Developmental Biology 328 (2009) 173–187

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Perspective

More constraint on *ParaHox* than *Hox* gene families in early metazoan evolution

Manon Quiquand [a], Nathalie Yanze [b], Jürgen Schmich [c], Volker Schmid [b], Brigitte Galliot [a,*,1], Stefano Piraino [c,1]

[a] Department of Zoology and Animal Biology, University of Geneva, 30 quai Ernest-Ansermet, CH-1211 Geneve 4, Switzerland
[b] Institute of Zoology, Biocenter/Pharmacenter, Basel University, Switzerland
[c] Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, Università del Salento, 73100 Lecce, Italy

ARTICLE INFO

Article history:
Received for publication 19 June 2008
Revised 14 January 2009
Accepted 14 January 2009
Available online 27 January 2009

Keywords:
Hox/*ParaHox* genes
ProtoHox gene
Pdx/*Xlox* ortholog
Cnidarians
Hydrozoans
Phylogenetic scoring
Metagrouping of gene families
Clytia
Turritopsis
Cladonema

ABSTRACT

Hox and *ParaHox* (*H*/*P*) genes belong to evolutionary-sister clusters that arose through duplication of a ProtoHOX cluster early in animal evolution. In contrast to bilaterians, cnidarians express, beside *PG1*, *PG2* and *Gsx* orthologs, numerous *Hox*-related genes with unclear origin. We characterized from marine hydrozoans three novel *Hox*-related genes expressed at medusa and polyp stages, which include a *Pdx*/*Xlox ParaHox* ortholog induced 1 day later than *Gsx* during embryonic development. To reconstruct *H*/*P* genes' early evolution, we performed multiple systematic comparative phylogenetic analyses, which identified derived sequences that blur the phylogenetic picture, recorded dramatically different evolutionary rates between *ParaHox* and *Hox* in cnidarians and showed the unexpected grouping of [*Gsx*–*Pdx*/*Xlox*–*PG2*–*PG3*] families in a single metagroup distinct from *PG1*. We propose a novel more parsimonious evolutionary scenario whereby *H*/*P* genes originated from a [*Gsx*–*Pdx*/*Xlox*–*PG2*–*PG3*]-related *ProtoHox* gene, the «posterior» and «anterior» *H*/*P* genes appearing secondarily. The ProtoHOX cluster would have contained the three *Gsx*/*PG2*, *Pdx*/*PG3*, *Cdx*/*PG9* paralogs and produced through tandem duplication the primordial HOX and ParaHOX clusters in the Cnidaria–Bilateria ancestor. The stronger constraint on cnidarian *ParaHox* genes suggests that the primary function of pre-bilaterian *H*/*P* genes was to drive cellular evolutionary novelties such as neurogenesis rather than axis specification.

Introduction

The ANTP-class *Hox* genes encode transcription factors that act as vectorial driving systems for patterning the animal body axes during early development. They arose via repeated gene duplication events that led to the formation of multiple evolutionarily-conserved gene families in bilaterians, with structural features conserved from protostomes to deuterostomes, namely their 60 amino acid long DNA-binding domain (named homeodomain, HD) and a clustered chromosomal organization. As a result *Hox* gene families form paralogous groups (PGs) distributed in a conserved order along the cluster indicating that this organization was already present in the bilaterian ancestor (McGinnis and Krumlauf, 1992). Moreover these multigenic complexes obey to the spatial colinearity rule whereby during early development the spatial domain of activity of one given gene along the anterior–posterior body axis in a given cell layer (neural tube, mesodermal derivatives) is related to its specific position within the cluster. Hence the most 3′ genes along the cluster are expressed anteriorly, the most 5′ genes posteriorly, and the PGs distribute in several classes according to their embryonic expression pattern: anterior (PG1, PG2, PG3), central (PG4–PG8) and posterior (PG9–PG13/14/15). Moreover, in vertebrates the physical arrangement of *Hox* genes on the chromosome also drives their progressive temporal activation during development, the expression of the most anterior genes preceding that of more posterior genes, a property named temporal colinearity (Kmita and Duboule, 2003). Invertebrate genomes usually contain a single cluster with a number of PGs varying among taxa, although in some phyla as nematodes and tunicates, the cluster got disintegrated (Ferrier and Holland, 2002; Seo et al., 2004). In contrast, tetrapod genomes typically contain four clusters as the HOX cluster underwent several rounds of duplication in the common vertebrate ancestor (Wagner et al., 2003; Duboule, 2007).

The ParaHOX cluster is the evolutionary sister of the HOX cluster but much simpler as it is unique and contains only three genes, *Gsx*/*Ind*, *Pdx*/*Xlox* (also named *IPF-1*, insulin promoter factor 1 in vertebrates) and *Cdx*/*Cad*, initially described as related to anterior, PG3 and posterior genes respectively (Brooke et al., 1998). Phylogenetic analyses indicate that the HOX and ParaHOX clusters originated from an ancestral ProtoHOX cluster by segmental tandem duplication (Garcia-Fernandez, 2005b). Moreover, the close relationship of *Hox* and *ParaHox* (*H*/*P*) genes strongly suggested that the hypothetical ProtoHOX cluster arose by repeated *cis*-duplications from a founder

* Corresponding author. Fax: +41 22 379 33 40.
E-mail address: brigitte.galliot@zoo.unige.ch (B. Galliot).
[1] These authors equally contributed to this work.

0012-1606/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2009.01.022

ProtoHox gene, itself originating from the duplication of a non-Hox *ANTP-class* gene possibly related to *evx* and *mox* (Brooke et al., 1998; Gauchat et al., 2000; Garcia-Fernandez, 2005b).

Cnidarians, whose origin predated the diversification of Bilateria, are organized along an oral–aboral axis, differentiate a neuro-muscular system as well as sensory organs, and therefore provide a unique evolutionary position to trace the core developmental processes at work in early eumetazoans as apical/anterior patterning (Bode et al., 1999; Galliot and Miller, 2000), axis specification (Gauchat et al., 2000; Hobmayer et al., 2000; Yanze et al., 2001; Ball et al., 2004; Finnerty et al., 2004; Lee et al., 2006; Rentzsch et al., 2006) or eye development (Kozmik et al., 2003; Stierwald et al., 2004). Moreover, cnidarians are complex animals, one class living exclusively as polyps (anthozoans, i.e. coral, sea anemone), whereas the three other classes (hydrozoans, cubozoans, scyphozoans), collectively named medusozoans, include in most species a parental medusa stage (Fig. 1A). Morphological and molecular evidences have recently depicted a new scenario on the evolutionary position of cnidarians, which are lately considered as triploblastic (Seipel and Schmid, 2005, 2006) and possibly bilaterian animals (Finnerty et al., 2004).

To unravel the composition of the ProtoHOX cluster at the time of its duplication and the evolutionary steps leading to the extant *H/P* complement, numerous ANTP-class homeobox genes were identified from evolutionarily-distant cnidarian species and insights were obtained from comparative genomic studies of Porifera, Cnidaria and higher Metazoa (Finnerty and Martindale, 1999; Gauchat et al., 2000; Finnerty et al., 2004; Chourrout et al., 2006; Kamm et al., 2006; Ryan et al., 2007; Larroux et al., 2007). Whereas phylogenetic analyses clearly identified the cnidarian ANTP-class genes belonging to *non-Hox* families (Gauchat et al., 2000; Kamm and Schierwater, 2006; Ryan et al., 2006), the case of cnidarian *H/P* gene families still remains ambiguous. The *Evx/Mox* families, proposed to represent the closest ancestors to *ProtoHox* genes (Gauchat et al., 2000; Minguillon and Garcia-Fernandez, 2003), were both identified in anthozoans and hydrozoans. But the cnidarian *Hox*-like genes (see Table 1) show a less supported affiliation with any of the bilaterian *Hox PGs* (Gauchat et al., 2000; Yanze et al., 2001; Kamm et al., 2006; Chourrout et al., 2006; Ryan et al., 2007). For example, several anthozoan and hydrozoan genes display some signature residues of the bilaterian PG1 family suggesting a common origin (Fig. S2). Similarly, three *Nematostella* genes with no counterpart in Medusozoa appear as *PG2*-related genes. More controversial is the situation of "posterior-like" genes that despite several PG9 signature residues (Fig. S2), exhibit a weak grouping with *PG9* genes, not confirmed when the dataset and/or the methodology were modified (Fig. 1B).

Among the *ParaHox* genes, the *cnox2/anthox2* genes isolated from hydrozoan, scyphozoan and anthozoan species were unambiguously recognized as *Gsx/Ind* homologs (Finnerty and Martindale, 1999; Gauchat et al., 2000; Hayward et al., 2001; Yanze et al., 2001), whereas *Cnox4* from *Eleutheria* (hydrozoan) was proposed as a *Cdx/cad* homolog (Gauchat et al., 2000; Kamm et al., 2006). Hence the identification in cnidarians of *Hox* (*PG1*, *PG2*) and *ParaHox* (*Gsx*, *Cdx-like*) orthologs (Fig. 1B) indicated that the tandem segmental duplication of the ancestral ProtoHOX cluster predated the Cnidaria–Bilateria divergence (Gauchat et al., 2000). However, the *Cnox4/Cdx* grouping is significantly supported only in small datasets with few "posterior" bilaterian H/P sequences (Kamm et al., 2006). Therefore, as for posterior *Hox* genes, a clear correlation between cnidarian and "posterior" bilaterian *ParaHox* genes is missing. Concerning the "central" *Pdx/Xlox ParaHox* gene, cnidarian orthologs were never identified to date although a HD distantly related to both Pdx/Xlox and Cdx families (therefore named Xlox/Cdx) has been found within the *Nematostella* genome (Chourrout et al., 2006).

The recent genomic analyses indicate a fragmented chromosomal organization of the cnidarian *H/P* genes. In the coral *Acropora formosa*, a physical linkage was first found between the *PG1*-related gene *Antp* and the *evx* homolog (Miller and Miles, 1993). More recently, studies using in silico genomic walks in *Nematostella* confirmed the existence of a linkage between *eve*, *PG1* and *PG2 Hox* genes (Chourrout et al., 2006; Kamm et al., 2006). They also showed the clustering of *Gsx* and *Xlox/Cdx* and the frequent repetitive duplication of homeogenes, those multiple copies remaining clustered in most cases. In *Hydra*, molecular analyses failed to find any clustering between *PG1* and *PG9*-like genes over a 250 kb scale (Gauchat et al., 2000), a result confirmed by genomic analyses performed in *Hydra* and *Eleutheria*, which revealed clustering of *H/P* genes only when duplicated (Chourrout et al., 2006; Kamm et al., 2006).

To provide a more complete picture of the early evolution of *H/P* genes, three hydrozoan species that present a full life cycle (alternating between the polyp and medusa stages) were screened for *H/P* genes: *Turritopsis dohrnii* (*Td*), an anthomedusa with an outstanding potential for life cycle reversal and morph rejuvenation (Piraino et al., 1996, 2004), *Cladonema radiatum* (*Cr*), an anthomedusa currently used as a model for lens eye differentiation (Stierwald et al., 2004; Suga et al., 2008) and *Clytia hemisphaerica* (*Ch*), a leptomedusa, whose early development is amenable to functional dissection (Momose and Houliston, 2007). Five novel H/P sequences were obtained, representing three distinct gene families, the *Pdx/Xlox ParaHox* family and two orphan cnidarian *Hox*-related families, *CnoxA* and *CnoxC*. We used these new sequences to reconsider the phylogeny of the *H/P* genes. As the size and the composition of the sampling was shown to dramatically influence the phylogenetic reconstruction (Wallberg et al., 2004), a systematic approach was followed to test the robustness of the nodes that define the *H/P* families within three distinct samplings, each of them being submitted to variations of their content. Thanks to this strategy, we were able i) to identify some cnidarian and bilaterian

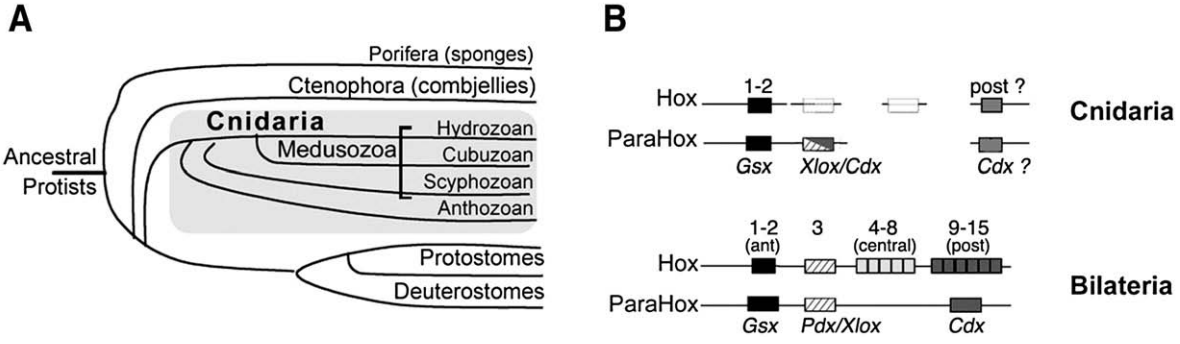


Fig. 1. Introductory schemes. (A) Scheme depicting the four distinct classes that compose the Cnidaria and its phylogenetic position. (B) Outline of recorded paralogous *Hox* and *ParaHox gene* families in extant Cnidaria and Bilateria. Empty boxes in cnidarians correspond to intermediate gene families not identified so far. (?) Homology between cnidarian and bilaterian posterior (post) or Cdx homeodomains was proposed on some shared signature residues but their grouping is weakly supported in phylogenetic analyses.

Table 1
Cnidarian genes representative of the Pre-Hox, ProtoHox, ParaHox and Hox gene families in anthozoans and medusozoans

Cnidarian H/P families	Anthozoa	Medusozoa[a]		Comments
		Hydrozoa	Scyphozoa	
Pre-Hox				
Evx/Eve	*Evx Nv, Eve Af*	*Evx Ssp*	Nd	Highly conserved family
Mox	*MoxA Nv, MoxB Nv, MoxC Nv, MoxD Nv*	*Cnox5 Hm = Mox Hm*	Nd	Highly conserved family, 04 copies in *Nematostella*
ProtoHox-related?				
CnoxA		*Cnox2 Pc*, **CnoxA Td**, **CnoxA Cr**		Unclear status: either orphan (datasetB), or grouping with Mox (datasetA), Gsx (datasetC) Hydrozoan-specific?
ParaHox				
Gsx	*Anthox2 Nv = Gsx, Cnox2 Am,*	*Cnox2 Cv, Cnox2 Hv, Cnox2 Ed, Cnox2 Ssp, Gsx Pc, Gsx/Cnox2 Hys*	*Scox2 Cx*	Divergence of the Gsx family between Anthozoa, Medusozoa and Bilateria
Pdx/Xlox	*XloxCdx-Nv, Antp-Nv*	***Pdx Td, Pdx Ch***	Nd	Highly conserved in Hydrozoa, highly derived in *Nematostella*
Cdx	Nd	*Cnox4 Ed*	Nd	Unclear status: grouping either Cdx (small datasets only) or with CnoxA (all datasets)
Hox				
PG1	*Anthox6 Nv = HoxA, Anthox6 Ms, AntpC Af (HoxB Nv = NVHD060)*	*Cnox1 Hv = Cnox4 Hm, Cnox1 Cv, Cnox5 Ed, Cnox1 Pc*	Nd	Divergence of the PG1 gene family between Anthozoa, Hydrozoa and Bilateria
PG2	*Anthox7 Nv = HoxC, Anthox8a Nv = HoxDa Anthox8b Nv = HoxDb*	Nd	Nd	Duplicated in *Nematostella*, Lost in Medusozoa?
PG9	Nd	Nd	Nd	See the PG9/Cdx-related cnidarian families
PG9/Cdx-related				
CnoxB	*Anthox1-Nv (HoxF)*	*Cnox4 Pc, HoxB Hm*	*Scox3 Cx*	
CnoxC		*Cnox3 Hv = Cnox1 Hm, HoxC2 Hm, HoxC3 Hm, Cnox1 Ed*, **CnoxC Ch**	Nd	Hydrozoan-specific?
CnoxD		*Cnox3-Ed CnoxD Hm*		Hydrozoan-specific?
CnoxE	*Anthox1a Nv (HoxE)*	Nd	*Scox1 Cx, Scox4 Cx*	

The novel genes reported here are written bold.
[a] There is currently no cubozoan Hox/ParaHox sequence available.

derived sequences that blur the phylogenetic picture, ii) to compare the respective evolutionary rates of *Hox* and *ParaHox* gene families and iii) to reconsider the metagrouping events, i.e. the clustering between the different *H/P* gene families.

Materials and methods

Origins and culture of hydrozoan species

Turritopsis dohrnii (*Td*) polyp colonies were collected from the Ionian sea (Schmich et al., 2007) and hosted as in (Piraino et al., 1996). *Clytia hemisphaerica* (*Ch*) colonies were collected from the Mediterranean sea and cultured according to (Carre and Carre, 2000; Stierwald et al., 2004) respectively. Animals were starved at least 3 days before mRNA extraction.

Cloning of the Pdx/Xlox, CnoxA and CnoxC genes

For the *Pdx/Xlox Td*, *PdxXlox Ch* and *CnoxA Td* genes, *Turritopsis* mRNA (900 ng) was extracted from polyps ($n = 50$), medusa buds ($n = 50$) and newly liberated medusae ($n = 100$) using a Quickprep mRNA extraction kit (GE Healthcare) and *Clytia* mRNA was extracted from planulae with Dynabeads Direct kit (Dynal). Homology PCR was performed with degenerated primers (Murtha et al., 1991) under standard conditions with low annealing temperature (37–40 °C). Those PCR fragments were then extended by RACE amplification and subcloned into the pCRII-TOPO vector (Invitrogen). For *CnoxC Ch* cloning, *Clytia* cDNA was produced with the Sensiscript Reverse Transcriptase (Qiagen) from 2 days-old larvae RNA (400 ng, RNeasy Qiagen). The PCR product (657 bp) was obtained by chance after 48 cycles (profile: 96 °C 45 s, 50 °C 3 min, 72 °C 3 min plus 6 s extension per cycle) with the degenerated primer (ckIckrttytgraaccadatytt) corresponding to the KIWFQNRR motif, which annealed on both sides. It was subsequently inserted into the pGEM-T Easy vector (Promega). The *CnoxA Cr* sequence was identified among *Cladonema* ESTs (National Institute of Genetics Mishima) kindly made available by Takashi Gojobori and Walter Gehring.

Expression analyses

In situ hybridization were performed on *Clytia* embryos generously provided by Evelyn Houliston as described in (Chevalier et al., 2006) with the following modifications: samples were incubated ON at 4 °C in anti-DIG-AP antibody 1/2000 (Roche). Detection was performed in NBT/BCIP solution containing 100 µg/ml NBT, 175 µg/ml BCIP, 100 mM Tris–HCl pH 8.5, 50 mM $MgCl_2$, 100 mM NaCl. After extensive washes, samples were mounted in Mowiol and observed with an Axioplan microscope (Zeiss). For semi-quantitative RT-PCR, *Turritopsis* and *Clytia* cDNAs were produced with the SuperScript First Strand Synthesis System (Invitrogen) from mRNAs isolated from various stages with the QuickPrep micro mRNA Purification kit (GE Healthcare). One microliter from each cDNA was PCR amplified first for 15 cycles (94 °C, 30 s; 65 °C, 30 s; 72 °C, 30 s; − 1 °C touch-down at each cycle), then for either 9, 14, 21 or 25 cycles (94 °C, 30 s; 50 °C, 30 s; 72 °C, 30 s) with the Pdx–TdF/–TdR (324 bp), CnoxA–TdF/–TdR (550 bp), CnoxC–ChF/–ChR (384 bp), Actin-F01/-R01 primers (sequences on request).

Phylogenetic analyses

The novel HD sequences were tested against the Swissprot, trEMBL and Genbank protein databases through the BLASTp program (www.expasy.org/tools/blast/). The sequences providing the best scores were selected, added to the currently available Hox-related cnidarian sequences and H/P sequences from slow-evolving species representing most bilaterian phyla (Table S1) and aligned with T-Coffee (Notredame et al., 2000). The newly cloned cnidarian HD sequences were assigned to the different H/P families using the maximum likelihood (PhyML) and Bayesian interference algorithms. The best model of protein evolution available within PhyML was determined using the ProtTest 1.3 program (Abascal et al., 2005). In both cases the JTT model of amino acid substitution was used. Bayesian phylogenetic analysis was performed using MrBayes version 3.1.2 under a mixed rate model of amino acid substitution (Ronquist and Huelsenbeck, 2003), assuming the presence of invariant sites and using a gamma distribution approximated by four different rate categories to model rate variation

across sites. Starting from random trees two independent runs of four incrementally heated Metropolis-coupled Markov chain Monte Carlo chains were simultaneously performed for 1,000,000 generations with trees being sampled every 100 generations. The likelihoods of the generations were scrutinized to estimate the beginning of the stationary phase. Those trees were used to create a consensus tree either after the first 100,000 generations (burn-in = 1000) for the bayesian trees generated from the datasetA (21 distinct alignments containing 34 to 50 HD sequences), or after the first 200,000 to 600,000 generations (burn-in = 2000–6000) for the Bayesian tree generated from the datasetB (16 alignments containing 85 to 189 HD sequences). Tree topologies were also reconstructed on the same datasets by Maximum Likelihood using the PhyML v2.4.4 program (Guindon and Gascuel, 2003). The JTT + I + Γ (JTT plus Gamma distributed rates plus invariant residues) model of amino-acid substitutions giving the highest likelihood value and a gamma shape distribution with eight discrete categories were used. Supports at nodes were assessed with the bootstrap method using 100 replicates. The gamma shape parameter and the proportion of invariant sites were estimated during the ML search.

Results

Cloning of two hydrozoan Pdx/Xlox homologs

Two partial cDNAs, respectively 464 bp and 310 bp long, encoding 96% identical HDs were identified from two hydrozoan jellyfish, *Turritopsis* and *Clytia*. These HDs showed highest identity rate with bilaterian Pdx/Xlox sequences, either from deuterostomes (69–71%) including vertebrates (Slack, 1995), amphioxus (Brooke et al., 1998), tunicates (Ferrier and Holland, 2002), hemichordates (Peterson, 2004), echinoderms (Hwang et al., 2003), or from protostomes as sipunculids (Ferrier and Holland, 2001b), molluscs (Canapa et al., 2005; Barucca et al., 2006), and annelids (Park et al., 2006; Frobius and Seaver, 2006). In contrast identity was lower with the *Nematostella* XloxCdx HD (66.7%, Fig. S1A). Moreover, both HDs possessed the H44 Pdx/Xlox signature residue (Hwang et al., 2003) (Fig. S1B), indicating that these two genes likely belonged to the *Pdx/Xlox ParaHox* group previously unmapped in cnidarians.

The hydrozoan CnoxA and CnoxC families

We also isolated from *Turritopsis* a full-length (838 bp) homeobox gene that we named *CnoxA Td*. The HD, located between positions 109 and 168, showed the highest identity (85%) with the orphan gene *cnox2 Pc* from *Podocoryne carnea* (Masuda-Nakagawa et al., 2000) (Fig. S1C). Concomitantly an additional *cnox2 Pc* homologue, *CnoxA Cr*, was isolated from *Cladonema*. In Neighbor-Joining analyses those CnoxA HDs, which did not contain any signature typical for any H/P family (Fig. S2), formed an orphan cnidarian family with the *cnox2 Pc* sequence (not shown). Finally we isolated from *Clytia* a homeogene whose HD sequence was closely related to the hydrozoan Cnox3 family, which shares some signature residues with the posterior Hox PGs (Fig. S2) but failed to show convincing relationships with any *H/P* bilaterian families in previous phylogenetic analyses (Gauchat et al., 2000). To prevent confusion with numbers defining Hox paralogs, we named this novel gene "*CnoxC Ch*". Therefore these three homeogenes (*CnoxA Td*, *CnoxA Cr*, *CnoxC Ch*) indeed are related to the *H/P* gene families but in the absence of any obvious affiliation, were considered as orphan cnidarian *H/P*-related genes.

Temporo-spatial regulation of Gsx, Pdx and CnoxC expression in developing Clytia

Transcripts of those novel genes were detected by RT-PCR at both medusa and polyp stages in *Clytia* and *Turritopsis* (Fig. S3), indicating that they are expressed and thus likely functional in those hydrozoan species. To detect whether these genes were regulated during development, gastrulating embryos and growing planulae from *Clytia* were hybridized to the *Gsx Ch*, *Pdx Ch* and *CnoxC Ch* riboprobes (Fig. 2). The *FoxQ2a Ch* riboprobe used in the same experiment (Fig. 2A) provided the anterior/aboral pattern at the gastrula and planula stages as previously described (Chevalier et al., 2006). The *Gsx Ch* expressing cells were exclusively detected in endodermal cells of embryos and planulae, as large spots in the posterior/oral region at the late gastrula stage then in the anterior/aboral half, extending subsequently towards the oral pole until it covers the whole endodermal region (Fig. 2B). This pattern is consistent with that described in *Podocoryne* (Yanze et al., 2001), although the initial posterior wave of expression had not been reported in *Podocoryne* where the detection level was weaker than in *Clytia*.

Interestingly *Pdx Ch* expression was turned on approximately 1 day later than *Gsx Ch* one: in fact *Pdx Ch* expressing cells were hardly detectable at the gastrula stage, but formed in 1 day old planula a spotty pattern that extended over both layers, leaving the anterior/aboral pole free of transcripts. In 3 days old planulae *Pdx Ch* expression was restricted to the endodermal region and rather diffuse (Fig. 2C). Similarly the *CnoxC Ch* expressing cells were also first detected in 1 day old planula, where they were clearly restricted to the tip of the anterior/aboral pole in the ectodermal layer. They subsequently extended in the endodermal region towards the posterior/oral pole until they formed an ubiquitous pattern that left free the posterior/oral pole (Fig. 2D). Hence the *Gsx Ch*, *Pdx Ch* and *CnoxC Ch* genes exhibit highly dynamic expression patterns during the early development of *Clytia*, suggesting that they play some role in the ongoing developmental processes. Moreover these three genes are also expressed at the medusa and polyp stages (Fig. S3).

Comparative phylogenetic analyses with three sampling strategies

To face the problem of the sampling influence on the clustering of the sequences and the solidity of the nodes (Wallberg et al., 2004), three distinct sampling strategies were applied to test the affiliation of the novel hydrozoan HD sequences (Fig. S4), two were based on the alignment of HD sequences, the first one restricted to ParaHox families (datasetA), the second inclusive for all H/P families (datasetB), and the third one was based on the alignment of full-length homeoprotein sequences (datasetC). The sequences from *Turritopsis* (*Pdx Td*, *CnoxA Td*), *Clytia* (*Pdx Ch*, *CnoxC Ch*) and *Cladonema* (*CnoxA Cr*) were aligned with the closest sequences identified in BLAST similarity search, the full set of cnidarian H/P sequences currently available ($n = 43$) and a set of previously aligned H/P HD sequences (Gauchat et al., 2000). DatasetA contained up to 50 sequences representing the *ParaHox*, *CnoxA* and *Mox* families (Fig. 3B, Fig. S4, Table S2), whereas datasetB that represented all H/P as well as *Evx* and *Mox* families was much larger, up to 189 sequences (Fig. 5A, Fig. S4, Table S3). Three non-Hox ANTP-class HD sequences (Msh Cv, NK-2 Hv, Cnox3 Pc) were used as outgroups in most analyses. Consensus trees representative of the topology obtained with datasets A and B are depicted in Fig. 3A and Fig. 4 respectively.

To test the stability of the nodes that define H/P families and identify the sequences that alter their robustness, datasets A and B were submitted to systematic variations of their composition, resulting in 21 distinct alignments for datasetA (Fig. 3B, Table S2) and 16 for datasetB (Fig. 5A, Table S3). Those alignments were tested in both PhyML and Bayesian analyses (74 trees) with replicates to provide bootstrap proportion (BP) and posterior probability (PP). Each H/P family (F) tested in a given alignment was scored according to the formula: $S^F = BP^F + [100 \times PP^F]$, where the possible maximal S^F value is 200. This approach allowed us to produce and compare up to 37 scores for each *H/P* gene family, depicted as graphs in Fig. 3 (datasetA, Table S2) and Fig. 5 (datasetB, Table S3).

Finally, in an attempt to recover more phylogenetic information to analyze the possible affiliation of cnidarian H/P-related sequences that displayed orphan positions in datasetA and datasetB trees, T-coffee alignments of 45 to 50 full-length homeoproteins were performed and tested in PhyML and Treefinder analyses (datasetC, Fig. S4). This approach confirmed the clustering of *H/P* gene families previously observed with HD sequences (Fig. S7), indicating that the background level of phylogenetic noise of those alignments did not affect the overall phylogenetic signal.

The hydrozoan Pdx/Xlox sequences display a high level of conservation

In datasetA the two hydrozoan Pdx sequences grouped with the bilaterian Pdx/Xlox/IPF-1 sequences forming a well supported group ($126 \leq S^{Pdx/Xlox} \leq 188$) when the two *Nematostella* sequences XloxCdx-Nv and Antp-Nv were not included (Fig. 3, Table S2). In the six alignments that included both XloxCdx-Nv and Antp-Nv, the Pdx/Xlox group was poorly supported ($12 \leq S^{Pdx/Xlox} \leq 122$); for example the A21 and A13 alignments that are identical except for these two sequences, display $S^{Pdx/Xlox}$ values of 179 and 13 respectively (Figs. 3, S5). The $S^{Pdx/Xlox}$ was drastically reduced in A13 because the Pdx/Xlox group was disintegrated in Bayesian analyses. The same phenomenon was observed in the alignments A10 and A12 where the $S^{Pdx/Xlox}$ dropped from 188 to 12. Except in those Bayesian trees, the Pdx/Xlox family always formed and the two new hydrozoan Pdx sequences always grouped in. Moreover XloxCdx-Nv and Antp-Nv also affected the Gsx score in A12, A13, A14 where $S^{Gsx} < 110$ (Fig. 3B), whereas the Mox, CnoxA and bilaterian Cdx scores remained unaltered (Fig. 3C).

The phylogenetic analyses performed on datasetB confirmed the affiliation of the hydrozoan Pdx sequences to the bilaterian Pdx/Xlox family with a significant support when XloxCdx-Nv and/or Antp-Nv were absent ($115 \leq S^{euPdx/Xlox} \leq 178$). When present the scores were significantly lower and more variable ($4 \leq S^{euPdx/Xlox} \leq 106$) (Figs. 5A,F, G, Table S4). For example the alignments B13 (Fig. 4) and B9 (Fig. S6) that differ only by the presence of Xlox/Cdx-Nv and Antp-Nv scored 166 and 68 respectively (Fig. 5A, yellow curve). As in datasetA, the $S^{Pdx/Xlox}$ was drastically reduced in trees showing a complete disintegration of the Pdx/Xlox family in Bayesian analyses (B7, B8). By contrast the variations in the bilaterian Pdx/Xlox sampling did not affect the stability of the Pdx group. However orphan bilaterian sequences could also exert a negative effect as in B7 where the addition of Lox1 Hm and Hox5 Dl to B5 alignment dramatically reduced $S^{euPdx/Xlox}$ from 93 down to 4 (Fig. 5A,F, Table S3).

The affiliation of XloxCdx was tested in 14 alignments, 8 from datasetA, 6 from datasetB. A similar behavior was recorded in both datasets, i.e. grouping together with the Pdx/Xlox family in all PhyML trees, but appearing in Bayesian analyses either as a Pdx/Xlox family member (24 trees), or as an orphan (4 trees) when the Pdx/Xlox family was actually disintegrated (Fig. 6). But XloxCdx-Nv, either alone or in combination with the Antp-Nv sequence, also affected the stability of the eumetazoan Cdx family as in 8 trees of datasetA where Cdx appeared paraphyletic, branching from the Pdx/Xlox family (see A9, A15 -> A17 in Bayesian analyses and A14 and A18 in both analyses, Table S2). In this latter case, XloxCdx-Nv branched at the root of Cdx within the Pdx/Xlox family. Similarly in datasetC, the clustering of XloxCdx-Nv at the root of the Cdx family was frequently observed (13/16 trees,

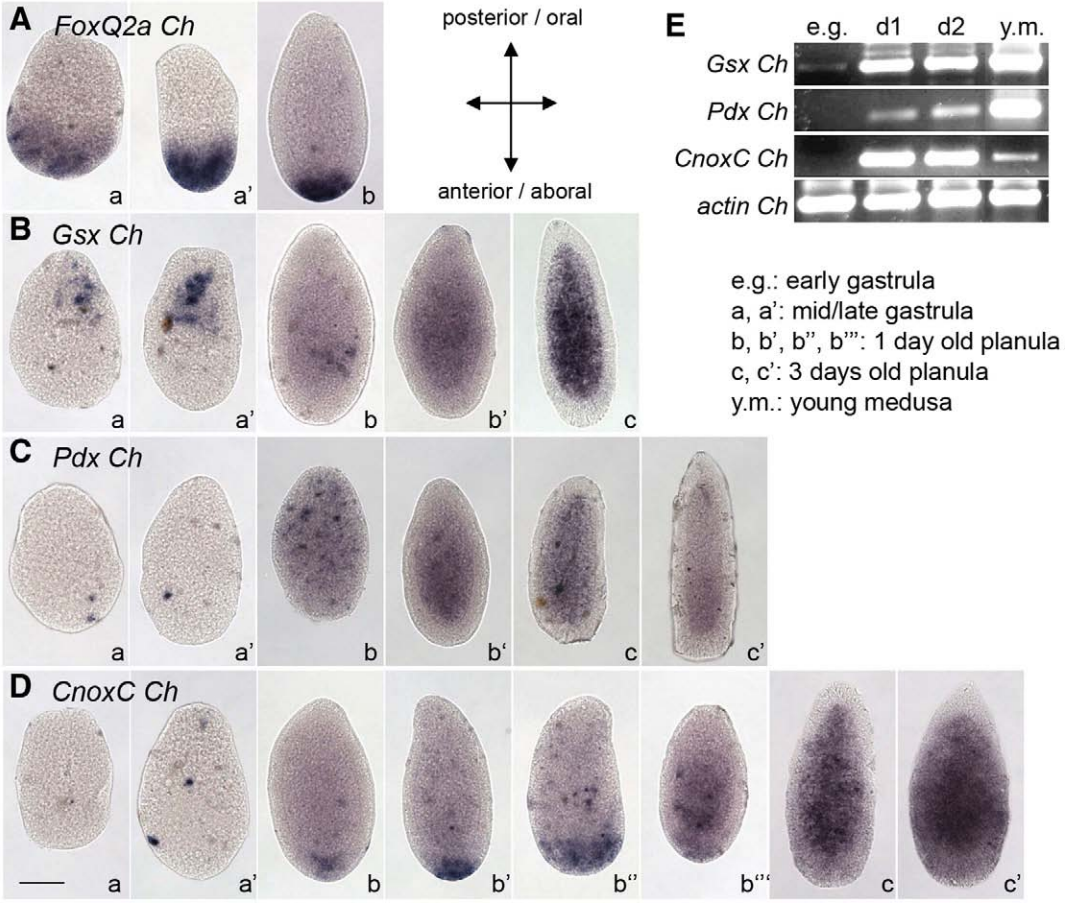


Fig. 2. Expression of *Gsx Ch*, *Pdx Ch* and *CnoxC Ch* in developing *Clytia*. (A–D) In situ hybridization patterns observed in mid/late gastrula (a panels), 1 day old (b panels) and 3 days old (c panels) planulae. Scale bar: 100 μm. (E) Expression detected by semi-quantitative RT-PCR in early gastrulae (e.g.), one (d1) and 2 days old (d2) planulae, in young medusa (y.m.).

81%), rarely with the Pdx family (2/16 trees, 12.5%). These results indicate that the presence of the two hydrozoan Pdx sequences enhanced the attraction of XloxCdx-Nv towards the Pdx family in analyses restricted to HD sequences. Hence, the systematic analysis of trees built from either large or small datasets provided clearcut evidences that *Pdx Td*, *Pdx Ch* and *XloxCdx-Nv* belong to the *Pdx/Xlox ParaHox* gene family, *XloxCdx-Nv* being a derived *Pdx/Xlox* representative.

Cnox4-Ed as a distantly-related hydrozoan Cdx gene

Previous analyses had shown that Cnox4-Ed and Cdx HD sequences share some signature residues (Fig. S2) and phylogenetic affinities (Gauchat et al., 2000; Chourrout et al., 2006; Kamm et al., 2006). In datasetA Cnox4-Ed that was included in 8 alignments (Table S2) was indeed frequently grouping with Cdx sequences (11/16 trees, 68.8%), although with low scores ($31 \leq S^{Cdx+Cnox4\text{-}Ed} \leq 104$) as supported by both methods in only three alignments (A8, A11, A21). This grouping was still observed in the presence of the Xlox/Cdx-Nv and Antp-Nv sequences, however with S values drastically lowered ($31 \leq S^{cdx+Cnox4\text{-}Ed} \leq 41$), as supported only with the PhyML method. In the corresponding Bayesian analyses Cnox4-Ed was also attracted towards the CnoxA family (3/16 trees, 18.7%) or orphan (2/16 trees, 12.5%). A similar behavior was noted in datasetC where Cnox4-Ed grouped with Cdx (12/23 trees, 52%), CnoxA (7/23 trees, 30%), or took an orphan position (4/23 trees, 17%). Finally in datasetB, the grouping of the Cnox4-Ed and Cdx sequences was less frequent (8/32 trees, 25%) likely prevented by the presence of the derived planarian Pnox6 Pn sequence, which favored the grouping of Cnox4-Ed with CnoxA (13/32 trees, 40.6%, Fig. 6, Table S3). These data confirm the frequent grouping of Cnox4-Ed with Cdx sequences in small datasets. In large datasets this grouping was more labile, altered by the influence of derived sequences.

Characterization of the hydrozoan CnoxA family

The three hydrozoan CnoxA sequences formed a robust node with high and stable S values in datasetA ($S^{CnoxA} \geq 166$, Table S2), datasetB ($S^{CnoxA} \geq 160$, Fig. 5E, Table S3) and datasetC (BP>80, Fig. S7). The

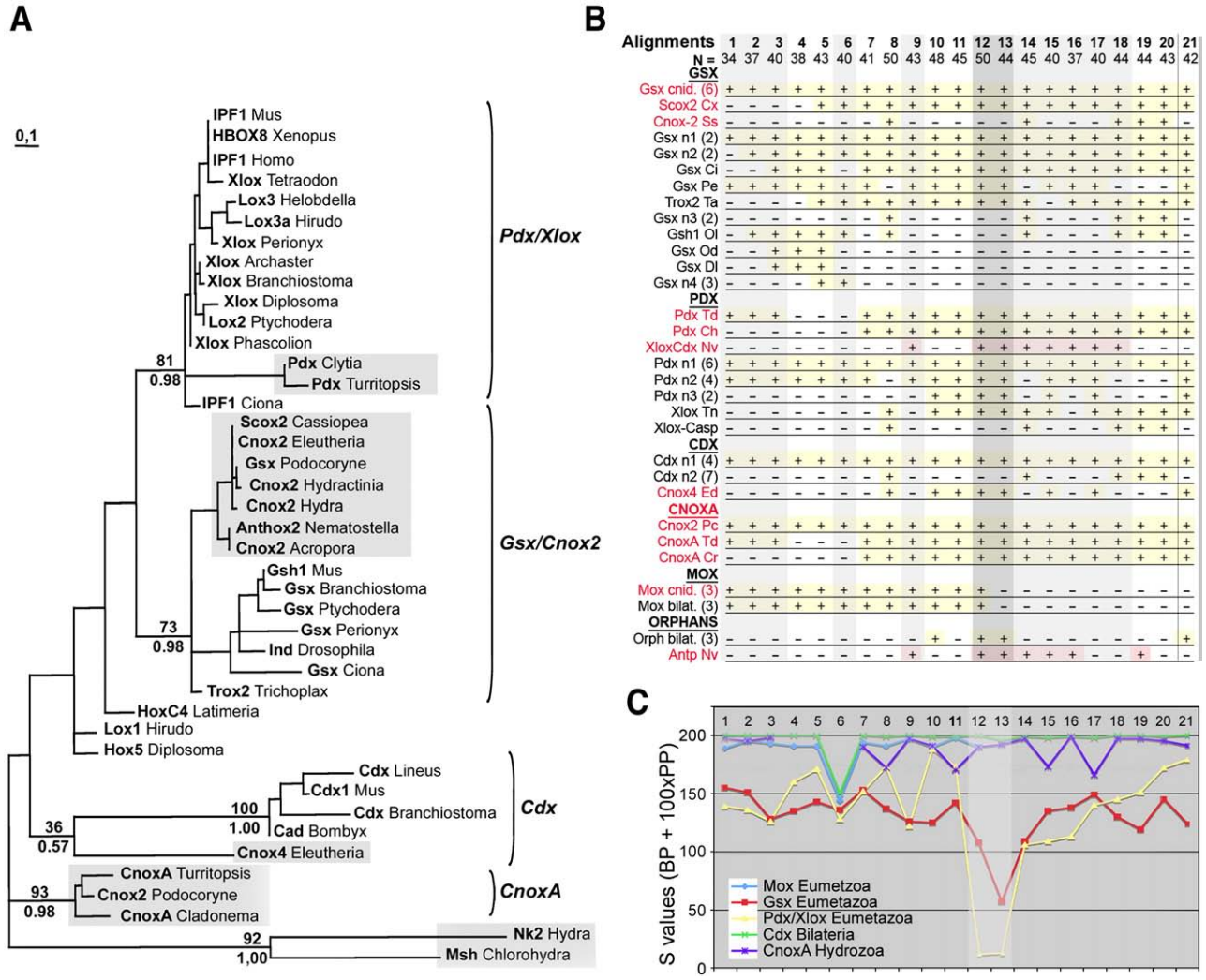


Fig. 3. Phylogenetic relationships between the cnidarian and bilaterian ParaHox HD sequences (datasetA). (A) Tree drawn using the PhyML program, corresponding to the alignment A21 (15 cnidarian and 27 bilaterian HD sequences, see in (B) also tested in Bayesian analysis. For each node, the BP (top) and PP (bottom) values obtained after 100 and 10,000 replicates respectively are indicated. Gene families significantly supported are indicated on the right, cnidarian sequences are boxed in grey. For species code and accession numbers see Table S1. (B) Composition of the 21 alignments of datasetA. Cnidarian gene names are written red. Bilaterian sequences of a given family were grouped, e.g. "Gsx n4 (3)" means that the 3 sequences forming the Gsx group number 4 were included (+) or not (−). See Table S2 for the detailed composition of each tree. N: total number of sequences; light grey shading: alignments where $50 < S^{Pdx/Xlox} < 150$; dark grey shading: alignments where $S^{Pdx/Xlox} < 50$; pink shading: presence of Antp-Nv and Xlox/Cdx-Nv sequences. (C) Graph showing the variations of S values for each gene family according to the alignment composition. The compiled S scores that represent the robustness of a given family, were deduced from BP and PP values of PhyML and bayesian analyses (see in Table S2). Alignment numbers are indicated at the top. Note that the Pdx and Gsx families were supported in both analyses in all alignments except when the Antp-Nv and Xlox/Cdx-Nv sequences were included together with two bilaterian orphans (A12, A13, light grey box, Fig. S5).

affiliation of *CnoxA* to one or the other H/P family was ambiguous as in datasetA *CnoxA* either appeared as an orphan H/P-related family (15/42 trees, 36%) or grouped with the Mox family (16/42 trees, 38%, Table S2), this latter grouping being supported by both methods in seven alignments ($107 \leq S^{CnoxA+Mox} \leq 131$) and not affected by the presence of Xlox/Cdx-Nv and Antp-Nv. More rarely the CnoxA and Cdx sequences grouped together (5/42 trees, 12%) with weak support ($21 \leq S^{CnoxA+Mox} \leq 74$) and in two alignments CnoxA + Mox + Cdx grouped together.

In datasetB the CnoxA family was also frequently orphan (27/32 trees, 84%) but displayed a position in the vicinity of the root in 13/16 Phyml trees, i.e. close to the Evx and Mox families, frequently associated with Cnox4-Ed as stated above. When tested in datasetC, the CnoxA sequences that are all possibly full-length (Fig. S1C) frequently grouped with the Gsx family (15/21 trees, 71.4%) with some bootstrap support (Fig. 6). In some trees this metagroup (CnoxA + Gsx) was associated with the Cdx (4/21, 19%) or Mox (3/17, 17.6%) families. CnoxA sequences also attracted Cnox4-Ed (4/22 trees, 18.2%) or derived cnidarian sequences (Cnox3-Ed, XloxCdx-Nv, 2/22 trees, 9%). This result suggested that CnoxA and Gsx proteins share some common residues outside the HD as the Pro-, His- or Cys-rich stretches upstream to the HD, or the K61 residue immediately downstream, also present in the Cnox4-Ed and Cdx sequences. It should be noted that CnoxA never grouped with the other "orphan" cnidarian families (CnoxB to CnoxE).

Characterization of the cnidarian CnoxB, CnoxC, CnoxD and CnoxE families

Six sequences, representing five distinct hydrozoan genes form the cnidarian *CnoxC* family (Fig. S1, Fig. 4). In datasetB this family displayed a high phylogenetic score ($S^{CnoxC} \geq 150$), insensitive to the variations introduced in the dataset (Fig. 5). Moreover, three additional cnidarian families, CnoxB, CnoxD and CnoxE, containing medusozoan sequences, formed in datasetB and datasetC trees (Fig. 4, Fig. S7, Table 1). In datasetB, Anthox1-Nv grouped together with CnoxB sequences in 12/18 trees (67%) whereas Anthox1a-Nv grouped with CnoxE in 6/28 trees (21.4%) (see MG1 and MG3 in Fig. 6), suggesting that *Anthox1-Nv* and *Anthox1a-Nv* are the anthozoan orthologs of the medusozoan *CnoxB* and *CnoxE* families respectively. The metagrouping of CnoxC together with CnoxB and Anthox1-Nv sequences (metagroup 2, MG2) was observed in 13/18 trees (72%) with significant support ($77 \leq S^{MG2} \leq 149$, Fig. 6). In datasetB, the clustering of these four families (CnoxB, CnoxC, CnoxD and CnoxE) was recorded in the vicinity of the PG9/Cdx sequences in 5/16 trees (31%) although without any supporting bootstrap value (MG9, Fig. 6). In datasetC, these "orphan" cnidarian families that frequently clustered together (as noted in 63% of the trees, not shown), also appeared to share a common origin with PG9 and Cdx families (Fig. S7B). Surprisingly in datasetB the metagrouping of the PG9 and Cdx bilaterian families was noted in only 5/28 trees (17%) with very low scores (≤ 10). Therefore, in the absence of any other significant metagrouping event, we consider these four cnidarian families as putative "PG9/Cdx" orthologs.

The Gsx and Pdx/Xlox families are more conserved than the PG1 and PG2 Hox families in eumetzoans

To evaluate the divergence of the Hox and ParaHox families between bilaterians and cnidarians, the eumetazoan phylogenetic scores obtained in the alignments of datasetB were systematically compared to the bilaterian and/or cnidarian ones (Figs. 5A–F). The mean values and standard deviations were calculated for each gene *H/P* gene family identified in eumetazoans ($\underline{S^{euF}}$), bilaterians ($\underline{S^{biF}}$), cnidarians ($\underline{S^{cnF}}$) and/or hydrozoans ($\underline{S^{hyF}}$) (Fig. 5G, Table S4). Concerning the eumetazoan PG1 family, the scores were low, all trees scoring < 102 but one (Fig. 5C, $\underline{S^{euPG1}} = 45 \pm 33$). In contrast the score of the bilaterian PG1 family were very high ($\underline{S^{biPG1}} = 186 \pm 6$) whereas the hydrozoan PG1 family that was supported by both methods in 12/15 alignments, also scored significantly higher than S^{euPG1} ($\underline{S^{hyPG1}} = 102 \pm 32$) (Fig. 5C,G, Table S3). Concerning the eumetazoan PG2 family, we similarly noted a large gap between the bilaterian and the eumetazoan scores ($\underline{S^{biPG2}} = 137 \pm 42$ versus $\underline{S^{euPG2}} = 66 \pm 42$). These significative differences between eumetazoan and bilaterian scores of the PG1 and PG2 families can only be explained by a strong divergence of the cnidarian sequences.

In contrast, the eumetazoan ParaHox families displayed higher and more stable scores: the eumetazoan Gsx family scored between 100 and 150 in 14/16 alignments ($\underline{S^{euGsx}} = 121 \pm 27$, Fig. 5B,G, Table S3) and as reported above, the eumetazoan Pdx family scored high and stable in the 9 alignments where the XloxCdx and Antp-Nv sequences were excluded ($\underline{S^{euPdx}} = 148 \pm 24$), still relatively well when taking into account all alignments ($\underline{S^{euPdx}} = 112 \pm 52$, Fig. 5A,G, Tables S3 and S4). In case of the Gsx family, where a large number of cnidarian sequences is available, the eumetazoan, bilaterian and cnidarian scores were close to each other ($\underline{S^{biGsx}} = 136 \pm 26$ and $\underline{S^{cnGsx}} = 151 \pm 33$). In summary the eumetazoan scores obtained by the Gsx and Pdx families were about twice higher than those recorded for the PG1 and PG2 families ($\underline{S^{euGsx}}$ (121) ~ $\underline{S^{euPdx}}$ (112/148) > $\underline{S^{euPG2}}$ (66) > $\underline{S^{euPG1}}$ (45), Fig. 5G). This quantitative analysis clearly measures higher amounts of phylogenetic signal in the cnidarian Gsx or the Pdx/Xlox ParaHox sequences compared to the PG1 and PG2 Hox ones, indicating a higher level of conservation for the ParaHox compared to the Hox families in cnidarians.

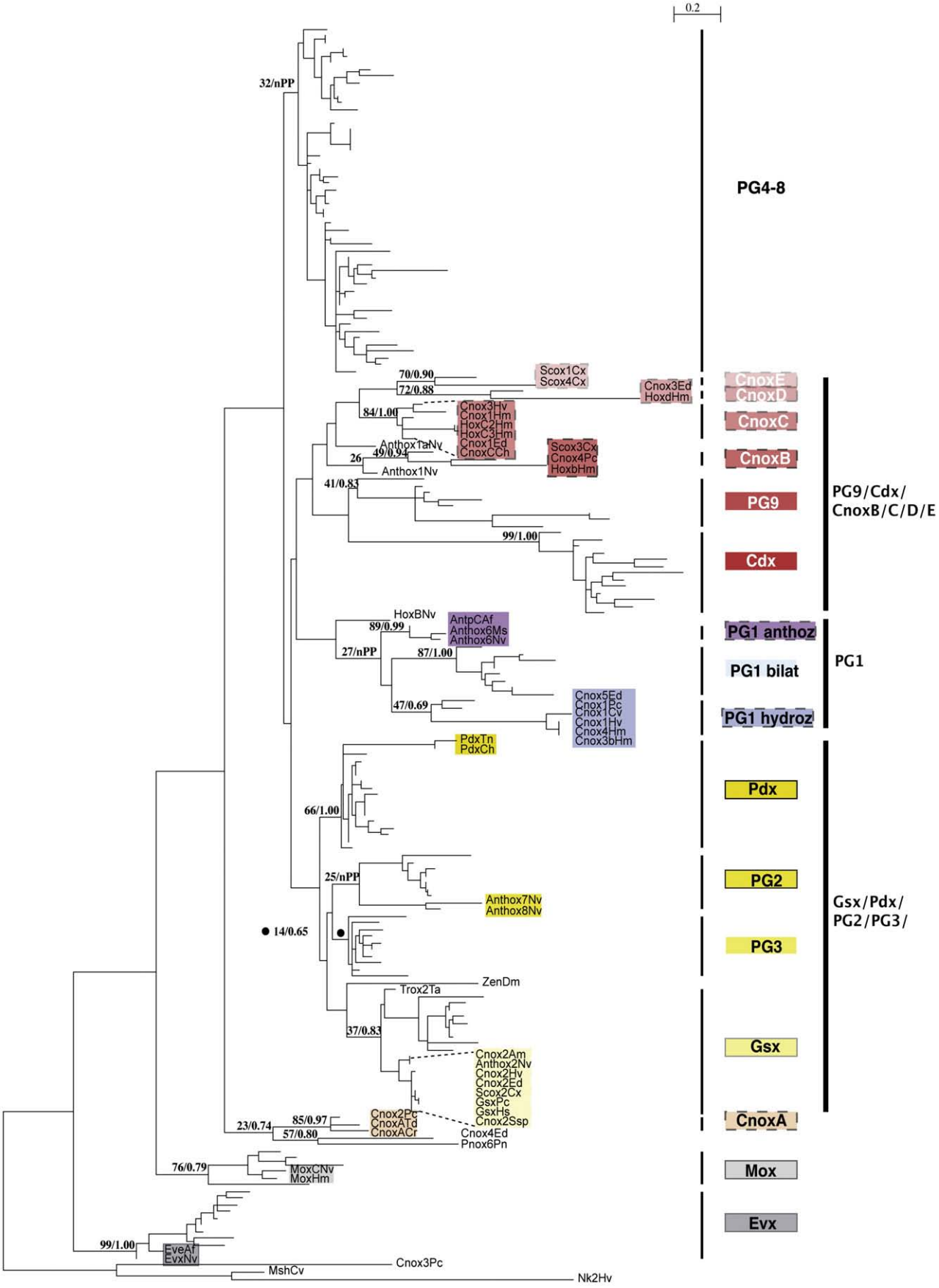
PG1 does not have any ParaHox counterpart

When the clustering of the *H/P* gene families, named "meta-grouping", was investigated in the trees generated from datasetB (Fig. 6 and not shown), we noticed that the "anterior" Hox families, i.e. PG1, PG2 and PG3 actually never clustered together but rather formed two distinct branches, PG1 on one side, PG2 and PG3 together on another side as depicted in Fig. 4. Moreover the Gsx and Pdx/Xlox ParaHox sequences never clustered with the PG1 ones but repeatedly with the PG2/PG3 metagroup as observed in 34.4% of the trees built on datasetB (Fig. 6), with the best scores recorded in the B1, B4, and B13 trees (57, 58, 63 respectively). This indicates firstly that the *PG1 Hox* family on one side and the *PG2/PG3* on the other side, should be considered as two distinct groups with separate origins; secondly that the *Gsx* and *Pdx/Xlox ParaHox* gene families should no longer be considered as *PG1*-related but rather as homologous to the *PG2/PG3 Hox* families.

Discussion

A systematic quantitative strategy to score the phylogenetic signals present in H/P sequences

With the increasing number of Hox-related HD sequences identified from cnidarian species (up to 43 by now), phylogenetic analyses repeatedly established that most cnidarian *Hox*-like genes display rather divergent sequences when compared to bilaterian ones, a feature that is not common within the ANTP-class (Gauchat et al., 2000; Ryan et al., 2006; Larroux et al., 2007). In fact there is a strong contrast between the pre-bilaterian *non-Hox* ANTP-class gene families (*Msx*, *NK2*, *Emx*, …) that are highly conserved, and the cnidarian *Hox*-like gene families that display limited support with any bilaterian Hox sequences. Therefore the phylogeny of cnidarian H/P sequences remained difficult to establish and numerous scenarios were proposed to describe the early evolution of *H/P* genes. Phylogenetic analyses applied to either large or restricted (Gauchat et al., 2000; Finnerty et al., 2004; Kamm et al., 2006)


0.2
32/nPP
70/0.90
72/0.88
Scox1Cx
Scox4Cx
Cnox3Ed
HoxdHm
84/1.00
Cnox3Hv
Cnox1Hm
HoxC2Hm
HoxC3Hm
Cnox1Ed
CnoxCCh
Anthox1aNv
49/0.94
26
Anthox1Nv
Scox3Cx
Cnox4Pc
HoxbHm
41/0.83
99/1.00
HoxBNv
89/0.99
AntpCAf
Anthox6Ms
Anthox6Nv
27/nPP
87/1.00
47/0.69
Cnox5Ed
Cnox1Bc
Cnox1Cv
Cnox1Hv
Cnox4Hm
Cnox3bHm
PdxTn
PdxCh
66/1.00
25/nPP
Anthox7Nv
Anthox8Nv
14/0.65
ZenDm
Trox2Ta
37/0.83
Cnox2Am
Anthox2Nv
Cnox2Hv
Cnox2Ed
Scox2Cx
GsxPc
GsxHs
Cnox2Ssp
Cnox2Pc
CnoxATd
CnoxACr
23/0.74
85/0.97
57/0.80
Cnox4Ed
Pnox6Pn
76/0.79
MoxCNv
MoxHm
99/1.00
EveAf
EvxNv
Cnox3Pc
MshCv
Nk2Hv
PG4-8
CnoxE
CnoxD
CnoxC
CnoxB
PG9
Cdx
PG9/Cdx/
CnoxB/C/D/E
PG1 anthoz
PG1 bilat
PG1 hydroz
PG1
Pdx
PG2
PG3
Gsx
CnoxA
Gsx/Pdx/
PG2/PG3/
Mox
Evx

sequence datasets, provided conflicting results, e.g. the presence of "posterior" *H/P* genes in cnidarians. In this study, we developed strategies where the phylogenetic signal of each *H/P* gene family was systematically quantified in three different samplings, a first one focused on the ParaHox sequences to assess the family identity of the novel sequences, a second one inclusive for the cnidarian Hox, ParaHox and Hox-related sequences to analyse the metagrouping of the *H/P* families and identify the phylogenetic relationships of five cnidarian "orphan" *Hox*-related gene families, and finally, a third dataset that contained selected full-length homeoprotein sequences to confirm the obtained results and detect when possible additional phylogenetic information. As *H/P* gene families presumably arose from *non-Hox* ANTP-class genes (Gauchat et al., 2000; Garcia-Fernandez, 2005b; Larroux et al., 2007), all trees were rooted with the non-Hox Msx, NK-2 sequences.

The discovery of hydrozoan Pdx/Xlox orthologs highlights the different constraints applied on Hox and ParaHox genes in their early evolution

These analyses reported two novel hydrozoan sequences, whose identification as cnidarian *Pdx/Xlox* orthologs was unambiguously proven in three samplings, including in alignments containing the largest numbers of sequences, as the increase in the number of homologous sequences is supposed to strengthen the significant nodes (Wallberg et al., 2004). Moreover, they also confirmed the presence of *Gsx*, *PG2* and *PG1* orthologs in cnidarians. Among the *H/P* families, the *Gsx* and *Pdx/Xlox ParaHox* families displayed the highest scores, indicating that cnidarian and bilaterian members of those families are more similar to each other than are the *PG1* and *PG2 Hox* family members. Therefore we assume that the *Gsx* and *Pdx/Xlox ParaHox* sequences have less evolved from the ancestral ParaHox sequences, suggesting that differential evolutionary rates applied to the cnidarian *ParaHox* and *Hox* gene families, the former ones being more constrained. Does it mean that *ParaHox* genes represent better the ancestral status of the *ProtoHox* genes? This question remains open and data from basal non-cnidarian species as ctenophora and placozoa are needed to tell whether it is a cnidarian specificity or not.

The Gsx-PG2-PG3-Pdx/Xlox metagroup does not include PG1

In addition this new phylogenetic picture did not show the expected affiliation between the *H/P* gene families; the PG2, PG3, Gsx and Pdx/Xlox grouped together but never exhibited in the 32 trees obtained from datasetB the linkage usually displayed (Ferrier and Holland, 2001a), i.e. *Gsx* related to *PG1/PG2* and *Pdx/Xlox* related to *PG3*. By contrast, the systematic testing of those three different samplings allowed us to sort out the bilaterian H/P sequences in four metagroups: *PG1*, [*Gsx–PG2–PG3–Pdx/Xlox*], [*PG4–PG8*], [*PG9–Cdx*]. The metagrouping of the [*Gsx–PG2–PG3–Pdx/Xlox*] gene families is reported here for the first time, indicating that, in contrast to the prevalent view, PG2 and PG3 are twin families actually not closely related to any of the other Hox families, specially not to PG1. The eumetazoan *PG1* family contains three distinct sub-families, bilaterian, anthozoan and hydrozoan, highlighting the early divergence between Anthozoa and Medusozoa. The *Evx* and *Mox* families, which both contain cnidarian homologs, rarely grouped together (Fig. 6) but frequently took positions close to the root of the tree in PhyML analyses, suggesting that they represent intermediates between the non-Hox and the H/P sub-classes of the ANTP-class, as previously proposed (Gauchat et al., 2000; Garcia-Fernandez, 2005a).

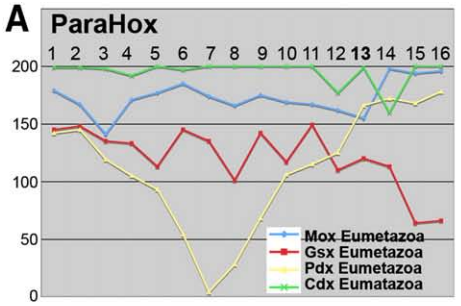
Phylogenetic analyses of the cnidarian H/P genes are obscured by some highly derived sequences

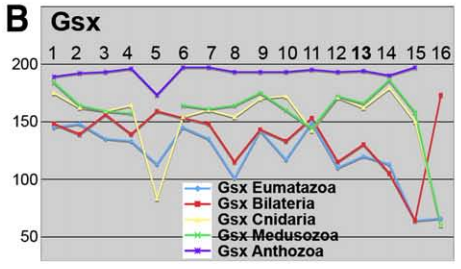
Concerning the ambiguous Xlox/Cdx *Nematostella* sequence (Chourrout et al., 2006), the presence of the H44 Pdx/Xlox signature residue together with the absence of Cdx signature residues in its HD, provide support for a *Pdx/Xlox* origin of this gene. Moreover the presence of the hydrozoan Pdx/Xlox sequences clearly enhanced the grouping of Xlox/Cdx towards the Pdx/Xlox family. In addition to the XloxCdx and Antp-Nv sequences from *Nematostella* but also Hox5 Dl or Lox1 Him from bilaterians were identified as derived sequences, which drastically altered the robustness of the Pdx/Xlox group, also affecting in some cases the robustness of other gene families. Therefore those highly derived sequences considerably obscure the phylogenetic picture of the cnidarian *H/P* genes and need to be systematically searched prior to drawing any conclusion. Additional derived sequences with negative effects were also identified in this study: Pnox6 Pn on the grouping of Cnox4-Ed with the Cdx family, Anthox8 Nv on the eumetazoan PG2 score, the urochordate Gsx sequences on the eumetazoan Gsx score. This work also shows that hydrozoans and anthozoans display strongly heterogenous evolutionary patterns: whereas the *Hox* families (*PG1*, *PG2*) exhibit a higher rate of conservation in anthozoans than in medusozoans, this is not the case for the *Pdx/Xlox* and *Cdx* families, more conserved in hydrozoans (Pdx/Xlox, Cnox4-Ed) than in *Nematostella* (XloxCdx). Consequently, data from both anthozoans and medusozoans are needed when considering cnidarians in evolution.

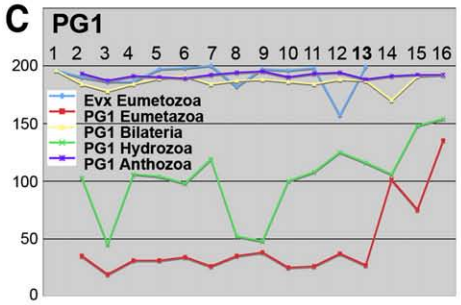
The "posterior" H/P gene families are highly derived in cnidarians

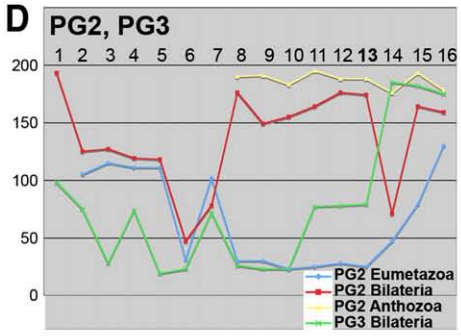
As anticipated, we did not observe any cnidarian sequences in any tree that would branch together with the central PG4 to PG8 groups, implying that those Hox paralogs arose after Cnidaria divergence. Concerning the posterior genes, it was previously reported that some *Hydra* (Cnox3 Hv) and *Nematostella* (Anthox1, Anthox1a) sequences share few signature residues with the posterior family but when the residues were given the same weight, clustering no longer appeared in phylogenetic analyses (Gauchat et al., 2000). Here, in datasets B and C, the cnidarian CnoxB, CnoxC, CnoxD and CnoxE HD sequences frequently grouped together with Cdx/PG9 (Fig. 4, Fig. 6), rarely with [PG2–Gsx/PG3–Pdx], never with PG1. This metagrouping analysis suggested a Cdx/PG9 origin for those families, without predicting from the data presented here the origin of the *CnoxB*, *CnoxC*, *CnoxD* and *CnoxE* ancestor(s); they could be derived from the *PG9/Cdx* ancestor of the ProtoHOX cluster, but also from *Cdx* of the primordial ParaHOX cluster or *PG9* of the primordial HOX cluster (Fig. 7). Hence "posterior" gene families, either *Hox*- or *ParaHox*-related, are highly divergent in cnidarians but can nevertheless be identified when numerous datasets are compared and when metagroupings are taken into consideration. Concerning the "posterior" *ParaHox* gene, the putative *Cdx* ortholog, the hydrozoan Cnox4-Ed, indeed more frequently grouped with Cdx in small datasets but also in large datasets that did not include the derived sequence Pnox-6. Nevertheless additional medusozoan Cnox4-Ed like sequences are required before concluding about the presence of genuine Cdx orthologs in cnidarians. Finally, the *CnoxA* cnidarian family is related neither to the "posterior" cnidarian families nor to the PG1 families, but exhibits some *Mox* and *Gsx* phylogenetic signal. Therefore, we propose that *CnoxA* has possibly arisen through duplication from more ancestral sequences than the extant H/P families, and might thus represent *ProtoHox* gene families as the *Gsx/Pdx/PG2/PG3* ancestor, likely arisen from *Evx/Mox* duplication (Fig. 7, right).

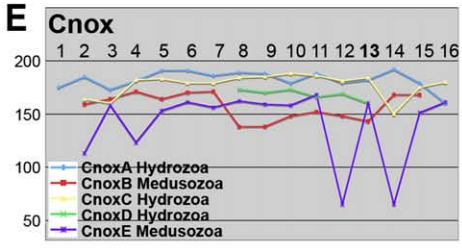
Fig. 4. Tree showing the phylogenetic relationships between cnidarian and bilaterian Hox and ParaHox HD sequences. This tree corresponds to the alignment B13 that contains 47 cnidarian and 140 bilaterian H/P sequences (see composition in Fig. 5F and Table S3). The robustness of the phylogenetic inference was tested through replicates, 100 in PhyML and 10,000 in Bayesian programs with BP and PP values noted on the nodes. Each gene family sits on a colored background, dashed outlined when cnidarian only, full when cnidarian and bilaterian, no outline when bilaterian only. Note the five Hox-related cnidarian families, CnoxA to CnoxE, on orange backgrounds. Except Pnox6 Pn, only the names of the cnidarian sequences are noted. The Nk2, Msh and Cnox3 Pc sequences were used as outgroups.











F

Alignments	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	B16
n =	116	176	178	180	180	178	184	188	189	188	186	188	187	89	99	85
PDX/Xlox																
Pdx Bilateria	12	12	12	15	15	12	15	15	15	15	15	15	15	10	10	7
Pdx Td	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Pdx Ch	+	+	+	+	+	+	+	+	+	+	+	+	+	–	–	–
XloxCdx Nv	–	–	–	+	+	+	+	+	+	–	–	–	–	–	–	–
GSX																
Gsx Bilateria	9	10	10	10	9	10	10	10	10	10	10	10	10	10	10	4
Trox2 Ta	+	+	+	+	–	+	+	+	+	+	+	+	+	+	+	+
Gsx Medusozoa	7	7	7	7	7	7	7	7	7	7	6	6	7	4	4	4
Gsx Anthozoa	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1
CDX																
Cdx Bilateria	12	13	13	13	13	13	13	13	13	13	13	13	13	4	4	4
Cnox4 Ed	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PG1																
PG1 Bilateria	8	8	8	8	8	8	8	8	8	8	8	8	8	5	5	5
PG1 Hydrozoa	–	6	6	6	6	6	6	6	6	6	6	6	6	5	5	5
PG1 Anthozoa	–	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2
PG2																
PG2 Eumetazoa																
PG2 Bilateria	2	7	7	7	7	7	7	7	7	7	7	7	7	3	3	3
Anthox7 Nv	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Anthox8 Nv	–	–	–	–	–	–	–	+	+	+	+	+	+	+	+	+
PG3 - PG8																
PG3 Bilateria	11	11	11	11	11	11	11	11	11	11	11	11	11	4	4	4
PG4-8 Bilateria	30	52	51	52	52	52	53	52	52	52	52	52	52	9	9	9
PG9																
PG9 Bilateria	–	8	8	8	8	8	8	8	8	8	8	8	8	6	6	6
CNOXA																
Cnox2 Pc	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CnoxA Td	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CnoxA Cr	+	+	+	+	+	+	+	+	+	+	+	+	+	–	–	–
CNOXB																
Scox3 Cx	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CnoxB Hm	–	–	–	–	–	–	–	+	+	+	+	+	+	–	–	–
Cnox4 Pc	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–
Anthox1 Nv	–	–	–	–	–	–	–	+	+	+	+	+	+	+	+	+
CNOXC																
Cnox3 Hv	–	+	+	+	+	+	+	+	+	+	+	+	+	–	–	–
Cnox1 Hm	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
HoxC2 Hm	–	–	–	–	–	–	–	+	+	+	+	+	+	–	–	–
HoxC3 Hm	–	–	–	–	–	–	–	+	+	+	+	+	+	–	–	–
Cnox1 Ed	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CnoxC Ch	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CNOXD																
Hoxd Hm	–	–	–	–	–	–	–	+	+	+	+	+	+	–	–	–
Cnox3 Ed	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CNOXE																
Scox1 Cx	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Scox4 Cx	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Anthox1a Nv	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–
ORPHANS																
Hox5Dl, Lox1Hm	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Pnox6 Pn	–	–	–	–	–	–	–	+	+	+	–	+	+	+	+	+
HoxB Nv	–	–	–	–	–	–	–	+	+	+	+	+	+	–	–	–
Antp Nv	–	–	–	–	+	+	+	–	+	+	–	–	–	–	–	–
MOX																
Mox Bilateria	3	3	3	3	3	3	3	4	4	4	4	4	4	3	3	3
Mox Cnidaria	3	3	3	3	3	3	3	2	2	2	2	2	2	3	3	3
EVX																
Evx Bilateria	6	9	9	9	9	9	9	9	9	9	9	9	9	–	–	–
Eve Cnidaria	2	2	2	2	2	2	2	2	2	2	3	3	2	–	–	–
OUTGROUPS																
Bilateria	–	–	10	–	–	–	–	–	–	–	–	–	–	–	8	10
NK2 Hv, Msh Cv	+	+	+	+	+	+	+	+	+	+	+	+	+	–	+	+
Cnox3 Pc	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Cnox3 Cv	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–

G

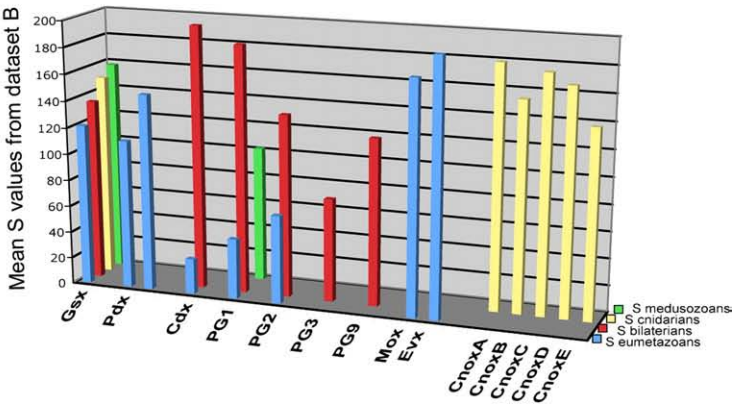


Fig. 5. Analysis of the robustness and the stability of the H/P family nodes. (A–E) Variations in the S values supporting (A) the *ParaHox* families in eumetazoans except *Cdx* restricted here to bilaterian sequences, (B) the *Gsx* family in eumetazoans, bilaterians, medusozoans and anthozoans, (C) the *PG1* family in eumetazoans, bilaterians, hydrozoans and anthozoans, (D) the *PG2* family in eumetazoans, bilaterians and anthozoans and the bilaterian *PG3* family, (E) the "orphan" cnidarian Hox-related families. The *Mox* (A) and *Evx* (C) families were included as control families that exhibit high and stable scores. See all BP and PP values in Table S3. (F) Composition in ParaHox, Hox and Hox-related sequences forming the 16 distinct alignments of datasetB that were tested in PhyML and Bayesien analyses. Left: Names of the H/P gene families (underlined) and cnidarian sequences (red). Numbers on yellow background indicate the number of sequences representing each gene family in a given alignment. n: total number of sequences included in each alignment that is numbered at the top. Note that the alignments B4 to B9 contain the XloxCdx sequence. (G) Graph aligning the means of the scores obtained by the various H/P families from datasetB (depicted in panels A–E); the blue bars correspond to the scores obtained by the H/P families in eumetazoans; red bars in bilaterians; yellow bars in cnidarians; green bars in medusozoans (Gsx and PG1 only). Two mean values are provided for the eumetazoan Pdx family, one taking into account all alignments ($N=16$, left) and a second one that considered only the alignments lacking the XloxCdx and Antp-Nv sequences ($N=9$, right). See in Table S4 the numerical values and standard deviations. Note the low mean values for the Cdx, PG1 and PG2 families in eumetazoans.

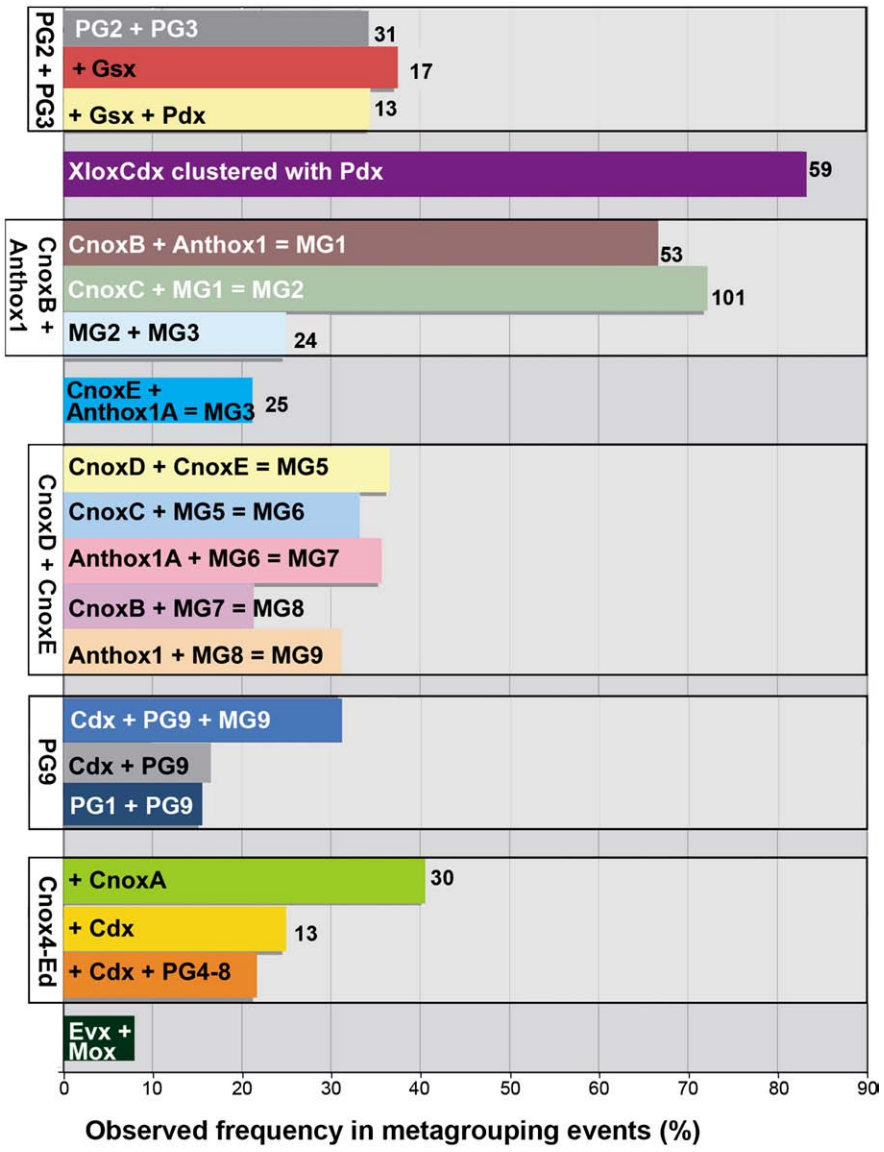
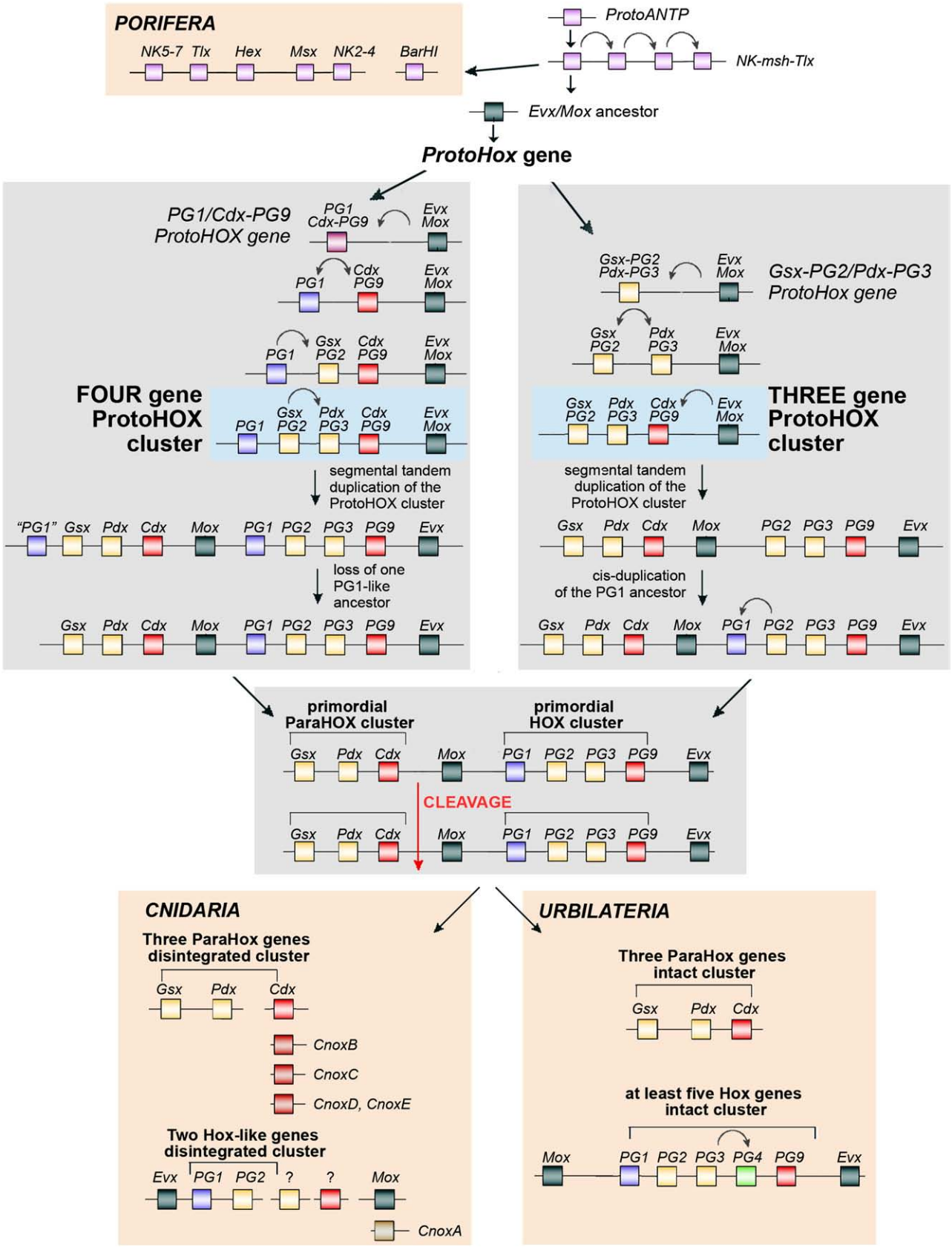


Fig. 6. Analysis of the metagrouping between the ParaHox, Hox and Hox-related families. Respective frequencies of the metagrouping events (MG) observed among the 32 trees built on alignments from datasetB. The percentages correspond to the number of trees where the indicated metagrouping event was observed over the number of expected events. The names of the gene families that appeared clustered are indicated on the corresponding bars. The mean values of the phylogenetic scores supporting each MG event are given on the right.

The various evolutionary models for the composition of the ancestral ProtoHOX cluster

The finding of *Hox* and *ParaHox* cognate genes in Cnidaria but not in Porifera (Larroux et al., 2007) supports the emergence of the ProtoHOX cluster after the divergence of Porifera and its segmental tandem duplication that led to the formation of the HOX and ParaHOX clusters in the Cnidaria–Bilateria common ancestor. An open question relates to the configuration of the ProtoHOX cluster: How many and what genes were originally included in this cluster ? Four hypothesis of *H*/*P* evolution were proposed since the discovery of the ParaHOX complex in amphioxus (Brooke et al., 1998; Garcia-Fernandez, 2005b). In *the A-3-C-P four-gene model*, the ancestral ProtoHOX cluster is composed by four genes, anterior, group3, central, posterior as founders of the main paralog groups (Brooke et al., 1998). This hypothesis is supported by the joint occurrence of anterior, intermediate (PG3, central) and posterior *Hox* genes in most bilaterians (Ferrier and Holland, 2001a); it assumes that the lack of the central *ParaHox* group in extant Metazoa would be due to an early loss from the primitive ParaHOX cluster and that cnidarians would have lost independently both PG3 and central genes. *The A-3-P three-gene model* is based on evolutionary parsimony (Finnerty and Martindale, 1999; Ferrier and Holland, 2001a): The apparent absence of central *ParaHox genes* in bilaterians and cnidarians, coupled with the lack of central *Hox* genes in cnidarians, suggested a simpler, three-gene composition of the original ProtoHOX cluster, including representatives of the anterior, PG3 and posterior genes. This model assumes that cnidarians, following the ProtoHOX cluster duplication, lost their *Hox* PG3 and the corresponding *ParaHox* (i.e., *Pdx*/*Xlox*). The radiation of Bilateria was paralleled by the origin of central paralogs by tandem duplication on the HOX cluster. *The A-P two-gene model* assumes that the ProtoHOX cluster contained only two Hox-related genes, one anterior and one posterior. According to this model, the absence of intermediate *Hox*/*ParaHox* genes in cnidarians is a plesiomorphic character that should be interpreted as an evidence of ancient divergence of cnidarians from bilaterians. After the ancestral ProtoHOX cluster duplication, the radiation of Bilateria was linked to the origin of PG3 on both HOX and ParaHOX clusters and PG4 to PG8 on the HOX cluster by independent tandem duplication from the anterior one (Baguna and Riutort, 2004; Garcia-Fernandez, 2005b). The recent

A-3 two-gene model takes into account the lack of posterior orthologs in cnidarians as well as the faster evolutionary rate of the posterior *Hox* paralogs compared to the anterior or central ones in bilaterians (Chourrout et al., 2006). Therefore this model also proposes that only two genes were present in the ProtoHOX cluster, one anterior and one PG3. Following trans-duplication in HOX and ParaHOX clusters, the



non-anterior genes would have appeared by separate tandem duplication events on each of the two clusters, independently in cnidarians and bilaterians. Finally, Ryan et al. (2007) dispute the twin cluster model, arguing that the HOX and ParaHOX clusters arose as the split of a cluster formed via repeated tandem duplications of individual genes rather than via the duplication of a ProtoHOX cluster. This hypothesis is based on the analysis of *Nematostella*, amphioxus and *Drosophila* sequences, from which they concluded that *PG3* representatives were lost in cnidarians and that *Gsx* formed as an independent lineage, as they observed the absence of grouping between *PG1*, *PG2* and *Gsx*. However the identification of genomic regions in *Nematostella* that are syntenic with the human *H*/*P* sequences clearly favors the hypothesis of a duplication that gave rise to the HOX and ParaHOX clusters prior to the divergence of Cnidaria (Hui et al., 2008).

The pristine composition of the ancestral Hox-like cluster

The results presented in this paper fully support a segmental tandem duplication of a ProtoHOX cluster but do not support any of the models proposed above as they provide two major changes in the evolutionary picture of the *H*/*P* genes: firstly the presence of *Pdx*/*Xlox* cnidarian orthologs, second the identification of a major and highly conserved metagroup, the *Gsx*/*Pdx*/*PG2*/*PG3*, which is clearly distinct from "anterior" and "posterior" *H*/*P* gene families, i.e. *PG1* and *Cdx*/*PG9* respectively. Moreover, this study proposes the *CnoxB* to *CnoxE* gene families as cnidarian representatives of the ancestral *Cdx*/*PG9* gene families. These new data favor two possible hypotheses.

In the first one, the *four-gene ProtoHOX cluster*, depicted in Fig. 7 (left), would originate from a *PG1–PG9*/*Cdx* ancestral *ProtoHox* gene, i.e. an *A*/*P* ancestor as previously postulated (Zhang and Nei, 1996). Hence, after three subsequent rounds of gene duplication, this protoHOX cluster would have contained, in addition to the *Evx*/*Mox* ancestor, one gene related to the "anterior" group providing a common *PG1* ancestor to cnidarians and bilaterians (although with no counterpart in the ParaHOX cluster), a second one related to group 2 (*PG2*/*Gsx*), a third one related to the group 3 (*PG3*/*Pdx*/*Xlox*) and a fourth one related to the posterior group (*PG9*/*Cdx*). The tandem duplication of this ProtoHOX cluster in the common cnidarian–bilaterian ancestor would have led to the formation of two highly similar clusters, the ParaHOX and HOX clusters.

However, we favor a second more parsimonious alternative scenario, the *three-gene ProtoHOX cluster* (Fig. 7, right), where the ancestral *ProtoHox* gene would have rather resembled the *Gsx*/*Pdx*/*PG2*/*PG3* gene families instead of *A*/*P* ancestors, the posterior ancestors being generated by *cis*-duplication at the subsequent stage either from the *Evx*/*Mox* ancestor, or from the *Gsx*/*Pdx*/*PG2*/*PG3* ancestor. Similarly, in the absence of any "anterior" representative in the ParaHOX cluster, it is tempting to speculate that the PG1 ancestor might have actually been absent from the ProtoHOX cluster, arising later onto the primordial HOX-cluster from the *PG2 cis*-duplication. As a consequence the ancestral ProtoHOX cluster would have contained only three genes in addition to the *Evx*/*Mox* ancestor, *Gsx*/*PG2*, *Pdx*/*PG3* and *Cdx*/*PG9*. The three-gene ProtoHOX cluster hypothesis is more parsimonious as it requires only three steps from the *Evx*/*Mox* ancestor up to the segmental duplication, whereas the four-gene ProtoHOX cluster hypothesis requires four steps. In both scenarios an additional step is subsequently needed, *cis*-duplication of the *PG1* ancestor in the three-gene ProtoHOX model, deletion of the *PG1*-like ancestor from the ParaHOX cluster in the four-gene ProtoHOX model (Fig. 7).

The subsequent evolution of these two primordial clusters would be identical whatever the initial hypothesis: the *PG3* got likely lost from the HOX cluster in the early evolution of Cnidaria but was maintained in the radiation of Bilateria, where submitted to an early wave of tandem duplications, it produced the central genes (*PG4* to *PG8*). In cnidarians the HOX and ParaHOX clusters underwent some parallel evolution: their clustered organisation became highly disintegrated (Chourrout et al., 2006; Kamm et al., 2006) and the "posterior" genes highly derived compared to the other paralogous groups, submitted to several duplication events, leading to the formation of the *CnoxB*, *CnoxC*, *CnoxD*, *CnoxE* families. In contrast, the posterior genes were maintained in Bilateria, the *PG9* ancestor being duplicated independently in Lophotrochozoa (*post1*, *post2*) and Chordata (*PG9–PG15*).

The firstly evolved H/*P genes rather supported cell differentiation novelties than axis specification*

Being likely absent from poriferans (Larroux et al., 2007), we assume that *H*/*P* genes arose early in eumetazoan evolution to be recruited into regulatory networks that allowed the differentiation of more complex anatomies. The expression patterns presented here (Fig. 2) are consistent with an early developmental function for *Pdx*, *Gsx* and *CnoxC* in *Clytia*: *Gsx* was detected first in gastrulae, present in endodermal cells of the posterior half. One day after fertilization, both *Gsx* and *Pdx* displayed a transient punctuated pattern, which transformed into a diffuse staining of the endoderm, more intense in case of *Gsx*. In *Podocoryne*, *Gsx* exhibits a very similar pattern except the earliest transient wave of expression at the posterior pole of gastrulae that was not reported (Yanze et al., 2001). In contrast *CnoxC* expression pattern was clearly different, with *CnoxC* transcripts first localized at the anterior pole in 1 day old planula before extending subsequently all along the axis but leaving free the posterior pole. Further studies should tell us whether *Pdx* and *Gsx* are chromosomally clustered in *Clytia*, potentially sharing some regulatory region that would be reminiscent of those controlling temporal colinearity in vertebrate Hox genes (Kmita and Duboule, 2003).

The currently available data indicate that the *H*/*P* genes at the time cnidarians diverged, were already involved in cell differentiation as myogenesis (Aerne et al., 1995; Yanze et al., 1999) and neurogenesis (Miljkovic-Licina et al., 2007) and recruited for some developmental processes as apical patterning, but not yet following the rules that lead to body axis specification in bilaterians (Gauchat et al., 2000; Kamm et al., 2006; Chourrout et al., 2006). Interestingly, the *Gsx*/*cnox2*/*Ind* gene family is likely involved in neurogenesis from cnidarians (Hayward et al., 2001; Miljkovic-Licina et al., 2007) to bilaterians, i.e. *Drosophila* and mouse (Weiss et al., 1998; Toresson and Campbell, 2001; Yun et al., 2003). Similarly we expect some key cellular function for the cnidarian *Pdx*/*Xlox*, possibly restricted to jellyfish anatomy. In fact, the functional and expression analyses performed in developing vertebrates and annelids highlighted a role for *Pdx*/*Xlox*/*IPF* and *Cdx* genes in cell and tissue differentiation rather than axis specification (Wysocka-Diller et al., 1995; Offield et al., 1996; Milewski et al., 1998; Melloul, 2004; Frobius and Seaver, 2006; Young and Deschamps, in press). Hence the stabilization of *ParaHox* genes in early animal

Fig. 7. A three-gene ProtoHOX cluster derived from a *Gsx*/*PG2-Pdx*/*PG3 ProtoHox* gene as the most parsimonious model to describe the early evolution of the *H*/*P* gene families. In the absence of *H*/*P* genes in poriferans, *ProtoHox* genes likely appeared after Porifera divergence as a result of a *cis*-duplication event of a *non-Hox* ANTP-class gene, possibly an *Evx*/*Mox* ancestor gene. According to the nature of this ancestral *ProtoHox* gene, we describe two possible scenarios. In the first case (left), the repeated *cis*-duplication from a *PG1*/*Cdx-PG9 ProtoHox* gene led to the formation of a four-gene ProtoHOX cluster containing *PG1*, *Gsx*/*PG2*, *Pdx*/*PG3* and *Cdx*/*PG9*. In the second case (right), three paralogs, corresponding to *Gsx*/*PG2*, *Pdx*/*PG3* and *Cdx*/*PG9*, arose from a *Gsx*/*PG2-Pdx*/*PG3 ProtoHox* gene. Subsequently the segmental tandem duplication of this ProtoHOX cluster led to the formation of the primordial ParaHOX and HOX clusters. The absence of *PG1*-related sequence among *ParaHox* genes can be interpreted in two ways, either secondarily lost (four-gene ProtoHOX model) or never present in the primordial ParaHOX cluster (three-gene ProtoHOX model). This work supports *CnoxB*, *CnoxC*, *CnoxD*, *CnoxE* gene families as derived from *Cdx*/*PG9*, and *Cnox4-Ed* gene as a possible *Cdx* ortholog. In cnidarians, the HOX and ParaHOX clusters got disintegrated while remaining intact in numerous bilaterian species. The emergence of the *PG4* to *PG8* paralogs is likely a more recent event that occurred after Cnidaria divergence.

evolution would have coincided with the establishment of evolutionarily-conserved regulatory networks driving cellular novelties such as neurogenesis or myogenesis, identified first in cnidarians and maintained in bilaterians.

Acknowledgments

We thank Brigitte Aeschbach and Elisa Giangrande for excellent technical assistance, Evelyn Houliston for helping us with the *Clytia* culture and providing *Clytia* embryos, Cédric Berney and Bastien Chopard for advices about phylogenetic analyses, Juan Montoya-Burgos for critical reading and discussions. The authors are grateful to Walter Gehring for kindly allowing the use of the CnoxA-Cr sequence from an ongoing *Cladonema* EST project, to the National Institute of Genetics (Mishima, Japan) and the Laboratory for DNA Data Analysis who carried out the *Cladonema* EST project under the direction of Takashi Gojobori. BG's laboratory is supported by the Swiss National Foundation, Canton of Geneva, Claraz Donation and Geneva Academic Society; SP's laboratory is supported by the European Commission (VthFP, Marie Curie Fellowship Program, contract HPMD-CT-2001-00099), MIUR (COFIN and FIRB projects), Administration of Lecce Province, CONISMA, and MarBEF European Network.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2009.01.022.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.
Aerne, B.L., Baader, C.D., Schmid, V., 1995. Life stage and tissue-specific expression of the homeobox gene cnox1-Pc of the hydrozoan *Podocoryne carnea*. Dev. Biol. 169, 547–556.
Baguna, J., Riutort, M., 2004. The dawn of bilaterian animals: the case of acoelomorph flatworms. BioEssays 26, 1046–1057.
Ball, E.E., Hayward, D.C., Saint, R., Miller, D.J., 2004. A simple plan—cnidarians and the origins of developmental mechanisms. Nat. Rev. Genet. 5, 567–577.
Barucca, M., Biscotti, M.A., Olmo, E., Canapa, A., 2006. All the three ParaHox genes are present in *Nuttallochiton mirandus* (Mollusca: polyplacophora): evolutionary considerations. J. Exp. Zool. B Mol. Dev. Evol. 306, 164–167.
Bode, H., Martinez, D., Shenk, M.A., Smith, K., Steele, R., Technau, U., 1999. Evolution of head development. Biol. Bull. 196, 408–410.
Brooke, N.M., Garcia-Fernandez, J., Holland, P.W., 1998. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature 392, 920–922.
Canapa, A., Biscotti, M.A., Olmo, E., Barucca, M., 2005. Isolation of Hox and ParaHox genes in the bivalve *Pecten maximus*. Gene 348, 83–88.
Carre, D., Carre, C., 2000. Origin of germ cells, sex determination, and sex inversion in medusae of the genus *Clytia* (Hydrozoa, leptomedusae): the influence of temperature. J. Exp. Zool. 287, 233–242.
Chevalier, S., Martin, A., Leclere, L., Amiel, A., Houliston, E., 2006. Polarised expression of FoxB and FoxQ2 genes during development of the hydrozoan *Clytia hemisphaerica*. Dev. Genes Evol. 216, 709–720.
Chourrout, D., Delsuc, F., Chourrout, P., Edvardsen, R.B., Rentzsch, F., Renfer, E., Jensen, M.F., et al., 2006. Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. Nature 442, 684–687.
Duboule, D., 2007. The rise and fall of Hox gene clusters. Development 134, 2549–2560.
Ferrier, D.E., Holland, P.W., 2001a. Ancient origin of the Hox gene cluster. Nat. Rev. Genet. 2, 33–38.
Ferrier, D.E., Holland, P.W., 2001b. Sipunculan ParaHox genes. Evol. Dev. 3, 263–270.
Ferrier, D.E., Holland, P.W., 2002. Ciona intestinalis ParaHox genes: evolution of Hox/ParaHox cluster integrity, developmental mode, and temporal colinearity. Mol. Phylogenet. Evol. 24, 412–417.
Finnerty, J.R., Martindale, M.Q., 1999. Ancient origins of axial patterning genes: Hox genes and ParaHox genes in the Cnidaria. Evol. Dev. 1, 16–23.
Finnerty, J.R., Pang, K., Burton, P., Paulson, D., Martindale, M.Q., 2004. Origins of bilateral symmetry: Hox and dpp expression in a sea anemone. Science 304, 1335–1337.
Frobius, A.C., Seaver, E.C., 2006. ParaHox gene expression in the polychaete annelid *Capitella* sp. I. Dev. Genes Evol. 216, 81–88.
Galliot, B., Miller, D., 2000. Origin of anterior patterning. How old is our head? Trends Genet. 16, 1–5.
Garcia-Fernandez, J., 2005a. Hox, ParaHox, ProtoHox: facts and guesses. Heredity 94, 145–152.
Garcia-Fernandez, J., 2005b. The genesis and evolution of homeobox gene clusters. Nat. Rev. Genet. 6, 881–892.
Gauchat, D., Mazet, F., Berney, C., Schummer, M., Kreger, S., Pawlowski, J., Galliot, B., 2000. 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.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.
Hayward, D.C., Catmull, J., Reece-Hoyes, J.S., Berghammer, H., Dodd, H., Hann, S.J., Miller, D.J., et al., 2001. Gene structure and larval expression of cnox-2Am from the coral *Acropora millepora*. Dev. Genes Evol. 211, 10–19.
Hobmayer, B., Rentzsch, F., Kuhn, K., Happel, C.M., von Laue, C.C., Snyder, P., Rothbacher, U., et al., 2000. WNT signalling molecules act in axis formation in the diploblastic metazoan Hydra. Nature 407, 186–189.
Hui, J.H., Holland, P.W., Ferrier, D.E., 2008. Do cnidarians have a ParaHox cluster? Analysis of synteny around a *Nematostella* homeobox gene cluster. Evol. Dev. 10, 725–730.
Hwang, S.P., Wu, J.Y., Chen, C.A., Hui, C.F., Chen, C.P., 2003. Novel pattern of AtXlox gene expression in starfish *Archaster typicus* embryos. Dev. Growth Differ. 45, 85–93.
Kamm, K., Schierwater, B., 2006. Ancient complexity of the non-Hox ANTP gene complement in the anthozoan *Nematostella vectensis*: implications for the evolution of the ANTP superclass. J. Exp. Zool. B Mol. Dev. Evol. 306, 589–596.
Kamm, K., Schierwater, B., Jakob, W., Dellaporta, S.L., Miller, D.J., 2006. Axial patterning and diversification in the cnidaria predate the Hox system. Curr. Biol. 16, 920–926.
Kmita, M., Duboule, D., 2003. Organizing axes in time and space; 25 years of colinear tinkering. Science 301, 331–333.
Kozmik, Z., Daube, M., Frei, E., Norman, B., Kos, L., Dishaw, L.J., Noll, M., et al., 2003. Role of Pax genes in eye evolution: a cnidarian PaxB gene uniting Pax2 and Pax6 functions. Dev. Cell 5, 773–785.
Larroux, C., Fahey, B., Degnan, S.M., Adamski, M., Rokhsar, D.S., Degnan, B.M., 2007. The NK homeobox gene cluster predates the origin of Hox genes. Curr. Biol. 17, 706–710.
Lee, P.N., Pang, K., Matus, D.Q., Martindale, M.Q., 2006. A WNT of things to come: evolution of Wnt signaling and polarity in cnidarians. Semin. Cell Dev. Biol. 17, 157–167.
Masuda-Nakagawa, L.M., Gröger, H., Aerne, B.L., Schmid, V., 2000. The HOX-like gene Cnox2-Pc is expressed at the anterior region in all life cycle stages of the jellyfish *Podocoryne carnea*. Dev. Genes Evol. 210, 151–156.
McGinnis, W., Krumlauf, R., 1992. Homeobox genes and axial patterning. Cell 68, 283–302.
Melloul, D., 2004. Transcription factors in islet development and physiology: role of PDX-1 in beta-cell function. Ann. N.Y. Acad. Sci. 1014, 28–37.
Milewski, W.M., Duguay, S.J., Chan, S.J., Steiner, D.F., 1998. Conservation of PDX-1 structure, function, and expression in zebrafish. Endocrinology 139, 1440–1449.
Miljkovic-Licina, M., Chera, S., Ghila, L., Galliot, B., 2007. Head regeneration in wild-type hydra requires de novo neurogenesis. Development 134, 1191–1201.
Miller, D.J., Miles, A., 1993. Homeobox genes and the zootype [letter]. Nature 365, 215–216.
Minguillon, C., Garcia-Fernandez, J., 2003. Genesis and evolution of the Evx and Mox genes and the extended Hox and ParaHox gene clusters. Genome Biol. 4, R12.
Momose, T., Houliston, E., 2007. Two oppositely localised frizzled RNAs as axis determinants in a cnidarian embryo. PLoS Biol. 5, e70.
Murtha, M.T., Leckman, J.F., Ruddle, F.H., 1991. Detection of homeobox genes in development and evolution. Proc. Nat. Acad. Sci. U. S. A. 88, 10711–10715.
Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205–217.
Offield, M.F., Jetton, T.L., Labosky, P.A., Ray, M., Stein, R.W., Magnuson, M.A., Hogan, B.L., et al., 1996. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Development 122, 983–995.
Park, B.J., Cho, S.J., Tak, E.S., Lee, B.E., Park, S.C., 2006. The existence of all three ParaHox genes in the clitellate annelid, *Perionyx excavatus*. Dev. Genes Evol. 216, 551–553.
Peterson, K.J., 2004. Isolation of Hox and Parahox genes in the hemichordate *Ptychodera flava* and the evolution of deuterostome Hox genes. Mol. Phylogenet. Evol. 31, 1208–1215.
Piraino, S., Boero, F., Aeschbach, B., Schmid, V., 1996. Reversing the life cycle: Medusae transforming into polyps and cell transdifferentiation in *Turritopsis nutricula* (Cnidaria, Hydrozoa). Biol. Bull. 190, 302–312.
Piraino, S., De Vito, D., Schmich, J., Bouillon, J., Boero, F., 2004. Reverse development in cnidarians. Can. J. Zool. 82, 1748–1754.
Rentzsch, F., Anton, R., Saina, M., Hammerschmidt, M., Holstein, T.W., Technau, U., 2006. Asymmetric expression of the BMP antagonists chordin and gremlin in the sea anemone *Nematostella vectensis*: implications for the evolution of axial patterning. Dev. Biol. 296, 375–387.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Ryan, J.F., Burton, P.M., Mazza, M.E., Kwong, G.K., Mullikin, J.C., Finnerty, J.R., 2006. The cnidarian–bilaterian ancestor possessed at least 56 homeoboxes. Evidence from the starlet sea anemone, *Nematostella vectensis*. Genome Biol. 7, R64.
Ryan, J.F., Mazza, M.E., Pang, K., Matus, D.Q., Baxevanis, A.D., Martindale, M.Q., 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.
Schmich, J., Kraus, Y., De Vito, D., Graziussi, D., Boero, F., Piraino, S., 2007. Induction of reverse development in two marine Hydrozoans. Int. J. Dev. Biol. 51, 45–56.
Seipel, K., Schmid, V., 2005. Evolution of striated muscle: jellyfish and the origin of triploblasty. Dev. Biol. 282, 14–26.
Seipel, K., Schmid, V., 2006. Mesodermal anatomies in cnidarian polyps and medusae. Int. J. Dev. Biol. 50, 589–599.
Seo, H.C., Edvardsen, R.B., Maeland, A.D., Bjordal, M., Jensen, M.F., Hansen, A., Flaat, M., et al., 2004. Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. Nature 431, 67–71.

Slack, J.M., 1995. Developmental biology of the pancreas. Development 121, 1569–1580.
Stierwald, M., Yanze, N., Bamert, R.P., Kammermeier, L., Schmid, V., 2004. The *Sine oculis*/Six class family of homeobox genes in jellyfish with and without eyes: development and eye regeneration. Dev. Biol. 274, 70–81.
Suga, H., Schmid, V., Gehring, W.J., 2008. Evolution and functional diversity of jellyfish opsins. Curr. Biol. 18, 51–55.
Toresson, H., Campbell, K., 2001. A role for Gsh1 in the developing striatum and olfactory bulb of Gsh2 mutant mice. Development 128, 4769–4780.
Wagner, G.P., Amemiya, C., Ruddle, F., 2003. Hox cluster duplications and the opportunity for evolutionary novelties. Proc. Natl. Acad. Sci. U. S. A. 100, 14603–14606.
Wallberg, A., Thollesson, M., Farris, J., Jondelius, U., 2004. The phylogenetic position of the comb jellies (Ctenophora) and the importance of taxonomic sampling. Cladistics 20, 558–578.
Weiss, J.B., Von Ohlen, T., Mellerick, D.M., Dressler, G., Doe, C.Q., Scott, M.P., 1998. Dorsoventral patterning in the *Drosophila* central nervous system: the intermediate neuroblasts defective homeobox gene specifies intermediate column identity. Genes Dev. 12, 3591–3602.
Wysocka-Diller, J., Aisemberg, G.O., Macagno, E.R., 1995. A novel homeobox cluster expressed in repeated structures of the midgut. Dev. Biol. 171, 439–447.
Yanze, N., Groger, H., Muller, P., Schmid, V., 1999. Reversible inactivation of cell-type-specific regulatory and structural genes in migrating isolated striated muscle cells of jellyfish. Dev. Biol. 213, 194–201.
Yanze, N., Spring, J., Schmidli, C., Schmid, V., 2001. Conservation of Hox/ParaHox-related genes in the early development of a cnidarian. Dev. Biol. 236, 89–98.
Young, T., Deschamps, J., in press. Hox, Cdx and antero-posterior patterning in the mouse embryo. Curr. Top. Dev. Biol.
Yun, K., Garel, S., Fischman, S., Rubenstein, J.L., 2003. Patterning of the lateral ganglionic eminence by the Gsh1 and Gsh2 homeobox genes regulates striatal and olfactory bulb histogenesis and the growth of axons through the basal ganglia. J. Comp. Neurol. 461, 151–165.
Zhang, J., Nei, M., 1996. Evolution of Antennapedia-class homeobox genes. Genetics 142, 295–303.