Zebrafish					
Viewpoints	**Left**	**Right**	**Anterior**	**Cluster**	**Posterior**
nfe2l2a	8.8%	91.2%	71.6%	27.1%	1.4%
hnrpa3	45.9%	54.1%	63.9%	33.4%	2.7%
hoxd4a	83.3%	16.7%			
hoxd10a	70.7%	29.3%			
hoxd11a	67.4%	32.6%			
hoxd13a	23.4%	76.6%			
evx2	26.8%	73.2%			
lnpa	20.7%	79.3%			
atp5g3a	62.4%	37.6%	0.0%	14.8%	85.2%

Amphioxus					
Viewpoints	**Left**	**Right**	**Anterior**	**Cluster**	**Posterior**
Meox	44.2%	55.8%	100.0%	0.0%	0.0%
Hnrpa	38.5%	61.5%	71.7%	25.6%	2.7%
Mtx2	30.9%	69.1%	41.2%	42.7%	16.1%
Hox2	24.8%	75.2%	24.8%	75.2%	0.0%
Hox5	38.5%	61.5%	18.8%	80.2%	1.0%
Hox6	44.1%	55.9%	24.7%	71.3%	4.0%
Hox7	51.6%	48.4%	21.6%	74.1%	4.3%
Hox9	46.9%	53.1%	14.1%	83.2%	2.6%
Hox11	62.1%	37.9%	0.0%	99.5%	0.5%
Hox13	65.7%	34.3%	1.7%	88.7%	9.5%
Hox15	74.2%	25.8%	5.3%	68.9%	25.8%
Evxa	66.1%	33.9%	11.6%	45.6%	42.8%
Lnp	73.0%	27.0%	6.4%	35.5%	58.1%
Gpatch8	50.6%	49.4%	12.0%	3.5%	84.5%

Supplementary Table 1
Contact distribution for each of the amphioxus viewpoints. The slashed cell correspond to those viewpoints where the exclusion of reads around the viewpoint overlap with more than 30% of the total length of the zebrafish HoxDa cluster.

Amphioxus		Zebrafish	
Eliminated Viewpoints	Correlation	Eliminated Viewpoints	Correlation
Gpatch88	0,93	*atp5g3a*	0,84
Lnp	0,96	*lnpa*	0,90
Evxa	0,95	*evx2*	0,81
Hox15	0,92	*hoxd13a*	0,89
Hox13	0,93	*hoxd11a*	0,90
Hox11	0,95	*hoxd10a*	0,92
Hox9	0,94	*hoxd4a*	0,93
Hox7	0,94	*hoxd3*	0,91
Hox6	0,95	*nfe2l2a*	0,89
Hox5	0,96	*hnrnpa3*	0,91
Hox2	0,95	*evx2, hoxd10a*	0,86
Mtx2	0,93	*lnpa, hoxd3*	0,92
Hnrnpa	0,94	*atp5g3a, nfe2l2a*	0,86
Meox	0,94	*hoxd13a, hnrnpa3*	0,88
Evxa, Hox13	0,94	*hoxd11a, hoxd4a*	0,91
Hox5, Meox	0,94	*evx2, hoxd11a, nfe2l2a, hnrnpa3*	0,85
Gpatch8, Hox9	0,88	*atp5g3a, hoxd13a, hoxd11a, hoxd10a*	0,86
Lnp, Hox2	0,95	*lnpa, hoxd13a, hoxd10a, hoxd3*	0,91
Hox11, Mtx2	0,93	*evx2, hoxd13a, hoxd3, nfe2l2a*	0,82
Hox15, Hox13, Hox7, Meox	0,82	*atp5g3a, hoxd11a, hoxd4a, hoxd3*	0,87
Gpatch8, Hox,11, Hox2, Mtx2	0,87	*atp5g3a, evx2, hoxd11a, hoxd10a, hoxd4a, nfe2l2a*	0,84
Lnp, Evxa, Hox9, Hox6	0,93	*lnpa, hoxd13a, hoxd4a, hoxd3, nfe2l2a, hnrnpa3*	0,77
Evxa, Hox11, Hox6, Hnrnpa	0,91	*lnpa, hoxd13a, hoxd10a, hoxd4a, hoxd3, hnrnpa3*	0,83
Hox15, Hox11, Hox7, Mtx2	0,89	*atp5g3a, evx2, hoxd13a, hoxd11a, hoxd3, nfe2l2a*	0,80
Evxa, Hox13, Hox6, Hox5, Mtx2, Meox	0,92	*evx2, hoxd13a, hoxd11a, hoxd4a, hoxd3, nfe2l2a*	0,82
Gpatch8, Hox11, Hox9, Hox6, Hox5, Hox2	0,90		
Lnp, Hox15, Hox11, Hox7, Hox2, Mtx2	0,86	KEY	
Evxa, Hox15, Hox9, Hox7, Mtx2, Hnrnpa	0,81	Number of viewpoints eliminated	
Lnp,Hox13,Hox9,Hox6,Hox2,Hnrnpa	0,88	1	
Gpatch8, Evxa, Hox13, Hox7, Hox6, Hox2, Mtx2, Hnrnpa	0,82	2	
Evxa, Hox15, Hox11, Hox9, Hox5, Hox2, Hnrnpa, Meox	0,89	4	
Lnp, Evxa, Hox11, Hox7, Hox6, Hox5, Mtx2, Hnrnpa	0,85	6	
Lnp, Hox15, Hox13, Hox11, Hox9, Hox7, Hox2, Mtx2	0,85	8	
Evxa, Hox15, Hox13, Hox9, Hox6, Hox5, Hox, 2Hnrnpa	0,82	10	
Lnp, Evxa, Hox13, Hox11, Hox9, Hox6, Hox5, Hox2, Mtx2, Meox	0,89		
Gpatch8, Evxa, Hox15, Hox13, Hox9, Hox7, Hox6, Hox5, Mtx2, Hnrnpa	0,83		
Lnp, Hox15, Hox13, Hox9, Hox7, Hox6, Hox5, Hox2, Mtx2, Hnrnpa	0,68		
Gpatch8, Lnp, Evxa, Hox13, Hox11, Hox9, Hox6, Hox5, Mtx2, Meox	0,88		
Gpatch8, Evxa, Hox15, Hox13, Hox11, Hox7, Hox6, Hox2, Mtx2, Hnrnpa	0,54		

Supplementary Table 2
Spearman's correlations of the virtual Hi-C jackknife resampling experiments

Dataset 1	Dataset 2	Pearson's correlation	Spearman's correlation
ESCells-HindIII-rep1	ESCells-HindIII-rep2	0,98	0,97
ESCells-HindIII-rep1	ESCells-NcoI	0,94	0,95
ESCells-HindIII-rep1	Cortex-HindIII-rep1	0,8	0,76
ESCells-HindIII-rep1	Cortex-HindIII-rep2	0,79	0,8
ESCells-HindIII-rep2	ESCells-NcoI	0,93	0,94
ESCells-HindIII-rep2	Cortex-HindIII-rep1	0,79	0,79
ESCells-HindIII-rep2	Cortex-HindIII-rep2	0,79	0,8
ESCells-NcoI	Cortex-HindIII-rep1	0,7	0,76
ESCells-NcoI	Cortex-HindIII-rep2	0,71	0,78
Cortex-HindIII-rep1	Cortex-HindIII-rep2	0,86	0,83

Supplementary Table 3
Correlation coefficients for the mouse HoxD locus between Hi-C datasets corresponding to different experimental replicates and tissues.

	Viewpoint	Read 5'->3'	Non-read 5'->3'
European amphioxus 4C	Hox2	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTGGCCCCCTGTCCCCCAGATC	CAAGCAGAAGACGGCATACGACGCCATAGCGTCCCT AGCCG
	Hox5	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTCGCCAACACTCTGCTATACATGATC	CAAGCAGAAGACGGCATACGAGCGCACGAGATAGTG GGTAC
	Hox6	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTATCTCACGTCAGAAAGCTGATC	CAAGCAGAAGACGGCATACGACATCTCGTTGAAATT GCGGAC
	Hox7	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTGTCGGGATGATACATGGTGACGATC	CAAGCAGAAGACGGCATACGAAGGAGACCCTTGCCA CCACC
	Hox9	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTCTTCATCTCAGGAAAACACGTGATC	CAAGCAGAAGACGGCATACGAATTACCCTACATGCC ACGGG
	Hox11	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTAGGCCTTTTTTTCAACACAAGGATC	CAAGCAGAAGACGGCATACGAGCGGAAAGAAAATCT ACGCG
	Hox13	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTCAAAAAGCGAACGGCCAGATC	CAAGCAGAAGACGGCATACGAGGATATCACTCCCAG TTCCTGG
	Hox15	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTCCGTCTCGGCCCCGGCGATC	CAAGCAGAAGACGGCATACGAATGGCGCAGGAGTGC CCCGC
	Evx2	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTGCGGTGCACAAGCAAAGATC	CAAGCAGAAGACGGCATACGATTACTGCCTTCACCC AGCAC
	Mtx2	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTTACTTGCCGACGCAGCGATC	CAAGCAGAAGACGGCATACGAGGTCAGGGGTTACTG GACGAG
	Lnpa	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTGTGTTGAACTGTATATTCCATGATC	CAAGCAGAAGACGGCATACGATGGACTTACTTTGAA GCGTG
	Gpatch8	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTTCCTGTGTATGAAAGTCCTGATC	CAAGCAGAAGACGGCATACGACTTTAACAGAACCTC GGCAC
	Meox	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTAACGCCTTGTTGCAAGTGATC	CAAGCAGAAGACGGCATACGAATGCTACGAGCTTTA CTAAACAGGAC
	Hnrnpa	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTTTTCCCATCCTTTGTAGGACGATC	CAAGCAGAAGACGGCATACGAGGGATACTCCGAGTA AGTTCTGC
	Nfe2l	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTTCCATCCCCTGTTGGGGATC	CAAGCAGAAGACGGCATACGATTGTGCAGTTCCTGA AGAATCTC
Zebrafish 4C	**Viewpoint**	**Read 5'->3' '**	**Non-read 5'->3'**
	Hnrnpa3	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTATTAGGCATGGTTGTTATACTGATC	CAAGCAGAAGACGGCATACGAGAAAGTGACATCACA TCTGCT
	Nfe2l2a	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTAAGCAAATGGATGACTGAGGATC	CAAGCAGAAGACGGCATACGAACTTGACTTTTTGCA ACCCAC
	Atp5g3a	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTGCACAATAACAGTGTTTATTTGATC	CAAGCAGAAGACGGCATACGATTTTCATTACAAGAC GCCGC
	Evx2	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTGGTTCCCAAATAAGCTGTCGATC	CAAGCAGAAGACGGCATACGACCTAATGTCTGAGTG CTAGCATG
	Lnpa	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTCCTTCACTGCTCTGTCACGTGATC	CAAGCAGAAGACGGCATACGACCGCTGTATCAGCTG ACAGTTGC
	HoxD4a	AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACGCT CTTCCGATCTTTGTTCATCATCATATGGGAGGATC	CAAGCAGAAGACGGCATACGACAGTTGACAG CAAGTAGGAGG
Enhancer cloning	**Enhancer**	**Fwd 5'->3'**	**Rev 5'->3'**
	1655	CACTTCCAGCCCACTTCCTC	CGGGTGCCAGGTTTATTCGA
	1739	GCTGAGATTTCCAAACAACCACA	GGGACACGGAGGTTGATAAGT
	1784	CCACCCGGAAATCTTTGTCC	TATGCGCTCTGAGATGACGG
	1801	TTCCGCATGCCTTACACACA	CCCGCGATATAAAGCCCAGT
	2473	AGTCACCCGTTAGATTCCCT	ACCATACGCTGCTTATCCATGA

Supplementary Table 4
Primers used in the 4C-seq and enhancer cloning experiments.

Supplementary Note

Vertebrate Hox gene deserts were ancestrally populated by a conserved array of neighboring genes.

In most jawed vertebrate species, the genomic arrangements of HoxA and HoxD neighborhoods show little similarity: many HoxA long-range cis-regulatory elements (CREs) are embedded in introns of neighboring genes, while HoxD long-range CREs are located in gene deserts [2,8]. Therefore, the long-range regulatory interactions of HoxA and HoxD could have evolved independently [9,10]. Here, we show that these differences are a derived situation and that the genomic organisations of the four Hox clusters neighboring regions were ancestrally very similar. First, out of the 15 gene families consistently found in the vicinity of the four clusters, 6 of them (*Nfe2*, *Hnrnpa*, *Cbx1-3-5*, *Calcoco*, *Atp5g* and *Creb*) have 3 to 4 members in the Hox regions of at least one gnathostome species, confirming that their close Hox linkages predate the WGDs (Fig. 1b). Second, 'desertification' process (erosion of flanking genes' exons [13] while non-coding introns containing CREs are maintained) has probably been rampant thanks to the functionally redundant genes copies generated by the two vertebrate WGD duplications. The genomes of slow evolving species such as elephant shark, still retain additional paralogs of Hox-neighboring genes that have been secondarily lost in mammals and teleosts (such as *Eve1* linked to HoxB, *Lnp2* linked to HoxC, and *Jazf2* linked to HoxD); in turn, the HoxD-linked *Atp5g3* is present in mammals but has been lost in elephant shark (Fig. 1b, [54]). The presence of Hox distal CREs within the introns of some of these lost paralogs further demonstrate that the coding regions of Hox neighboring genes were selectively lost, generating gene deserts in the process. In elephant shark, two of the four conserved distal regulatory islands (I and II) of mouse HoxD posterior desert are located inside the introns of the additional *Jazf2* copy, in an arrangement almost identical to that typically found in HoxA clusters (Supplementary Figure 2). In fact, a pseudogenic relict of a *Jafz2* exon is still detectable in mouse (Supplementary Figure 2). Furthermore, part of the HoxD global control region, the CsB element, has a paralogous copy in HoxA posterior

side inside the intron 4 of the gene *Hibadh*, implying that a paralog of this gene was also present in the HoxD posterior neighborhood (Supplementary Figure 1, [12]). Thus, HoxD 'gene desert' was originally occupied by a *Hibadh2* and a *Jazf2* genes, showing that HoxA and D genomic organisations are in fact almost identical. In sum, these results show that the syntenic organisation of the four clusters was originally very similar and the close linkage of most Hox neighboring genes predate the WGDs.

Reconstruction of the syntenic arrangement around the Hox cluster of the pre-WGD vertebrate ancestor.

Comparisons between the four paralogous clusters in different vertebrates species allow us to infer that several Hox-neighboring genes were already present around the single cluster of the pre-WGD vertebrate ancestor. We recovered 6 genes that have at least 3 paralogous copies (*Nfe2*, *Hnrnpa*, *Cbx1-3-5*, *Calcoco*, *Atp5g* and *Creb*). Five other genes only have one or two copies in vertebrates (*Copz*, *Skap*, *Smug1*, *Mtx2* and *Lnp*) but their locations and orientations in amphioxus are almost identical to those of vertebrates, and therefore their syntenic arrangement was already present in the chordate ancestor (Fig. 1b). In the case of three other jawed vertebrate genes with less than 3 copies, *Hibadh*, *Jafz* and *Ube2z*, a conserved tight linkage with their closest vertebrate neighbor is also present in at least one additional species outside vertebrates, indicating that they belong to ancestral microsyntenic pairs that predate the origin of vertebrates (Supplementary Figure 3). Finally, for the two remaining jawed vertebrate neighboring genes, *Snx10-11* and *Ttll6*, we have not found any data that could help to infer their status in the pre-WGD vertebrate ancestor.

Most chordate Hox neighboring genes belong to the same ancestral Hox macrosyntenic linkage group.

The majority of chordate Hox neighboring genes, including the conserved anterior genes and the vertebrate and amphioxus specific posterior ones (*Agps*,

Copz, *Nfe2*, *Skap*, *Smug1*, *Mtx2*, *Evx*, *Lnp*, *Hibadh*, *Calcoco*, *Jafz*, *Atp5g*, *Ube2z*, *Creb*, *Nifk*, *Mdh1b*, *Dlgap1-4*, *Ubp1-Tfcp2*, *Asb1* and *Gpatch8*) are tightly linked to at least another chordate neighboring gene in at least one non-chordate species (Supplementary Figure 3). This indicates that they all belong to the same ancestral large macrosyntenic linkage group that also contained the Hox cluster (or the putative Hox 'ghost' locus, in the case of placozoans [55]) as well as additional related ANTP homeobox genes (such as *Evx*, *Mnx*, *Dlx* and *Meox*), and some other genes showing conserved Hox macrosyntenic associations in different non vertebrate species (*Elac1*, *Mcm6*, *Tmem169*, *Nhej1*, *Smarcc*, *Myd98*, Supplementary Figure 3). In general, these inferred ancestral linkages between different neighboring genes and the Hox cluster and between neighboring genes with other neighboring genes show an extremely patchy phylogenetic distribution: most of these syntenic associations have been heavily reshuffled in different animal lineages and gene order and orientations are highly variable. Thus, although all these genes were probably located in the same Hox chromosome in the bilaterian ancestor, in most phyla their linkage has only been maintained at the macrosyntenic level.

Supplementary References

54. Ravi, V. *et al.* Elephant shark (Callorhinchus milii) provides insights into the evolution of Hox gene clusters in gnathostomes. *Proc Natl Acad Sci U S A* **106**, 16327-32 (2009).

55. Mendivil Ramos, O., Barker, D. & Ferrier, D.E. Ghost loci imply Hox and ParaHox existence in the last common ancestor of animals. *Curr Biol* **22**, 1951-6 (2012).