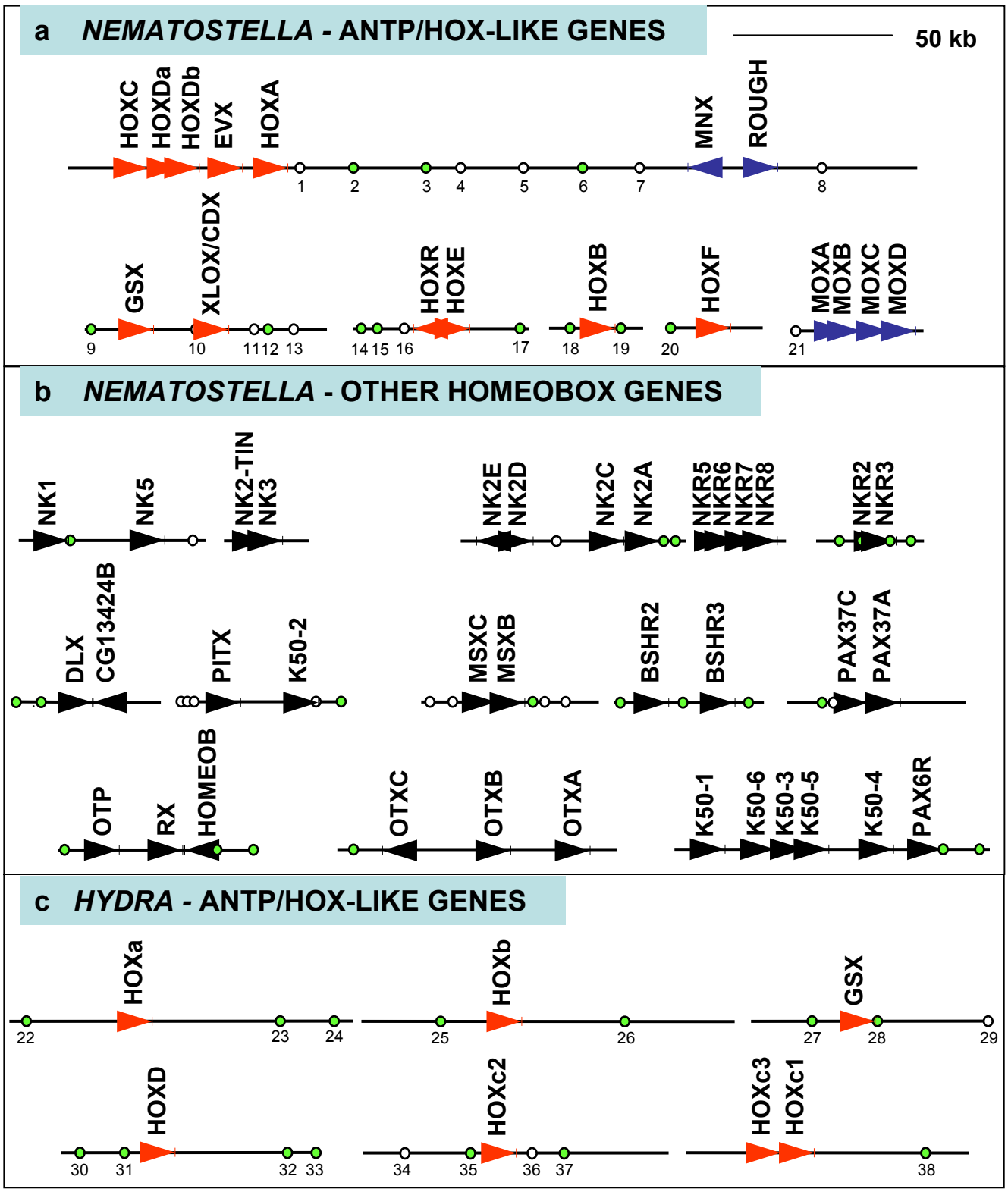
Table 1 | Classification of homeobox genes detected in the genomes of *Nematostella vectensis* and *Hydra magnipapillata*, based on phylogenetic analyses. In *Nematostella*, a significant subset of genes could be allocated to known classes and superfamilies but diverged from known homeobox gene groups, while three genes could not be allocated to known classes (see also SOM). The *Hydra* genome shows a substantial reduction of the homeobox gene complement, and all identified gene groups are also present in *Nematostella.*

a


PP < 0.5
BP = 53
HOX 4-8
HOX 9-14
PP = 1.0
BP = 96
PP = 0.99
XLOX
BP = 98
HOX 3
PP = 0.87
BP = 96
HOX 2
PP < 0.50
BP = 71
GSX
0.1
HOX 1
PP = 1.0
BP = 99
CDX
PP = 1.0
BP = 98
PP = 1.0
BP = 98

b


CDX
PP = 1.0
BP = 99
PP = -
BP = 94
PP = 1.0
BP = 93
PP = 1.0
BP = 93
Nv Hox C
Nv Hox Da
Nv Hox Db
PP = 1.0
BP = 97
Nv Xlox/Cdx
HOX 3
HOX 1
XLOX
PP = 0.72
BP = 69
HOX 2
Nv Hox related
HOX 9-14
PP = 1.0
BP = 96
PP = 0.83
BP = 88
GSX
Nv Gsx
Nv Hox B
Nv Hox A
PP = -
BP < 50
Nv Hox F
Nv Hox E
PP = 0.97
BP = 92
HOX 4-8
PP = -
BP = 64
0.1


a NEMATOSTELLA - ANTP/HOX-LIKE GENES
50 kb
HOXC
HOXDa
HOXDb
EVX
HOXA
MNX
ROUGH
1
2
3
4
5
6
7
8
GSX
XLOX/CDX
HOXR
HOXE
HOXB
HOXF
MOXA
MOXB
MOXC
MOXD
9
10
11
12
13
14
15
16
17
18
19
20
21
b NEMATOSTELLA - OTHER HOMEOBOX GENES
NK1
NK5
NK2-TIN
NK3
NK2E
NK2D
NK2C
NK2A
NKR5
NKR6
NKR7
NKR8
NKR2
NKR3
DLX
CG13424B
PITX
K50-2
MSXC
MSXB
BSHR2
BSHR3
PAX37C
PAX37A
OTP
RX
HOMEOB
OTXC
OTXB
OTXA
K50-1
K50-6
K50-3
K50-5
K50-4
PAX6R
c HYDRA - ANTP/HOX-LIKE GENES
HOXa
HOXb
GSX
22
23
24
25
26
27
28
29
HOXD
HOXc2
HOXc3
HOXc1
30
31
32
33
34
35
36
37
38

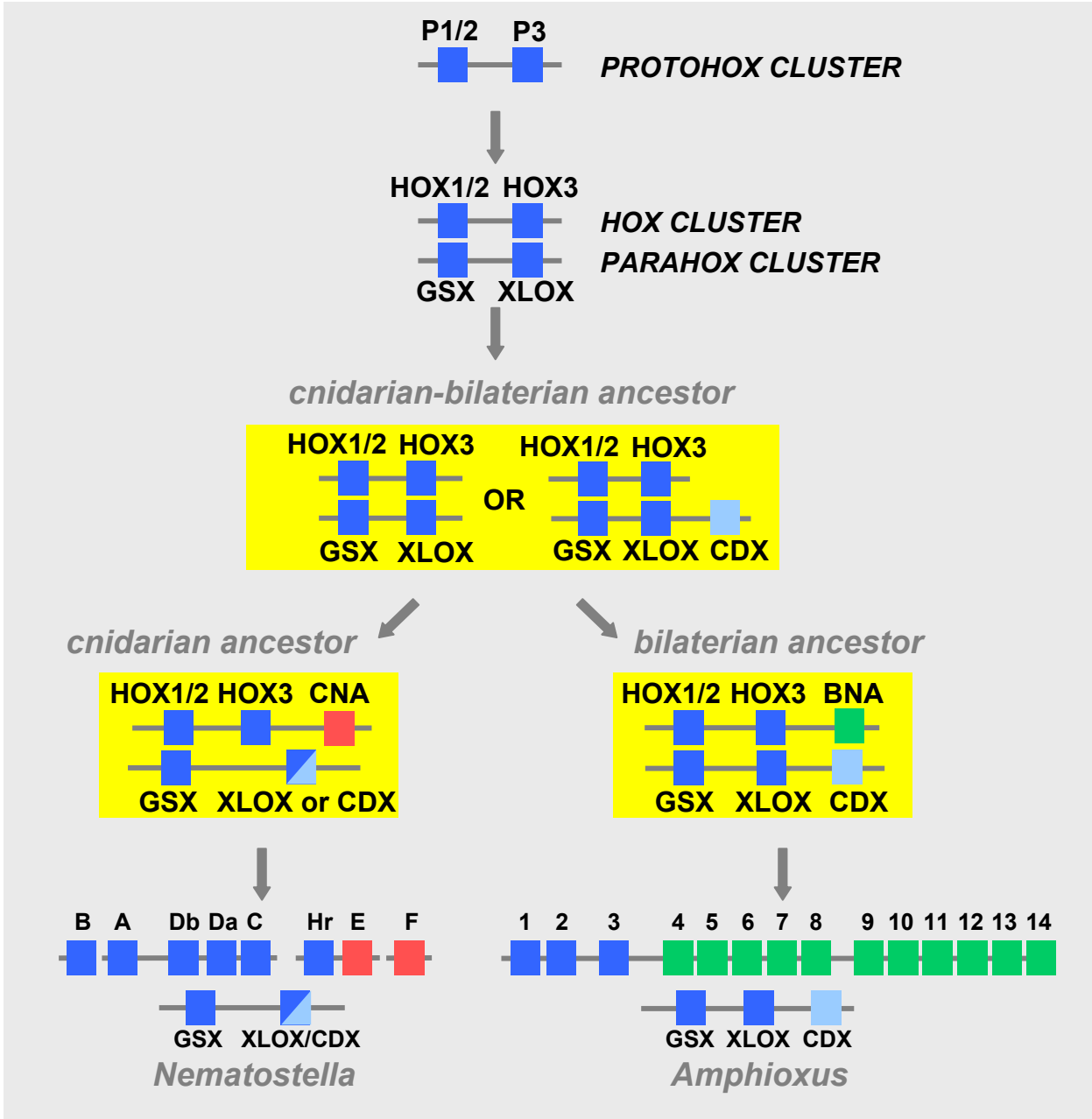

P1/2 P3
PROTOHOX CLUSTER
HOX1/2 HOX3
GSX XLOX
HOX CLUSTER
PARAHOX CLUSTER
cnidarian-bilaterian ancestor
HOX1/2 HOX3
GSX XLOX
OR
HOX1/2 HOX3
GSX XLOX CDX
cnidarian ancestor
HOX1/2 HOX3 CNA
GSX XLOX or CDX
bilaterian ancestor
HOX1/2 HOX3 BNA
GSX XLOX CDX
B A Db Da C Hr E F
GSX XLOX/CDX
Nematostella
1 2 3 4 5 6 7 8 9 10 11 12 13 14
GSX XLOX CDX
Amphioxus

TABLE1:

CLASS/SUBCLASS	SUPERFAMILY	# GENE GROUPS (WITHOUT HOX/PARAHOX)			# GENES (WITH HOX/PARAHOX)	
		known	*Nematostella*	*Hydra*	*Nematostella*	*Hydra*
ANTP/HOX-LIKE	HOX	-	-	-	8	6
	PARAHOX	-	-	-	2	1
	EXT-HOX	2	2	1	5	1
	EHG-BOX	3	3		3	
	UNPLACED					
ANTP/NK-LIKE	*PLACED*	20	18	8	35	11
	UNPLACED				25	
PRD/PAIRED-LIKE Q50	*PLACED*	15	9	3	9	5
	UNPLACED				8	3
PRD/PAIRED-LIKE K50	*PLACED*	4	3	3	5	5
	UNPLACED				6	1
PRD/PAX S50		4	3	1-2	5	3
POU		6	4	2	5	2
LIM		6	6	3-4	6	4
TALE		5	5	3	7	4-5
SIX		3	3	2	5	2
PROX		1				
CUT		3	1		1	
ZFH		2				
HNF1		2	1		1	
UNPLACED					3	4
TOTAL		76	58	26-28	139	52-53