Available online at www.sciencedirect.com
ScienceDirect
Developmental Biology 300 (2006) 49–62
DEVELOPMENTAL BIOLOGY
www.elsevier.com/locate/ydbio

# Sea urchin Forkhead gene family: Phylogeny and embryonic expression

Qiang Tu, C. Titus Brown, Eric H. Davidson, Paola Oliveri *

*Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA*

Received for publication 31 May 2006; revised 14 September 2006; accepted 15 September 2006
Available online 22 September 2006

**Abstract**

Transcription factors of the Forkhead (Fox) family have been identified in many metazoans, and play important roles in diverse biological processes. Here we define the set of *fox* genes present in the sea urchin genome, and survey their usage during development. This genome includes 22 *fox* genes, only three of which were previously known. Of the 23 *fox* gene subclasses identified in vertebrate genomes, the *Strongylocentrotus purpuratus* genome has orthologues of all but four (E, H, R and S). Phylogenetic analysis suggests that one *S. purpuratus fox* gene is equally related to *foxA* and *foxB* of vertebrates; this gene defines a new class. Two other genes appear to be specific to the sea urchin, with respect to the genomes so far sequenced. *Fox* genes orthologous with those of vertebrates but lacking in arthropod or nematode genomes may be deuterostome-specific (subclasses I, J1, J2, L1, M and Q1), while the majority are pan-bilaterian. All but one of the *S. purpuratus fox* genes (*SpfoxQ1*) are expressed during embryogenesis, most in a very specific temporal and spatial manner. The sea urchin *fox* genes clearly execute many different regulatory functions, and almost all of them participate in the process of embryonic development.


*Keywords:* Forkhead; Fkh; Fox; Sea urchin; Genome; Embryo

## Introduction

Transcription factors are encoded by regulatory genes which lie at the nodes of the gene networks governing biological phenomena. To understand the architecture of the gene regulatory networks underlying embryonic development, their components must be identified. To this end all genes encoding DNA-binding transcription factors in the genome of *Strongylocentrotus purpuratus* were annotated, and their spatial and temporal embryonic expression surveyed. Here we focus on the *forkhead* (*fkh*)/*fox* gene family. Other families of regulatory genes are discussed elsewhere in this collection (Arnone et al., this issue; Howard-Ashby et al., (2006a),(this issue); Materna et al., 2006; Rizzo et al., 2006).

Forkhead transcription factors are widely represented in the animal kingdom, and are characterized by the highly conserved fkh DNA-binding domain, 110 amino acids in length (Kaufmann and Knochel, 1996). The structure of the fkh domain alone and of the fkh/DNA complex has been determined for many members of the family (Clark et al., 1993; Kaufmann et al., 1994; Stroud et al., 2006). This domain folds into a helix–turn–helix structure created by a core of three α helices capped at one end by three antiparallel β strands, and accompanied by two large loops called 'wings' (Clark et al., 1993; Li and Tucker, 1993), from which derives the term "winged helix". The fkh domain binds to DNA in a sequence-specific manner (Gajiwala and Burley, 2000; Kaufmann et al., 1994; Li and Tucker, 1993; Marsden et al., 1998; Pierrou et al., 1994). The founding member of this family, *forkhead*, was studied in *Drosophila* (Weigel et al., 1989). To date, more than 900 proteins have been identified as members of this gene family (see Interpro database (Mulder et al., 2005)). A unified nomenclature has been proposed for these factors based on phylogenetic analysis (Kaestner et al., 2000). The Forkhead or Fox transcription factors are divided into 23 classes, each identified by a letter, and a number is used to distinguish members within the same organism belonging to the same class (for update see http://www.biology.pomona.edu/fox.html; Mazet et al., 2003).

Genetic and functional studies have shown Fox transcription factors to be involved in the regulation of many biological processes. They play important roles in embryonic development, cell fate specification and cell differentiation, and morphogenesis, as well as in regulation of cell cycle and metabolism, and as effectors of signal transduction and chromatin structure (Carlsson and Mahlapuu, 2002; Pohl and Knochel, 2005). Genomic surveys

* Corresponding author. Fax: +1 626 583 8351.
*E-mail address:* poliveri@caltech.edu (P. Oliveri).

0012-1606/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2006.09.031

and embryo expression studies for this gene family have been conducted in many animal species, including human (Pierrou et al., 1994), *Drosophila* (Lee and Frasch, 2004), amphioxus (Mazet et al., 2003), *Ciona* (Yagi et al., 2003), cnidarians (Magie et al., 2005) and sponges (Adell and Muller, 2004).

Three *forkhead* genes have been already characterized in the sea urchin (David et al., 1999; Harada et al., 1996; Oliveri et al., 2006; Ransick et al., 2002). Here we describe the identification of all the *forkhead/fox* genes in the newly sequenced *Strongylocentrotus purpuratus* genome obtained by the Human Genome Sequencing Center (Baylor College of Medicine; http://www.hgsc.bcm.tmc.edu/projects/seaurchin). We also present a detailed phylogenetic analysis of these genes with reference to the *forkhead* genes of 7 other animal species; and we characterize their spatial and temporal expression profiles during the 72 h of embryonic development.

## Material and methods

### *Genome analysis*

The list of all known Forkhead family transcription factors in metazoans was obtained according to InterPro database entry IPR001766. The sequences of these transcription factors were downloaded from UniProt database (Wu et al., 2006) and a Forkhead reference database was constructed. Initially, this database was used to search the unassembled sea urchin genome traces (2× coverage) by TBLASTN (Altschul et al., 1997). The identified traces (cutoff of *e* value was 1*e*–3) were merged into contigs when possible, and then were used to search back against the reference database by BLASTX to remove non-significant hits (cutoff *e* value was 1*e*–6). Later the reference database was used to search against the draft assembly (6× coverage) and the Glean3 gene model predictions (http://www.hgsc.bcm.tmc.edu/projects/seaurchin/).

In the initial analysis, performed on the genome traces (2× coverage), we obtained a total of 24 distinct *forkhead* sequences showing various degrees of conservation to known *fox* genes. Using these 24 sequences we cloned large cDNAs fragments by PCR and library screening.

We also searched Glean3 gene models by HMMER software (Eddy, 1998) with forkhead domain models from the Pfam database (version 17) (Finn et al., 2006), using a cutoff of 1*e*–10. In searches carried out with the global alignment model (Pfam ls) against the Glean3 gene models, all forkhead genes could be identified in the assembled genome. *SpfoxO* is the only exception, in that its forkhead domain was divided into two scaffolds, and could be identified only using the local alignment model (Pfam fs). In the genome draft assembly there are two copies of *SpfoxQ2*, which are identical only in ORF and cannot be distinguished from our cloned cDNA. They most likely represent the two *SpfoxQ2* alleles. Two allelic sequences are clearly present for *SpfoxJ2*, *SpfoxO* and *SpfoxM*.

All the genes previously identified, and only these genes, were found in the Glean3 gene models. The exon structures of matched contigs were predicted by Genscan software(Burge and Karlin, 1997) and later confirmed by the Glean3 gene models.

The isolated cDNAs and transcriptome data (Samanta et al., in press) (http://www.systemix.org/sea-urchin/) were used to annotate the *forkhead* genes in the sea urchin genome.

### *Cloning*

Using the genomic data we designed primers to amplify large fragments of coding sequence. The PCR was performed using the Expand High Fidelity PCR System (Roche) with the conditions suggested by the manufacturer. Each amplified fragment was purified using the QIAquick PCR purification Kit (Qiagen), then subcloned in pGEM-T vector (Promega) and sequenced to confirm the fragment identity.

Those fragments which were less than 500 bp were used as probes to screen cDNA arrayed libraries (Rast et al., 2000) of 9.5 h, 40 h and 72 h according to the peak expression levels determined by QPCR (see below). Partial or complete cDNA sequences were thus obtained for each forkhead factor.

### *Phylogenetic analysis*

Predicted forkhead factor protein sequences from whole genomes of *Homo sapiens*, *Fugu rubripes*, *Drosophila melanogaster*, *Anopheles gambiae*, and *Caenorhabditis elegans* were downloaded from the Ensembl database (Hubbard et al., 2005). Sequences of *Ciona intestinalis* (annotation release 1.0) was downloaded from JGI (http://genome.jgi-psf.org/Cioin2/Cioin2.home.html). We searched these protein sequences using HMMER and Pfam forkhead domain models as described above, in order to identify all genes encoding forkhead transcription factors in these genomes. For the genomes annotated by Ensembl, the different transcription/translational forms of the same gene were removed. Then all forkhead domains were extracted by searching with Pfam and SMART databases (Letunic et al., 2006). From this we obtained a complete and non-redundant dataset of forkhead domains for each different genome. We also obtained forkhead domain sequences of *Nematostella vectensis* from Magie et al. (2005).

Multiple alignments were performed by ClustalW (Thompson et al., 1994) and phylogenetic analysis was performed using MEGA3 (Kumar et al., 2004), by Neighbor-Joining method using the "pairwise deletion" setting. In addition, we constructed a maximum parsimony tree using PHYLIP (Felsestein, 2005) program with randomizing the input order of sequences, and a maximum likelihood tree using PHYML program (Guindon et al., 2005) with JTT substitution model and BIONJ as the starting tree. The bootstrap values have been obtained from sampling 1000 times in MEGA3 and 100 times in the other two methods.

Phylogenetic analysis was carried out using the 8-genome forkhead domains dataset together with sequences from the Chordata Fox Nomenclature Committee website (http://www.biology.pomona.edu/fox.html) (Kaestner et al., 2000), which have been classified accordingly to the current nomenclature.

### *Quantitative PCR (QPCR) assays*

Primers were designed starting from single reads, short assembled genomic sequences, or cDNAs, using Primer3 (Rozen and Skaletsky, 2000). The features of the amplicons we chose were as previously described by Rast et al. (2000). A list of primer pairs used can be found in Table 1 of Supplemental Material. The primers designed accordingly to the initial set of sequences were later corrected for maximum efficiency of amplification when mismatches were identified compared to a better genomic assembly sequence and/or cDNA sequences. Here we consider only the data derived from the best set of primers. QPCRs were performed to measure the expression levels of each gene during embryonic development. Total RNA was extracted from embryos of 0, 9.3, 18, 24, 30, 48 and 70 h post fertilization using the RNeasy micro kit (Qiagen) accordingly to the conditions suggested by the manufacturer. RNA corresponding to 700 embryos was then reverse-transcribed using TaqMan Reverse Transcription Reagents (Applied Biosystems) in 100 µl reactions. Each PCR reaction was performed in triplicate, using cDNA corresponding to 2 embryos/reaction. Ubiquitin amplifications were carried out on the same samples, for use as internal controls by which the absolute prevalence in numbers of transcripts for each gene can be calculated (Oliveri et al., 2002; Oliveri and Davidson, 2004).

### *Whole-mount in situ hybridization (WMISH)*

Whole-mount *in situ* hybridization probes > 500 bp were designed according to the cloned cDNA sequences (see Table 1 of Supplemental Material). The hybridizations were performed as previously described (Minokawa et al., 2004).

## Results

### *Identification of forkhead/fox transcription factors in the Strongylocentrotus purpuratus genome*

Transcription factors are easily identifiable by sequence homology because of the high degree of conservation of their

DNA-binding domains and sometimes other functional domains. In order to identify all the *fox* genes encoded in the sea urchin genome we constructed a reference database, using InterPro entry IPR001766. At the time the analysis was done, this database included 567 forkhead sequences from different metazoans. We used the reference database to search the sea urchin genome by homology, as described in Materials and methods. A total of 19 new *fox* genes was identified. Together with the three already known, there are thus 22 *fox* genes encoded in the sea urchin genome (Table 1).

### *Phylogenetic analysis*

The classification of each sea urchin *fox* gene was determined by phylogenetic analysis. We searched six fully sequenced genomes, and identified 42 forkhead family members in *H. sapiens*, 45 in *F. rubripes*, 27 in *C. intestinalis*, 18 in *D. melanogaster*, 18 in *A. gambiae*, and 15 in *C. elegans*. According to Magie et al. (2005), there are 15 *fox* genes in the genome of the cnidarian *N. vectensis*, and these were also used. Sequences from the Chordata Fox Nomenclature Committee website (Materials and methods) were used as an official classification source. Each sequence in our 8-genome dataset was assigned to the appropriate class by phylogenetic relationship to the data set from the classification source. To exclude any redundancy with the Nomenclature dataset a 8-genome only phylogenetic analysis was then performed. The sea urchin *fox* gene affiliations can be seen in an abbreviated phylogenetic tree based on the *fox* genes of human and fly in Fig. 1, and the more complete eight genome analysis is shown in Fig. 1 of Supplemental Material.

Table 1.
*fox* genes in *S. purpuratus*

| Gene [a] | Genbank accession | cDNA length [b] | Exon [c] | Glean3 model [d] | Reference |
|---|---|---|---|---|---|
| *SpFoxA* | DQ459376 | / | 1 | 06676 | (Oliveri et al., 2006) |
| *SpFoxAB-Like* | DQ286736 | 500 | 1 | 22846 | |
| *SpFoxB* | NM_214632 | / | 1 | 04551 | (David et al., 1999) |
| *SpFoxC* | DQ286740 | 1071 | 1 | 24139 | |
| *SpFoxD* | DQ286738 | 1102 | 2 | 14418 [e] | |
| *SpFoxF* | DQ286741 | 604 | 2 | 00975 | |
| *SpFoxG* | DQ286739 | 556 | 1 | 09771 | |
| *SpFoxI* | DQ286747 | 955 | 1 | 23894 | |
| *SpFoxJ1* | DQ286742 | 2198 | 3 | 27969 | |
| *SpFoxJ2* | DQ286737 | 3151 | 8 | 07644 [f] | |
| *SpFoxK* | DQ286748 | 1820 | 7 | 25010 | |
| *SpFoxL1* | DQ286750 | 1294 | 1 | 15719 | |
| *SpFoxL2* | DQ286745 | 1118 | 1 | 14809 | |
| *SpFoxM* | DQ286752 | 1797 | 6 | 25590 [g] | |
| *SpFoxN1/4* | DQ286753 | 1914 | 5 | 02320 | |
| *SpFoxN1/4* | DQ286754 | 827 | 5 | 02320 | |
| *SpFoxN2/3* | DQ286744 | 1022 | 9 | 15243 | |
| *SpFoxO* | DQ286746 | 1632 | 2 | 09178 [h] | |
| *SpFoxP* | DQ286749 | 1878 | 13 | 09876 | |
| *SpFoxQ1* | DQ286751 | 1224 | 1 | 19345 | |
| *SpFoxQ2* | DQ286735 | 1265 | 1 | 12384 [i] | |
| *SpFoxX* | DQ286743 | 2333 | 1 | / [j] | |
| *SpFoxY* | AF517552 | / | 7 | 10403 | (Ransick et al., 2002) |

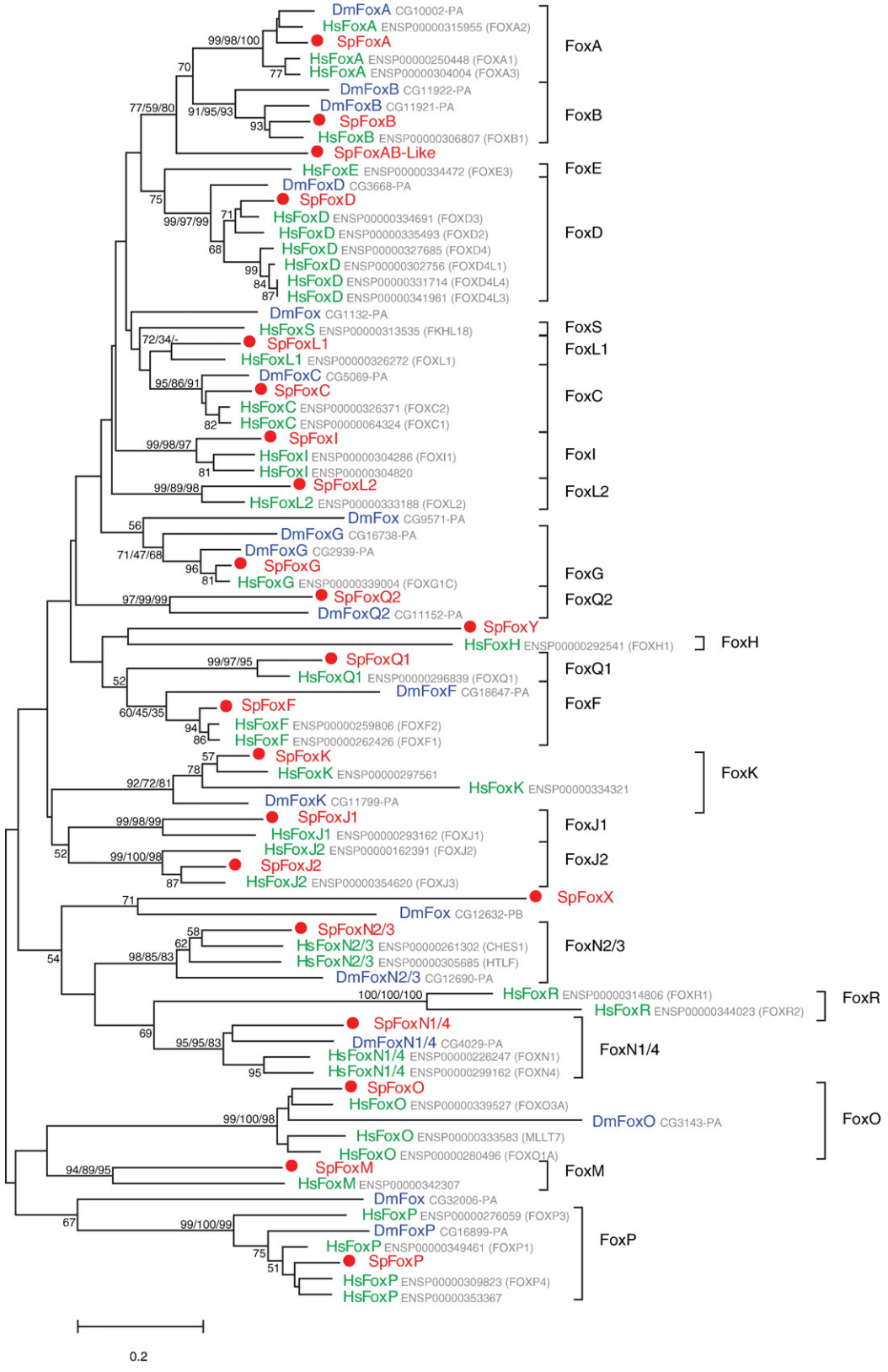
[a] Gene name as it appears in Genbank and in the annotation of the sea urchin genome.
[b] Length in bp of the complete or partial cDNA isolated in this study.
[c] Number of exons identified by comparison of the cDNA fragment with the genomic sequence.
[d] Glean3 model(s) corresponding to the gene in the first column.
[e] *SpfoxD* has an allelic form which is resolved in Glean3_27648.
[f] *SpfoxJ2* has an allelic form which is resolved in Glean3_11635, and part of coding sequence is resolved in another model Glean3_07262.
[g] *SpfoxM* has an allelic form which is resolved in Glean3_25738.
[h] *SpfoxO* has an allelic form which is resolved in Glean3_28698, and part of coding sequence is resolved in another model Glean3_27018.
[i] *SpfoxQ2* has two allelic models, the other model is Glean3_19002.
[j] *SpfoxX* has been identified in the traces and was not present in the first assembly, thus no Glean3 model is available.

Almost all forkhead classes exist in the *S. purpuratus* genome, and each class is represented by a single member (Fig. 1). Classes E, H, R and S are missing from the sea urchin genome, however. The existence of a clear *foxE* class member in a cnidarian (Magie et al., 2005) suggests that this *fox* gene class was present in ancestral bilaterians but was lost in the lineage leading to sea urchins. On the other hand, the *foxH* class so far appears confined to chordates, and the newly identified classes *foxR* (http://www.biology.pomona.edu/fox.html) and *foxS* (Heglind et al., 2005) seem to be present only in vertebrate genomes (for summary see Fig. 2 of Supplemental Material).

Of the 22 sea urchin *fox* genes, 19 fall into known classes. The assignments are sustained by the high bootstrap values in the phylogenetic analysis performed with three different methods (Fig. 1). Concerning the fox classes, all the three methods give same topology with the only exception of the *foxL1* class. The bootstrap values of basal nodes of most classes are higher that 80% in all the trees generated. Exceptions are *foxF*, *foxG* and *foxL1* classes that generally show low bootstrap values. The two sea urchin genes that we named *SpfoxX* and *SpfoxY* did not group to any established class (see Supplemental Fig. 1) and group with different fox members in each of the three different methods used. *SpfoxY* was previously isolated from *S. purpuratus*, and described as *foxC-like* (Ransick et al., 2002), while in the closely related species *H. pulcherrimus* this gene was called *foxFQ-like* (Hibino et al., 2006). However, our analysis showed that it is impossible to assign *SpfoxY* to any of the canonical classes with high confidence, and we conclude that it is specific to the sea urchin genome. The phylogenetic analysis also identifies a new *fox* gene class that occupies a position basal to the A and B classes; hence we term this class "AB-like". The new class is populated by an *S. purpuratus* gene and by *Nematostella fox3*. The phylogenetic relationships of this new class were further examined in two ways. First, we built a phylogenetic tree using a dataset composed only of the *foxA*, *foxB* and *foxAB-like* genes from the eight genomes, with the *foxG* class as an outgroup (Fig. 2). Second, we constructed a phylogenetic tree using a larger dataset consisting of the first 200 sequences retrieved from BLASTP after querying with *SpfoxAB-like*, plus the eight genome *foxA*, *foxB* and *fox AB-like* classes (data not shown). The two different trees gave the same result, and for simplicity in Fig. 2 we show only the first. *SpfoxAB-like* unequivocally groups with *Nematostella fox3* and no other members of this class are identified in the genomes

we used, nor elsewhere in the Genebank database. This indicates that *foxAB-like* is an ancient class which may have been lost independently several times.

We also searched for other known conserved domains outside the Fkh domain using the 7-genomes dataset (excluded *N. vectensis*). This analysis shows that all the members of the *foxA* class contain a "Fork-head N" domain (Pfam PF08430) located at the N-terminal of the DNA-binding domain; most of the *foxK* genes, including the sea urchin one, are characterized by the presence of "FHA" domain (Pfam PF00498) at



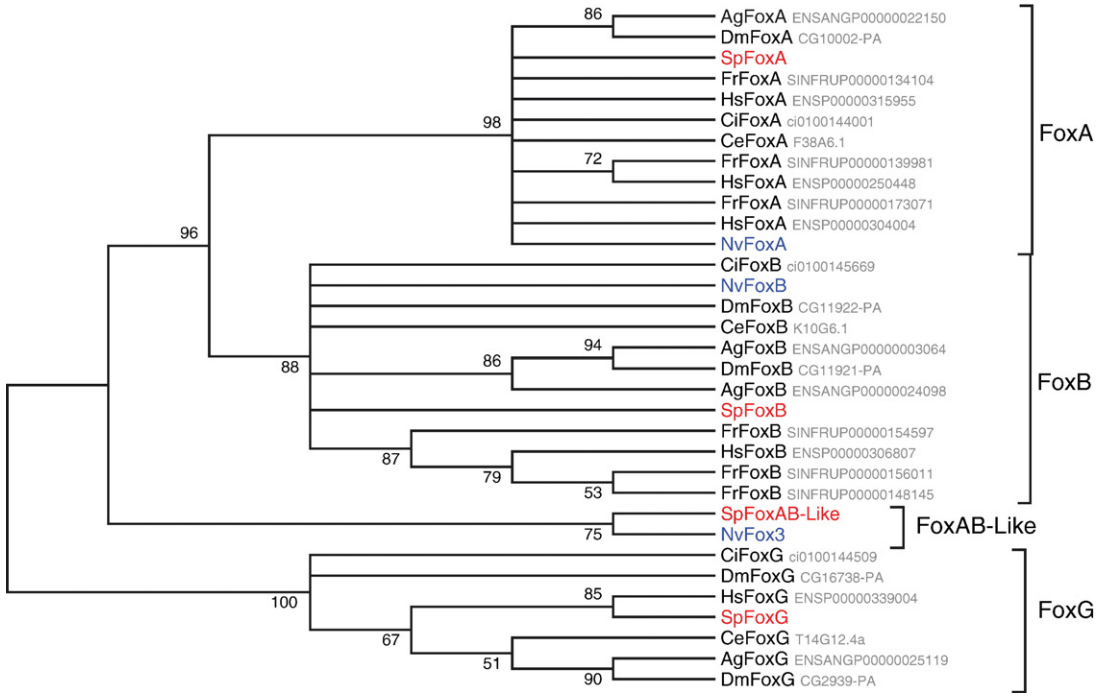



Fig. 2. Phylogenetic tree of forkhead domains of classes A, B and AB-like and G as outgroup. The sequences were collected from 8 full sequenced genomes: *Strongylocentrotus purpuratus* (Sp), *Homo sapiens* (Hs), *Fugu rubripes* (Fr), *Ciona intestinalis* (Ci), *Drosophila melanogaster* (Dm), *Anopheles gambiae* (Ag), *Caenorhabditis elegans* (Ce), and *Nematostella vectensis* (Nv). The tree was constructed by neighbor-joining algorithm and the percent bootstrap values bigger than 50% for1000 re-samplings are showed for each branch. The branches with bootstrap value smaller than 50% are condensed. The naming rules of sequence IDs are the same as described in Fig 1. Sequences from *S. purpuratus* are in red and sequences from *N. vectensis* are in blue. The sea urchin *SpfoxAB-like* and the cnidarian *Nvfox3* characterize a new fox class not identifiable by the comparison only between ecdysozoan and chordate *fox* genes.

the N-terminal of the Fkh domain; and the *foxP* genes present an hypothetical zinc finger domain detected by ScanRegExp method in Interpro ("ZINC_FINGER_C2H2_1", PS00028) also in position N-terminal of the Fkh domain. These other conserved domains are present in the sea urchin gene belonging to the described classes. All other classes of Fox factors present only the Fkh domain as known functional domain.

The structures of all the *S. purpuratus fox* genes were resolved, by application of the cloned cDNA sequences to the genome assembly and the transcriptome data (Samanta et al., in press). An unusual feature of the *fox* genes is that half of them (11) are encoded in single exons. This character is shared with *Drosophila* and human *fox* genes of the classes B and C, and with *Drosophila* classes A and G. Introns have appeared in other *fox* genes in one clade or another: for example, *SpfoxD* is encoded in two exons as is the *foxD* gene in *Drosophila*, while the six Human *foxD* genes are all encoded by a single exon; on the other hand, the two exon structure of *SpfoxF* is conserved with the human gene but not with the *Drosophila* orthologue. *SpfoxP* has the most complex gene structure, with 13 exons (Table 1). This seems to be a class character, as the four Human *foxP* genes and *Drosophila foxP* display diverse numbers of exons. Similarly, *fox* genes of classes K, N1/4, N2/3 and O are all conserved at the sequence level between sea urchin, *Drosophila* and human, yet nonetheless have different numbers of exons. Multiple introns provide the opportunity of alternative splicing, and indeed we isolated two alternatively spliced forms of *SpfoxN1/4* cDNA (Table 1).

### *Spatial and temporal expression patterns*

To determine the usage and thus the possible roles of these transcription factors, the expression pattern of each of the 19 previously uncharacterized *fox* genes was studied by both quantitative PCR (QPCR) and whole-mount *in situ* hybridization (WMISH). For the QPCR analysis specific primers were

Fig. 1. Neighbor-joining phylogenetic tree of all forkhead domains identified in *Strongylocentrotus purpuratus*, *Drosophila melanogaster* and *Homo sapiens* genomes. The classes of forkhead box (fox) factors are indicated on the right. The bootstrap values are given in percent and are shown only if >50%. For branch points at the base of each forkhead class bootstrap values obtained with different methods are indicated. They are separated by slashes and are presented in the following order: first, neighbor-joining using MEGA3, second, maximum-parsimony using PHYLIP and third, maximum-likelihood using PHYML. Dashed line indicates different tree topology and thus no bootstrap value. Sequences are color coded: red indicates *S. purpuratus* (Sp); blue, *D. melanogaster* (Dm); green *H. sapiens* (Hs). We assigned an identification name (ID) to each sequence in the tree as follows: the first two letters are the species abbreviations; the following letters give the *fox* gene class. Letters in small gray font provide the genome ID of the sequence derived from Ensembl. Sequences that do not show clear relationship to any known forkhead class are named just 'Fox' plus the Ensembl ID. Note that all *fox* gene classes identified in the sea urchin *fox* gene set are represented by a single orthologous *S. purpuratus* gene. The E, H, R and S classes are represented only by human genes. Three *S. purpuratus* genes do not belonging to any known vertebrate classes: these are *SpfoxAB-like*, *SpfoxX* and *SpfoxY.*

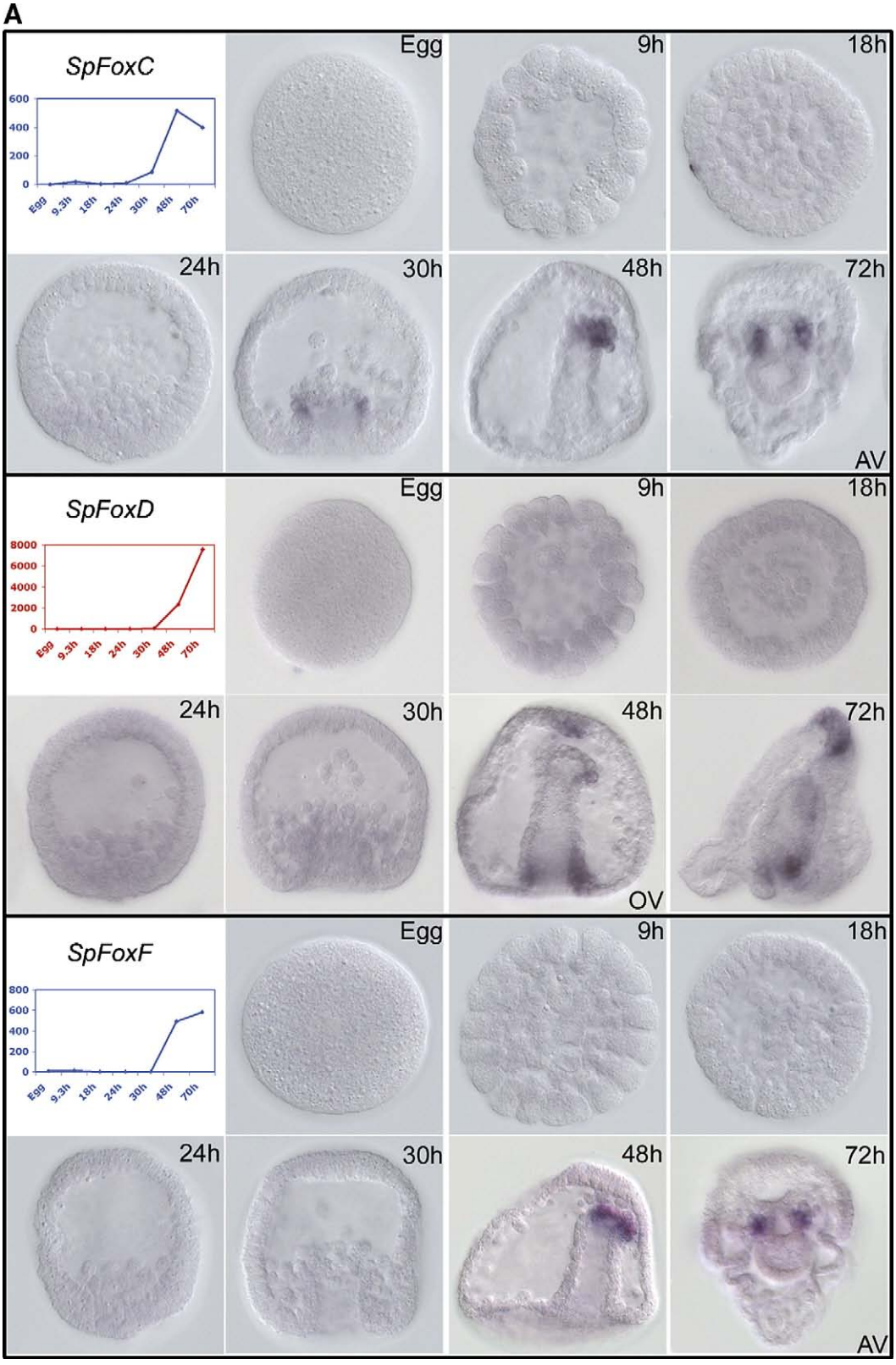



Fig. 3. Temporal and spatial expression of newly identified members of the *fox* gene family during sea urchin embryo development. Graphs show the temporal expression profile for each gene revealed by QPCR and expressed in number of molecules per embryo. The genes are classified by their expression levels: red, high >5000 copies/embryo, green, medium 1000–5000 copies/embryo; blue, low level <1000 copies/embryo; and black, no expression <20 copies/embryo. Spatial expressions are revealed by WMISH. The expression data of each gene, as named on top of the graph, are showed at different time point as indicated in each panel. All the embryos after 24 h are presented in lateral view unless differently specified. APV, apical plate view; OV, oral ectoderm view; VV, vegetal view; AV, anal ectoderm view.

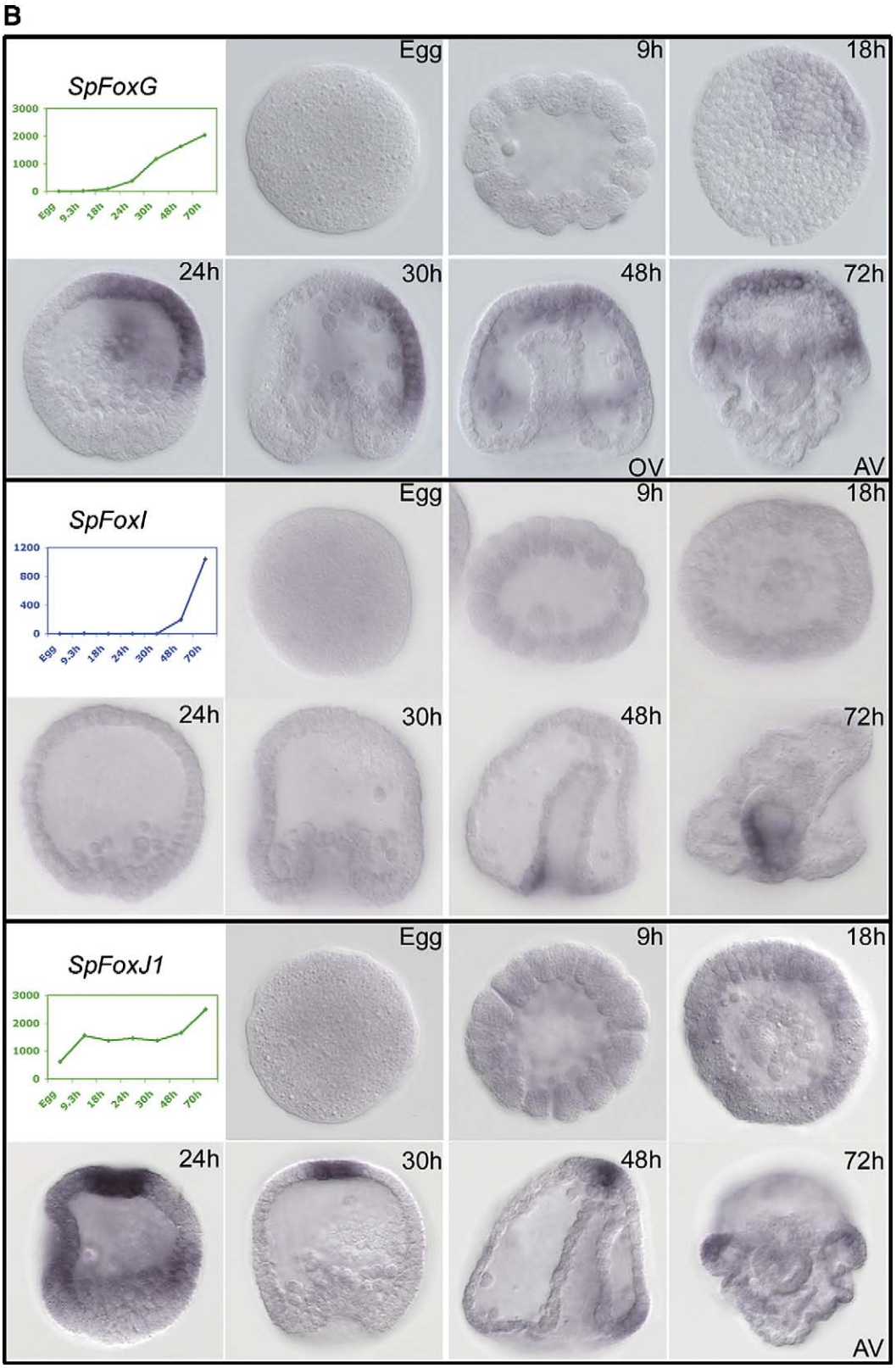



Fig. 3 (*continued*).

designed as described in Materials and methods (see also Supplemental Material). As controls, in five cases two sets of primers were synthesized against two different parts of the gene: the QPCR results were perfectly identical. Transcript abundance (Fig. 3) was determined for seven representative developmental stages, unfertilized egg (0 h), cleavage (9.3 h); blastula (18 h); mesenchyme blastula (24 h); early gastrula (30 h), late gastrula (48 h) and pluteus larva (72 h). Remarkably, all but one of these

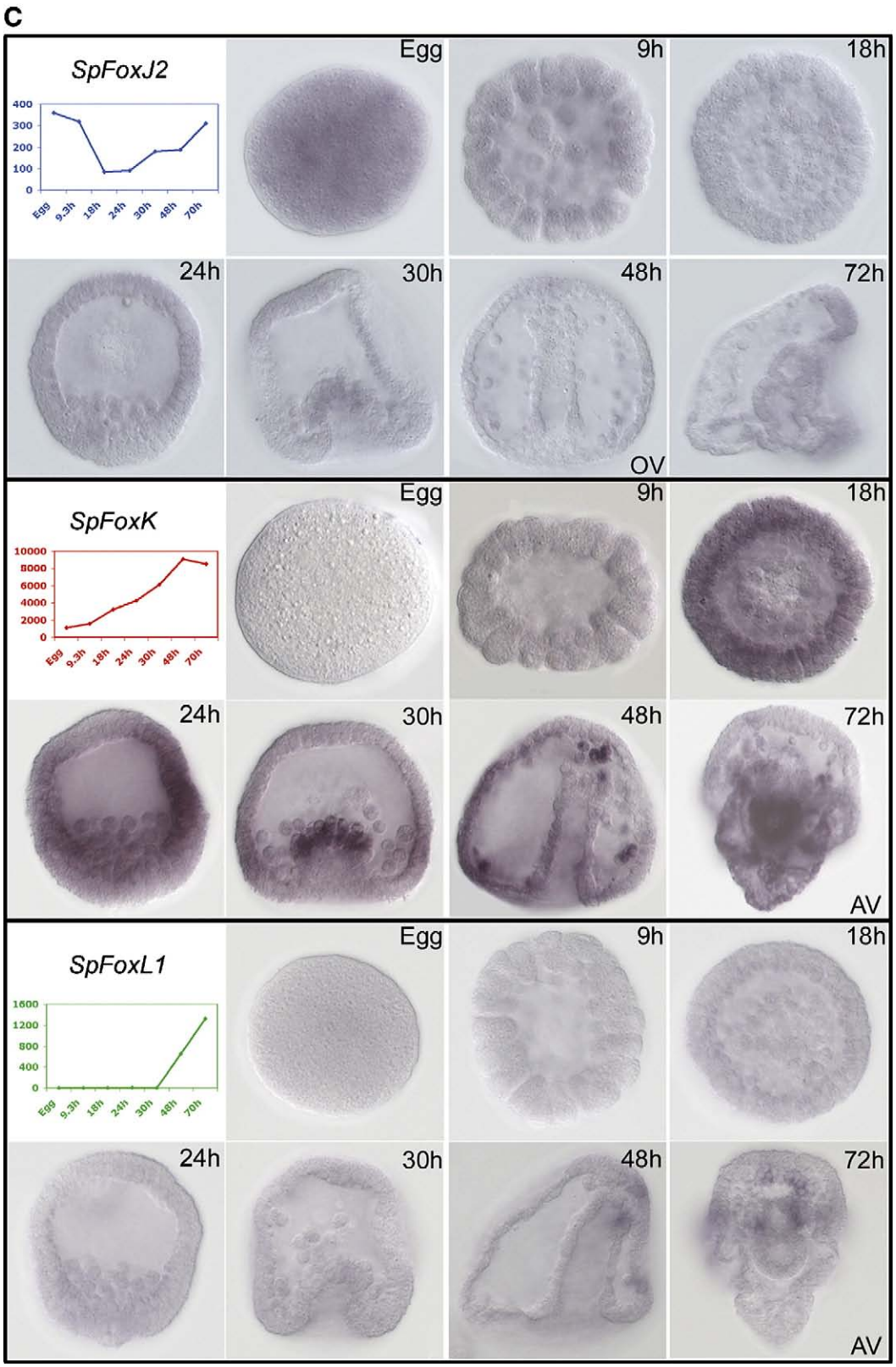



Fig. 3 (*continued*).

regulatory genes are expressed during development. The exception is *SpfoxQ1* (<20 transcripts/embryo at each time point). Five genes are expressed at high levels (more than 5000 copies/embryo, red graphs in Fig. 3): these genes are *SpfoxD*, *SpfoxK*, *SpfoxL2*, *SpfoxO* and *SpfoxQ2*. The rest of the *fox* genes, including the three already known, display medium to

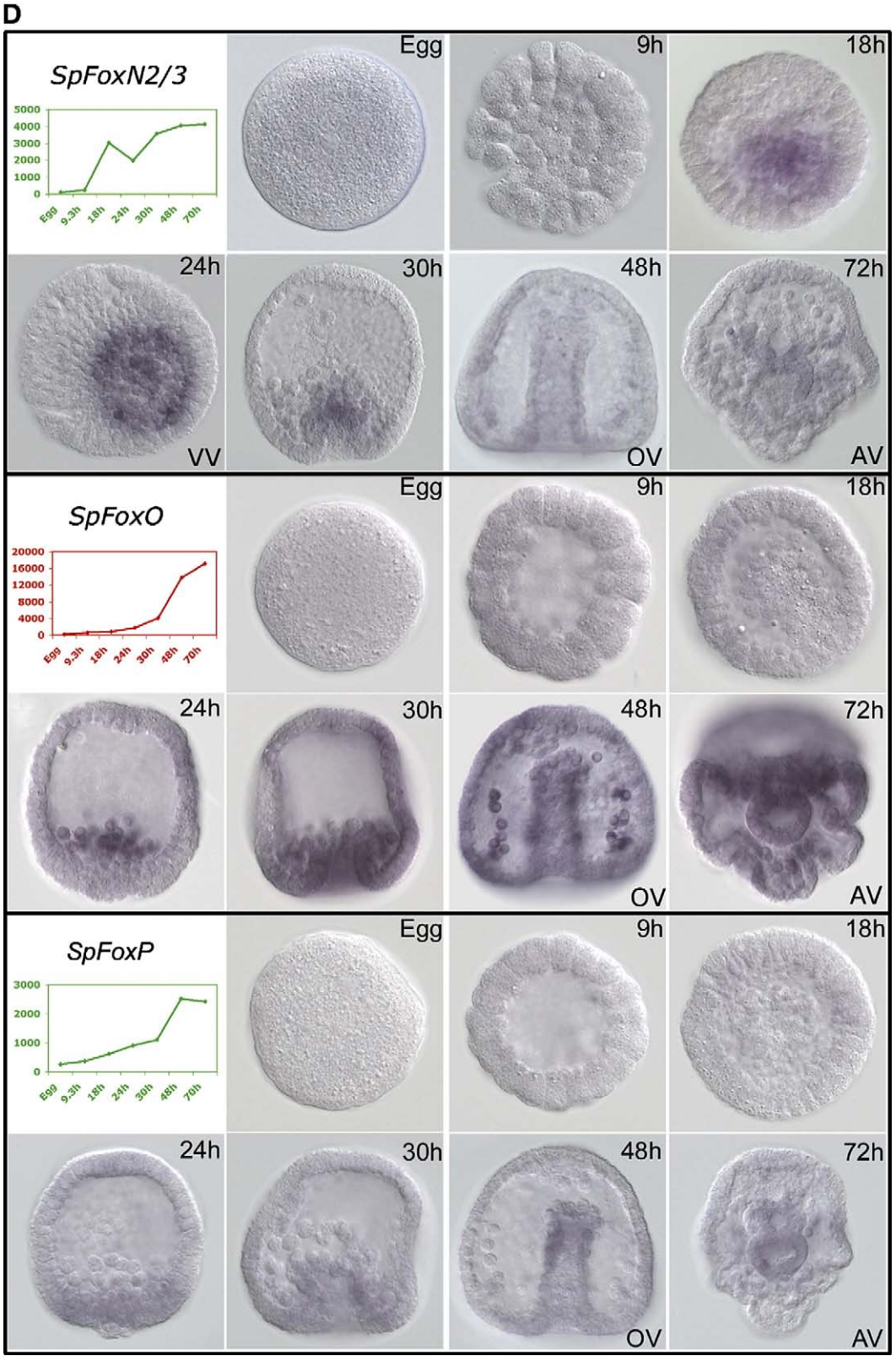



Fig. 3 (*continued*).

low abundance transcripts (400–4000 copies/embryo; blue and green graphs in Fig. 3) at some time(s) in development. Only one gene, *SpfoxX*, is exclusively maternal. Its transcripts have disappeared by blastula stage. Two other genes, *SpfoxJ2* and *SpfoxN1/4*, have maternal expression, but their transcripts increase by 30 h of development. The remaining sea urchin *fox* genes are all activated zygotically. SpfoxJ1, *SpfoxK* and *SpfoxQ2* start their expression very early in development

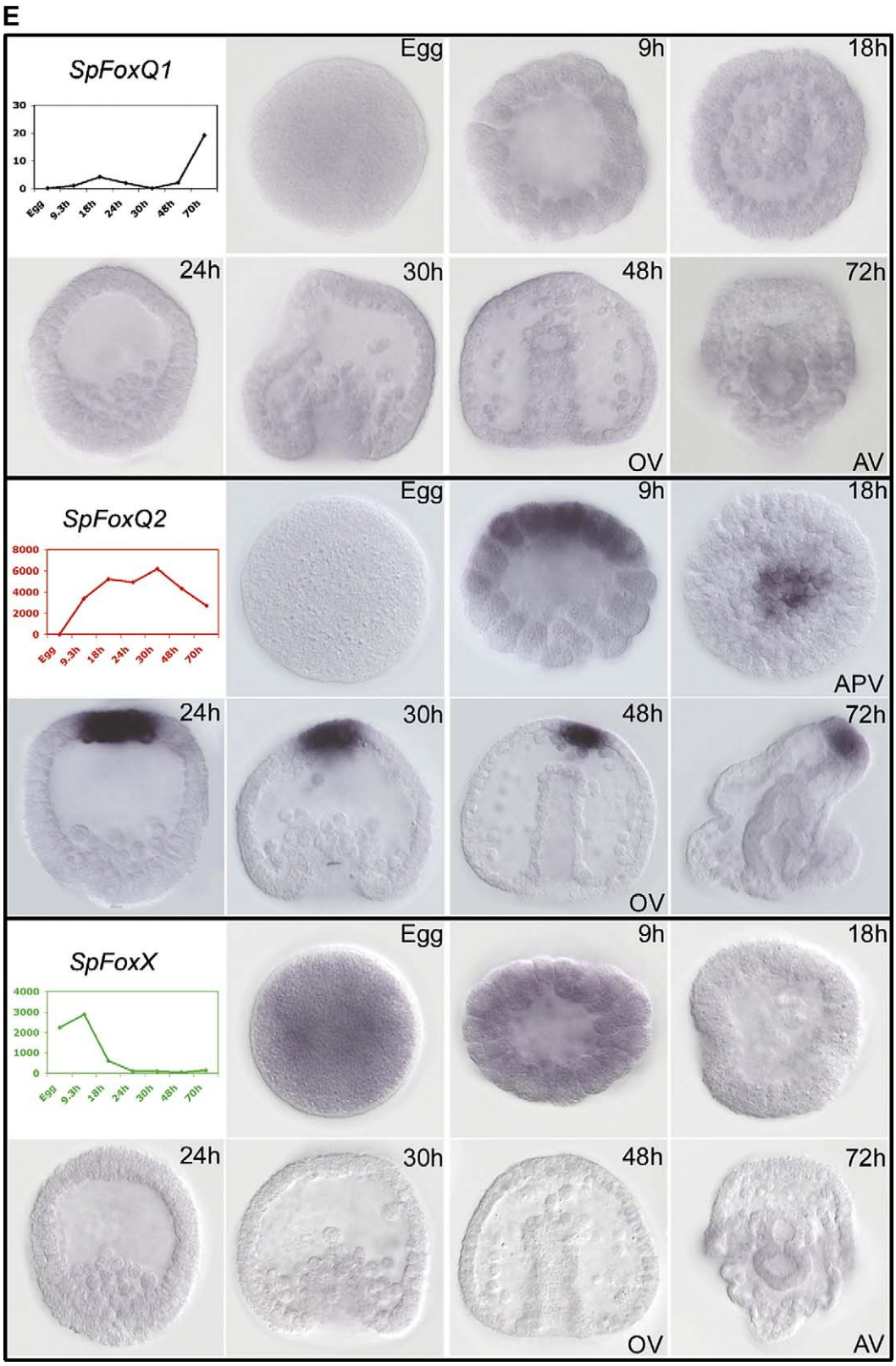



Fig. 3 (*continued*).

(9 h). The onset of expression for SpfoxG, *SpfoxM*, *SpfoxN2/3* and *SpfoxP* is at blastula stage, along with *SpfoxA* (Oliveri et al., 2006) and *SpfoxY* (Ransick et al., 2002). The *foxO* and *foxB* genes (Minokawa et al., 2004) start to be expressed at mesenchyme blastula. These early expressed genes are likely to function in embryonic specification processes. Expression of the rest of the *fox* genes can be detected only at later stages of development (after 30 h), during gastrulation and/or in the larva,

and they are possibly involved in cell differentiation. Not one of the genes displays a constant expression profile; all are developmentally regulated in specific ways.

For WMISH, DIG-labeled probes were transcribed from large (>500 bp) cDNA fragments (Table 1). For each experiment, *SpfoxA*, known to be expressed in the endoderm and oral ectoderm (Oliveri et al., 2006), was used as a positive control. The WMISH data shown in Fig. 3 are in excellent agreement with the temporal expression profiles determined by QPCR. In the following the expression patterns are considered by domain of transcription, though many genes show a dynamic expression pattern and at different stages are expressed in different cell types, a feature shared with the majority of transcription regulators of all other families (Howard-Ashby et al., (2006b)).

*Genes expressed in skeletogenic micromere/primary mesenchyme cells (PMCs)*

*SpfoxK*, *SpfoxN2/3* and *SpfoxO* are localized in this cell type. *SpfoxK* is one of the most highly expressed *fox* transcription factors. The amount of transcript increases steadily until a peak of about 9000 copies per embryo is reached at late gastrula stage, when this gene is expressed strongly in the PMCs. *SpfoxN2/3* is transiently expressed in the micromere lineage only at blastula stage, and *SpfoxO* starts its expression in PMCs at mesenchyme blastula and continues to be expressed in these cells until pluteus stage.

*Genes expressed in secondary mesenchyme cells (SMCs)*

Seven genes are expressed in the mesodermal cells. *SpfoxC*, *SpfoxF* and *SpfoxL1* begin to be expressed in this cell type late during gastrulation, and their transcripts are localized in the coelomic pouches. *SpfoxN2/3* is the earliest *fox* gene expressed in SMCs identified in this study. Its expression begins at mesenchyme blastula and shows a concentric pattern, staining the small micromeres at the center and a ring of the veg2 SMC, while the gene is not active in a torus of internally located secondary mesenchyme cells (Fig. 3). By late gastrula *SpfoxN2/3* transcripts cannot be detected in SMCs. *SpfoxJ2*, *SpfoxK* and *SpfoxO* are expressed in mesodermal cells by early gastrula. At late gastrula stage *SpfoxK* probe stains a group of cells budding off from the tip of the archenteron, while *SpfoxO*-positive cells are scattered in the blastocoel and possibly embedded in the aboral ectoderm. These are likely to be blastocoelar and pigment cells, two distinct populations of SMC.

*Genes expressed in endoderm*

The endoderm is the locus of expression of *SpfoxD*, *SpfoxI* and *SpfoxP*. The first two are activated at late gastrula and their transcripts are localized to the hindgut. *SpfoxI* shows a peculiar asymmetrical expression pattern with staining localized to the aboral side of the hindgut. *SpfoxD* on the other hand seems to be expressed at higher levels on the oral side of the hindgut. It is also expressed in a small patch of cells in the oral side of the foregut. *SpfoxP* is expressed in the vegetal plate at 24 h, and by the time gastrulation starts its expression is confined to the foregut and at pluteus stage the midgut.

*A gene expressed in the aboral ectoderm*

SpfoxK is the only regionally expressed gene to display aboral ectoderm transcripts. Initially (18 h) it is expressed throughout the whole ectoderm, excluding the apical plate, but later (30 h–48 h) expression becomes restricted to the aboral ectoderm.

*Genes expressed in the oral ectoderm and ciliated bands*

Four genes are expressed in the oral ectoderm territory, of which *SpfoxG* is the earliest. It is strongly expressed at 24 h in the oral ectoderm, but a faint staining with the same pattern can be observed already in some embryos at 18 h. The oral pattern persists until early gastrula and by the end of gastrulation *SpfoxG* expression is restricted to the ciliated bands. *SpfoxJ1*, *SpfoxL1* and *SpfoxO* are all expressed in the ciliated bands by the end of gastrulation and in larval stages. *SpfoxJ1* is transiently transcribed at mesenchyme blastula stage in one of the two ectodermal fields. *SpfoxL1* is also expressed in the cells surrounding the mouth of the pluteus.

*Genes expressed in the apical plate*

*SpfoxD*, *SpfoxJ1*and *SpfoxQ2* are expressed in all or a subset of cells derived from what is initially the animal pole of the embryo. The onset of expression differs considerably for each of these three genes. *SpfoxQ2* is expressed in all mesomeres at cleavage stage. By blastula stage it begins to be localized in the apical plate cells where it continues to be expressed until larval stage. *SpfoxJ1* is expressed in this territory starting from 24 h. The expression of *SpfoxJ1* at late gastrula seems to be reduced to the oral half of the apical plate, and a similar pattern is observed for *SpfoxD*.

*Ubiquitously expressed genes*

Seven genes have ubiquitous expression patterns at some times during development. The three maternal genes mentioned (i.e., *SpfoxJ2*, *SpfoxN1/4* and *SpfoxX*) are expressed ubiquitously in their maternal and early zygotic phases of expression. Similarly, *SpfoxJ1* is initially expressed everywhere, and then becomes localized to specific cell types (see above). *SpfoxAB-like*, *SpfoxL2* and *SpfoxM* are all zygotic genes and are expressed at different levels, respectively low, high and medium (Fig. 3 of Supplemental Material) but no localization in any cell types at any developmental time point is detected by WMISH.

From these data we may conclude that the sea urchin *fox* genes have widely variable expression patterns, and no bias in spatial or temporal expression pattern is observed for this gene family. Overall, 95.5% of the *fox* genes are utilized sometime in the first 72 h of development of the sea urchin. There is no correlation between level of expression and spatial localization.

## Discussion

*Evolutionary cooption of fox genes*

By combining phylogenetic analysis with the expression data in this paper, the evolutionary cooption of *fox* genes is revealed. Cooption is among the dominant mechanisms by which novel

regulatory architecture for development arises (reviewed by Davidson, 2001), and this process has affected utilization of the ancient *fox* gene family to a major extent. The usefulness of the global view afforded by a complete data set that includes all *S. purpuratus fox* genes is here enhanced by the fact that all but one of these genes (*SpfoxQ1*) is expressed in the embryo.

There are clear examples of *fox* genes belonging to panbilaterian families which display extremely conserved, dedicated assignment to given developmental pathways. The cardinal case is that of *SpfoxA*, which executes specific roles in the gene regulatory network controlling endoderm specification (Oliveri et al., 2006), and orthologues of which are required for endoderm development across the Bilateria. Another example in our dataset is the newly discovered *SpfoxQ2* gene, expressed in the embryonic apical plate. This is an early specified neurogenic territory of the sea urchin embryo (Yaguchi et al., 2006). In *Drosophila* a *foxQ2* ortholog (*fd102C*) is expressed in the anterior-most part of the embryo and later in the brain hemispheres (Lee and Frasch, 2004), and a similar expression pattern for a *foxQ2* orthologue is seen in amphioxus (Yu et al., 2003). But if instead of homologous utilization we focus on expression patterns that cannot be conserved, because they occur in clade-specific developmental contexts, we see that at least seven of the 16 regionally expressed *fox* genes have obviously been coopted to new functions during echinoderm or even echinoid evolution.

The mesoderm of the sea urchin embryo proper differs from that of all adult bilaterian body plans in that it consists exclusively of mesenchymal cell types. The only exception is the pluripotential cells of the small coelomic pouches, which are set aside for post-embryonic development of the adult body plan, and which play no role in development of the embryo itself. Coelomic organization in the embryo and larva, and its ultimately trimeric form, are in any case shared characters of the echinoderm–hemichordate superclade within the deuterostomes. Mesenchymal pigment cells of various kinds are common in bilaterians, but the particular regulatory pathway by which they arise in *S. purpuratus* (Ransick et al., 2002) is not universal even in echinoderms: in starfish the embryo does not make pigment cells at all, and the key regulatory gene required for their specification in sea urchins is not expressed in the equivalent domain of the embryo (Hinman and Davidson, unpublished observations). The secondary mesenchyme as a whole (an inclusive term including blastocoelar cells, pigment cells, and a few esophageal muscle cells) arises in a way that is again probably a shared character of echinoderms and hemichordates. The prominent skeletogenic (primary) mesenchyme of sea urchin embryos is an even more confined character, as only the thin-spined euechinoids generate a skeletogenic lineage from precociously specified micromeres. Thus *fox* genes that occur widely in the Bilateria, but that appear to participate in the specific developmental processes by which these mesodermal cell types arise, are all likely to have been coopted to these respective functions in the echinoderm–hemichordate, the echinoderm, or the echinoid lineages.

In this dataset *SpfoxB*, *SpfoxK*, *SpfoxN2/3* and *SpfoxO* are all expressed at one time or another specifically in skeletogenic mesenchyme cells. *SpfoxN2/3* is expressed so early that it is likely to function in the unique echinoid specification pathway for this cell type. *SpfoxC*, *SpfoxF*, and *SpfoxL1* are expressed in the coelomic pouches; *SpfoxK* in blastocoelar cells, and *SpfoxO* in pigment cells. All of these genes belong to panbilaterian families. Therefore, all of them are strong candidates for having undergone regulatory cooption events which targeted their transcription to the domains in which we detected their transcripts.

There is of course another mechanism by which novel developmental pathways may arise, and that is the clade-specific duplication of a gene followed by its paralogous divergence, and incorporation as a new gene in a new process. The *fox* gene family affords two possible cases of this as well. *SpfoxJ2* is apparently deuterostome-specific (it is absent in ecdysozoans, though there are no complete lophotrochozoan genomes to exclude the possibility that, the gene is also present in that group). This gene is expressed in the secondary mesenchyme and it is an allowable hypothesis that the evolution of the gene was part of the process by which the secondary mesenchyme developmental regulatory pathway evolved. The most striking example is *SpfoxY*. This gene is found only in sea urchins among genomes so far sequenced, and it is expressed precociously, and with exquisite specificity, in the small micromeres (Ransick et al., 2002). The small micromere lineage is a peculiarity of the echinoids, in which it contributes ultimately to the coelomic pouches. It will be very interesting to determine the mode of function and phylogenetic distribution of *foxY* genes within echinoderms.

### *New fox gene candidates for specific developmental functions*

Several of the observations in Fig. 3, taken together with data on *fox* gene function in other animals, suggest specific regulatory functions. For example, as remarked above, *SpfoxN2/3* is likely to participate in specification of the skeletogenic lineage. Another gene expressed later on during the construction of the skeleton, is *SpfoxO*. In other systems *foxO* is a target for signaling mediated by P13K/PDK1/PKB and initiated by insulin-like factors (for review, see Brunet et al., 2001; Kops and Burgering, 1999). It is known that in the sea urchin inhibition of P13K blocks skeletogenesis (Bradham et al., 2004), and *SpfoxO* is therefore a likely candidate to function as the transcriptional mediator downstream of the P13K signaling required in biomineralization.

The oral ectoderm-specific gene *SpfoxG* belongs to a class of genes known to function as inhibitors of TGF-β signaling. In mouse, FoxG1 stops the signal transduction cascade either by acting on FoxH protein or by binding to Smad factors (Dou et al., 2000; Rodriguez et al., 2001). In the sea urchin two members of the TGF-β family, Nodal and BMP2/4, are essential in oral–aboral ectoderm specification (Duboc et al., 2004). This suggests the nature of SpfoxG function in oral ectoderm specification.

Finally, genes of the M class are known to play a role in control of cell cycle. In other organisms *foxM* is target of cdk/cyclin and its expression promotes the transcription of genes essential for the progression of the cell cycle (for review, see Carlsson and

Mahlapuu, 2002). The non-localized expression of *SpfoxM* is consistent with this same role in the sea urchin embryo.

Overall, the main impression from this work is the tremendous diversity in the uses of this ancient and highly conserved class of regulatory genes. There are almost as many spatial and temporal patterns of expression in the embryo as there are *fox* genes in the genome.

## Acknowledgments

We thank Leha Vega for helping with the QPCR and Dr. T. Minokawa for the initial help with WMISH. We are also grateful to Dr. J. Nam for the critical support of phylogenetic analysis.

This research was supported by NIH grant HD37105 and by the Office of Science (BER), US Department of Energy, grant DE-FG02-03ER63584. P. Oliveri was supported by the Camilla Chandler Frost Fellowship.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2006.09.031.

## References

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

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Arnone, M.I., Rizzo, F., Annunziata, R., Cameron, R.A. and Martínez, P., this issue. *Hox* and *ParaHox* genes in *S. purpuratus*: colinearity rules. Dev. Biol. doi:10.1016/j.ydbio.2006.07.037.

Bradham, C.A., Miranda, E.L., McClay, D.R., 2004. PI3K inhibitors block skeletogenesis but not patterning in sea urchin embryos. Dev. Dyn. 229, 713–721.

Brunet, A., Datta, S.R., Greenberg, M.E., 2001. Transcription-dependent and -independent control of neuronal survival by the PI3K-Akt signaling pathway. Curr. Opin. Neurobiol. 11, 297–305.

Burge, C., Karlin, S., 1997. Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78–94.

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

Clark, K.L., Halay, E.D., Lai, E., Burley, S.K., 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature 364, 412–420.

David, E.S., Luke, N.H., Livingston, B.T., 1999. Characterization of a gene encoding a developmentally regulated winged helix transcription factor of the sea urchin *Strongylocentrotus purpuratus*. Gene 236, 97–105.

Davidson, E.H., 2001. Genomic regulatory systems. Development and Evolution. Academic Press, San Diego, CA.

Dou, C., Lee, J., Liu, B., Liu, F., Massague, J., Xuan, S., Lai, E., 2000. BF-1 interferes with transforming growth factor beta signaling by associating with Smad partners. Mol. Cell. Biol. 20, 6201–6211.

Duboc, V., Rottinger, E., Besnardeau, L., Lepage, T., 2004. Nodal and BMP2/4 signaling organizes the oral–aboral axis of the sea urchin embryo. Dev. Cell 6, 397–410.

Eddy, S.R., 1998. Profile hidden Markov models. Bioinformatics 14, 755–763.

Felsestein, J., 2005. PHYLIP (Phylogenetic Interference Package) version 3.6. distributed by the author. Department of Genome Sciences. University of Washington, Seattle.

Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S.R., Sonnhammer, E.L., Bateman, A., 2006. Pfam: clans, web tools and services. Nucleic Acids Res. 34, D247–D251.

Gajiwala, K.S., Burley, S.K., 2000. Winged helix proteins. Curr. Opin. Struct. Biol. 10, 110–116.

Guindon, S., Lethiec, F., Duroux, P., Gascuel, O., 2005. PHYML Online—A web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33, W557–W559.

Harada, Y., Akasaka, K., Shimada, H., Peterson, K.J., Davidson, E.H., Satoh, N., 1996. Spatial expression of a forkhead homologue in the sea urchin embryo. Mech. Dev. 60, 163–173.

Heglind, M., Cederberg, A., Aquino, J., Lucas, G., Ernfors, P., Enerback, S., 2005. Lack of the central nervous system- and neural crest-expressed forkhead gene Foxs1 affects motor function and body weight. Mol. Cell. Biol. 25, 5616–5625.

Hibino, T., Ishii, Y., Levin, M., Nishino, A., 2006. Ion flow regulates left–right asymmetry in sea urchin development. Dev. Genes Evol. 216, 265–276.

Howard-Ashby, M., Materna, S.C., Brown, C.T., Chen, L., Cameron, R.A., Davidson, E.H., 2006a. Identification and characterization of homeobox transcription factor genes in *S. purpuratus*, and their expression in embryonic development. Dev. Biol. 300, 74–89.

Howard-Ashby, M., Materna, S.C., Brown, C.T., Tu, Q., Oliveri, P., Cameron, R.A., Davidson, E.H., 2006b. High regulatory gene use in sea urchin embryogenesis: implication for bilaterian development and evolution. Dev. Biol. 300, 27–34.

Howard-Ashby, M., Materna, S.C., Brown, C.T., Chen, L., Cameron, R.A. and Davidson, E.H., this issue. Other transcription factors in *S. purpuratus*. Dev. Biol.

Hubbard, T., Andrews, D., Caccamo, M., Cameron, G., Chen, Y., Clamp, M., Clarke, L., Coates, G., Cox, T., Cunningham, F., Curwen, V., Cutts, T., Down, T., Durbin, R., Fernandez-Suarez, X.M., Gilbert, J., Hammond, M., Herrero, J., Hotz, H., Howe, K., Iyer, V., Jekosch, K., Kahari, A., Kasprzyk, A., Keefe, D., Keenan, S., Kokocinsci, F., London, D., Longden, I., McVicker, G., Melsopp, C., Meidl, P., Potter, S., Proctor, G., Rae, M., Rios, D., Schuster, M., Searle, S., Severin, J., Slater, G., Smedley, D., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Storey, R., Trevanion, S., Ureta-Vidal, A., Vogel, J., White, S., Woodwark, C., Birney, E., 2005. Ensembl. Nucleic Acids Res. 33, D447–D453.

Kaestner, K.H., Knochel, W., Martinez, D.E., 2000. Unified nomenclature for the winged helix/forkhead transcription factors. Genes Dev. 14, 142–146.

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

Kaufmann, E., Hoch, M., Jackle, H., 1994. The interaction of DNA with the DNA-binding domain encoded by the *Drosophila* gene fork head. Eur. J. Biochem. 223, 329–337.

Kops, G.J., Burgering, B.M., 1999. Forkhead transcription factors: new insights into protein kinase B (c-akt) signaling. J. Mol. Med. 77, 656–665.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

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

Letunic, I., Copley, R.R., Pils, B., Pinkert, S., Schultz, J., Bork, P., 2006. SMART 5: domains in the context of genomes and networks. Nucleic Acids Res. 34, D257–D260.

Li, C., Tucker, P.W., 1993. DNA-binding properties and secondary structural model of the hepatocyte nuclear factor 3/fork head domain. Proc. Natl. Acad. Sci. U. S. A. 90, 11583–11587.

Magie, C.R., Pang, K., Martindale, M.Q., 2005. Genomic inventory and expression of Sox and Fox genes in the cnidarian *Nematostella vectensis*. Dev. Genes Evol. 215, 618–630.

Marsden, I., Jin, C., Liao, X., 1998. Structural changes in the region directly adjacent to the DNA-binding helix highlight a possible mechanism to explain the observed changes in the sequence-specific binding of winged helix proteins. J. Mol. Biol. 278, 293–299.

Materna, S.C., Howard-Ashby, M., Gray, R., Davidson, E.H., 2006. The C2H2

zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development. Dev. Biol. 300, 108–121.

Mazet, F., Yu, J.K., Liberles, D.A., Holland, L.Z., Shimeld, S.M., 2003. Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. Gene 316, 79–89.

Minokawa, T., Rast, J.P., Arenas-Mena, C., Franco, C.B., Davidson, E.H., 2004. Expression patterns of four different regulatory genes that function during sea urchin development. Gene Expression Patterns 4, 449–456.

Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McDowall, J., Mitchell, A., Nikolskaya, A.N., Orchard, S., Pagni, M., Ponting, C.P., Quevillon, E., Selengut, J., Sigrist, C.J., Silventoinen, V., Studholme, D.J., Vaughan, R., Wu, C.H., 2005. InterPro, progress and status in 2005. Nucleic Acids Res. 33, D201–D205.

Oliveri, P., Davidson, E.H., 2004. Gene regulatory network analysis in sea urchin embryos. Methods Cell Biol. 74, 775–794.

Oliveri, P., Carrick, D.M., Davidson, E.H., 2002. A regulatory gene network that directs micromere specification in the sea urchin embryo. Dev. Biol. 246, 209–228.

Oliveri, P., Walton, K., Davidson, E.H., McClay, D.R., 2006. Repression of mesodermal fate by foxa, a key endoderm regulator of the sea urchin embryo. Development 133, 4173–4181.

Pierrou, S., Hellqvist, M., Samuelsson, L., Enerback, S., Carlsson, P., 1994. Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending. EMBO J. 13, 5002–5012.

Pohl, B.S., Knochel, W., 2005. Of Fox and Frogs: Fox (fork head/winged helix) transcription factors in *Xenopus* development. Gene 344, 21–32.

Ransick, A., Rast, J.P., Minokawa, T., Calestani, C., Davidson, E.H., 2002. New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. Dev. Biol. 246, 132–147.

Rast, J.P., Amore, G., Calestani, C., Livi, C.B., Ransick, A., Davidson, E.H., 2000. Recovery of developmentally defined gene sets from high-density cDNA macroarrays. Dev. Biol. 228, 270–286.

Rizzo, F., Fernandez-Serra, M., Squarzoni, P., Archimandritis, A., Arnone, M.I., 2006. Identification and developmental expression of the ets gene family in the sea urchin (*Strongylocentrotus purpuratus*). Dev. Biol. 300, 35–48.

Rodriguez, C., Huang, L.J., Son, J.K., McKee, A., Xiao, Z., Lodish, H.F., 2001. Functional cloning of the proto-oncogene brain factor-1 (BF-1) as a Smad-binding antagonist of transforming growth factor-beta signaling. J. Biol. Chem. 276, 30224–30230.

Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132, 365–386.

Samanta, M.P., Tongprasit, W., Istrail, S., Cameron, R.A., Tu, Q., Davidson, E.H. and Stolc, V., in press. A high-resolution transcriptome map of the sea urchin embryo. Science

Stroud, J.C., Wu, Y., Bates, D.L., Han, A., Nowick, K., Paabo, S., Tong, H., Chen, L., 2006. Structure of the forkhead domain of FOXP2 bound to DNA. Structure 14, 159–166.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Weigel, D., Jurgens, G., Kuttner, F., Seifert, E., Jackle, H., 1989. The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo. Cell 57, 645–658.

Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Mazumder, R., O'Donovan, C., Redaschi, N., Suzek, B., 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res. 34, D187–D191.

Yagi, K., Satou, Y., Mazet, F., Shimeld, S.M., Degnan, B., Rokhsar, D., Levine, M., Kohara, Y., Satoh, N., 2003. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*: III. Genes for Fox, ETS, nuclear receptors and NFkappaB. Dev. Genes Evol. 213, 235–244.

Yaguchi, S., Nakajima, Y., Wang, D., Burke, R.D., 2006. Embryonic expression of engrailed in sea urchins. Gene Expression Patterns.

Yu, J.K., Holland, N.D., Holland, L.Z., 2003. AmphiFoxQ2, a novel winged helix/forkhead gene, exclusively marks the anterior end of the amphioxus embryo. Dev. Genes Evol. 213, 102–105.