Available online at www.sciencedirect.com

SCIENCE DIRECT®

Molecular Phylogenetics and Evolution 39 (2006) 468–477

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

Expansion of the SOX gene family predated the emergence of the Bilateria

Muriel Jager [a], Eric Quéinnec [a], Evelyn Houliston [b], Michaël Manuel [a,*]

[a] *Équipe Évolution et Développement, UMR 7138 "Systématique, Adaptation, Evolution" CNRS/UMPC/MNHN/IRD/ENS Bâtiment B, 7ième étage, Université P et M Curie, 9 quai saint Bernard, 75005 Paris, France*

[b] *UMR 7009 "Laboratoire de Biologie du Développement" Observatoire océanologique de Villefranche-sur-mer, CNRS/Université P et M Curie, 06230 Villefranche-sur-mer, France*

Received 17 August 2005; revised 15 November 2005; accepted 9 December 2005
Available online 24 January 2006

Abstract

Members of the SOX gene family are involved in regulating many developmental processes including neuronal determination and differentiation, and in carcinogenesis. So far they have only been identified in species from the Bilateria (deuterostomes and protostomes). To understand the origins of the SOX family, we used a PCR-based strategy to obtain 28 new sequences of SOX gene HMG domains from four non-bilaterian Metazoa: two sponge species, one ctenophore and one cnidarian. One additional SOX sequence was retrieved from EST sequences of the cnidarian species *Clytia hemisphaerica.* Unexpected SOX gene diversity was found in these species, especially in the cnidarian and the ctenophore. The topology of gene relationships deduced by Maximum Likelihood analysis, although not supported by bootstrap values, suggested that the SOX family started to diversify in the metazoan stem branch prior to the divergence of demosponges, and that further diversification occurred in the eumetazoan branch, as well as later in calcisponges, ctenophores, cnidarians and vertebrates. In contrast, gene loss appears to have occurred in the nematode and probably in other protostome lineages, explaining their lower number of SOX genes.

Keywords: Cnidaria; Ctenophora; Development; Duplication; Evolution; HMG domain; Multigenic family; Nervous system; Phylogeny; Porifera; SOX genes

1. Introduction

The regulation of developmental processes in metazoans involves a great number of multigenic families, including a diversity of transcription factors families (such as homeodomain—HD—or T-domain containing proteins and bHLH proteins), various families of signalling secreted proteins (e.g., Wnt family, TGF-β), and membrane receptors (e.g., Notch). Strikingly, gene diversity in most of these families is specific to the Metazoa, with homologous proteins from the few available complete genomes of non-metazoan eukaryotes (notably in the taxa Fungi and Plantae) being either weakly diversified or having diversified independently (Gauchat et al., 2000; Ledent et al., 2002; Prud'homme et al., 2002), or being absent (Manuel et al., 2004). It has remained unclear until recently, however, whether multigenic families were already diversified in the last common metazoan ancestor, or whether they diversified in the bilaterian branch, where a marked increase in morphological complexity occurred.

Recent studies have shown that homologues of many regulatory genes acting in the development of the Bilateria are also present in cnidarians (Finnerty and Martindale, 1999; Gauchat et al., 2000; Yates, 2000), some of them also occurring in sponges, the basal-most living metazoans (Adell et al., 2003; Hoshiyama et al., 1998; Manuel and Le Parco, 2000; Manuel et al., 2004; Seimiya et al., 1994). Furthermore, analyses of EST datasets from cnidarian species (Kortschak et al., 2003; Yang et al., 2004) have revealed

[*] Corresponding author. Fax: +33 1 44 27 32 58.
E-mail address: michael.manuel@snv.jussieu.fr (M. Manuel).

1055-7903/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2005.12.005

unexpected genetic complexity, including the presence of a number of orthologues of genes so far known only from vertebrate species (and absent from the genomes of fly and nematode). These studies indicate that the molecular toolkit for development of a multicellular animal has an ancient origin, in some cases dating back to the common branch of Metazoa, and that acquisition of multicellularity was linked to genetic diversification.

Existing patterns of gene diversity result from a complex series of evolutionary events including both duplications in particular lineages and gene losses. Examples of duplications are well documented at various levels of the phylogeny, including within particular phyla, for example in the vertebrates (e.g., Hugues, 1999; Ledent et al., 2002), and in sponges (Manuel and Le Parco, 2000). Gene losses have been described for example in the bHLH family (Ledent et al., 2002), in the Wnt family (Prud'homme et al., 2002), in nuclear receptors (Bertrand et al., 2004) and from analyses of complete genome data (Krylov et al., 2003); with the genome of *Caenorhabditis elegans* offering an extreme example (Babenko and Krylov, 2004; Blaxter, 1998). Shared gene losses have started to be used as phylogenetic characters for reconstructing the phylogeny of Bilateria (Copley et al., 2004).

The SOX family is one of the many multigenic families acting in the regulation of development in the Bilateria. Its founding member was SRY, the mammalian testis determining factor (Gubbay et al., 1990). SOX proteins are transcription factors that contain a DNA-binding domain known as the HMG (high-mobility group) domain, comprising 79 amino acids (Gubbay et al., 1990). This domain is highly conserved in all SOX proteins (Wegner, 1999). The SOX family is one of many families within the HMG domain superfamily, others being the TCF and MATA groups (which, like SOX proteins, contain only one HMG domain), and the HMG/UBF group (which contain multiple HMG domains) (Laudet et al., 1993; Soullier et al., 1999). Outside the HMG domain, SOX sequences are variable, although some other conserved domain can be identified among particular SOX groups (Bowles et al., 2000; Uchikawa et al., 1999). In vertebrates and in *Drosophila*, SOX proteins are involved in a variety of developmental processes such as germ layer formation, organ development, and cell type specification. Some of them are also implied in carcinogenesis (Dong et al., 2004). They are expressed in most tissues and cell types during at least one stage of development in vertebrates (Wegner, 1999). A subset of SOX genes is implicated in neural development in Bilateria (Wegner, 1999) including *Xenopus* (Sasai, 2001) and *Drosophila* (Buescher et al., 2002; Nambu and Nambu, 1996; Overton et al., 2002; Sanchez-Soriano and Russell, 1998), and the mollusc *Patella* (Le Gouar et al., 2004). The role of some SOX genes in neural development could thus be inherited from an ancestor of the Bilateria.

Until now, SOX genes have been identified and studied only in bilaterian lineages. Complete SOX genes repertories are available from the completely sequenced genomes of chordates (*Ciona intestinalis*, Leveugle et al., 2004; *Fugu rubripes*, Koopman et al., 2004; *Homo sapiens*, *Mus musculus*, Shepers et al., 2002) and non-chordates (*Caenorhabditis elegans* and *Drosophila melanogaster*). Based on analyses of the primary sequence and structural indicators as intron–exon organisation, the SOX family has been divided into eight groups (A–H) (Bowles et al., 2000; Koopman et al., 2004). Only a single gene was found for most groups in the invertebrates (*C. elegans*, *D. melanogaster*, and *C. intestinalis*), in contrast to multiple genes in the vertebrates, leading to the suggestion that SOX genes diversified during vertebrate evolution and genome expansion (Koopman et al., 2004).

In the present study, we looked for homologues of SOX genes in four species of non-bilaterian metazoans to gain insights about the early evolution of the family. We chose two sponge species, *Sycon raphanus* (Calcispongia) and *Ephydatia muelleri* (Demospongiae), one ctenophore (*Pleurobrachia pileus*) and one hydrozoan cnidarian (*Clytia hemisphaerica*). The respective phylogenetic position of Cnidaria and Ctenophora remains debated but it is clear that they both belong to the Eumetazoa (the clade comprising all non-sponge metazoans), and are excluded from the Bilateria (Collins, 1998; Kim et al., 1999; Manuel et al., 2003; Zrzavy et al., 1998). There is still uncertainty about the phylogenetic status of sponge but ribosomal RNA phylogenies most frequently branch the calcareous sponges (Calcispongia) as the sister-group of Eumetazoa, rather than of the other sponges (Cavalier-Smith et al., 1996; Manuel et al., 2003; Medina et al., 2001; Zrzavy et al., 1998). Among many morphological characters, the eumetazoans are distinguished from sponges by their possession of nerve cells and a nervous system. Since the SOX genes are candidates to be involved in the origin of the nervous system, we have sought to understand the evolution of gene diversity in this multigenic family at the base of the metazoan tree. Degenerate oligonucleotide primers were used to amplify multiple partial HMG box sequences from the four species, and sequences were compared to available data from the Bilateria through phylogenetic analyses, suggesting a scenario for the molecular evolution of the SOX family in the Metazoa, which involves early gene duplications prior to the bilaterian split.

2. Materials and methods

2.1. Specimen collection and nucleic acid extraction

Species names, collecting locations, and starting material for the four sampled taxa are listed in Table 1. All material was carefully inspected to avoid contamination from other metazoan species. *Clytia hemisphaerica* polyps and medusas are cultured in the laboratory, and embryos are produced by in vitro fertilization, reducing the risks of contamination. Adults and embryos from *Pleurobrachia pileus* and *Clytia hemisphaerica* were let unfed for at least 24 h to allow prey digestion, and were carefully observed under a

Table 1
Informations concerning the organisms used in this study

Species name	Phylum	Collecting location	Starting material
Pleurobrachia pileus	Ctenophora	Villefranche sur mer, France	Adults, gastrulae, cyddipids
Clytia hemisphaerica	Cnidaria	Villefranche sur mer, France	Adults, planulae, cDNA library (adults and embryonic stages)
Ephydatia muelleri	Demospongiae (Porifera)	Belgium	Young adult sponges after gemmulation
Sycon raphanus	Calcispongia (Porifera)	Marseille, France	cDNA library (adults + embryonic stages)

microscope, then washed thoroughly with sterile sea water before RNA extraction. *Ephydatia muelleri* gemmules were cultured in sterile freshwater. RNA extraction and PCR amplification were performed in an isolated room, equipped with UV light.

Total RNA was extracted from samples ground to powder in liquid nitrogen, using the "RNeasy Mini Kit" (Qiagen). Reverse transcription of cDNA from total RNA extracts (5–10 μg) was performed using MMLV-RT (RT-P CR kit, Stratagen) and an oligo(dT) primer (5′ GAGAG AACTAGTCTCGAGT(x18) 3′). *Sycon raphanus* adult sponges, containing embryos, were prepared for cDNA library construction as explained in Kruse et al. (1997).

2.2. PCR amplification, cloning and sequencing

Two sets of degenerate oligonucleotide primers were designed for PCR amplification of the HMG box of SOX genes on the basis of an amino-acid alignment of SOX HMG domain sequences from various Bilateria. The primers were chosen from regions highly conserved between SOX genes and divergent in other HMG domains. First round amplification (40 cycles, annealing temperature 45 °C) used SOX1 (5′ MGNCCNATGAAYGCNT TYATG 3′), and SOX1rev (5′ TTNCKNCKNGGNCK RTAYTT 3′), corresponding to amino acid sequences RPMNAFM and KYRPRR, respectively. This primer set amplifies a fragment of 218 bp, containing 177 bp (59 aa) of informative sequence, corresponding to aa 12–70 of the HMG domain. A second round of nested PCR amplification (30 cycles, annealing temperature 50 °C) used SOX2 (5′ AAYCCNAARATGCAYAAYWSNGA 3′) and SOX2rev (5′ TARTCNGGRTGYTCYTTCATRTG 3′), corresponding to the amino acid sequences NPKMHNSE and HMKEHPDY, respectively. This second primer set amplifies a fragment of 137 bp, containing 90 bp (30 aa) of informative sequence, corresponding to aa 33–62 of the HMG domain. PCR conditions were as described in Jager et al. (2003). To avoid cross contamination between the four species, each species was treated separately in the following order: *Pleurobrachia pileus*, *Clytia hemisphaerica*, *Sycon raphanus* and *Ephydatia muelleri*. Separate sets of PCR solutions were used for each species. A negative control (lacking DNA matrix) was performed during each PCR, to confirm the absence of contaminating DNA.

PCR products were cloned as described in Jager et al. (2003). Plasmids were sequenced in our laboratory on an ALF Express (Pharmacia) automated sequencer, or were sent to Genome Express (Grenoble, France) for sequencing. Each identified sequence was designated by a species-specific prefix (*Che* for *C. hemisphaerica*, *Ppi* for *P. pileus*, *Emu* for *E. muelleri*, *Sra* for *S. raphanus*), followed by SOX or HMG (for non-SOX HMG box sequences), and a number corresponding to the order of identification (which thus does not correspond to orthology) (Wegner, 1999). All new sequences have been deposited into the GenBank database (see Accession Nos. in Table S1, Supplementary materials).

2.3. EST data from clytia hemisphaerica

HMG domain containing sequences were also searched by Blast (tblastn) on an unpublished collection of 10,000 ESTs from *Clytia hemisphaerica*. The ESTs were sequenced by the Genoscope (Evry, France) from a cDNA library derived from a mixture of medusae, larvae and embryos constructed by BioSystems (in plasmid vector Express 1). The new SOX sequence from the ESTs (*CheSOX10*) has been deposited in the GenBank database (Table S1).

2.4. Phylogenetic analyses

In addition to sequences from this work, the alignment includes the SOX genes sampling of Koopman et al. (2004), to which we added SOX genes sequences from *Ciona intestinalis* (Leveugle et al., 2004), sequences from various invertebrates and sequences from *Hydra magnipapillata* ESTs, obtained by Blast search in the GenBank database and CnidBase (http://cnidbase.bu.edu/).

A representative sampling of non-SOX HMG box families (as outgroups) was selected from Soullier et al. (1999). We also included some sequences of the HMG domain of the CIC (capicua homolog) protein family from Leveugle et al. (2004), sequences from *Hydra magnipapillata* ESTs, and sequences from *Monosiga brevicollis* (Choanoflagellata) obtained from Blast search of Choanobase (http://mcb.berkeley.edu/labs/king/choano/). The alignment, made by eye with the Bioedit package, was unambiguous over the whole length of the HMG domain, and contained no gaps (Figure S1). The alignment comprises 59 positions, corresponding to the amino-acid sequence obtained for most genes from this work, except *SraHMG5*, *CheSOX8* and *CheSOX9* for which the sequences were only 30 aa long and the lacking residues were scored as missing data.

Phylogenetic analysis was carried out from the amino-acid alignment by the Maximum-Likelihood (ML) method using the PhyML program (Guindon and Gascuel, 2003), with the JTT (Jones et al., 1992) model of amino-acid substitution. A Neighbour-Joining tree was used as the input tree to generate the ML tree. A gamma distribution with four discrete categories was used in the ML analyses. The gamma shape parameter and the proportion of invariant sites were optimised during the ML search. The statistical significance of the nodes was assessed by bootstrapping (200 replicates).

3. Results

3.1. Multiple SOX genes in non-bilaterian metazoa

PCR fragments were amplified by using degenerate primers selected within the SOX gene HMG box from four non-bilaterian species (Table 1 and Table S1). After cloning and sequencing, 120 clones sequenced from ctenophore (*Pleurobrachia pileus*) yielded 13 distinct sequences named *PpiSOX1–PpiSOX13* (GenBank Accession Nos.: AY769217–AY769229). From the hydrozoan cnidarian *Clytia hemisphaerica*, 110 clones yielded nine distinct sequences (*CheSOX1–CheSOX9*, GenBank Accession Nos.: AY769230–AY769238). The analysis of a collection of 10,000 ESTs from *C. hemisphaerica* led us to identify an additional sequence (*CheSOX10*, GenBank Accession No.: DQ138605), and an EST identical to the *CheSOX5* sequence previously obtained by PCR. In *Ephydatia muelleri* (freshwater demosponge), 100 clones amplified from young adult cDNA revealed only three distinct sequences (*EmuSOX1–EmuSOX3*, GenBank Accession Nos.: AY769244–AY769246) and in the calcareous sponge *Sycon raphanus* 100 clones generated from an adult/embryos cDNA library yielded five distinct sequences (*SraSOX1–SraSOX3*, GenBank Accession Nos.: AY769239–AY769241; *SraHMG4* and *SraHMG5*, GenBank Accession Nos.: AY769243 and AY769242, respectively).

All sequences from this work have been translated and included in the alignment shown in Figure S1 (Supplementary material), together with SOX and non-SOX HMG domains from various bilaterians. In two cases, small groups of amino-acid sequences from a single species were highly similar (*PpiSOX2*, *PpiSOX12*, and *PpiSOX13*; *SraSOX1*, *SraSOX2*, and *SraSOX3*). We considered these to be true sequences and not the result of PCR or sequencing artefacts because of the extent of variation between the corresponding nucleotide sequences: there were 14 nucleic acid differences between *PpiSOX2* and *PpiSOX12*, 11 between *PpiSOX2* and *PpiSOX12*, and 13 between *PpiSOX12* and *PpiSOX13*. There were five nucleic acid differences between *SraSOX1* and *SraSOX2*, four nucleic acid differences between *SraSOX1* and *SraSOX3* and seven nucleic acid differences between *SraSOX2* and *SraSOX3*. Nevertheless, it is not possible to exclude completely the possibility that these sequences would result from PCR artefacts.

We cannot claim to have identified all the SOX genes in these animals, due to possible uneven representation of cDNAs in the starting materials and/or bias originating from the primer sets. It is also possible that some of the sequence differences in this region represent alleles or pseudogenes rather than distinct coding sequences, but this seems unlikely except possibly for the two sets of similar sequences discussed above.

In any case, the large number of sequences recovered in the cnidarian and ctenophore species was surprising, exceeding the number of SOX genes reported for *C. intestinalis*, *D. melanogaster* and *C. elegans*—6, 8 and 5 genes, respectively.

3.2. Phylogenetic analysis of the SOX gene family

The dataset comprising the new SOX sequences from non-Bilateria, and three additional sequences recovered from a Blast search in *Hydra magnipapillata* ESTs, as well as a number of available sequences from Bilateria, was analysed using Maximum Likelihood (ML), a model-based probabilistic method (see Section 2) (Fig. 1). The tree was rooted using a representative sampling of the various known non-SOX HMG domain families (Soullier et al., 1999). The availability of only the main part of the HMG domain for the four non-bilaterian species obviously limits the amount of phylogenetic information useful for such analyses; however, there are a high number of informative sites in the HMG domain (47 out of 59 positions in the alignment at the level of the SOX family). Soullier et al. (1999) have shown that the signal present inside the SOX gene HMG domain is strong enough to generate the same phylogenetic structure as that reconstructed from whole coding sequences. In consequence of the shortness and large number of sequences, the overall degree of robustness of the tree, judged from bootstrap (BP) values, is low, i.e., most of the nodes have a BP below 50%.

Our first goal with this analysis was to determine whether the new sequences from non-bilaterians belonged or not to the SOX family. We observed that the ML tree (Fig. 1) contains a clade (labelled “SOX Family” in the figure) including all of the SOX sequences from the Bilateria and 29 of our 31 new sequences. Within this clade, three genes from *S. raphanus* (*SraSOX1*, *2* and *3*) and one human gene (*HSASOX30*) fall in a basal and divergent position, so that their membership of the SOX family is questionable. Of the two remaining gene from *S. raphanus*, *SraHMG5* does not belong to this clade, and *SraHMG4* branches as its sister-group. When using alternative tree reconstruction methods (distance Neighbour-Joining, and Maximum Parsimony with a reduced dataset), or different samplings of outgroup sequences, this clade was always retrieved, but the position of *SraHMG4* was variable, branching in some analyses more distantly among the outgroup sequences (results not shown).

Careful inspection of the alignment (Figure S2) for diagnostic residues also supports the inclusion of genes *SraSOX1*, *2* and *3* from *Sycon raphanus*, but not *SraHMG4*, in the SOX family. Although no single residue (between aa 12 and 70) in the HMG domain is diagnostic of the SOX family, the combination V(12), W(13), H(29),

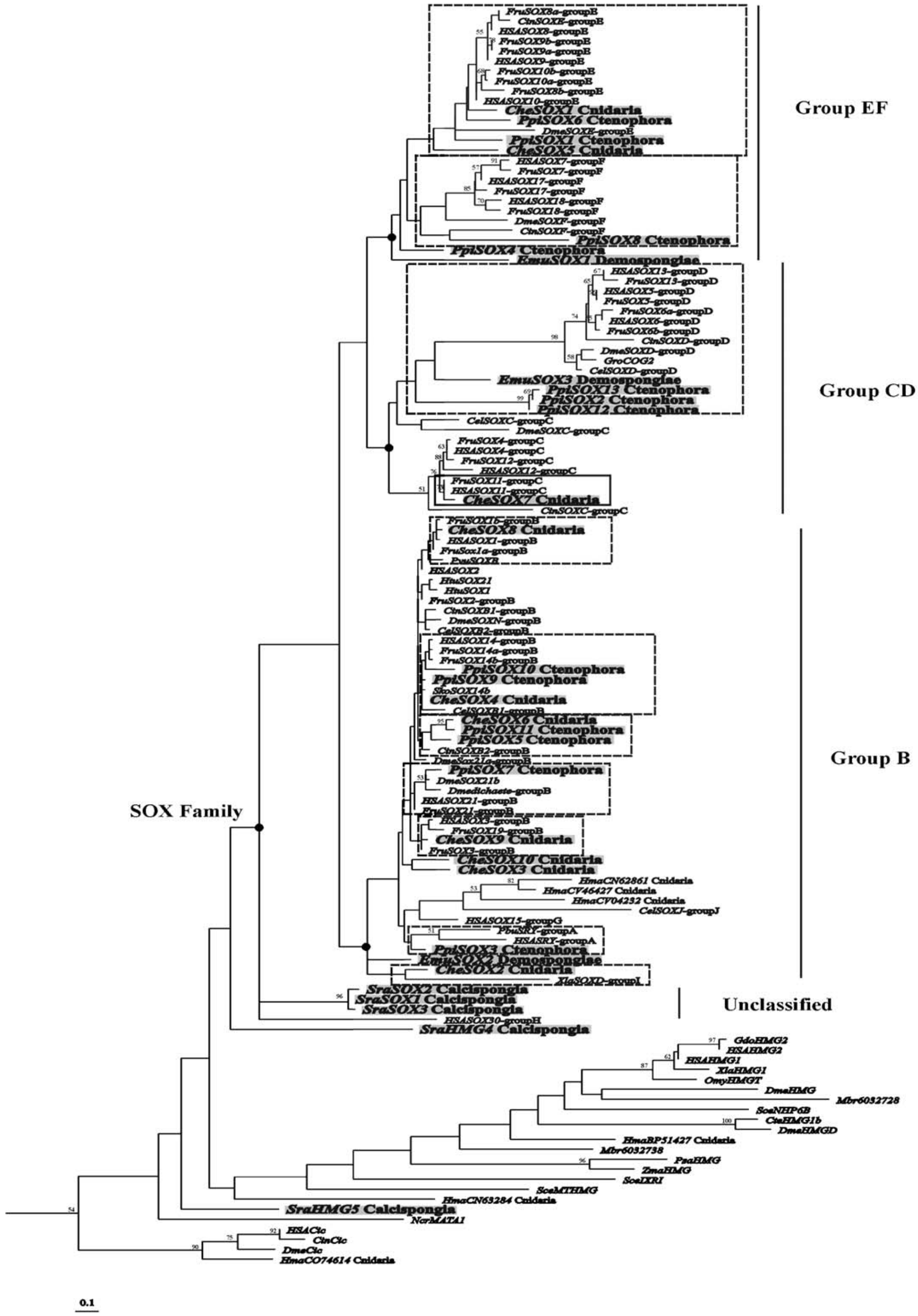

Fig. 1. Phylogenetic analysis of SOX HMG domains using Maximum Likelihood method (ML). The tree was computed from an amino-acid alignment of partial HMG domains (59 amino-acids, representing positions 12–70 of the HMG domain, excepted for *CheSOX8*, *CheSOX9* and *SraHMG5*: 30 amino-acids). The tree is rooted arbitrarily on three TCF sequences, which were not included in the figure to simplify the presentation. The tree likelihood was $\log L = -6221.21879$. Bootstrap proportions (200 replicates) are indicated on the branches when >50%. Each gene is preceded by the initial of the genus name and the two first letters of the species epithet: *Cel*, *Caenorhabditis elegans*; *Cin*, *Ciona intestinalis*; *Cte*, *Chironomus tentans*; *Dme*, *Drosophila melanogaster*; *Gdo*, *Gallus domesticus*; *Gro*, *Globodera rostochiensis*; *Hma*, *Hydra magnipapillata*; *HSA*, *Homo sapiens*; *Htu*, *Heliocidaris tuberculata*; *Hvu*, *Hydra vulgaris*; *Fru*, *Fugu rubripes*; *Mbr*, *Monosiga brevicollis*; *Omy*, *Oncorhynchus mykis*; *Ncr*, *Neurospora crassa*; *Pbu*, *Petrogale burbidgei*; *Pvu*, *Patella vulgata*; *Psa*, *Pisum sativum*; *Sce*, *Saccharomyces cerevisiae*; *Sko*, *Saccoglossus kowalevskii*; *Xla*, *Xenopus laevis*; *Zma*, *Zea mays*. Genes from *Pleurobrachia pileus* (*Ppi*), *Clytia hemisphaerica* (*Che*), *Sycon raphanus* (*Sra*) and *Ephydatia muelleri* (*Emu*) are in bold and in gray boxes. The group to which each *SOX* gene belongs according to previous works (Bowles et al., 2000; Koopman et al., 2004; Leveugle et al., 2004) is indicated. Orthology relationships are indicated by boxes. Full line box indicates node with bootstrap higher than 50% whereas broken line boxes specify nodes with BP lower than 50%.

N(30), E(55), R(58) can be considered diagnostic since the 70 bilaterian SOX HMG domains examined were found to contain all (59/70) or most (6/70 with five residues, 4/70 with four residues, and 1/70 with three residues) of these residues. In contrast, most non-SOX HMG examined had none of these six residues, although some had one, and *NcrMATA1* had two (H(29) and N(30)).

Turning to our new non-bilaterian sequences, *SraSOX1*, *2* and *3* contain five out of six residues of this diagnostic combination (lacking only R(58)). On this basis we consider these sponge genes to belong to the SOX family. In contrast, *SraHMG4* differs from SOX proteins at all these diagnostic residues except H(29) and N(30), supporting its exclusion from the SOX family. These arguments are consistent with the topology obtained from phylogenetic analyses.

All remaining non-bilaterian sequences identified in our study contain the full diagnostic combination of six residues, with the exceptions of *EmuSOX1* and *PpiSOX9* (five residues), *PpiSOX4* (four residues) and *PpiSOX8* (only two residues). For the latter sequence, there is contradiction between the branching of the sequence in the tree and the information from diagnostic residues, so that the affiliation of this gene to the SOX family remains unclear.

Among the SOX family (Fig. 1), three main monophyletic gene subgroups are apparent that, in line with previous studies, we call group B (used here in an extended sense, see Section 4), group CD and group EF. The human gene *HSASOX30*, and three genes from *S. raphanus* (*SraSOX1*, *SraSOX2*, and *SraSOX3*) are positioned outside these three groups at the base of the SOX family in a paraphyletic arrangement. The remaining new SOX genes from non-bilaterians are distributed between groups B, CD and EF.

None of these three groups is supported by bootstrap values, but a careful inspection of the alignment for the presence of diagnostic residues (Figure S2) allows a critical evaluation of the main branchings observed in the tree among the SOX family. Most bilaterian genes belonging to group EF are characterized by the possession of leucine residues at positions 21, 28 and 33. All sequences from non-bilaterians that fall into group EF in the tree (*CheSox1* and *5*, *PpiSox1*, *4*, *6* and *8*, and *EmuSox1*) have these three diagnostic leucines. In addition, *CheSox1* and *PpiSox6* have the peptide LADQY (21–25) which characterizes group E genes, among group EF.

For group CD, we found only one diagnostic residue (isoleucine at position 21), also found however in a few SOX sequences outside group CD (e.g., *CelSoxJ*, *PpiSOX7*, *DmeSOX21b*, *Dmedichaete*, *EmuSOX2*, and *CheSOX2*). All sequences from non-bilaterian eumetazoans that fall into group CD in the tree (*CheSox7* and *PpiSox2*, *12* and *13*) have this isoleucine, but the sponge sequence *EmuSox3* does not. The affiliation of *EmuSox3* to group CD may thus be considered more doubtful. The *CheSox7* sequence includes the peptides IERRKIMEQSPDM (16–28), and RLKHMXDY (60–68), which characterize most group C genes, within group CD.

For group B, two diagnostic residues can be proposed: K/R (27) and K (57) (as noticed also by Koopman et al., 2004). Again, all non-bilaterian sequences falling into group B in the tree (*CheSox2*, *3*, *4*, *6*, *8*, *9*, *10*, *PpiSox3*, *5*, *7*, *9*, *10*, *11*, and *EmuSox2*) have these diagnostic residues. In addition, *CheSox4*, *6*, *PpiSox5*, *9*, *10*, *11* have the peptide MAQEN (21–25) found in most bilaterian group B genes. We are cautious with respect to the attribution of *CheSox8* and *CheSox9* to group B in the tree (Fig. 1), because the sequences are shorter than other sequences in the dataset (30 aa instead of 59 aa). However, both *CheSox8* and *CheSox9* contain group B diagnostic residue K (57).

The branching pattern of non-bilaterian sequences within groups B, CD and EF is complex in the ML tree (Fig. 1), and as mentioned above must be treated with caution in the absence of branch support. Each one of these groups comprises one sequence from the demosponge *Ephydatia*. Within group EF, *EmuSOX1* branches in a basal position, as *EmuSOX2* within group B, and within group CD, *EmuSOX3* branches as the sister-group of Bilateria group D. There is no instance where the position of a sequence from Porifera would suggest orthology with a particular SOX gene from the Bilateria or from Cnidaria or Ctenophora.

Hypotheses of orthology for our non-bilaterian genes, and their statistical support, are indicated by boxes in Fig. 1 (full line: group supported by a bootstrap value >50%; broken line: unsupported groups), and recapitulated in Table 2. Clustering of several genes from the same non-bilaterian species occurs four times, for the three ctenophoran genes *PpiSOX2*, *12* and *13* (among group CD), the two genes *CheSOX3* and *CheSOX10* from *Clytia hemisphaerica*, the three genes from *Hydra magnipapillata* (*Hma*, among group B), and for the three unclassified calcareous sponge genes *SraSOX1*, *2* and *3*. The remaining non-bilaterian genes are dispersed within groups B, CD and EF. An orthology relationship between one cnidarian and one ctenophoran gene is suggested in one instance among Group B, with a high bootstrap value (*CheSOX6* + *PpiSOX11*). We identified 11 putative orthology groups containing gene(s) from Bilateria and from Cnidaria and/or Ctenophora (Fig. 1 and Table 2): two among group EF, two among group CD (of which the grouping of *CheSOX7* with *SOX11* genes from *H. sapiens* and *F. rubripes* is supported by a bootstrap value of 73%), and seven among group B.

4. Discussion

4.1. Unexpected SOX gene diversity in four non-bilaterian species

For the first time, HMG domain sequences from non-bilaterian phyla (Calcispongia, Demospongiae, Ctenophora and Cnidaria) have been included in a phylogenetic study of the SOX family. A first striking finding was the high number of SOX gene sequences recovered from both the cnidarian and the ctenophore representatives (10 and 13 from

Table 2
Orthologies and paralogies for non-bilaterian SOX genes, deduced from PhyML analysis

Species and genes	Orthologies with Bilateria genes—SOX group	Orthologies between Cnidaria and Ctenophora genes	Paralogies
Pleurobrachia pileus			
PpiSOX1	Bilateria Group E—Group EF	*CheSOX1* and *5*	*PpiSOX6*
PpiSOX2	Bilateria Group D—Group CD		*PpiSOX12* and *13* (BP = 99%)
PpiSOX3	Mammalian *SRY SOX* genes—Group B		
PpiSOX4	Bilateria—Group EF	*CheSOX1* and *5*	*PpiSOX1*, *6* and *8*
PpiSOX5	*CinSOXB2*—Group B	*CheSOX6*	*PpiSOX11*
PpiSOX6	Bilateria Group E—Group EF	*CheSOX1* and *5*	*PpiSOX1*
PpiSOX7	Vertebrate *SOX21* genes *Drosophila dichaete* and *SOX21b* (BP = 57%)—Group B		
PpiSOX8	Bilateria Group F—Group EF		
PpSOX9	Bilateria *SOX14* genes—Group B	*CheSOX4*	*PpiSOX10*
PpSOX1O	Bilateria *SOX14* genes—Group B	*CheSOX4*	*PpiSOX9*
PpiSOX11	*CinSOXB2*—Group B	*CheSOX6* (BP = 95%)	*PpiSOX5*
PpiSOX12	Bilateria Group D—Group CD		*PpiSOX2* and *13* (BP = 99%)
PpiSOX13	Bilateria Group D—Group CD		*PpiSOX2* and *12* (BP = 99%)
Clytia hemisphaerica			
CheSOX1	Bilateria group E—Group EF	*PpiSOX1* and *6*	*CheSOX5*
CheSOX2	*XlaSOXD*—Group B		
CheSOX3			*CheSOX10*
CheSOX4	Bilateria *SOX14* genes—Group B	*PpiSOX9* and *10*	
CheSOX5	Bilateria Group E—Group EF	*PpiSOX1* and *6*	*CheSOX1*
CheSOX6	*CinSOXB2*—Group B	*PpiSOX11* (BP = 95%); *PpiSOX5*	
CheSOX7	Vertebrate *SOX11* genes (BP = 73%)—Group CD		
CheSOX8	Bilateria *SOX1* genes and *Patella PvuSOXB*—Group B		
CheSOX9	Vertebrate *SOX3* and *SOX19* genes—Group B		
CheSOX1O			*CheSOX3*
Sycon raphanus			
SraSOX1			*SraSOX2* and *3* (BP = 96%)
SraSOX2			*SraSOX1* and *3* (BP = 96%)
SraSOX3			*SraSOX1* and *2* (BP = 96%)
Ephydatia muelleri			
EmuSOX1	Bilateria Group EF	*CheSOX1* and *5*; *PpiSOX1*, *4*, *6*, and *8*	
EmuSOX2	Bilateria Group B	*CheSOX3*, *4*, *6*, *8*, *9* and *10*; *PpiSOX3*, *5*, *7*, *9*, *10* and *11*	
EmuSOX3	Bilateria Group D—GroupCD	*PpiSOX2, 12* and *13*	

C. hemisphaerica and *P. pileus*, respectively). Given that most of these sequences probably represent true paralogues (see Section 3), the SOX family seems to be significantly more diverse in these two non-bilaterian species than it is in *D. melanogaster* (8 genes), *C. elegans* (5 genes) or *C. intestinalis* (6 genes) (Koopman et al., 2004). Among the completely sequenced bilaterian genomes, only the vertebrates have a higher number of SOX genes (e.g., 20 genes in *H. sapiens* and *M. musculus*, 24 genes in *F. rubripes*). Our data thus refute the idea that non-bilaterians should have fewer genes in relation to their morphological simplicity (e.g., Knoll and Carroll, 1999), and are consistent with EST analysis of the scleractinian *Acropora millepora*, which revealed "paradoxical" high gene complexity (Kortschak et al., 2003). They also concord with the high gene diversity found in the Wnt family in the sea anemone *Nematostella vectensis* (Kusserow et al., 2005). In contrast, a relatively low number of SOX sequences were found in each of the two sponge species. This may reflect lower *SOX* gene diversity, but may also be partly due to restricted sampling or PCR primer bias and will require confirmation.

4.2. Low number of characters and branch statistical robustness

Lack of branch support in the phylogenetic analyses is a general problem for gene families containing only short conserved regions. It is known that the bootstrap method, as an empirical estimator of robustness, does not perform well for datasets comprising a low number of characters (in our case, 59 characters of which 47 are informative). For example, a recent study based on simulations of character evolution on fixed topologies (Alfaro et al., 2003) showed that many correct nodes were attributed less than 50% of bootstraps, when the dataset comprised only 50 characters. Accordingly, boostrap values <50% in our tree do not necessarily invalidate the corresponding branches. In addition, this is not a problem specific to the bootstrap method (e.g., in Alfaro et al., 2003, Bayesian posterior probabilities of correct nodes also tend to be low for low number of characters). This intrinsic limitation must be kept in mind when inferring orthologies and paralogies from tree topology, with such gene datasets.

4.3. Classification of the SOX gene family

Our phylogenetic analyses basically confirmed the classification of SOX genes proposed in previous studies (Bowles et al., 2000; Koopman et al., 2004; Leveugle et al., 2004; Soullier et al., 1999; Wright et al., 1993), but it should be noted that only group D (a group comprising only sequences from Bilateria) is supported by a significant bootstrap value (98%). To simplify the classification, we incorporated genes classified as group A (= *SRY)*, groups G, I, and J in previous studies into an extended group B. According to previous analyses of bilaterian SOX genes (Katoh and Miyata, 1999; Stevanovic et al., 1993; Wegner, 1999) group A genes are derived *SOX3* paralogues, but their position in our tree (Fig. 1) does not confirm this conclusion. Groups G, I, and J were created for single gene sequences (*HSASOX15*, *XlaSOXD*, and *CelSOXJ*, respectively) that did not fit well in previous classification schemes (Bowles et al., 2000; Wegner, 1999); they branch in a basal position among extended group B in our tree (Fig. 1). Within group CD, group C appears as paraphyletic and thus may not actually represent an orthology group.

4.4. History of the SOX gene family

Previous to this study, *SOX* genes had been sampled only from bilaterian taxa. The substantial amount of new data from several non-bilaterian species allow us to propose an evolutionary scenario concerning the early history of the SOX gene family. It should be borne in mind that this scenario remains speculative due to the lack of support for most nodes in the trees, based on bootstrap values.

Since SOX genes (but not other HMG class genes) are absent from the genomes of eukaryote taxa close to the Metazoa such as the ascomycete fungi *S. cerevisiae* and *S. pombe*, the SOX family probably derived from a duplication of a single ancestral HMG box containing gene in the common branch of Metazoa, or possibly even earlier in the Choanozoa (= Metazoa + their sister group the Choanoflagellata). We were unable to find any SOX genes in the ESTs from the choanoflagellate *Monosiga* (http://mcb.berkeley.edu/labs/king/choano/), but their existence cannot be excluded in the absence of a complete choanoflagellate genome. We conclude that SOX genes are likely to be part to the molecular toolkit associated with the acquisition of multicellularity.

The presence of one demosponge gene in each one of the three main SOX family groups (labelled B, CD, and EF) suggests that the SOX family started to diversify within the metazoan stem branch, prior to the divergence of poriferan lineages. The tree topology shown in Fig. 1 suggests that at least two duplications occurred at this stage, from which the three main SOX gene groups originated. These three groups are further supported by the presence of diagnostic residues, also found in the non-bilaterian sequences. The three genes from calcisponge branching basally to the three main groups (*SraSOX1*, *2*, and *3*) may indicate further initial diversity than that found in other metazoan genomes (and missing data or subsequent losses for other species), or more likely represent artefactual basal branching resulting from later sequence divergence.

Within the Eumetazoa, our results contrast with the situation observed for genes of the Wnt family (Kusserow et al., 2005). While most Wnt groups of orthology previously identified among bilaterian genes have a well-supported cnidarian representative, orthology between bilaterian and ctenophoran/cnidarian *SOX* genes is more difficult to assess. For example, among the 20 *SOX* genes identified in the human genome, 12 (*HSASOX8*, *9* and *10*, *HSASOX7*, *17* and *18*, *HSASOX11*, *HSASOX1*, *HSASOX14*, *HSASOX21*, *HSASOX3*, *HSASRY*) are closely related to a ctenophoran or cnidarian gene in the tree topology (Fig. 1), but in only one case (*HSASOX11* and *CheSOX7*) is this relation statistically supported. However, the clustering of some of the cnidarian and ctenophore genes, but no sponge gene, close to particular bilaterian groups of orthology, may suggest that a further phase of *SOX* gene duplications occurred in the common branch of Eumetazoa.

The later history of the SOX family involved further diversification in some phyla, as exemplified by the many duplications unique to the vertebrates, and possibly by the duplications in the ctenophore lineage between *PpiSOX2*, *12* and *13*, in the cnidarian lineage between *CheSOX3* and *CheSOX10* and between the three *H. magnipapillata SOX* genes and in the calcareous sponge lineage between *SraSOX1*, *SraSOX2* and *SraSOX3* (provided that these sequences are not alleles). The latter case is reminiscent of the diversification found for the homeobox gene classes Sycox and NK-2 in *Sycon raphanus* (Manuel and Le Parco, 2000). In contrast, at least the nematode lineage experienced gene losses: group E and group F each contain genes from non-bilaterians, vertebrates, *Ciona*, and *Drosophila*, but no *C. elegans* gene. That the nematode has apparently lost both of these groups is not surprising given the high number of documented gene losses in this lineage (Babenko and Krylov, 2004; Blaxter, 1998).

4.5. Perspectives

We were originally stimulated to search for *SOX* genes in non-bilaterians because of their involvement in neurogenesis in Bilateria. In *Drosophila* and in vertebrates, *SOX* genes (notably many group B genes) have been shown to play key roles in the specification of the neurectoderm and in neuronal differentiation (Le Gouar et al., 2004; Nagai, 2001; Sasai, 2001; Wegner, 1999). The occurrence of *SOX* genes in sponges implies that the primitive function(s) of this gene family was not related to the nervous system, because sponges are not considered to have one (see for example Bergquist, 1978; Brusca and Brusca, 2003). Future expression and functional studies of *SOX* genes in Porifera, Ctenophora and Cnidaria are required to understand the evolution of function in this important family of regulator genes.

Acknowledgments

We thank Pierrette Lamarre for technical help. We are grateful to Evelyn Richelle-Maurer who provided the *Ephydatia* gemmules, and to Michel Veuille for useful advice and discussion. This work was supported by a grant from the French Ministry of Research (ACI jeunes chercheurs) and a grant from the GIS "Institut de la Génomique Marine." We thank the Genoscope (Evry) for sequencing the ESTs from *Clytia hemisphaerica*.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2005.12.005.

References

Adell, T., Grebenjuk, V.A., Wiens, M., Muller, W.E., 2003. Isolation and characterization of two T-box genes from sponges, the phylogenetically oldest metazoan taxon. Dev. Genes Evol. 213, 421–434.

Alfaro, M.E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap? A simulation study comparing the performance of bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255–266.

Babenko, V.B., Krylov, D.M., 2004. Comparative analysis of complete genomes reveals gene loss, acquisition and acceleration of evolutionary rates in Metazoa, suggests a prevalence of evolution via gene acquisition and indicates that the evolutionary rates in animals tend to be conserved. Nucleic Acids Res. 32, 5029–5035.

Bergquist, P.R., 1978. Sponges. Hutchinson, London.

Bertrand, S., Brunet, F.G., Escriva, H., Parmentier, G., Laudet, V., Robinson-Rechavi, M., 2004. Evolutionary genomics of nuclear receptors: from twenty-five ancestral genes to derived endocrine systems. Mol. Biol. Evol. 21, 1923–1937.

Blaxter, M., 1998. *Caenorhabditis elegans* is a nematode. Science 282, 2041–2046.

Bowles, J., Schepers, G., Koopman, P., 2000. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev. Biol. 227, 239–255.

Brusca, R.C., Brusca, G.J., 2003. Invertebrates, second edition Sinauer Associations, Sunderland, MA.

Buescher, M., Hing, F.S., Chia, W., 2002. Formation of neuroblasts in the embryonic central nervous system of *Drosophila melanogaster* is controlled by *SoxNeuro*. Development 129, 4193–4203.

Cavalier-Smith, T., Allsopp, M.T.E.P., Chao, E.E., Boury-Esnault, N., Vacelet, J., 1996. Sponge phylogeny, animal monophyly, and the origin of the nervous system: 18S rRNA evidence. Can. J. Zool. 74, 2031–2045.

Collins, A.G., 1998. Evaluating multiple alternative hypotheses for the origin of Bilateria: an analysis of 18S rRNA molecular evidence. Proc. Natl. Acad. Sci. USA 95, 15458–15463.

Copley, R.R., Aloy, P., Russel, R.B., Telford, M., 2004. Systematic searches for molecular synapomorphies in model metazoan genomes give some support for Ecdysozoa after accounting for the idiosyncrasies of *Caenorhabditis elegans*. Evol. Dev. 6, 164–169.

Dong, C., Wilhelm, D., Koopman, P., 2004. Sox genes and cancer. Cytogenet. Genome Res. 105, 442–447.

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.

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. USA 97, 4493–4498.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P., Lovell-Badge, R., 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature 346, 245–250.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by Maximum Likelihood. Syst. Biol. 52 (5), 696–704.

Hoshiyama, D., Suga, H., Iwabe, N., Koyanagi, M., Nikoh, N., Kuma, K., Matsuda, F., Honjo, T., Miyata, T., 1998. Sponge Pax cDNA related to Pax-2/5/8 and ancient gene duplications in the Pax family. J. Mol. Evol. 47, 640–648.

Hugues, A., 1999. Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. J. Mol. Evol. 48, 565–576.

Jager, M., Hassanin, A., Manuel, M., Le Guyader, H., Deutsch, J., 2003. MADS-box genes in *Ginkgo biloba* and the evolution of the *AGAMOUS* family. Mol. Biol. Evol. 20, 842–854.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. CABIOS 8 (3), 275–282.

Katoh, K., Miyata, T., 1999. A heuristic approach of maximum likelihood method for inferring phylogenetic tree and an application to the mammalian SOX-3 origin of the testis-determining gene SRY. FEBS Lett. 463, 129–132.

Kim, J., Kim, W., Cunningham, C.W., 1999. A new perspective on lower metazoan relationships from 18S rDNA sequences. Mol. Biol. Evol. 16, 423–427.

Knoll, A.H., Carroll, S.B., 1999. Early animal evolution: emerging views from comparative biology and geology. Science 284, 2129–2137.

Koopman, P., Schepers, G., Brenner, S., Venkatesh, B., 2004. Origin and diversity of the SOX transcription factor gene family: genome-wide analysis in *Fugu rubripes*. Gene 328, 177–186.

Kortschak, R.D., Samuel, G., Saint, R., Miller, D.J., 2003. EST analysis of the cnidarian *Acropora millepora* reveals extensive gene loss and rapid sequence divergence in the model invertebrates. Curr. Biol. 13, 2190–2195.

Kruse, M., Muller, I.M., Muller, W.E., 1997. Early evolution of metazoan serine/threonine and tyrosine kinases: identification of selected kinases in marine sponges. Mol. Biol. Evol. 14, 1326–1334.

Krylov, D.M., Wolf, Y.I., Rogozin, I.B., Koonin, E.V., 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. Genome Res. 13, 2229–2235.

Kusserow, A., Pang, K., Sturm, C., Hrouda, M., Lentfer, J., Schmidt, H.A., Technau, U., Von Haeseler, A., Hobmayer, B., Martindale, M.Q., Holstein, T.W., 2005. Unexpected complexity of the Wnt gene family in a sea anemone. Nature 433 (7022), 156–160.

Laudet, V., Stehelin, D., Clevers, H., 1993. Ancestry and diversity of the HMG box superfamily. Nucleic Acids Res. 21, 2493–2501.

Ledent, V., Paquet, O., Vervoort, M., 2002. Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol. 3, 1–18.

Le Gouar, M., Guillou, A., Vervoort, M., 2004. Expression of a SoxB and a Wnt2/13 gene during the development of the mollusc *Patella vulgata*. Dev. Genes Evol. 214, 250–256.

Leveugle, M., Prat, K., Popovici, C., Birnbaum, D., Coulier, F., 2004. Phylogenetic analysis of *Ciona intestinalis* gene superfamilies supports the hypothesis of successive gene expansions. J. Mol. Evol. 58, 168–181.

Manuel, M., Le Parco, Y., 2000. Homeobox gene diversification in the calcareous sponge, *Sycon raphanus*. Mol. Phylogenet. Evol. 17, 97–107.

Manuel, M., Borchiellini, C., Alivon, E., Le Parco, Y., Vacelet, J., Boury-Esnault, N., 2003. Phylogeny and evolution of calcareous sponges: monophyly of calcinea and calcaronea, high level of morphological homoplasy, and the primitive nature of axial symmetry. Syst. Biol. 52, 311–333.

Manuel, M., Le Parco, Y., Borchiellini, C., 2004. Comparative analysis of Brachyury T-domains, with the characterization of two new sponge sequences, from a hexactinellid and a calcisponge. Gene 340, 291–301.

Medina, M., Collins, A.G., Silberman, J.D., Sogin, M.L., 2001. Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. Proc. Natl. Acad. Sci. USA 98, 9707–9712.

Nagai, K., 2001. Molecular evolution of Sry and Sox gene. Gene 270, 161–169.

Nambu, P.A., Nambu, J.R., 1996. The Drosophila fish-hook gene encodes a HMG domain protein essential for segmentation and CNS development. Development 122, 3467–3475.

Overton, P.M., Meadows, L.A., Urban, J., Russell, S., 2002. Evidence for differential and redundant function of the Sox genes *Dichaete* and *SoxN* during CNS development in *Drosophila*. Development 129, 4219–4228.

Prud'homme, B., Lartillot, N., Balavoine, G., Adoutte, A., Vervoort, M., 2002. Phylogenetic analysis of the *Wnt* gene family insights from Lophotrochozoan members. Curr. Biol. 12, 1395–1400.

Sanchez-Soriano, N., Russell, S., 1998. The *Drosophila* Sox-domain protein Dichaete is required for the development of the central nervous system midline. Development 125, 3989–3996.

Sasai, Y., 2001. Roles of Sox factors in neural determination: conserved signaling in evolution? Int. J. Dev. Biol. 45, 321–326.

Seimiya, M., Ishiguro, H., Miura, K., Watanabe, Y., Kurosawa, Y., 1994. Homeobox-containing genes in the most primitive metazoa, the sponges. Eur. J. Biochem. 221, 219–225.

Shepers, G.E., Teasdale, R.D., Koopman, P., 2002. Extent, homology, and nomenclature of the mouse and human *Sox* transcription factor gene families. Dev. Cell 3, 167–170.

Soullier, S., Jay, P., Poulat, F., Vanacker, J.M., Berta, P., Laudet, V., 1999. Diversification pattern of the HMG and SOX family members during evolution. J. Mol. Evol. 48, 517–527.

Stevanovic, M., Lovell-Badge, R., Collignon, J., Goodfellow, P.N., 1993. SOX3 is an X-linked gene related to SRY. Hum. Mol. Genet. 12, 2013–2018.

Uchikawa, M., Kamachi, Y., Kondoh, H., 1999. Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. Mech. Dev. 84, 103–120.

Wegner, M., 1999. From head to toes: the multiple facets of Sox proteins. Nucleic Acids Res. 27, 1409–1420.

Wright, E.M., Snopek, B., Koopman, P., 1993. Seven new members of the Sox gene family expressed during mouse development. Nucleic Acids Res. 21, 744.

Yang, Y., Cun, S., Xie, X., Lin, J., Wei, J., Yang, W., Mou, C., Yu, C., Ye, L., Lu, Y., Fu, Z., Xu, A., 2004. EST analysis of gene expression in the tentacle of *Cyanea capillata*. FEBS Lett. 538, 183–191.

Yates, J.R., 2000. Mass spectrometry, from genomics to proteomics. TIG 16, 1–5.

Zrzavy, J., Mihulka, S., Kepka, P., Bezdek, A., Tietz, D., 1998. Phylogeny of the Metazoa based on morphological and 18S ribosomal DNA evidence. Cladistics 14, 249–285.