Dev Genes Evol (2005) 215: 618–630
DOI 10.1007/s00427-005-0022-y

ORIGINAL ARTICLE

Craig R. Magie · Kevin Pang · Mark Q. Martindale

Genomic inventory and expression of *Sox* and *Fox* genes in the cnidarian *Nematostella vectensis*

Received: 7 June 2005 / Accepted: 23 August 2005 / Published online: 29 September 2005

Abstract The *Sox* and *Forkhead* (*Fox*) gene families are comprised of transcription factors that play important roles in a variety of developmental processes, including germ layer specification, gastrulation, cell fate determination, and morphogenesis. Both the *Sox* and *Fox* gene families are divided into subgroups based on the amino acid sequence of their respective DNA-binding domains, the high-mobility group (HMG) box (*Sox* genes) or Forkhead domain (*Fox* genes). Utilizing the draft genome sequence of the cnidarian *Nematostella vectensis*, we examined the genomic complement of *Sox* and *Fox* genes in this organism to gain insight into the nature of these gene families in a basal metazoan. We identified 14 *Sox* genes and 15 *Fox* genes in *Nematostella* and conducted a Bayesian phylogenetic analysis comparing HMG box and Forkhead domain sequences from *Nematostella* with diverse taxa. We found that the majority of bilaterian *Sox* groups have clear *Nematostella* orthologs, while only a minority of *Fox* groups are represented, suggesting that the evolutionary pressures driving the diversification of these gene families may be distinct from one another. In addition, we examined the expression of a subset of these genes during development in *Nematostella* and found that some of these genes are expressed in patterns consistent with roles in germ layer specification and the regulation of cellular behaviors important for gastrulation. The diversity of expression patterns among members of these gene families in *Nematostella* reinforces the notion that despite their relatively simple morphology, cnidarians possess much of the molecular complexity observed in bilaterian taxa.

Keywords Sox · Forkhead · *Nematostella* · Gastrulation

Communicated by D. Weisblat

C. R. Magie · K. Pang · M. Q. Martindale (✉)
Kewalo Marine Laboratory, Pacific Biomedical Research Center,
University of Hawai'i,
41 Ahui St.,
Honolulu, HI 96813, USA
e-mail: mqmartin@hawaii.edu
Tel.: +1-808-5397326
Fax: +1-808-5994817

Introduction

Gastrulation, or the process through which an embryo internalizes the germ layers that will form the various structures of the adult animal, is the primary morphogenetic event during early development. This transition from a monolayered blastula to a multilayered gastrula is accomplished through a variety of cellular mechanisms in different metazoans, although there are definite similarities across phyla in the genes involved in regulating this process. An examination of the changes in expression and regulation of these "gastrulation genes" in diverse taxa promises to provide insight into the evolution of various gastrulation mechanisms. Additionally, since many of these genes are members of multigene families, their investigation in multiple taxa can also help identify the principles underlying the evolution of gene families and their recruitment into discrete developmental events. For studies in either of these categories, data from basal metazoans are critical to provide a means of comparison to the more derived taxa.

The phylum Cnidaria (which includes sea anemones, jellyfish, and corals) is just such an out-group of "primitive" metazoans. As the likely sister group to the Bilateria (Collins 1998; Medina et al. 2001), cnidarians are perfectly placed to provide insight into the evolution of gastrulation. They are traditionally considered to be diploblastic (i.e., have only two germ layers rather than the three present in bilaterian metazoans) and radially symmetric. While cnidarians exhibit relatively simple body-plan organization, they exhibit diverse gastrulation strategies. Indeed, all of the mechanisms of gastrulation found in Bilateria taxa are present in cnidarians (Tardent 1978). There are four major clades of cnidarians: the basal anthozoans (sea anemones and corals) and the three medusazoan clades, Scyphozoa, Cubozoa, and Hydrozoa. Among the Cnidaria, anthozoans seem to be the most relevant for comparison to bilaterian taxa because of their basal position and simple life history. The starlet sea anemone, *Nematostella vectensis*, has recently emerged as an important cnidarian model system for use in studies aimed at inferring character states

ancestral to the evolution of Bilateria (Fritzenwanker and Technau 2002; Hand and Uhlinger 1992). In this paper, we report the temporal and spatial expression of members of the *Forkhead* (*Fox*) and *Sox* gene families in *Nematostella*. Both of these gene families are involved in regulating numerous developmental processes in other organisms, including mesendodermal patterning and cellular behaviors necessary for gastrulation (Carlsson and Mahlapuu 2002; Pevny and Lovell-Badge 1997; Wegner 1999).

The *Sox* gene family is composed of transcription factors related to the mammalian *SRY* genes. They have been identified in many taxa, with 20 family members identified in the mouse and human genomes (Schepers et al. 2002). They are members of the high-mobility group (HMG) superfamily, which also includes TCF, MATA, and HMG/UBF proteins (Laudet et al. 1993; Soullier et al. 1999). Their defining characteristic is the presence of the highly conserved 79 amino acid DNA-binding HMG box (Gubbay et al. 1990). All Sox proteins seem capable of binding to the same primary sequence, (A/T)(A/T)CAA(A/T)G (Harley et al. 1994). Outside of the HMG box, Sox sequences diverge widely, and specificity in their activity is thought to result from their temporal and spatial expression, as well as tissue-specific combinatorial interactions with other transcription factors and cofactors (Kamachi et al. 1999, 2000; Wilson and Koopman 2002). Sox proteins have been implicated in a variety of developmental processes, including development of the central nervous system, neural crest specification, gastrulation, and mesendodermal patterning (Heeg-Truesdell and LaBonne 2004; Tam et al. 2003; Zhang et al. 2004).

Members of the *Fox* gene family are also transcription factors, defined by the presence of a 110 amino acid winged helix domain, also known as the Forkhead/HNF-3 domain (Kaufmann and Knochel 1996; Weigel and Jackle 1990). In humans, the *Fox* gene family consists of 39 members, although invertebrates seem to have considerably fewer (Carlsson and Mahlapuu 2002). *Drosophila melanogaster*, for example, has 17 members (Lee and Frasch 2004). Similar to the *Sox* family, they have been implicated in a diverse array of developmental processes, including eye development, epithelial organization in the

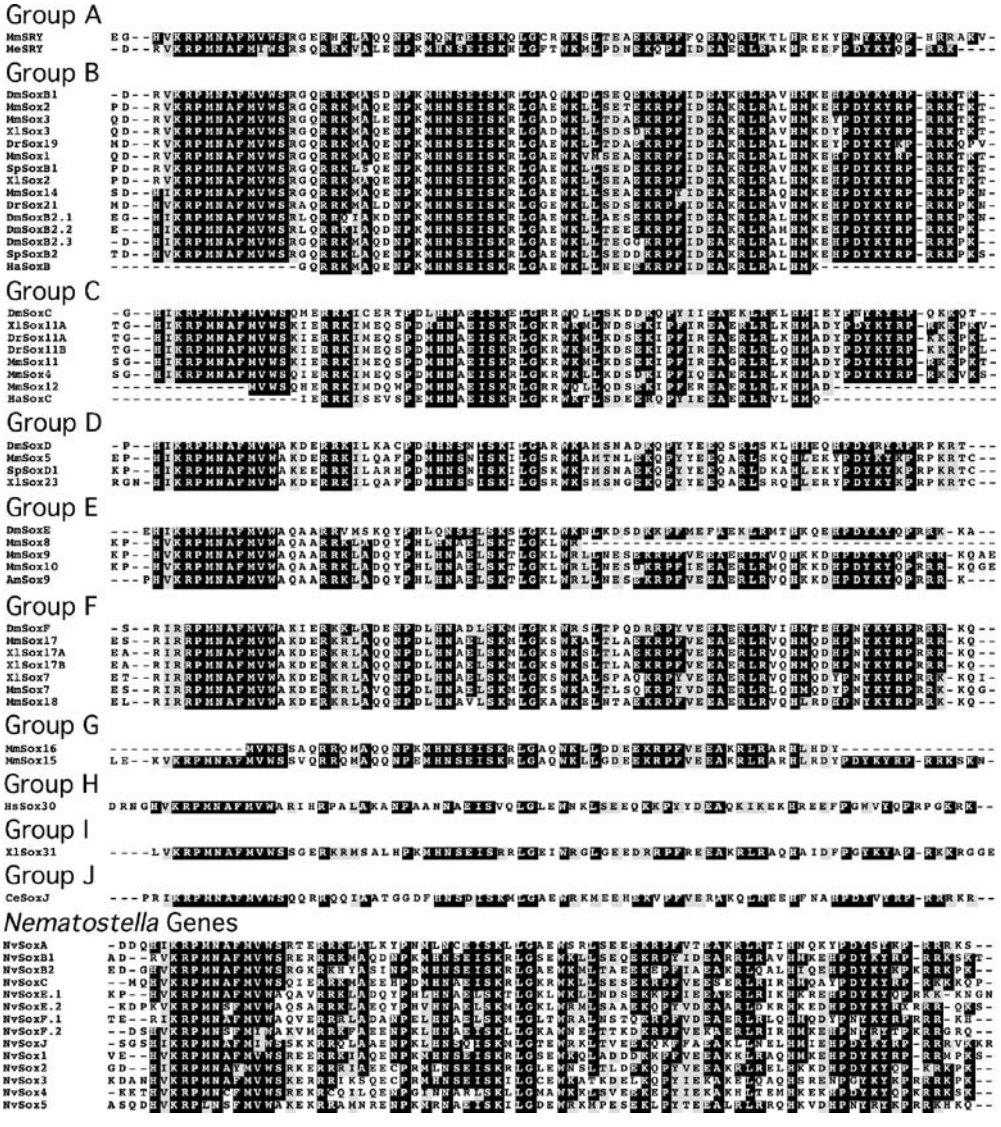


Fig. 1 Alignment of the HMG box from Sox protein sequences used in this study. Boxshade alignment of the HMG box domain from Sox proteins belonging to the various groups in the gene family. The 14 *Nematostella* genes reported in this study are listed at the *bottom*. Taxa represented are as follows: *Alligator mississippiensis* (*Am*), *Caenorhabditis elegans* (*Ce*), *Drosophila melanogaster* (*Dm*), *Danio rerio* (*Dr*), *Haliotis asininia* (*Ha*), *Homo sapiens* (*Hs*), *Macropus eugenii* (*Me*), *Mus musculus* (*Mm*), *Nematostella vectensis* (*Nv*), *Strongylocentrotus purpuratus* (*Sp*), *Xenopus laevis* (*Xl*)

lung, maintenance of neural crest precursor cells, as well as gastrulation and axial patterning (Carlsson and Mahlapuu 2002). Fox transcription factors are targets of a number of signaling pathways, including the TGFβ/Smad pathway downstream of Nodal involved in endoderm specification in vertebrates (Hoodless et al. 2001; Yamamoto et al. 2001). Interestingly, *Sox* genes (specifically *Sox17*) are also involved in endoderm specification in vertebrates as downstream targets of Nodal signaling (Alexander and Stainier 1999), and in *Xenopus*, *Foxa1* and *Foxa2* are direct transcriptional targets of *Sox17* (Sinner et al. 2004). In addition, Sox and Fox functions are linked during the development of the neural crest, as expression of *FoxD3* in *Xenopus* neural crest requires *Sox10* (Honore et al. 2003).

Here we report the genomic inventory and phylogenetic analysis of the multiple *Sox* and *Fox* genes present in *N. vectensis*, along with expression data for a subset of them. Our analysis of the evolution of these gene families as aided by the recent completion of a genome sequencing initiative for *Nematostella* indicates that at least some of the members of each of these families are expressed in a manner consistent with a role in gastrulation and/or germ layer specification, providing insight into the evolution of both these processes.

Materials and methods

Identification of *Sox* and *Fox* genes

Degenerate primers were designed to the most highly conserved regions of bilaterian orthologs for *Sox* and *Fox* genes. Gene fragments were obtained by PCR amplification from genomic DNA and embryonic cDNA. PCR fragments

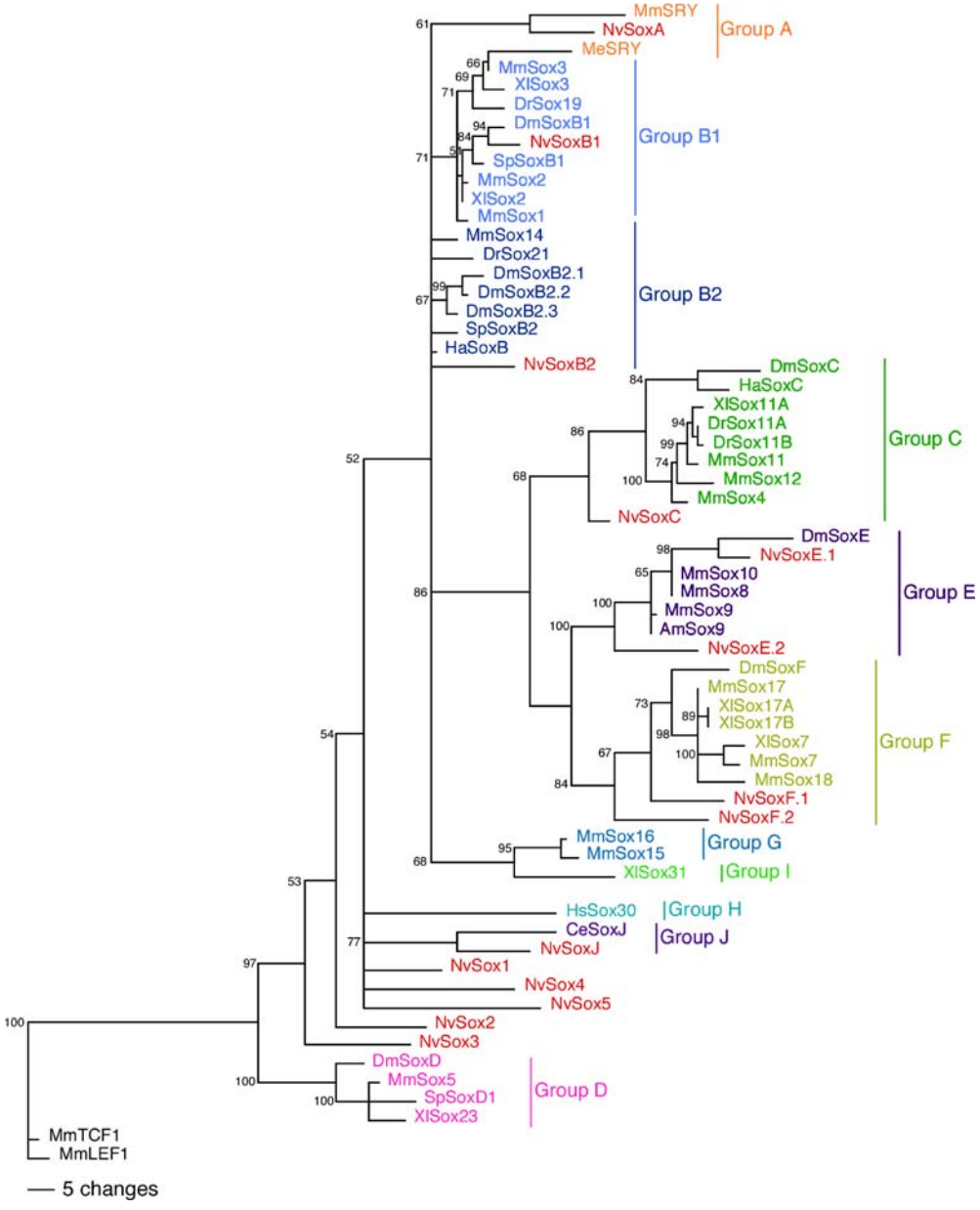


Fig. 2 Phylogenetic relationship between *Nematostella Sox* genes and those from other organisms. Bayesian phylogram (Mr.Bayes v3.0b4) produced from the comparison of the HMG box domains of the Sox proteins aligned in Fig. 1. This topology represents the 50% majority-rule consensus tree resulting from 9,001 trees generated. The posterior probability for each node is indicated adjacent to the node, and the branch length representing five changes is indicated by the scale bar at the *bottom*. *Mus musculus* TCF1 and LEF1 are included as out-group sequences (Bowles et al. 2000). Previously reported Sox family groups A–I are largely supported, with clear *Nematostella* orthologs in many cases, although we could not resolve group B2. The *Nematostella* sequences *NvSox1*, *NvSox2*, *NvSox3*, *NvSox4*, and *NvSox5* do not align with any particular Sox group

were cloned into the pGEM-T easy plasmid vector (Promega) and sent to Gene Gateway, LLC, for sequencing. Sequences from authentic clones were used to design nested sets of nondegenerate primers wi-th annealing temperatures between 68 and 70°C for rapid amplification of cDNA ends (RACE). Both 3′-RACE and 5′-RACE were performed using the SMART RACE cDNA Amplification Kit (BD Biosciences Clontech). *Sox* genes identified by this method (and GenBank accession numbers) are as follows: *NvSoxB1* (DQ173695), *NvSoxB2* (DQ173696), *NvSoxE.1* (DQ173697), *NvSoxF.1* (DQ173698), *NvSox1* (DQ173692), *NvSox2* (DQ173693), and *NvSox3* (DQ173694). *Fox* genes identified by this method (and GenBank accession numbers) are as follows: *NvFoxA* (AY465175; Martindale et al. 2004), *NvFoxB* (DQ173688), *NvFoxC* (DQ173689), *NvFoxD.1* (DQ173690), *NvFoxE* (DQ173691), and *NvFox1* (DQ173687).

Subsequent to the initiation of our PCR survey, the Joint Genome Institute (DOE) made available trace files from the *N. vectensis* genome-sequencing project. In silico searches of the *Nematostella* genome were accomplished by performing a BLAST search against the *N. vectensis* trace archive using the amino acid sequence of either an HMG box (*Sox* genes) or Forkhead domain (*Fox* genes). The resulting sequence fragments were downloaded into MacVector 8.0 (Accelrys) and assembled using AssemblyLIGN (Accelrys). *Sox* genes identified by this method (with representative trace identifiers) are as follows: *NvSoxA* (573062011), *NvSoxC* (595435966), *NvSoxE.2* (567604228), *NvSoxF.2* (578421981), *NvSoxJ* (557 583743), *NvSox4* (573066226), and *NvSox5* (558243222). *Fox* genes identified by this method (with representative trace identifiers) are as follows: *NvFoxD.2* (573154754), *NvFoxL2* (559524830), *NvFoxN* (560471737), *NvFoxO* (557608368), *NvFox2* (558523303), *NvFox3* (568704718), *NvFox4* (557 660630), *NvFox5* (578453495), and *NvFox6* (600265481).

Phylogenetic analysis

Amino acid sequences for the HMG box and Forkhead domains were aligned in MacVector 8.0 (Accelrys) via the ClustalW alignment tool and corrected by hand for obvious errors. A Bayesian phylogenetic analysis was performed on these alignments with Mr.Bayes v3.0b4 (Huelsenbeck and Ronquist 2001) using the Whelan and Goldman (2001) model of protein evolution, with 1,000,000 generations sampled every 100 generations and four chains. A "consensus tree" was produced with PAUP*4.0b10 (Swofford 2002) from the last 9,001 trees representing 900,100 stationary generations. Posterior probabilities and branch lengths were calculated from this "consensus."

Gene expression

Embryos from various stages were fixed in fresh ice-cold 3.7% formaldehyde with 0.2% glutaraldehyde in 1/3× seawater for 90 s and then postfixed in 3.7% formaldehyde in 1/3× seawater at 4°C for 1 h. Fixed embryos were rinsed five times in PBS buffer plus 0.1% Tween 20 (PTw) and once in deionized water, and transferred to 100% methanol for storage at −20°C. Early embryos were removed from the jelly of the egg mass by treating with freshly made 2% cysteine in 1/3× seawater (pH 7.4–7.6) for 10–15 min. Planula and polyp stages were relaxed in 7% $MgCl_2$ in 1/3× seawater for 10 min prior to fixation. In situ hybridization using 1–2 kb digoxigenin-labeled riboprobes were performed to determine the spatial and temporal distribution of transcripts as previously described (Finnerty et al. 2003). Probe concentration ranged from 0.05 to 1.00 ng/ml, and hybridizations were performed at 65°C for 20–44 h. Probe detection was achieved by incubation with an antidigoxigenin antibody conjugated to alkaline phosphatase (Roche). Subsequently, the presence of alkaline phosphatase was detected by a colorimetric detection reaction using the substrate NBT–BCIP. Specimens were photographed on a Zeiss Axioplan with a Nikon Coolpix 990 digital camera.

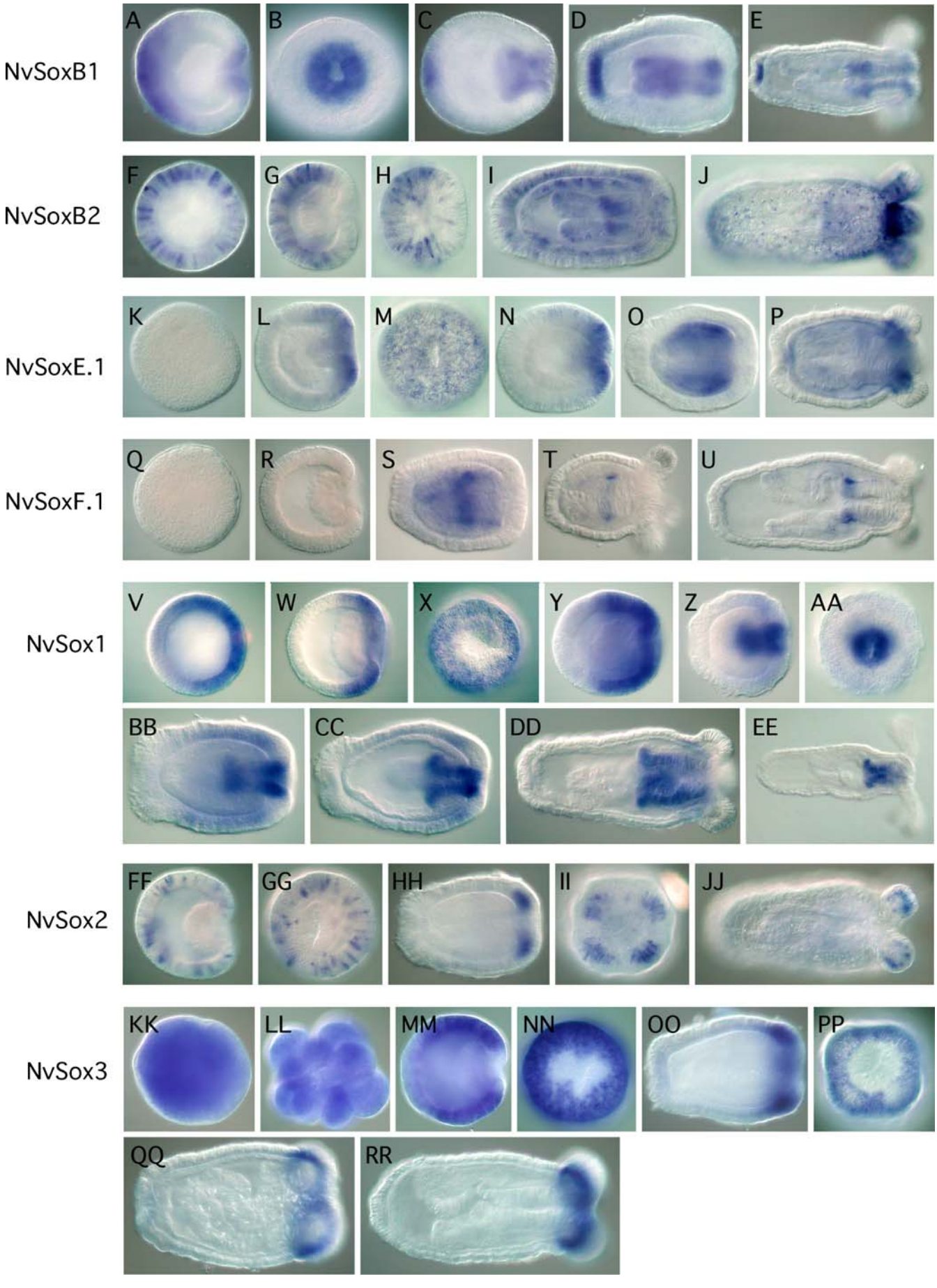
Results

Identification of *NematostellaSox* genes

Our investigations into the *Sox* gene family in *Nematostella* utilized both a degenerate PCR approach and an in silico screening of the *Nematostella* genome as it was sequenced. These analyses identified 14 *Sox* genes. An alignment of the 79 amino acid HMG box of the proteins encoded by these genes is shown in Fig. 1. A comparison with the HMG box in Sox proteins from other organisms indicates a high degree of conservation. A comparison of the consensus HMG box sequence of *Nematostella* Sox proteins with that of the other Sox proteins used in this study indicates 62% identity.

Orthology of *NematostellaSox* genes

Sox proteins have been categorized into groups based on phylogenetic analyses of the amino acid sequence of their HMG box domains. Orthologs of many Sox groups have been found in both vertebrates and invertebrates, although members of groups G, H, and I have only been found in vertebrates to date, and group J is unique to *Caenorhabditis elegans*. To determine which groups are represented among proteins encoded by *Nematostella Sox* genes, we performed a Bayesian phylogenetic analysis (see "Materials and methods") using the sequences shown in Fig. 1. The resulting consensus tree is shown in Fig. 2. As expected, our analysis recovered most of the Sox groups previously reported (Bowles et al. 2000), although we could recover only very weak support for group B2. Some of the *Nematostella* sequences can be unambiguously assigned to a particular group, while others are not so clearly aligned. *NvSoxA* clusters with the group A gene *MmSRY*, although the other group A sequence, *MeSRY*, clusters elsewhere with the group B1 genes by this analysis. This result is consistent with prior phylogenetic analyses of the Sox


NvSoxB1
A
B
C
D
E
NvSoxB2
F
G
H
I
J
NvSoxE.1
K
L
M
N
O
P
NvSoxF.1
Q
R
S
T
U
NvSox1
V
W
X
Y
Z
AA
BB
CC
DD
EE
NvSox2
FF
GG
HH
II
JJ
NvSox3
KK
LL
MM
NN
OO
PP
QQ
RR

◀ **Fig. 3** Expression patterns for a subset of the *Nematostella Sox* genes. In situ expression patterns at various developmental stages for *NvSox* genes. All images are oriented with the oral pole to the right, except **b**, **m**, **x**, **aa**, **gg**, **ii**, **nn**, and **pp**, which are oral views. **a**–**e** *NvSoxB1* expression at the gastrula stage can be seen orally surrounding the blastopore (**a**, **b**), as well as broadly in the aboral half (**a**). Later expression is restricted to a spot at the aboral pole and the pharyngeal ectoderm (**c**–**e**). **f**–**j** *NvSoxB2* is expressed in a salt-and-pepper pattern throughout development. **k**–**p** *NvSoxE.1* exhibits salt-and-pepper expression in the endoderm during gastrulation (**l**, **m**), and broad endodermal expression is maintained through polyp formation (**o**–**p**). **q**–**u** *NvSoxF.1* is not expressed in the early embryo (**q**, **r**). Following gastrulation, broad endodermal expression can be seen (**s**), which is later restricted to a narrow band surrounding the pharynx (pharyngeal endoderm) (**t**, **u**). **v**–**ee** *NvSox1* is initially expressed in the ectoderm in the oral half of the embryo through the gastrula stage (**v**–**aa**). Following gastrulation, its expression is maintained in the pharyngeal ectoderm through polyp formation (**z**–**ee**). **ff**–**jj** *NvSox2* is expressed in a salt-and-pepper pattern early (**ff**, **gg**), with later expression restricted to the tentacle buds (**hh**–**jj**). **kk**–**rr** *NvSox3* exhibits strong maternal expression (**kk**, **ll**). During gastrulation, its expression becomes gradually restricted to the oral half of the embryo, surrounding the blastopore (**mm**, **nn**), with further restriction to the bases of the tentacle buds during polyp formation (**oo**–**rr**)

family, which found that group A was not monophyletic and that *MeSRY* was more closely related to group B genes than was *MmSRY* (Bowles et al. 2000). *NvSoxB1* and *NvSoxB2* are likely group B genes, with *NvSoxB1* aligning most closely with *DmSoxB1* and *SpSoxB1*, and *NvSoxB2* with group B2. *NvSoxC* aligns with the group C clade, although the posterior probability for its node (68) is somewhat low. *NvSoxE.1* and *NvSoxE.2* are clearly group E genes, and there is good support for the group F orthology of *NvSoxF.1* and *NvSoxF.2*. *NvSoxJ* aligns most closely with *CeSoxJ*, which is the sole member of that group. We found no clear evidence for *Nematostella* members of groups D, G, H, or I, although, except for group D, these groups were each represented by only one or two vertebrate sequences and, in the case of groups H and I, have only recently been defined in vertebrates (Bowles et al. 2000). *NvSox1*, *NvSox2*, *NvSox3*, *NvSox4*, and *NvSox5* could not be assigned clear orthology with particular Sox groups based on this analysis.

Nematostella Sox gene expression patterns

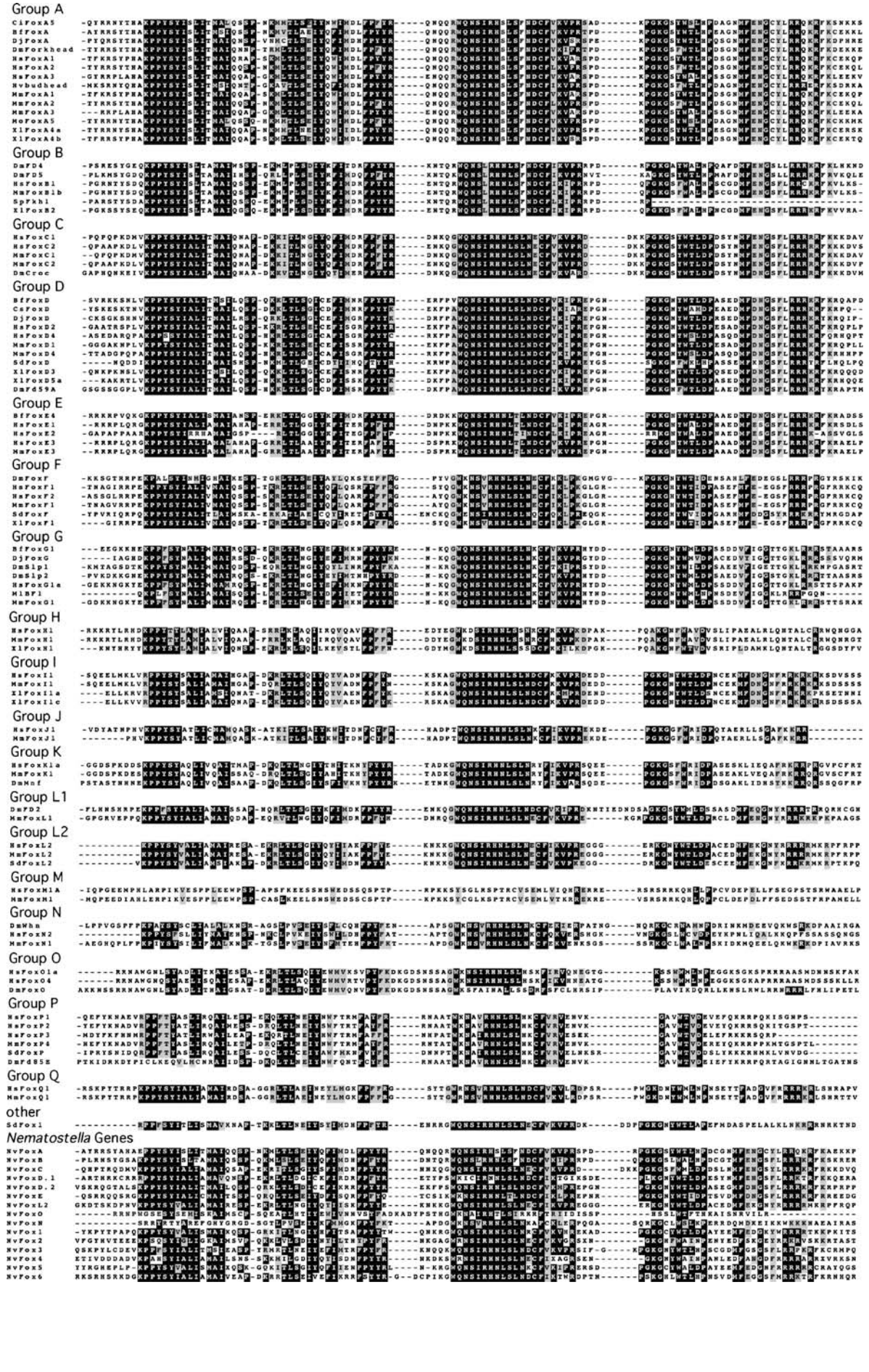
To gain some insight into the developmental processes regulated by *Sox* genes in *Nematostella*, we performed in situ hybridizations with a subset of the *Sox* genes identified (Fig. 3). We observe a variety of expression patterns across the gene family. *NvSoxB1* expression at the gastrula stage is seen broadly in the aboral half of the embryo, as well as a separate domain surrounding the blastopore (Fig. 3a,b). Through the planula stage and polyp formation, the aboral expression becomes limited to a spot at the aboral pole corresponding to the apical tuft, while the oral expression is restricted to the pharyngeal ectoderm (Fig. 3c–e). *NvSoxB2* exhibits a salt-and-pepper pattern of expression in the early embryo (Fig. 3f–h). *NvSoxB2* expression remains confined to individual cells scattered throughout both the ectoderm and endoderm through polyp development (Fig. 3i,j). *NvSoxE.1* is not expressed maternally (Fig. 3k) but begins to be expressed orally during gastrulation (Fig. 3l–n). Expression is maintained in the endoderm through polyp formation, at particularly high levels in the endoderm underlying the tentacle buds (Fig. 3o,p). *NvSoxF.1* is not expressed prior to gastrulation (Fig. 3q,r). During the planula stage, expression is turned on broadly in pharyngeal endoderm and becomes restricted to a ring surrounding the pharynx (Fig. 3s). During polyp formation, this endodermal expression becomes restricted to a ring within the pharyngeal endoderm at the base of the pharynx, near the junction between the pharynx and the first two directive mesenteries (Fig. 3t,u). During the early stages of development, *NvSox1* is expressed in the ectoderm in the oral half of the embryo (Fig. 3v–y). Following gastrulation, this expression becomes restricted to the pharyngeal ectoderm (Fig. 3z,aa), where it persists through polyp formation (Fig. 3bb–ee). *NvSox2* exhibits a salt-and-pepper pattern of expression during the early stages of development (Fig. 3ff, gg), which becomes restricted to the tentacle buds during polyp formation (Fig. 3hh–jj). *NvSox3* expression is unique among the genes we examined in that it is highly expressed maternally (Fig. 3kk,ll). As gastrulation proceeds, its expression is gradually restricted to the oral ectoderm (Fig. 3mm,nn). During polyp formation, expression becomes further restricted to the ectoderm at the bases of the tentacle buds (Fig. 3oo–rr).

Identification of *Nematostella Fox* genes

Similar to the *Sox* gene family, *Fox* genes are divided into a number of subgroups based on the sequence of their Forkhead domains (Mazet et al. 2003). An alignment of the Forkhead domains of Fox proteins from a variety of organisms, arranged by group, is shown in Fig. 4. The 15 *Nematostella Fox* genes identified in this study through degenerate PCR as well as in silico screening of the genome are aligned at the bottom.

Orthology of *Nematostella Fox* genes

To determine which Fox groups are represented among proteins encoded by *Nematostella Fox* genes, we performed a Bayesian phylogenetic analysis similar to our analysis of Sox sequences (see "Materials and methods") using the sequences shown in Fig. 4. The resulting consensus tree is shown in Fig. 5. Our analysis was able to confirm most of the Fox groups previously reported (Adell and Muller 2004; Mazet et al. 2003). In contrast to the *Sox* gene family, however, the majority of Fox groups did not have a clear *Nematostella* representative. *NvFoxA*, which was previously identified as *Nv-forkhead*, is clearly orthologous to group A Fox sequences, consistent with published results (Fritzenwanker et al. 2004; Martindale et al. 2004). *NvFoxB* clusters with group B, while *NvFox3* is basal to the clade comprised of groups A and B. *NvFoxC* is


Group A
Group B
Group C
Group D
Group E
Group F
Group G
Group H
Group I
Group J
Group K
Group L1
Group L2
Group M
Group N
Group O
Group P
Group Q
other
Nematostella Genes

◀ **Fig. 4** Alignment of the Forkhead domain from Fox protein sequences used in this study. Boxshade alignment of the Forkhead domain from Fox proteins used in this study, arranged by group. The 15 *Nematostella* genes identified are listed at the bottom. Taxa represented are as follows: *Branchiostoma floridae* (*Bf*), *Ciona intestinalis* (*Ci*), *Ciona selvatgi* (*Cs*), *Dugesia japonica* (*Dj*), *Drosophila melanogaster* (*Dm*), *Homo sapiens* (*Hs*), *Hydra vulgaris* (*Hv*), *Mnemiopsis leidyi* (*Ml*), *Mus musculus* (*Mm*), *Nematostella vectensis* (*Nv*), *Suberites domuncula* (*Sd*), *Xenopus laevis* (*Xl*)

the sole *Nematostella* ortholog in group C. *NvFoxD.1* and *NvFoxD.2* align with group D. *NvFoxE* is most similar to group E genes, and *NvFoxL2* falls within group L2. While the amino acid sequence of the Forkhead domains of *NvFoxN* and *NvFoxO* are more divergent than the other *Nematostella Fox* genes we identified, they do cluster strongly with groups N and O, respectively, suggesting that they are genuine *Fox* genes. *NvFoxO* is particularly unusual in that it possesses a stop codon within the Forkhead domain that truncates the domain at the start of the second wing. We could identify no *Nematostella* orthologs of groups F, G, H, I, J, K, L1, M, P, or Q. *NvFox1*, *NvFox2*, *NvFox3*, *NvFox4*, *NvFox5*, and *NvFox6* could not be assigned orthology with particular Fox groups based on this type of analysis. *NvFox1* and *NvFox5* are most similar to each other, however, as are *NvFox2* and *NvFox4*, suggesting *Nematostella*-specific gene duplications.

Nematostella Fox gene expression patterns

To gain some insight into the developmental processes regulated by *Fox* genes in *Nematostella* and thereby get some indication of how conserved their functions are likely

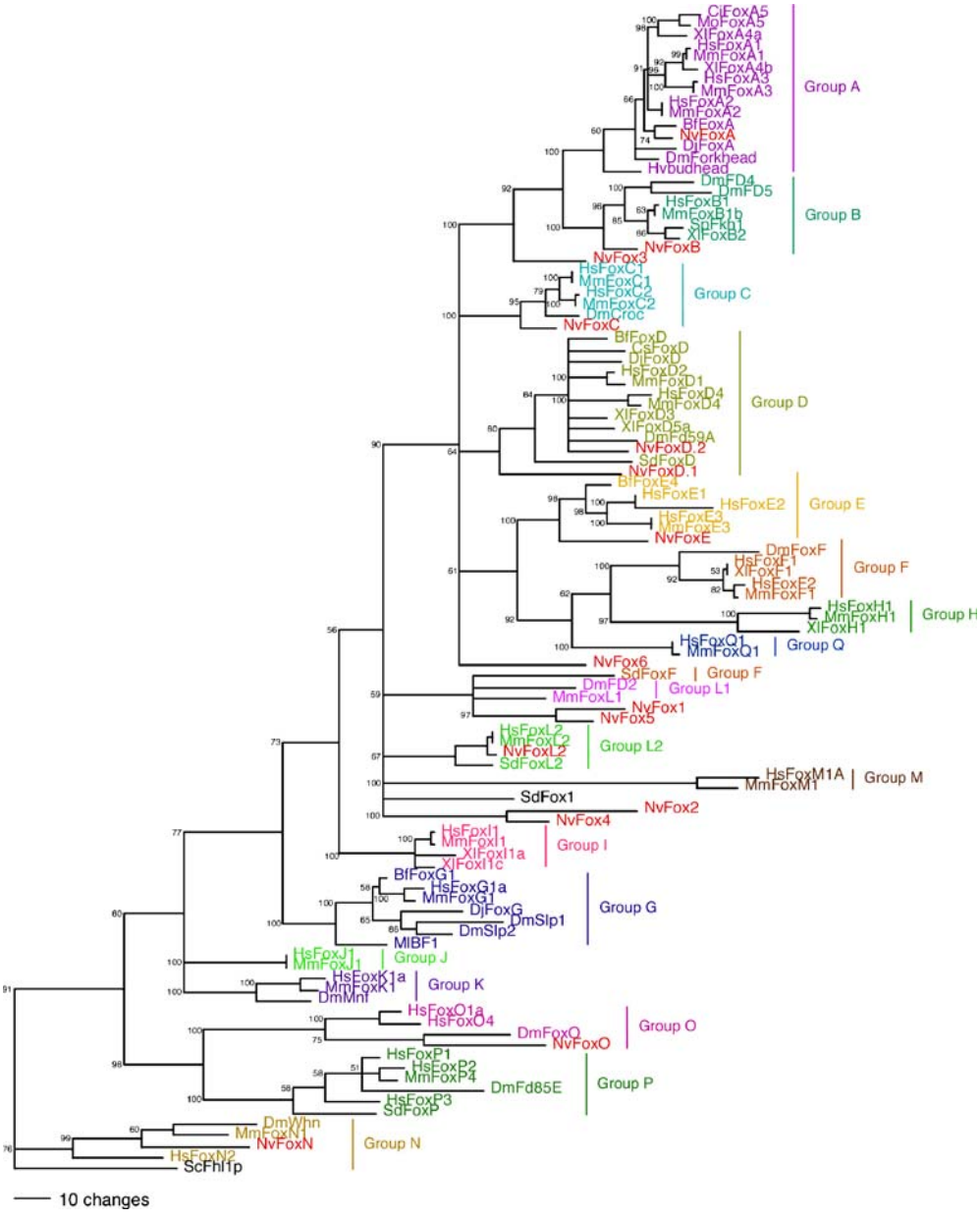


Fig. 5 Phylogenetic relationship between *Nematostella Fox* genes and those from other organisms. Bayesian phylogram produced from the comparison of the Forkhead domains of the Fox proteins aligned in Fig. 4. This topology represents the 50% majority-rule consensus tree resulting from 9,001 trees generated. The posterior probability for each node is indicated adjacent to the node, and the branch length representing ten changes is indicated by the scale bar at the *bottom*. *Saccharomyces cerevisiae* Fhl1p is included as an out-group sequence (Lee and Frasch 2004). Previously reported Fox family groups A–Q are largely supported, with clear *Nematostella* orthologs in many cases. The *Nematostella* sequences *NvFox1*, *NvFox2*, *NvFox3*, *NvFox4*, *NvFox5*, and *NvFox6* do not align with any particular Fox group. There is also evidence for *Nematostella*-specific gene duplications: the gene pairs *NvFox2* and *NvFox4* as well as *NvFox1* and *NvFox5* are more similar to each other than to any other Fox sequences

to be, we performed in situ hybridizations with a subset of the *Fox* genes we identified (Fig. 6). The expression pattern of *NvFoxA* has been reported previously and is expressed surrounding the blastopore during gastrulation and later in the pharynx and pharyngeal mesenteries. *NvFoxB* expression is found around the blastopore during gastrulation (Fig. 6a–f). During the planula larval stage, expression remains in a ring at the oral pole in the pharyngeal ectoderm (Fig. 6g–k), as well as in a subset of the more aboral pharyngeal cells (i.e., arrows in Fig. 6g–i,n). This pharyn-

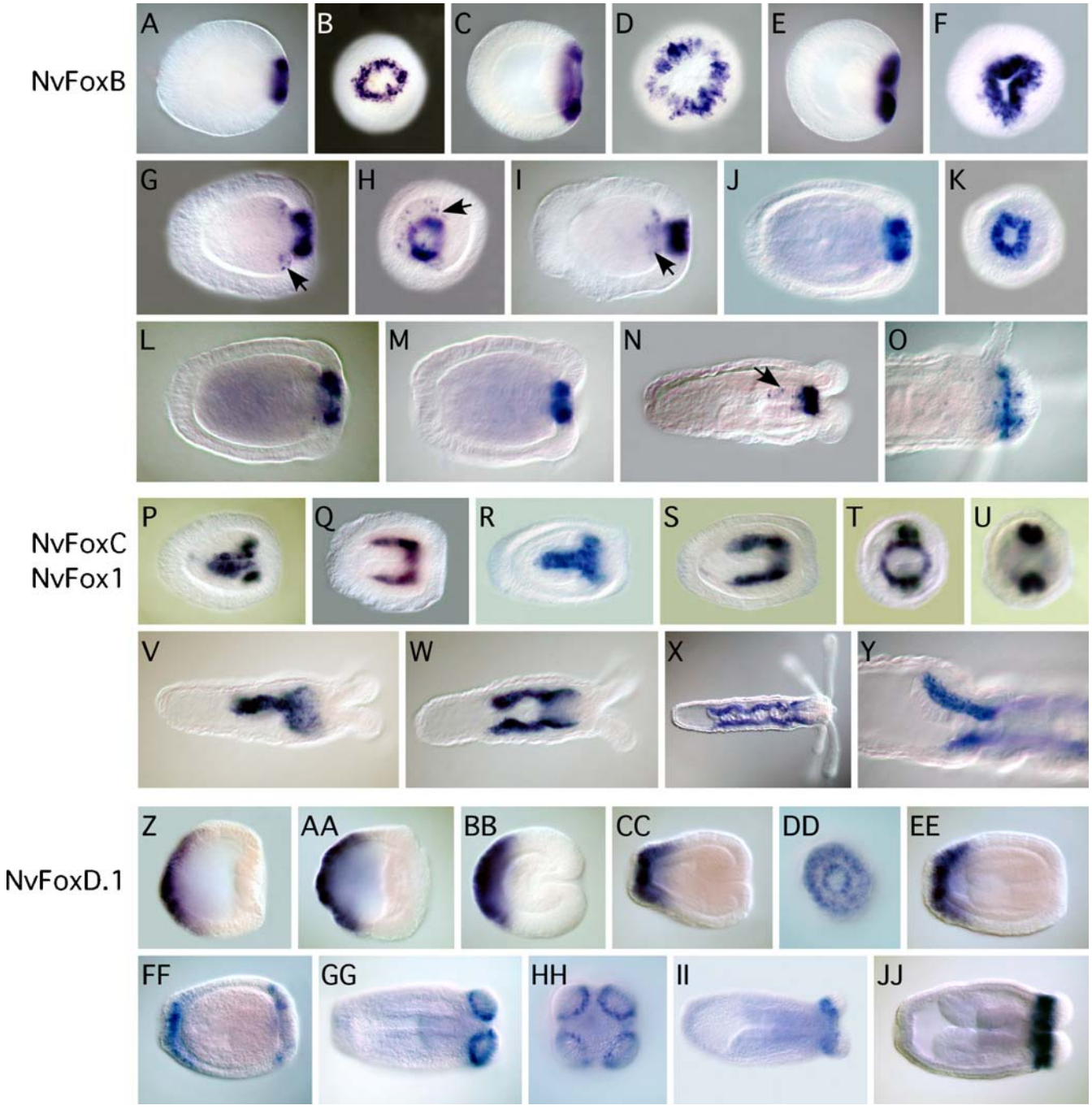


Fig. 6 Expression patterns for a subset of the *Nematostella Fox* genes. In situ expression patterns at various developmental stages for *NvFox* genes. All images are oriented with the oral pole to the right, except **b**, **d**, **f**, **h**, **k**, **t**, **u**, **dd**, and **hh**, which are oral views. **a**–**o** *NvFoxB* expression tightly rings the blastopore at the gastrula stage (**a**–**f**). Expression persists at the oral pole in the planula (**g**–**k**), with some internalized cells retaining expression (e.g., *arrows* in **g**, **h**, **i**, **n**). This oral expression remains through metamorphosis in a narrow domain in the pharyngeal ectoderm (**l**–**o**). **p**–**y** *NvFoxC* and *NvFox1* show identical expression patterns. They are not expressed during early embryonic stages. In the planula, expression can be seen in the endoderm in the presumptive mesenteries (**p**–**u**). This expression persists during polyp development (**t**–**y**). **z**–**jj** *NvFoxD.1* is expressed at the aboral pole during early development, through the planula stage (**z**–**ee**). At metamorphosis, the aboral expression recedes and new expression at the bases of the developing tentacles can be seen (**ff**–**jj**)

geal expression is maintained through polyp formation (Fig. 6l–n) and in the juvenile polyp, where it persists in a band around the oral end of the pharynx (Fig. 6o). *NvFoxC* and *NvFox1* exhibit identical expression patterns. Expression is first seen in the pharyngeal endoderm (but not body wall endoderm) of the first two directive mesenteries in the planula larva (Fig. 6p,q). Expression is maintained in these two mesenteries through development of the polyp (Fig. 6r–w) and persists even in the juvenile polyp (Fig. 6x,y), although no expression is seen in the other six mesenteries. In contrast to the oral expression of *NvFoxA* and *NvFoxB*, *NvFoxD.1* is expressed in the aboral third of the embryo during gastrulation (Fig. 6z–cc). This aboral expression is maintained through the planula stage (Fig. 6dd,ee). Upon development of the polyp, however, the aboral expression is shut off, and expression comes up at the bases of the developing tentacle buds and gradually increases in intensity (Fig. 6ff–jj).

Discussion

The *Sox* and *Fox* gene families are important regulators of a variety of developmental processes. Members of both are involved in germ layer specification, with *Fox* genes having a prominent role in mesoderm development (Carlsson and Mahlapuu 2002), while both *Sox* and *Fox* genes have roles in endodermal patterning (Shivdasani 2002; Tam et al. 2003). Additionally, both are involved in morphogenetic processes such as gastrulation and neural crest development (Heeg-Truesdell and LaBonne 2004). We have examined these gene families in the cnidarian *N. vectensis* to gain insight into their ancestral roles and the nature of their evolution.

Six of the ten groups of *Sox* genes previously identified in other organisms have orthologs in *Nematostella*, indicating that despite the animal's morphological simplicity, the *Nematostella* genome possesses much of the complexity observed in bilaterian taxa. This has also recently been observed for the *Wnt* gene family, in which *Nematostella* possesses orthologs in 11 of the 12 *Wnt* subfamilies observed in chordates, while the *D. melanogaster* genome possesses orthologs in seven subfamilies and *C. elegans* has only five *Wnt* genes (Kusserow et al. 2005). This observation supports the notion that the genomic structure of the last common ancestor of Cnidaria and Bilateria was more complex than previously appreciated and highlights an important role for gene loss in the evolution of some lineages.

In contrast to the situation for the Sox family, there are only clear *Nematostella* orthologs for 8 of the 19 previously identified Fox groups, although six *Nematostella* Fox sequences do not show orthology with any group (see Fig. 5). Given the involvement of Fox transcription factors in morphogenetic processes such as gastrulation and neural crest development, it has been suggested that the diversification of the *Fox* gene family may have been driven by an increasing complexity in body plan organization (Carlsson and Mahlapuu 2002). This hypothesis is supported by the correlation between *Fox* gene number and anatomical complexity in metazoans with fully sequenced genomes (4 in *Saccharomyces cerevisiae*, 15 in *C. elegans*, 20 in *D. melanogaster*, and 39 in *Homo sapiens*; Carlsson and Mahlapuu 2002). Our analysis of the *Nematostella* genome reveals 15 *Fox* genes, which is consistent with this view. We also observe that some *Nematostella* Fox sequences are more similar to each other than Fox sequences from other species, suggesting *Nematostella*-specific duplications (Fig. 5).

Given that the *Sox* and *Fox* gene families are involved in many of the same developmental processes, we might expect the evolutionary pressures driving the diversification of these gene families to be similar. Our finding that the majority of bilaterian Sox groups have *Nematostella* orthologs, while only a minority of Fox groups are represented, suggests, however, that they are under distinct evolutionary pressures. Since not all of the Sox groups are represented in *Nematostella*, a trivial explanation for this could be that the situation for the *Sox* and *Fox* families is actually comparable, but the presence of fewer groups in the *Sox* family masks its subsequent diversification. Certainly, the number of *Sox* and *Fox* genes in *Nematostella* is similar. Alternatively, perhaps the nature of the involvement of these genes in the processes they control (i.e., the types of targets they regulate, etc.) results in differing evolutionary pressures between the families. Future investigations into the details of Sox and Fox function in *Nematostella* relative to other organisms should provide insight into this issue.

Sox gene expression patterns

While it is difficult to argue the function of a gene based on its expression, comparison of the *Nematostella Sox* gene expression patterns with orthologs in other organisms results in some interesting observations. Group B *Sox* genes, important regulators of neural development (Taguchi et al. 2002), have been divided into B1 and B2 subgroups based on the observation that B1 proteins are generally transcriptional activators, while B2 proteins are typically repressors (Uchikawa et al. 1999). Our phylogenetic analysis did not reveal a clear distinction between these groups, although the B1 sequences did cluster more closely with one another than with any of the B2 sequences (Fig. 2). *NvSoxB2* is a likely B2 group gene, and its salt-and-pepper expression pattern in the early embryo, which persists to the early polyp stages (Fig. 3f–j), may be indicative of a role in specifying neural cell fates since a pattern of this sort is suggestive of the neural net present in some cnidarians (Girosi et al. 2005; Grimmelikhuijzen and Westfall 1995). More markers of neural development in *Nematostella* are required to address whether the cells expressing *NvSoxB2* are neuronal in nature, however.

NvSoxB1 is most similar to *DmSoxB1* and *SpSoxB1*. *DmSoxB1* is expressed in the neurogenic ectoderm in the early *Drosophila* embryo and later in the central nervous system (CNS, where it is important for proper CNS

development (Cremazy et al. 2000). *SpSoxB1* is important in ectodermal differentiation in the sea urchin and is initially expressed in all blastomeres during urchin development. Downregulation of *SpSoxB1* expression in the endodermal micromeres is required for the proper nuclear localization of β-catenin, activation of TCF/LEF-dependent transcription, and specification of endodermal and mesenchyme fates (Kenny et al. 2003). Similarly, during early development in *Nematostella*, nuclear localization of β-catenin is observed in endodermal cells prior to and during gastrulation and plays an important role in the specification of endodermal fate, as LiCl treatment of *Nematostella* embryos results in a greater number of cells exhibiting nuclear β-catenin accumulation and hyperproliferation of endoderm at gastrulation (Wikramanayake et al. 2003). The aboral, ectodermal expression of *NvSoxB1*, along with the lack of expression in endodermal cells, is consistent with the notion that it may act through a mechanism similar to that of *SpSoxB1* in endodermal determination. The pharyngeal expression of *NvSoxB1* in *Nematostella* can be seen as a molecular indication of the ectodermal origin of the epithelium lining the lumen of the pharynx (Finnerty et al. 2004) and is consistent with a role in the ectodermal differentiation and/or repression of endodermal fate in those cells.

NvSoxF.1 is a member of the group F *Sox* genes, which in mammals are involved in vasculogenesis, hair follicle development, and endothelial cell organization (Hosking et al. 2001; Pennisi et al. 2000). *NvSoxF.1* is expressed in cells in the endodermal lining of the pharynx (Fig. 3ll,mm). These cells exhibit a squamous morphology that contrasts with the columnar organization of the pharyngeal ectoderm (Magie et al., unpublished data). Perhaps this morphology, which is reminiscent of the squamous morphology of endothelial cells, indicates an ancestral role for group F *Sox* genes in regulating epithelial cells of this type.

Fox gene expression patterns

Similar to what we have observed for members of the *Sox* family, i.e., *NvSox1* and *NvSoxB1*, some members of the *Nematostella Fox* gene family exhibit complementary oral and aboral expression patterns: in the early embryo, *NvFoxA* (Fritzenwanker et al. 2004; Martindale et al. 2004) and *NvFoxB* (Fig. 6v–aa) are expressed orally, while *NvFoxD.1* (Fig. 6a–c) is expressed aborally, although expression shifts to the oral pole at metamorphosis. The significance of this complementary pattern of expression of various family members is unclear, although its presence among both the *Fox* and *Sox* gene families is striking. Future functional studies of these genes will be necessary to address the basis of this phenomenon.

NvFoxB expression in the early embryo is restricted to the cells surrounding the blastopore and remains so throughout gastrulation. Expression can also be seen in a few individual cells after their internalization (arrows in Fig. 6g,h,i,n), although not in the majority of cells. The *NvFoxB* expression surrounding the blastopore is consistent with a role for *NvFoxB* in regulating the cell-shape changes associated with gastrulation. At the polyp stage, *NvFoxB* is expressed in a ring surrounding the mouth at the oral end of the pharyngeal ectoderm. This is highly reminiscent of the expression pattern observed for the *Hydra vulgaris Fox* gene *budhead*, which is expressed in an endodermal ring in the hypostome, a structure containing the mouth located just above the tentacles (Martinez et al. 1997). While *NvFoxB* is a member of a different group than *budhead* (group B vs. group A for *budhead*), our phylogenetic analysis indicates a close linkage for groups A and B, and the previously published expression pattern of *NvFoxA* indicates that it is also expressed in a ring within the pharyngeal ectoderm (Fritzenwanker et al. 2004; Martindale et al. 2004). The ectodermal expression of the *Nematostella* genes, however, is distinct from the endodermal *budhead* expression in *Hydra* (Martinez et al. 1997).

In vertebrates, the group C *Fox* genes *Foxc1* and *Foxc2* are crucial regulators of mesoderm development. They are expressed in the presomitic mesoderm and are required for proper somitogenesis (El-Hodiri et al. 2001; Iida et al. 1997; Kume et al. 1998). Cnidarians do not possess mesoderm. Instead, the *Nematostella* group C ortholog *NvFoxC* is expressed in the pharyngeal endoderm both surrounding the pharynx and adjacent to the pharyngeal mesenteries. This endodermal expression of *NvFoxC* supports the idea that the mesoderm of bilaterians arose from endoderm in a diploblastic ancestor (Martindale et al. 2004). An alternative hypothesis to explain the origin of mesoderm is that diploblasts such as cnidarians were primitively triploblastic but lost mesoderm resulting in the recruitment of mesodermal genes to other functions. The *NvFoxC* expression pattern we observe cannot distinguish between these possibilities. Sampling more taxa both within the Cnidaria as well as at the base of the Bilateria will be necessary to reconstruct the ancestral condition prior to the cnidarian–bilaterian divergence, which will be crucial to understanding the origin of mesoderm.

Foxd1, a mouse group D *Fox* gene, is expressed in the kidney and is involved in the formation of a tubular epithelium from mesenchymal cells (Hatini et al. 1996). *NvFoxD.1* expression at the bases of the tentacles during metamorphosis may indicate an ancestral role for members of this *Fox* group in the morphogenesis of tubular epithelia.

Overall, the *Fox* genes we have examined expression for in *Nematostella* define distinct domains along the oral/aboral axis, with *NvFoxC* and *NvFox1* expressed in the pharyngeal mesenteries, *NvFoxA* in the pharyngeal ectoderm, *NvFoxB* in the mouth/oral end of the pharynx, and *NvFoxD.1* in the tentacle buds. This raises the intriguing possibility that this gene family is involved in specifying fate along the oral/aboral axis, although future functional studies are required to determine which aspects of axial identity are under the control of this important family of transcription factors.

References

Adell T, Muller WE (2004) Isolation and characterization of five Fox (Forkhead) genes from the sponge *Suberites domuncula*. Gene 334:35–46

Alexander J, Stainier DY (1999) A molecular pathway leading to endoderm formation in zebrafish. Curr Biol 9:1147–1157

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

Carlsson P, Mahlapuu M (2002) Forkhead transcription factors: key players in development and metabolism. Dev Biol 250:1–23

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

Cremazy F, Berta P, Girard F (2000) Sox neuro, a new *Drosophila* Sox gene expressed in the developing central nervous system. Mech Dev 93:215–219

El-Hodiri H, Bhatia-Dey N, Kenyon K, Ault K, Dirksen M, Jamrich M (2001) Fox (forkhead) genes are involved in the dorsoventral patterning of the *Xenopus* mesoderm. Int J Dev Biol 45:265–271

Finnerty JR, Paulson D, Burton P, Pang K, Martindale MQ (2003) Early evolution of a homeobox gene: the parahox gene Gsx in the Cnidaria and the Bilateria. Evol Dev 5:331–345

Finnerty JR, Pang K, Burton P, Paulson D, Martindale MQ (2004) Origins of bilateral symmetry: Hox and dpp expression in a sea anemone. Science 304:1335–1337

Fritzenwanker JH, Technau U (2002) Induction of gametogenesis in the basal cnidarian *Nematostella vectensis* (Anthozoa). Dev Genes Evol 212:99–103

Fritzenwanker JH, Saina M, Technau U (2004) Analysis of forkhead and snail expression reveals epithelial–mesenchymal transitions during embryonic and larval development of *Nematostella vectensis*. Dev Biol 275:389–402

Girosi L, Ramoino P, Diaspro A, Gallus L, Ciarcia G, Tagliafierro G (2005) FMRFamide-like immunoreactivity in the sea-fan *Eunicella cavolini* (Cnidaria: Octocorallia). Cell Tissue Res 320:331–336

Grimmelikhuijzen CJ, Westfall JA (1995) The nervous systems of cnidarians. EXS 72:7–24

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

Hand C, Uhlinger KR (1992) The culture, sexual and asexual reproduction, and growth of the sea anemone *Nematostella vectensis*. Biol Bull 182:169–176

Harley VR, Lovell-Badge R, Goodfellow PN (1994) Definition of a consensus DNA binding site for SRY. Nucleic Acids Res 22:1500–1501

Hatini V, Huh SO, Herzlinger D, Soares VC, Lai E (1996) Essential role of stromal mesenchyme in kidney morphogenesis revealed by targeted disruption of Winged Helix transcription factor BF-2. Genes Dev 10:1467–1478

Heeg-Truesdell E, LaBonne C (2004) A slug, a fox, a pair of sox: transcriptional responses to neural crest inducing signals. Birth Defects Res Part C Embryo Today 72:124–139

Honore SM, Aybar MJ, Mayor R (2003) Sox10 is required for the early development of the prospective neural crest in *Xenopus* embryos. Dev Biol 260:79–96

Hoodless PA, Pye M, Chazaud C, Labbe E, Attisano L, Rossant J, Wrana JL (2001) FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. Genes Dev 15:1257–1271

Hosking BM, Wyeth JR, Pennisi DJ, Wang SC, Koopman P, Muscat GE (2001) Cloning and functional analysis of the Sry-related HMG box gene, Sox18. Gene 262:239–247

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Iida K, Koseki H, Kakinuma H, Kato N, Mizutani-Koseki Y, Ohuchi H, Yoshioka H, Noji S, Kawamura K, Kataoka Y, Ueno F, Taniguchi M, Yoshida N, Sugiyama T, Miura N (1997) Essential roles of the winged helix transcription factor MFH-1 in aortic arch patterning and skeletogenesis. Development 124:4627–4638

Kamachi Y, Cheah KS, Kondoh H (1999) Mechanism of regulatory target selection by the SOX high-mobility-group domain proteins as revealed by comparison of SOX1/2/3 and SOX9. Mol Cell Biol 19:107–120

Kamachi Y, Uchikawa M, Kondoh H (2000) Pairing SOX off: with partners in the regulation of embryonic development. Trends Genet 16:182–187

Kaufmann E, Knochel W (1996) Five years on the wings of fork head. Mech Dev 57:3–20

Kenny AP, Oleksyn DW, Newman LA, Angerer RC, Angerer LM (2003) Tight regulation of SpSoxB factors is required for patterning and morphogenesis in sea urchin embryos. Dev Biol 261:412–425

Kume T, Deng KY, Winfrey V, Gould DB, Walter MA, Hogan BL (1998) The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus. Cell 93:985–996

Kusserow A, Pang K, Sturm C, Hrouda M, Lentfer J, Schmidt HA, Technau U, von Haeseler A, Hobmayer B, Martindale MQ, Holstein TW (2005) Unexpected complexity of the Wnt gene family in a sea anemone. Nature 433:156–160

Laudet V, Stehelin D, Clevers H (1993) Ancestry and diversity of the HMG box superfamily. Nucleic Acids Res 21:2493–24501

Lee HH, Frasch M (2004) Survey of forkhead domain encoding genes in the *Drosophila* genome: classification and embryonic expression patterns. Dev Dyn 229:357–366

Martindale MQ, Pang K, Finnerty JR (2004) Investigating the origins of triploblasty: 'mesodermal' gene expression in a diploblastic animal, the sea anemone *Nematostella vectensis* (phylum, Cnidaria; class, Anthozoa). Development 131:2463–2474

Martinez DE, Dirksen ML, Bode PM, Jamrich M, Steele RE, Bode HR (1997) Budhead, a fork head/HNF-3 homologue, is expressed during axis formation and head specification in hydra. Dev Biol 192:523–536

Mazet F, Yu JK, Liberles DA, Holland LZ, Shimeld SM (2003) Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. Gene 316:79–89

Medina M, Collins AG, Silberman JD, Sogin ML (2001) Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. Proc Natl Acad Sci U S A 98:9707–9712

Pennisi D, Bowles J, Nagy A, Muscat G, Koopman P (2000) Mice null for sox18 are viable and display a mild coat defect. Mol Cell Biol 20:9331–9336

Pevny LH, Lovell-Badge R (1997) Sox genes find their feet. Curr Opin Genet Dev 7:338–344

Schepers GE, Teasdale RD, Koopman P (2002) Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. Dev Cell 3:167–170

Shivdasani RA (2002) Molecular regulation of vertebrate early endoderm development. Dev Biol 249:191–203

Sinner D, Rankin S, Lee M, Zorn AM (2004) Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes. Development 131:3069–3080

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

Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (* and other methods), version 4. Sinauer Associates, Sunderland, MA

Taguchi S, Tagawa K, Humphreys T, Satoh N (2002) Group B sox genes that contribute to specification of the vertebrate brain are expressed in the apical organ and ciliary bands of hemichordate larvae. Zoolog Sci 19:57–66

Tam PP, Kanai-Azuma M, Kanai Y (2003) Early endoderm development in vertebrates: lineage differentiation and morphogenetic function. Curr Opin Genet Dev 13:393–400

Tardent P (1978) 5 Entwicklungsperioden. In: Seidel F (ed) Morphogenese der Tiere: Enleitung zum Gesamtwerk Morphogenetische Arbeitsmethoden und Begriffssysteme (Colenterata, Cnidaria). Fischer-Verlag, Jena, pp 199–222

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

Weigel D, Jackle H (1990) The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? Cell 63:455–456

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699

Wikramanayake AH, Hong M, Lee PN, Pang K, Byrum CA, Bince JM, Xu R, Martindale MQ (2003) An ancient role for nuclear beta-catenin in the evolution of axial polarity and germ layer segregation. Nature 426:446–450

Wilson M, Koopman P (2002) Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. Curr Opin Genet Dev 12:441–446

Yamamoto M, Meno C, Sakai Y, Shiratori H, Mochida K, Ikawa Y, Saijoh Y, Hamada H (2001) The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior–posterior patterning and node formation in the mouse. Genes Dev 15:1242–1256

Zhang C, Basta T, Hernandez-Lagunas L, Simpson P, Stemple DL, Artinger KB, Klymkowsky MW (2004) Repression of nodal expression by maternal B1-type SOXs regulates germ layer formation in *Xenopus* and zebrafish. Dev Biol 273:23–37