Available online at www.sciencedirect.com
ScienceDirect
Developmental Biology 300 (2006) 108–120
DEVELOPMENTAL BIOLOGY
www.elsevier.com/locate/ydbio

The C_2H_2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development

Stefan C. Materna, Meredith Howard-Ashby, Rachel F. Gray, Eric H. Davidson *

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Received for publication 25 May 2006; revised 4 August 2006; accepted 16 August 2006
Available online 22 August 2006

Abstract

The C_2H_2 zinc finger is one of the most abundant protein domains and is thought to have been extensively replicated in diverse animal clades. Some well-studied proteins that contain this domain are transcriptional regulators. As part of an attempt to delineate all transcription factors encoded in the *Strongylocentrotus purpuratus* genome, we identified the C_2H_2 zinc finger genes indicated in the sequence, and examined their involvement in embryonic development. We found 377 zinc finger genes in the sea urchin genome, about half the number found in mice or humans. Their expression was measured by quantitative PCR. Up to the end of gastrulation less than a third of these genes is expressed, and about 75% of the expressed genes are maternal; both parameters distinguish these from all other classes of regulatory genes as measured in other studies. Spatial expression pattern was determined by whole mount *in situ* hybridization for 43 genes transcribed at a sufficient level, and localized expression was observed in diverse embryonic tissues. These genes may execute important regulatory functions in development. However, the functional meaning of the majority of this large gene family remains undefined.

Keywords: Zinc finger; Zf-c2h2; Transcription factor; Genome; Sea urchin; Development

Introduction

Zinc finger motifs are of particular interest in developmental biology because they occur in some prominent transcriptional regulators. Though there are more than seventy classes of zinc-binding motifs listed in the PFAM database, the specific transcriptional regulators fall mainly in the C_2H_2 zinc finger class, in which the zinc atom is complexed by two cysteines and two histidines, and in their structural relatives the C_4 zinc finger class, in which the zinc is complexed with four cysteines. The latter group consists mainly of nuclear hormone receptors and GATA factors (Krishna et al., 2003). In most animal genomes that have been sequenced, C_2H_2 zinc fingers are among the more abundant protein domains. This applies in particular to mammalian genomes, in which C_2H_2 zinc finger genes have been highly multiplied (Lander et al., 2001; Rubin et al., 2000). The C_2H_2 zinc finger genes usually far outnumber the zinc fingers of the C_4 type, of which most genomes contain only a few dozen. A prominent exception is *Caenorhabditis elegans*, where the nuclear hormone receptors have undergone extensive multiplication (Reece-Hoyes et al., 2005) and outnumber C_2H_2 zinc finger genes. Although sequence-specific DNA-binding proteins can be found in other groups of zinc fingers, specific transcription factors are rare, and these proteins are often part of the basal transcription apparatus or DNA repair machinery.

C_2H_2 zinc finger proteins are commonly viewed as transcriptional regulators, but they may be widely used for RNA binding. This is exemplified by the first known zinc finger transcription factor, *Xenopus* TFIIIa, which binds specifically to both DNA and RNA (Lu et al., 2003). Possibly just as typical for genes with higher numbers of zinc fingers is the *Xenopus xfin* gene, which codes for a protein with 37 zinc fingers. It is localized in the cytoplasm and has been shown to bind to RNA (Andreazzoli et al., 1993). Transcriptional regulatory activity has not been demonstrated for this protein. In addition to DNA and RNA binding, zinc finger domains may also be used for protein–protein interactions (Laity et al., 2001). Several examples are known, including well-known transcriptional regulators. For example, two of the five conserved zinc fingers of the

* Corresponding author. Fax: +1 626 793 3047.
E-mail address: Davidson@caltech.edu (E.H. Davidson).

0012-1606/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2006.08.032

Gli family transcription factor encoded by the *cubitus interruptus* gene are needed for specific interaction with another factor, converting the protein into its active form (Croker et al., 2006). C_2H_2 zinc finger genes are thought to account for 30%–50% of all transcription factors in metazoan genomes (Adams et al., 2000; Ruvkun and Hobert, 1998). However, except for a minority which is clearly orthologous to known regulatory factors, there are no canonical criteria which suffice to distinguished those C_2H_2 proteins that are dedicated sequence-specific transcription factors from those that bind RNA or perform other functions.

Regulatory genes of the nuclear hormone receptor and GATA classes in the sea urchin genome have been characterized by Howard-Ashby et al. (2006a; this issue). Here, in order to encompass the major remaining class of transcription factors, we identify all C_2H_2 zinc finger genes predicted by the genomic sequence, and determine their activity during development.

Materials and methods

Identification of zinc finger genes

Zinc finger genes were identified in the contig assembly by searching for the C_2H_2 zinc finger motif. We built a calibrated Hidden–Markov model from the PFAM seed alignment (PF00096, www.sanger.ac.uk/Software/Pfam/) and searched the sea urchin genome with *hmmsearch* (http://hmmer.wustl.edu/) accepting only domains with an *E*-value <0.1. After release of the gene predictions, the identified zinc finger genes were mapped onto the GLEAN models (Sea Urchin Sequencing Consortium, submitted) by finding near perfect matches to the calculated QPCR amplicon. On genes that did not perfectly match a GLEAN model, a BLAST search was performed against the remainder of the GLEAN gene predictions. The results were manually inspected and associations validated. Less than perfect matches are due to incorporation of sequences of different haplotypes in the scaffold assembly. Presumed zinc finger genes that did not match any GLEAN model were searched by BLAST against novel predictions from the whole genome tiling array (Samanta et al., submitted), identifying additional gene models for genes that are expressed in early development. Gene models were aligned with the contigs using the *spidey* genomic mapping program (Wheelan et al., 2001), and the match was validated through manual inspection.

Phylogenetic analysis

For identification of orthologous genes, we obtained the set of C_2H_2 zinc finger containing protein sequences of *Homo sapiens*, *Mus musculus*, *Ciona intestinalis*, *C. elegans*, *Drosophila melanogaster*, and *Nematostella vectensis*. Sequences of *Nematostella* were obtained from *Stellabase* (www.stellabase.org), *C. elegans* sequences from *Wormbase* (www.wormbase.org, WSWS156), and all others from *Ensembl* (www.ensembl.org, v.37—February 2006) by motif search for Interpro domain IP:007087. To obtain a nonredundant set of proteins, we kept only the longest protein and discarded shorter isoforms for any given gene. A BLASTP search was performed for each sea urchin protein against this set of C_2H_2 zinc finger proteins. Good hits were confirmed by manual inspection. For such genes the zinc finger region together with surrounding conserved sequence was excised and aligned using the *mafft* alignment program (Katoh et al., 2005) using 1000 iterations. Phylogenetic analysis was conducted using the neighbor-joining method with the MEGA program (Kumar et al., 2004). The calculated distance was Poisson-corrected, gaps were pairwise deleted, and 1000 iterations were used for calculating bootstrap values.

Transcriptional profiling

We performed transcriptional profiling using quantitative PCR (QPCR). QPCR is a comparative method, in which the accumulation of PCR product is monitored for a gene of interest and in the same sample for a given standard, through the use of a double strand-specific fluorescent dye. By choosing a threshold and determining $C_t\Delta$ (the difference in cycle number at which each PCR reaction crosses the threshold) the initial prevalence of a gene can be calculated, since the cycle difference is proportional to the abundance in the original reaction mix (Wong and Medrano, 2005).

For primer design, the sometimes short stretch of sequence containing the zinc finger domains was extended using a BLASTX search against the NCBI database of non-redundant proteins "*nr*" (www.ncbi.nlm.nih.gov). QPCR primers lying within these regions were obtained using the standalone version of *primer*3 (Rozen and Skaletzky, 2000). Primers were chosen to yield an amplicon of between 110 and 140 base pairs. Primers were tested for specificity on genomic DNA by QPCR. It was assumed that all genes dealt with were single copy. In this case, given equal amplification efficiency, all PCR products should accumulate to a given threshold at roughly the same cycle (C_t). Primer pairs that did not produce an acceptable C_t value (one cycle more or less compared to the mean C_t) were not used for transcriptional profiling and were redesigned. A more exact determination of primer efficiency was conducted for a representative set of primer pairs using serial dilutions (Wong and Medrano, 2005), and it confirmed the initial findings. We therefore generally assumed an amplification efficiency of 1.95. The presence of a single specific band was confirmed by gel electrophoresis.

Embryos were grown and harvested at fertilization, and at 6, 12, 18, 24, 36, and 48 h postfertilization. RNA was isolated with the Qiagen RNeasy Mini-Kit. RT reactions were performed with ABI (Foster City, USA) TaqMan cDNA synthesis kit according to manufacturer's instructions. QPCR was conducted on an ABI 7900 HT with ABI SYBR-Green reaction mix, using the following program: 1× (95°C–10 min,) 40× (60°C–30 s, 95°C–1 min). At the end of each program a dissociation curve was collected to confirm that only one product accumulated during the reaction. A no-template control also assured that no primer–dimers had formed. The RNA copy number was determined by calculating the $C_t\Delta$ for a given zinc finger gene with respect to the poly-ubiquitin gene, which was assumed to be represented by 88,000 transcripts per embryo during the developmental stages examined (Nemer et al., 1991). On each plate, each primer–cDNA combination was run in triplicate. The experiment was repeated once with the same cDNA and twice with cDNA from a second animal. For data analysis, the C_ts of wells that obviously did not amplify were omitted. $C_t\Delta$ was calculated for the triplicates of the four independent runs. The mean of the averages from the four runs was used to calculate the number of transcripts per embryo. Error bars were calculated from the standard deviation on the mean.

Whole mount in situ *hybridization*

In order to identify the spatial domain of expression for the higher expressed genes, we conducted *in situ* hybridization with digoxigenin-labeled antisense probes. From the general assumption that about 10 RNA molecules per cell are needed for sufficient staining, it follows that a minimum of several hundred molecules per embryo is needed to obtain a clear stain since spatially restricted regions like the endoderm contain no fewer than 60 cells prior to gastrulation.

For successful *in situ* hybridization using these methods, the probe should be a minimum of 600 base pairs in length. We attempted to obtain primer pairs by using either conserved sequence that is recognizably located within one exon, or, if no sequence of sufficient length could be obtained, by assuming the GLEAN gene models. Templates for *in situ* probes were amplified from cDNA using primers tailed with *Sp6* and *T7* promotors or subcloned into the pGEM-T Easy vector which contains *Sp6* and *T7* promotor sites adjacent to the multiple cloning site. Sequencing confirmed the identity of the gene. After digoxigenin labeling through *in vitro* transcription with Roche *Sp6* or *T7* polymerase, probes were run on a denaturing gel, confirming the size of the transcript. Whole mount *in situ* hybridizations were carried out according to Minokawa et al. (2004).

Results

Identification of C_2H_2 zinc finger-containing genes

We initially set out to identify all transcription factor genes using a BLAST-based approach, searching the trace archive of

the sea urchin genome project for domains commonly associated with gene regulatory activity (Howard-Ashby et al., 2006a). Identified traces were clustered by BLASTX against our own reference database. However, this method failed for C_2H_2 zinc finger-containing genes. Due to their tendency to contain multiple zinc finger domains, and the high degree of structural conservation between domains, trace sequences could not be binned unambiguously. Any two zinc finger domains on average are about 40% identical at the protein sequence level. The amino acids that are important for structural integrity, and a very conserved linker of six to seven amino acids that frequently connects zinc finger domains, are responsible for this high level of identity (Knight and Shimeld, 2001).

By searching the genome for C_2H_2 zinc finger motifs with a Hidden–Markov model (see Materials and methods), we identified 377 genes, of which most corresponded to predicted GLEAN gene models (Sea Urchin Sequencing Consortium, submitted). Of these 377 genes, 17 match a gene model that emerged through inspection of the whole genome tiling array results (Samanta et al., submitted). For 50 of the genes that were not among the original gene model predictions, we were able to obtain a primer pair with which an authenticated zinc finger amplicon could be generated from genomic DNA. Transcriptional profiling revealed that ten of these genes are transcribed, indicating that they are functional genes. The conserved domain structure of the remaining genes indicates that they too are more likely than not functional genes, and we have therefore included them in our list of identified zinc finger genes. Thus, though provisionally, we conclude that the sea urchin genome contains 377 C_2H_2 zinc finger genes.

Rapidly expanding gene families often contain a high number of pseudogenes (Glusman et al., 2001). However, these usually decay rapidly, accumulating stop codons and small deletions (Zhang and Gerstein, 2004). If generated through retrotransposition events, they consist of only one exon. Although single exon genes are common in our data set, they generally have long open reading frames >1 kb. We cannot exclude that the list of 377 C_2H_2 zinc finger genes includes some pseudogenes, but consider this unlikely.

The sea urchin genome contains more zinc finger genes than found in other invertebrates. There are 326 C_2H_2 zinc finger genes in the *Drosophila* genome (Chung et al., 2002), 198 in the urochordate *C. intestinalis* (Miwata et al., 2006), and 211 in *C. elegans* (Reece-Hoyes et al., 2005). In contrast to all other species, *C. elegans* contains more *nuclear hormone receptor*-like genes than C_2H_2 zinc finger genes. In the current version of the *N. vectensis* genome, we identified 170 zinc finger genes. The *Nematostella* genome, however, is at this point still in a provisional state. A search for C_2H_2 zinc finger genes in a non-redundant set of mouse proteins, applying our criteria, yielded 731 genes. A similar search in the human genome identified 764 genes. These numbers are consistent with other current estimates (Shannon et al., 2003; Tupler et al., 2001). The number of zinc finger genes in vertebrates far surpasses the number of zinc finger genes in the sea urchin genome, but this is a vertebrate not a chordate feature.

Properties of sea urchin zinc finger genes

The PFAM search identified over 3000 individual zinc finger domains, making this one of the most prominent protein domains in the sea urchin genome (Materna et al., 2006). Generally C_2H_2 zinc finger domains appear in tandem. Often they are separated by only a few amino acids the sequence of which is frequently TGEKPY/F, as has been described in other genomes (Laity et al., 2001). A few exceptions exist in which the domains are dispersed throughout the entire coding region. In the sea urchin genome, the median number of zinc finger domains per gene is eight.

The zinc finger motifs of about 10% of all identified zinc finger genes appear to be extremely closely related at the nucleotide level, giving them a repetitive structure consisting of exact repeated sequence units of >40 continuous base pairs (the entire zinc finger domain is 66 base pairs long). This causes obvious problems for primer design and transcriptional profiling by QPCR. However, where working primer pairs could be obtained, our results show that some of these genes are indeed expressed (e.g., *Sp-z410*, *Sp-z265*). We conclude that these genes are not artifacts of the assembly.

In most zinc finger genes, the C_2H_2 zinc finger motif is the only recognizable domain. Infrequently zinc finger genes contain a second kind of domain (about 16%), mostly other types of zinc binding domains, like the BED zinc finger. Two sea urchin zinc finger genes also encode homeodomains (*Sp-SmadIP1, Sp-atbf15*) (Howard-Ashby et al., 2006b). In mammalian genomesm, zinc finger factors frequently have repressive function, which is attributed to the presence of a Krüppel-associated-box (KRAB) domain. Almost half of all human and mouse C_2H_2 zinc finger factors contain this domain (Urrutia, 2003). A search for the KRAB domain (PF01352) was negative in the sea urchin genome; similarly no convincing SCAN domains (PF02023) (Williams et al., 1999) were found. Although the possibility remains that distant relatives of these domains are present in the sea urchin genome, this finding is consistent with the hypothesis that they arose after the divergence of the tetrapod lineage. About a third of all C_2H_2 zinc fingers in the *Drosophila* genome contain a ZAD domain which we could not identify in the sea urchin genome either (PF07776). It is thought to fulfill similar functions as the KRAB domain (Chung et al., 2002) and supposed to be dipteran-specific. The lack of a recognized, repressive domain in sea urchin zinc finger factors might suggest that a novel taxon-specific domain remains to be found in these proteins.

Transcriptional profiling

To determine the set of C_2H_2 zinc finger genes that is active in early development, we performed transcriptional profiling by QPCR. Expression of a total of 324 genes for which specific primer pairs could be obtained was monitored at seven time points between fertilization and late gastrulation (48 h post-fertilization). We consider genes to be transcribed significantly if they are represented by at least 200 transcripts per embryo. This excludes genes that even if expressed in as few as 40 cells in the

800-cell late gastrula will be represented by only 5 mRNAs per cell, and if expressed ubiquitously, 0.25 mRNAs per cell. According to this classification, only 112 of the genes are significantly transcribed, i.e., 35% of all zinc finger genes tested. A colorimetric summary of the expression dynamics for these 112 genes is displayed in Fig. 1 (individual time courses can be viewed at our website, http://sugp.caltech.edu). The overall transcriptional level of expressed zinc finger genes is relatively low, except for a few cases (red and yellow in Fig. 1). Typically, at peak levels of expression, there are but several hundred molecules per embryo. More than 95% of genes we classify as transcribed were also identified as transcribed in the whole genome tiling array analysis (Samanta et al., submitted). Furthermore, the present QPCR analysis of three previously known C_2H_2 zinc finger genes, viz. *Sp-Z12/z151* (Wang et al., 1995), *Sp-z13* (previously *Sp-krl*; Howard et al., 2001), and *Sp-blimp/krox/z51* (Livi and Davidson, 2006) was consistent with previously determined transcriptional profiles.

Eighty-three genes, that is, 74% of significantly expressed genes, are represented in maternal RNA, indicating a strong bias toward oogenetic transcription for zinc finger genes. Twenty-seven genes (24%) are expressed maternally only and are not reused during early development (Fig. 2). Thirty genes (27%) are expressed maternally and continue to be transcribed throughout early development, albeit showing fluctuations in transcript level (e.g., *Sp-spalt/z54*, *Sp-rreb/z48*, Fig. 1). Twenty-six maternal genes (23%) initiate a second expression phase after falling below our threshold of 200 molecules and are likely to have a zygotically regulated activation in addition to maternal expression (e.g., *Sp-atbf1/z30* or *Sp-z50*, Fig. 1). Twenty-nine genes (26%) are not maternally expressed and begin to be transcribed only in the course of development. Together, the "maternal only" and "constantly expressed" genes account for more than half of all expressed genes (51%, Fig. 2). Only four genes are activated for the first time at 6 h, but genes which are maternally expressed may well be transcribed at this point too. A plurality of the expressed genes, discounting those that are "constantly expressed", begin zygotic expression in the 24–36 h window (Fig. 2). This corresponds to the time prior to and overlapping with gastrulation. By 48 h, only five additional genes have begun to be transcribed.

In situ *hybridization*

We attempted to perform *in situ* hybridization for all the more highly expressed genes, and succeeded in collecting spatial expression data for 43 genes (summarized in Table 1). Localized expression was found for 22 genes during at least one phase of development, while 21 genes were observed to be expressed ubiquitously throughout. Patterns of expression for genes utilized

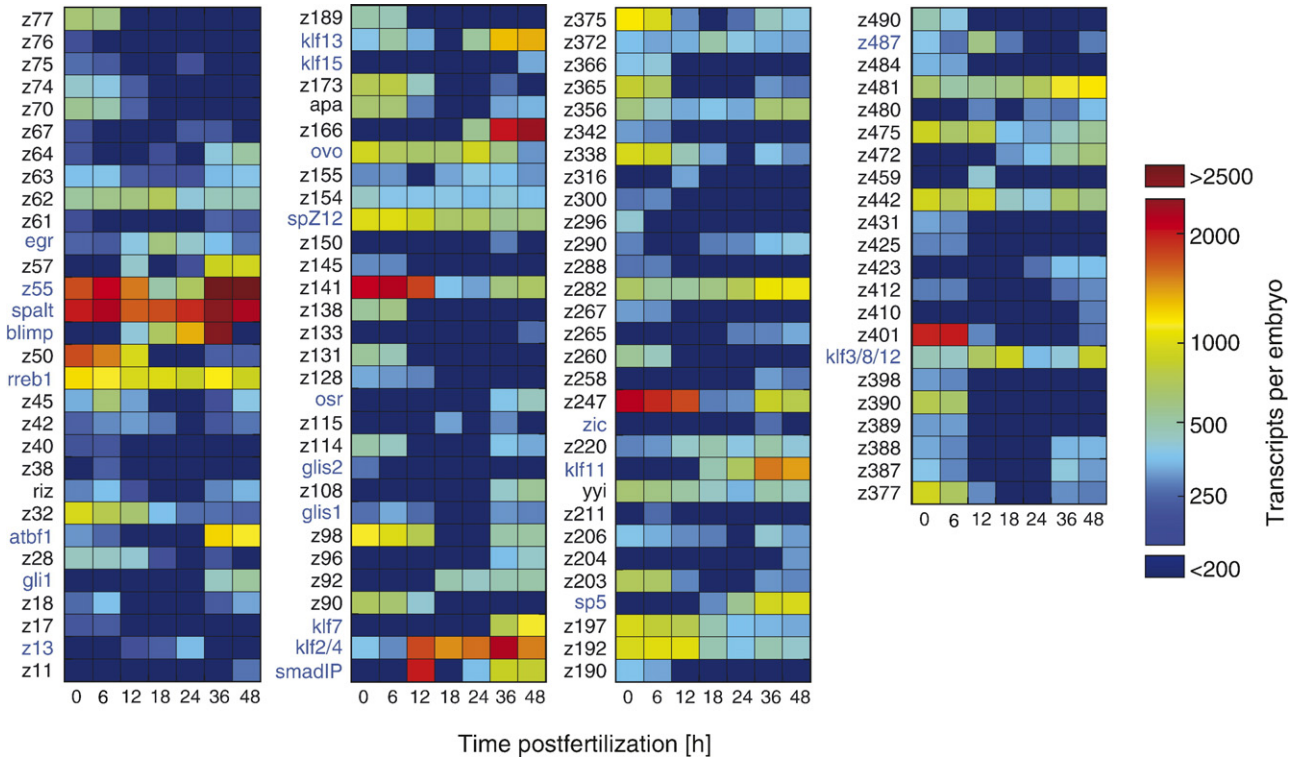

Fig. 1. Summary of transcriptional profiling data for 112 expressed genes. For each gene, the expression level was determined by QPCR at seven time points from fertilization to 48 h, the late gastrula stage. The color at each time point corresponds to the expression level in transcripts per embryo as indicated in the key. Zinc finger genes whose expression exceeds 200 transcripts per embryo at one or more time points were considered to be expressed, and only these genes are included in this figure. The majority are not expressed in this period of development (see text). Most zinc finger genes are expressed at low levels, such that they are represented by no more than several hundred transcripts per embryo at their peaks. Seventy-five percent of expressed zinc finger genes are expressed maternally. *Sp-blimp*, *Sp-z13* (previously *Sp-krl*) and *Sp-SpZ12* are previously known genes (*op. cit.*). *Sp-atbf1* and *Sp-smadIP* also contain homeobox domains in addition to the zinc finger domains (see Howard-Ashby et al., 2006a). Names of genes referred to in the text are colored blue.

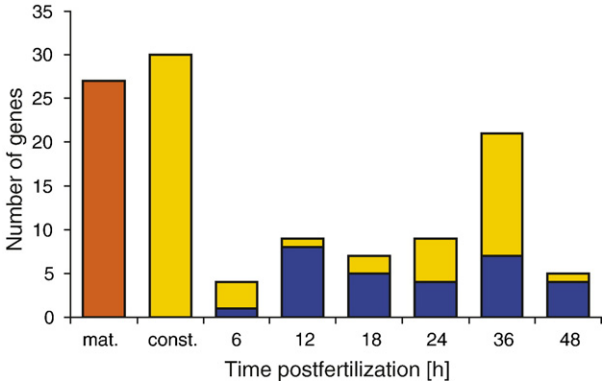

Fig. 2. Stages of initial activation of C_2H_2 zinc finger genes. Genes are classified by the start of zygotic expression (6 to 48 h postfertilization). Genes that are only zygotically expressed are depicted in blue, while genes that are maternally and zygotically expressed are yellow. Genes that are maternal and expressed throughout early development (allowing for some variation in expression level) are captured by the "constant" category ("const."). Genes that are maternal and fall below our threshold of 200 transcripts per embryo, i.e., are not expressed, but later have a phase of zygotic expression, are classified by the start of zygotic expression. Genes that are maternal only and not expressed zygotically are depicted in orange ("mat."). The 6 h category represents genes that display a clear increase in the transcript number at this point. Although "constant genes" may begin to be transcribed as well at this or at later times, unless there is net transcript accumulation this will not be visible.

specifically in the endomesodermal domains are shown in Fig. 3, and for genes transcribed detectably in ectodermal and apical domains in Fig. 4.

A highly confined spatial expression pattern is presented by *Sp-osr/z121*. It is expressed during gastrulation and is localized in what will become the midgut, possibly only at the midgut–hindgut boundary (Fig. 3D). Many other genes are also expressed in particular regions of the archenteron (Figs. 3H, J, L, P, T), though some are also expressed in ectodermal or apical domains. *Sp-klf13/z188* is expressed in the endodermal region at swimming blastula stage (Figs. 3Q, R), but becomes localized to the ectoderm after the beginning of gastrulation (Fig. 3S). The *Sp-z13* gene, previously known as *Sp-krl* (Howard et al., 2001), displays a particularly dynamic pattern of expression. It is first transcribed at ~7 h postfertilization, in the micromeres. During the 12 to 18 h period, its transcripts disappear from the cells of the skeletogenic lineage and become localized first to the endomesoderm and then in the mesoderm proper (Figs. 3A, B). After ingression of the primary mesenchyme cells, expression disappears from the mesodermal tissue and extends to the endodermal region (Fig. 4C).

Genes that are expressed in mesodermal tissues include *Sp-z166* which is localized in the vegetal plate after the ingression of primary mesenchyme cells (Fig. 4O). Similarly, after an initial phase of ubiquitous expression, *Sp-ovo/z157* is transcribed in the vegetal plate including the mesoderm at swimming blastula stage (Figs. 3M, N).

Expression of several genes localizes to different regions of the ectoderm. This includes genes expressed in the apical region (Figs. 4A–D, F) and genes expressed in either the oral or aboral ectoderm (e.g., Figs. 4H–K, M–T). *Sp-z487* is a gene that is expressed only transiently in embryonic development. It is initially expressed ubiquitously, and then becomes restricted to the ectoderm only, without showing any oral or aboral bias (Fig. 4L).

Identification of zinc finger orthologues

Due to the rapid evolution of zinc finger genes, it is often difficult to recognize orthology between different proteins even where it does exist. We identified zinc finger genes that belong to orthologous groups of highly conserved and well studied genes (Table 2). Even for these groups, only the zinc finger region can be aligned confidently, sometimes including some flanking sequence. The number of zinc finger domains per gene is known to be variable even between fairly closely related zinc fingers (Shannon et al., 2003). However, within the orthologous gene sets presented in this section, the number of zinc finger domains is invariant. All genes described here for which we identified sea urchin homologues are known transcription factors, involved in various developmental processes.

Zinc fingers of the gli *and* zic *families*

Gli (Glioma associated oncogene) proteins are known to play important roles in development, for example in promoting neuronal differentiation (Mayor and Aybar, 2001). They are transcriptional mediators of the hedgehog (Hh) signaling cascade. Gli proteins contain five zinc finger domains that are highly conserved. Human and mouse genomes both contain three *gli* genes. *Drosophila* contains one, the segment polarity gene *cubitus interruptus*, as does *C. elegans* (*Tra-1*). Two groups of *gli-similar* genes (*glis*) have been described. Mouse and human contain two genes of the *glis1* subfamily to which the *Drosophila* gene *lmd* also belongs. The *glis2* subfamily contains one group each of mouse, human and fly (*Dm-sug*) genes. According to the phylogenetic tree in Fig. 5, the sea urchin contains one gene from each group (*Sp-gli/z22*, *Sp-glis1/z113*, *Sp-glis2/z107*) indicating that these subgroups had already appeared before the divergence of the vertebrate lineage. Transcriptional profiling shows that *Sp-gli* is transcribed at 36 h postfertilization. The two *glis* genes are expressed maternally and are transcribed at low levels in the embryo.

Human and mouse genomes contain at least five *zic* (zinc finger genes of the cerebellum) genes. In *Drosophila* only one such gene can be found (*odd-paired/Dm-opa*). Ascidians contain at least two genes in this group, *Ci-macho1* and *Ci-zicL*, that seem to have diverged in this lineage (Yamada et al., 2003). *zic* genes in vertebrates are located in a cluster, underlining the close relationship of these genes. The Zic proteins, like the proteins of the *gli* family, are important regulators of neural development and interact with these (Aruga, 2004). A phylogenetic tree (Fig. 6) in which the *gli* genes are used as the outgroup identifies only one sea urchin *zic* gene (*Sp-zic/z244*). The sea urchin *zic* orthologue begins to be expressed at around 18 h postfertilization, and its transcripts are localized in the neurogenic apical plate (Figs. 4C, D).

Zinc fingers of the krüppel-like/Sp1 *family*

The genes of the *krüppel-like/Sp1* family encode a diverse group of transcriptional regulators. In mammals, the genes of this family are dispersed throughout the entire genome. Sp1 was

Table 1
Summary of spatial expression patterns

Name	Exp. start	Mat.	*In situ* result					
			7 h	12 h	18 h	24 h	36 h	48 h
z13	12		EM, MPS	SM	SM	E	–	ND
z30	36	M	–	–	–	–	FG	–
z48	c	M	UBIQ	–	MPS	SM	–	ND
z54/spalt	c	M	UBIQ	UBIQ	UBIQ	MPS, E, SM	FG, MG, HG	FG, MG, HG
z55	c	M	UBIQ	–	–	OE, ABO	FG, MG, HG, OE	FG, MG, HG, OE
z60/egr	c	M	UBIQ	UBIQ	UBIQ	OE, ABO	A	A
z67	24	M	–	–	–	MPS	B	–
z81/smadIP	12		–	ABO and OE	–	SM	A, FG	A, FG
z85/klf2/4	12	M	–	ABO or OE	ABO or OE	ABO or OE	ABO and/or OE	–
z86/klf7	18		–	–	–	ABO or OE	–	–
z92	18		UBIQ	UBIQ	UBIQ	MPS, E, SM	OE	UBIQ
z121/osr	36		–	–	–	–	MG, B	ND
z133	36		–	–	–	A	A	A
z141	6	M	UBIQ	UBIQ	UBIQ	MPS, E, SM	B	–
z157/ovo	c	M	UBIQ	UBIQ	A, MPS, E, SM	MPS, E, SM	–	–
z166	24		–	–	–	SM	OE or ABO	ND
z173	36	M	–	ABO or OE	ABO or OE	–	–	ND
z188/klf13	c	M	–	–	E, A	E, A, OE or ABO	A, OE	A, OE
z199/sp5	18		–	–	ABO or OE	ABO or OE	B, OE	HG
z204	48		–	–	–	–	B	B
z244 zic	36		–	A	A	A	A	A
z487	c	M	UBIQ	UBIQ	ABO, OE	–	–	–
z18	c	M	UBIQ	–	–	–	–	–
z28	c	M	UBIQ	–	–	–	–	–
z32	c	M	UBIQ	–	–	–	–	–
z38	6		UBIQ	–	–	–	–	–
z45	6	M	UBIQ	–	–	–	–	–
z62	c	M	UBIQ	UBIQ	UBIQ	UBIQ	UBIQ	ND
z65	48	M	UBIQ	–	–	–	–	ND
z70	n	M	UBIQ	–	–	–	–	–
z74	n	M	UBIQ	–	–	–	–	ND
z77	36	M	UBIQ	–	–	–	–	–
z90	n	M	UBIQ	–	–	–	–	ND
z98	24	M	UBIQ	–	–	–	–	–
z114	24	M	UBIQ	UBIQ	UBIQ	UBIQ	UBIQ	ND
z197	c	M	UBIQ	UBIQ	UBIQ	UBIQ	UBIQ	UBIQ
z212	c	M	UBIQ	UBIQ	UBIQ	UBIQ	UBIQ	ND
z214	18		–	–	UBIQ	UBIQ	–	ND
z247	c	M	UBIQ	–	–	–	–	–
z338	36		UBIQ	–	–	–	–	–
z401	48	M	UBIQ	–	–	–	–	–
z425	n	M	UBIQ	–	–	–	–	ND
z442	c	M	UBIQ	UBIQ	UBIQ	–	–	–

Genes that show a localized expression are listed in the upper half, genes with ubiquitous expression in the lower half. For each gene, the start of zygotic expression ("Exp. Start") is given (c—constant expression, n—no zygotic expression) and whether or not it is maternal (M—maternal expression). EM—endomesoderm, EC—ectoderm, MPS—micromeres/PMC/skeletogenic tissue, A—apical ectoderm/apical plate, SM—secondary/mesenchyme/mesoderm, E—endoderm, OE—oral ectoderm, ABO—aboral ectoderm, B—blastopore, FG—foregut, MG—midgut, HG—hindgut, UBIQ—ubiquitous.

one of the first mammalian transcription factors to be cloned and named by its sequence similarity to the *Drosophila* gap gene *krüppel*. While this gene contains four zinc fingers, the genes of the *krl/Sp1* family only contain three, which are linked by a highly conserved linker that frequently connects two adjacent zinc finger domains (Kaczynski et al., 2003). The human genome contains eight genes of the *Sp1* and 16 genes of the *krl* subfamily. All *Sp1* genes are transcriptional activators, whereas some *krl* genes are repressors and can counteract Sp1-mediated gene activation (Lomberk and Urrutia, 2005; Urrutia, 2003). In mice, *krl/Sp1* genes are expressed in a wide variety of tissues controlling various processes in development (e.g., *klf1* and *klf2* are involved in erythropoeisis and blood vessel and lung development) or controlling cell growth and proliferation. Several of them are known as tumor suppressor genes (e.g., *klf4* and *klf7*). Phylogenetic analysis (Fig. 7) reveals that the sea urchin genome contains six genes of the *klf* subfamily (*Sp*-klf2/4/z85, *Sp-klf3/8/12/z400*, *Sp-klf7/z86*, *Sp-klf11/z214*, *Sp-klf13/z188*, *Sp-klf15/z174*) and three genes of the *Sp1* subfamily (*Sp-sp2/z168*, *Sp-sp5/z199*, *Sp-sp8/z177*). All sea urchin *klf* genes and *sp5* are expressed in the early embryo. *Sp-klf2/4*, *Sp-klf7* and *Sp-sp5* are localized in the ectoderm (Figs. 4I–K, Q–T)

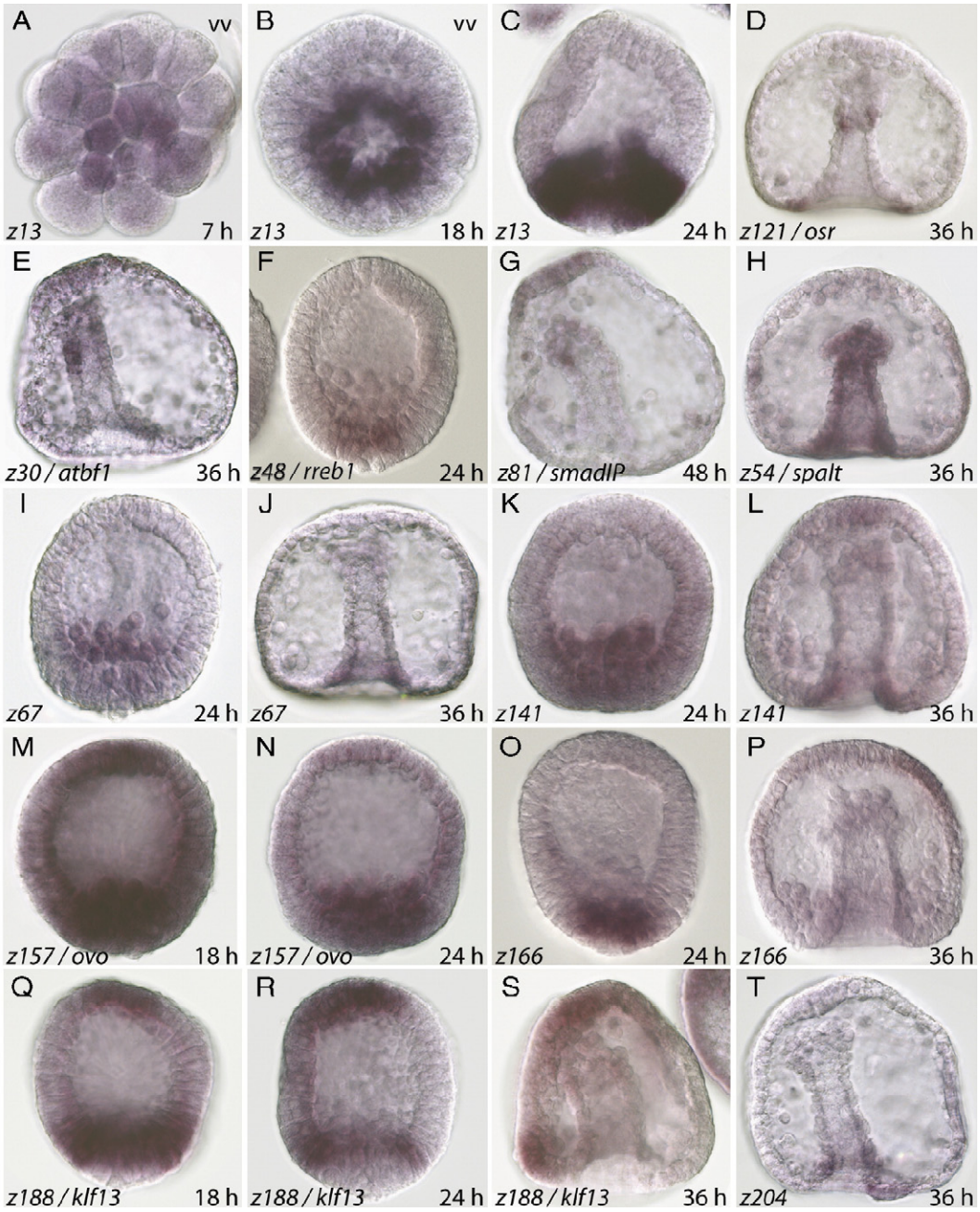

Fig. 3. Endomesodermal expression of C_2H_2 zinc finger genes. Whole mount *in situ* hybridizations are shown at stages listed in the lower right corner of each panel. A full summary of spatial expression patterns is given in Table 1. Gene names are given in the lower left corners. Embryos were recorded in lateral view unless stated otherwise (vv: vegetal view).

and *Sp-klf13* is expressed in the ectoderm and the endoderm (Figs. 3Q–S).

A sea urchin zinc finger gene described previously as *Sp-krl* (Howard et al., 2001) does not seem to be a close relative of the genes of the *krl/Sp* family identified by phylogenetic analysis (Supplementary Fig. S1). In fact, it is among the zinc finger genes for which no unequivocal homologue can be identified. In our data set this gene is designated *Sp-z13*.

Zinc fingers of the snail *family*

The *snail* gene family encodes a highly conserved group of transcriptional repressors. In most genomes, this family comprises at least one member of both the *snail* and the *scratch* subfamilies.

Snail is often involved in epithelium-mesenchyme transitions, for example in neural crest formation in mice (Manzanares et al., 2001). In vertebrates *snail* is thought to have given rise to two additional genes, *snail2/slug* and *snail3* (Manzanares et al., 2004). In *Drosophila* the three members of the family are *snail*, which provides spatial control of gene expression along the embryonic dorso-ventral axis, *escargot*, and *worniu*, which are employed in formation of the central nervous system (CNS). *Ciona* and *C. elegans* have one *snail* gene each. A clearly orthologous *snail* gene has even been found in *N. vectensis*

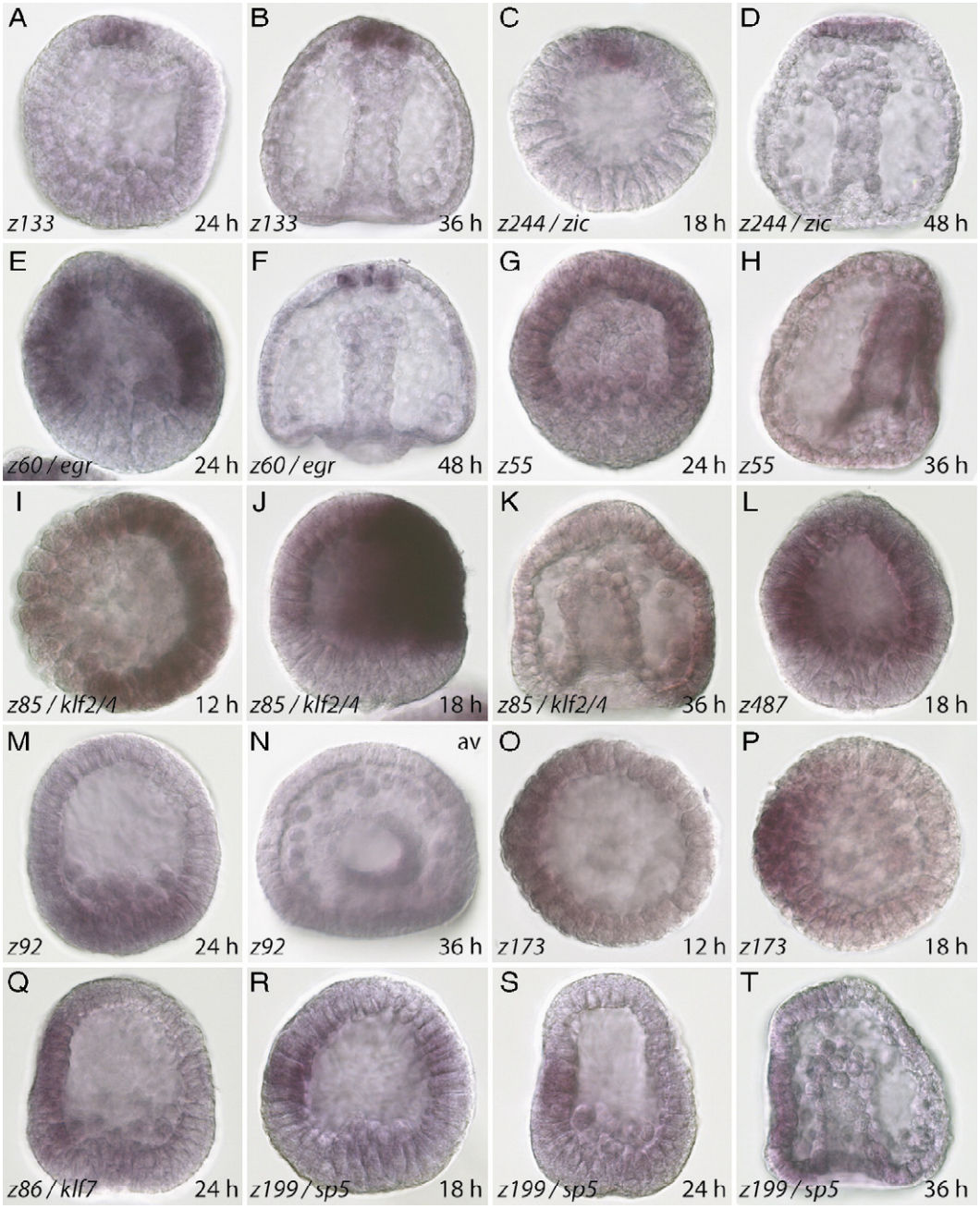

Fig. 4. Ectodermal expression of C_2H_2 zinc finger genes. This includes genes with expression in the apical plate (e.g., panels A–D), and oral and aboral ectoderm. A full summary of spatial expression patterns is given in Table 1. Embryos were recorded in lateral view.

(Corbo et al., 1997). The sea urchin contains one *snail* gene (*Sp-snail/z88*) that is most closely related to vertebrate Snail1 and Snail2 (Fig. 8).

The genes of the *scratch* subfamily in mice and fly are employed in neural crest formation and CNS development. In contrast to mammals, which contain two or more of these genes, the sea urchin genome contains only one homologue of *scratch* (*Sp-scratch/z213*). *Drosophila* contains two but they seem to be paralogous and they fall outside the mouse/human/sea urchin group. We identified an additional gene of the *scratch* subfamily in the sea urchin, that clusters with a novel transcript from *Drosophila* (Bootstrap value of 99), which we named *Sp-scratchX* (*Sp-scratchX/z191*). This may indicate an ancestral duplication of the scratch gene, one of which was lost in the vertebrate lineage.

Zinc finger genes of the spalt*,* egr*,* ovo*, and* odd *families*

spalt genes were first identified in *Drosophila,* where they are involved in diverse processes such as homeotic specification of the embryonic termini, wing patterning and sensory organ development. In vertebrates, four *spalt* genes can be found that are involved in similarly diverse processes ranging from limb development to development of the nervous system and organs like kidney and heart (Sweetman and Münsterberg, 2006). They

Table 2
Comparison of number of C_2H_2 zinc finger genes in sea urchin and other species

Gene family	S.p.	H.s.	M.m.	C.i.	D.m.	C.e.	N.v.
ZF total	377	764	731	198	326	211	170
zic	1	5	6	2	1	1	n.d.
gli	3	5	5	2	3	1	2
klf/sp1	9	24	24	n.d.	n.d.	n.d.	n.d.
snail	1	3	3	1	4	1	1
scratch	2	3	2	0	3	1	0
egr/krox	1	4	4	0	1	1	1
ovo	1	2	3	1	1	1	0
spalt	1	4	3	0	2	1	0
odd	1	2	2	1	3	2	0

S.p.—*Strongylocentrotus purpuratus*, H.s.—*Homo sapiens*, M.m.—*Mus musculus*, C.i.—*Ciona intestinalis*, D.m.—*Drosophila melanogaster*, C.e.—*Caenorhabditis elegans*, N.v—*Nematostella vectensis*, n.d.—not determined. The genome of *Nematostella* is still in a provisional state; absence as indicated in this table, may not mean absence from this genome.

contain three or four pairs of zinc finger domains that are located in the C-terminal half of the protein. The sea urchin contains one gene of this family, which we named *Sp-spalt* (*Sp-spalt/z54*, Supplementary Fig. 2A). This gene is expressed

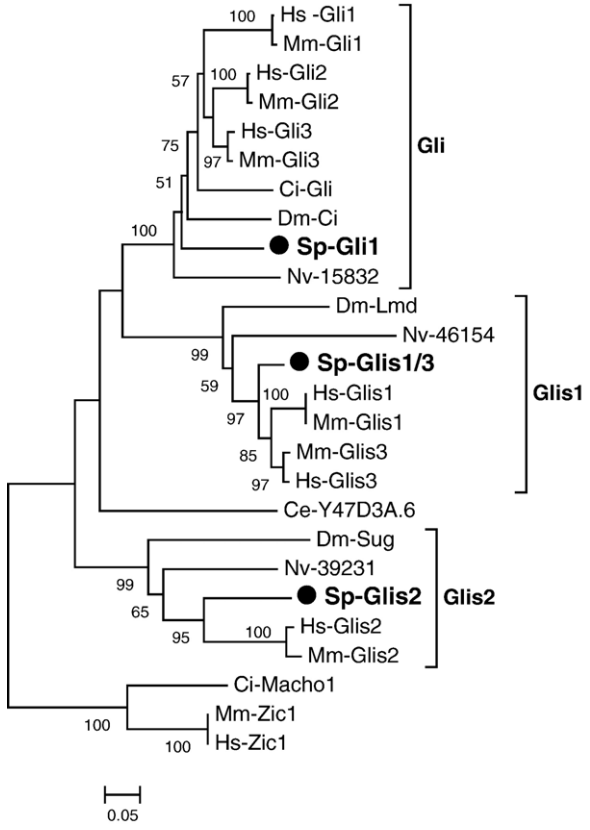

Fig. 5. Phylogenetic tree of Gli and Glis proteins, based on alignment of the zinc finger regions. Sea urchin proteins can be found in all three major classes (Gli, Glis1, Glis2), supported by high bootstrap values (>95). Only values higher than 50% are shown. Proteins of the Zic family were used as the outgroup. The scale bar indicates an evolutionary distance of 0.05 amino acid substitutions per position. Ce—*Caenorhabditis elegans*, Ci—*Ciona intestinalis*, Dm—*Drosophila melanogaster*, Hs—*Homo sapiens*, Mm—*Mus musculus*, Nv—*Nematostella vectensis*, Sp—*Strongylocentrotus purpuratus*.

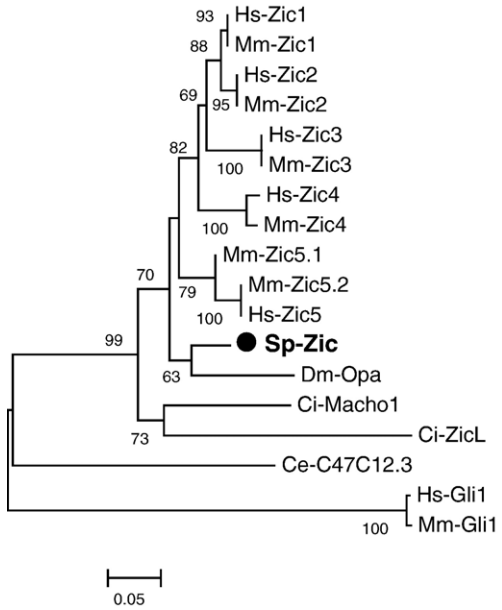

Fig. 6. Phylogenetic tree of Zic proteins, based on alignment of the zinc finger regions. Only one sea urchin Zic sequence could be identified. Gli proteins were used as the outgroup. Annotations are as in Fig. 5.

throughout embryonic development and its transcripts are localized to the endoderm (Fig. 3H).

The genes of the *egr/krox* family in mammals are mainly involved in brain development, with *egr1/krox-24* being expressed in the sensory cortex (Herdegen and Leah, 1998). Recently, a role in learning has been invoked for this gene. In *Drosophila*, this gene is expressed in the epidermis, and mutants show defects in recognition of myotubules by epidermal cells. Although the structure of Egr1/Krox-24 has been determined, and its interaction with DNA is one of the best understood, the binding sites seem to show considerable variability and not many direct targets have been identified. The sea urchin contains one member of this family (*Sp-egr/z60*), which seems to be most closely related to *egr1* (Supplementary Fig. 2B). The gene is expressed throughout early development. Whole mount *in situ* hybridization shows localization of the transcript to the ectoderm beginning with late blastula stage (Figs. 4E, F).

The *Drosophila ovo* gene is primarily expressed in the female germ line. Loss of *ovo* activity leads to sterility in female flies. In mice two *ovo-like* genes are known. Both seem to be involved in spermatogenesis (Dai et al., 1998). *ovol-1* is expressed in a range of different tissues such as kidney and epidermis, and mutants show defects in hair morphogenesis. In sea urchin, we identified one gene of this group (*Sp-ovo/z157,* Supplementary Fig. 2C). *Sp-ovo* is expressed constantly during embryonic development.

In *Drosophila*, *odd-skipped* (*odd*) was originally identified as a pair-rule gene because mutations at this locus lose portions of odd numbered segments in the embryo. Two closely related genes, *bowl* and *sob*, were identified, of which Bowl has an important function in the development of the terminal segments and the gut. Two orthologues are known in mammals. Mouse *osr1* is expressed in embryonic mesoderm and at later stages in the branchial arms and limb buds (Wang et al., 2005). It has

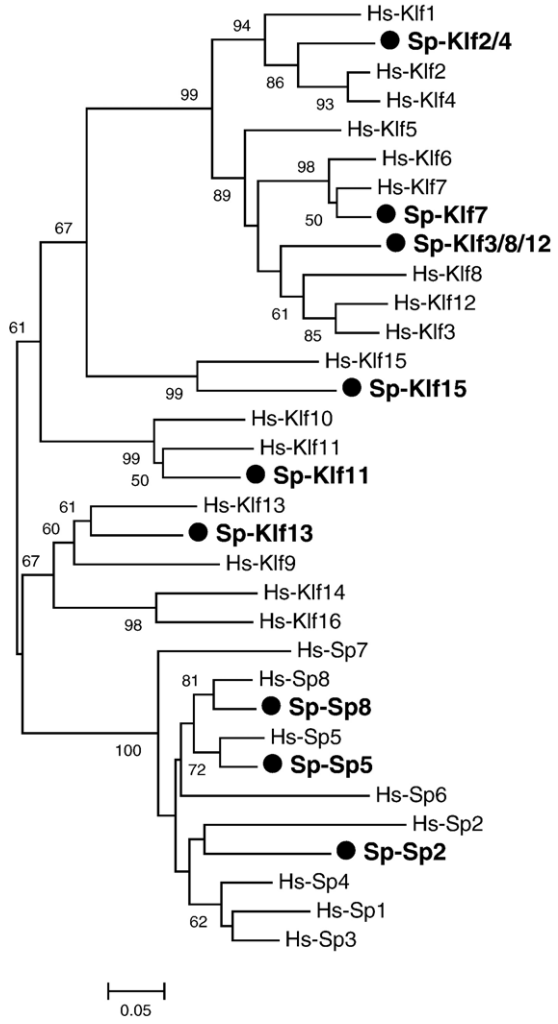

Fig. 7. Phylogenetic tree of human and sea urchin Klf and Sp1-like proteins based on alignment of the zinc finger regions. In the sea urchin, six Klf sequences and three Sp1-like sequences could be identified. Annotations are as in Fig. 5.

also been shown to be involved in heart and kidney development. *osr2* is also expressed in limbs and kidneys but mice deficient for *osr2* show no defects in these organs. We identified one *odd-like* gene in the sea urchin (*Sp-osr/z121*, Supplementary Fig. 2D). This gene is expressed late in embryonic development.

Discussion

The zinc finger genes thus far identified account for at least 1.5% of the total genes in the sea urchin genome (Sea Urchin Sequencing Consortium, submitted). This result parallels findings for other genomes in which zinc finger-containing genes are among the most abundant classes (Adams et al., 2000; Lander et al., 2001; Tupler et al., 2001). Yet the meaning of this large set of genes remains largely enigmatic. Only a minor fraction of the *Strongylocentrotus purpuratus* C_2H_2 zinc finger genes are orthologous to known regulatory genes of human or *Drosophila*. As detailed above, these include members of the *zic*, *gli*, *snail*, and *krl/Sp1* families, all of which are known for their roles in development. The relatively low number of clearly identifiable orthologues is consistent with results of Knight and Shimeld (2001), who report that only about 25% of fly, worm, and human C_2H_2 zinc finger genes can be assigned to orthologous groups.

Several features of the large class of genes defined by the presence of sequence encoding C_2H_2 zinc finger domains distinguish them from most classes of DNA-recognizing regulatory genes, and these features are unlikely to be independent. First, the lack of extensive interspecific conservation, such as permits unequivocal assignment of sequence orthology, contrasts sharply with what is seen for virtually every other prominent class of regulatory gene. In the *S. purpuratus* genome for example, except for rare stragglers almost every gene encoding homeodomain factors (Howard-Ashby et al., 2006b), ets factors (Rizzo et al., 2006), forkhead factors (Tu et al., 2006), bHLH factors, nuclear hormone receptors, and factors of many other smaller groups (Howard-Ashby et al., 2006a) can be related to one or another known regulatory gene subfamily in other animal genomes. Partly the problems in orthology assignment of zinc finger genes are due to internal structural aspects, which confound the algorithms that order phylogenetic similarity: there is a generally high level of sequence identity within zinc finger regions, due to structural requirements of the zinc finger structure itself. On average 40% of amino acids in the zinc finger region are identical (Knight and Shimeld, 2001). An additional difficulty is posed by the internally repetitive modular structure of many zinc finger proteins. However, other classes of

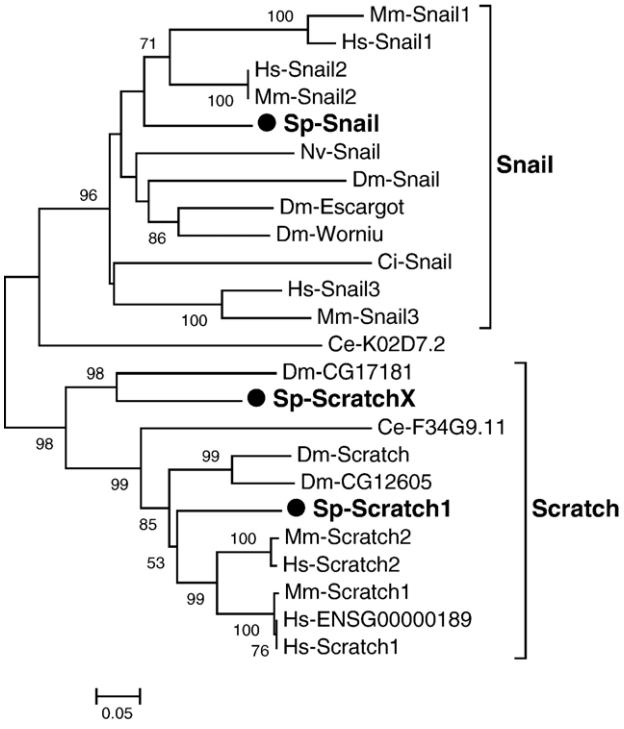

Fig. 8. Phylogenetic tree of proteins of the Snail family based on alignment of the zinc finger regions. The Snail subfamily clearly separates from the Scratch subfamily. The sea urchin genome encodes one Snail and one Scratch orthologue. A second protein clusters with a novel protein from *Drosophila* within the Scratch subfamily, but is clearly separated from other Scratches. We named this protein Sp-ScratchX. Annotations are as in Fig. 5.

regulatory genes such as those encoding the bHLH proteins are subject to processes such as domain shuffling (Morgenstern and Atchley, 1999), but these nonetheless retain identifiable orthologous relationships. The example of those zinc finger genes which can be assigned to orthology groups show that the general difficulty of doing this for most of the zinc finger genes is not just due to their canonical internal properties.

A second distinctive feature of the sea urchin zinc finger genes is that they appear to have been evolving rapidly and in a clade-specific way. This is the most obvious explanation for their lack of orthology with the zinc finger genes of other animal genomes. Their number per genome is extraordinarily flexible: Table 2 shows that there are almost twice as many zinc finger genes in *S. purpuratus* as in *Ciona*, another deuterostome; slightly more than in *Drosophila*; and half as many as in mammalian genomes. Rapid evolutionary change in zinc finger genes is their prominent characteristic. A good example is given by a cluster of zinc finger genes in the human genome on chromosome 19 (Shannon et al., 2003). Genes in this cluster were shown to be duplications of each other but the number of zinc finger domains they contain varies widely between seven and eighteen per gene. Another example, as we report here, is that no *S. purpuratus* zinc finger gene possesses either a KRAB or SCAN domain though these occur frequently in mammalian zinc finger genes. The expanded sea urchin zinc finger gene family, like those of other animals, is among the more evolutionarily flexible, lineage-specific families to be found in this genome.

Thirdly, the zinc finger genes are used differently in development than any other family of regulatory genes, or the set of all such genes (Howard-Ashby et al., 2006b). Up to the 48 h late gastrula stage 75 to 78% of this total set is significantly transcribed, while the expression measurements we report here, which were carried out with the same technology, showed that out of 324 zinc finger genes assayed only 112 are expressed by 48 h. Most of the expressed zinc finger genes exceeded the threshold set arbitrarily for significant expression by several fold, at least at one or more time points, and most in the non-expressed category are well below this threshold. For example, in the whole genome tiling array analysis of the embryo transcriptome, expression of only an additional 58 of the 377 zinc finger genes was detected, and these genes are usually represented by but 50 to 100 transcripts in the whole embryo (800 cells at 48 h). The large fraction of silent or essentially silent zinc finger genes is not the only difference. In the set of total regulatory genes, only 20% of the 192 that are significantly expressed are represented in maternal RNA of the unfertilized egg (Howard-Ashby et al., submitted), while about 75% of the 112 expressed zinc finger genes are maternally expressed. The high maternal utilization of zinc finger genes is emphasized by the observation that 27 of the zinc finger genes are only expressed maternally (i.e., up to 48 h of development), while the corresponding number for genes encoding all other types of transcription factors is but two (Howard-Ashby et al., submitted). We cannot exclude that these maternal RNAs are not fully processed and, hence, nonfunctional (Davidson, 1986). In order to show full maturity, the maternal transcripts would have to be cloned and sequenced. However, it is difficult to imagine that zinc finger genes produce such transcripts while all other transcription factors do not (Howard-Ashby et al., submitted).

These three major attributes of the zinc finger gene family, which distinguish it from the regulatory gene set as a whole, can be interpreted in alternative ways. The most likely explanation is simply that C_2H_2 zinc finger domains do not per se constitute evidence sufficient to assume regulatory gene function, so that the comparison is to some extent between apples and oranges. On the other hand, some zinc finger proteins are most certainly transcription factors and the prevailing view is that even though they are in most genomes poorly annotated, they generally possess DNA-binding capability and are involved in transcriptional regulation (Knight and Shimeld, 2001; Krishna et al., 2003; Miwata et al., 2006). In addition to the orthologues of known zinc finger regulatory genes identified in this work, the specific expression patterns summarized in Table 1 and shown in Figs. 3 and 4 are typical of bona fide regulatory gene products in their spatial specificity, their relatively low copy number, and their dynamic quality. These genes are very likely to execute important regulatory functions in specific tissues of the embryo. Nonetheless, there is accumulating evidence that zinc finger domains are employed in processes other than DNA binding. Many of the C_2H_2 zinc finger domains may belong to proteins the primary function of which is RNA binding or protein–protein interactions (Introduction; *op. cit.*). Relative prevalence of RNA binding function could help to account for the high representation of zinc finger sequences in the maternal transcript stockpile. Perhaps such other, nontranscriptional functions are very heavily represented in the majority set of C_2H_2 zinc finger genes that remain silent during embryogenesis, while the ones that do encode transcription factors are utilized more or less as are other regulatory genes (Howard-Ashby et al., submitted).

An alternative is that most zinc finger genes do indeed have regulatory function, but that they are clade—specifically specialized to execute the regulatory functions required to generate clade-specific features of the organism. Since they appear to be specific additions to most genomes, they may be involved in processes that are not shared between organisms, and these are primarily the processes of the terminal stages of development (Davidson and Erwin, 2006). This would explain why they are poorly represented in the canonical gene regulatory networks of the early-to mid-stage embryo. Even the set of zinc finger genes of mice and humans seems to have diverged significantly and this difference has been imagined to contribute to the differences between these species (Shannon et al., 2003). The flexibility and diversification of zinc finger gene use in the tip of the iceberg so far known to us are reminiscent of terminal differentiation processes, where for example alternative splicing and deployment of paralogous gene relatives are often most extreme. Though crystal structures indicate that three zinc fingers often suffice for recognition of sequence-specific DNA binding sites (Choo et al., 1997), many of the zinc finger genes of the sea urchin have a much higher number of zinc fingers (the same is true for zinc fingers in the human genome; data not shown). Different domains may be used alternately for recognizing

different binding sites, as in the transcription factor CTFC, an 11 zinc finger protein (Filippova et al., 2002). Alternative splicing may indeed also contribute versatility in the highly specific utilization of zinc finger regulators. As a rule, evolutionarily expanded gene families display reduced numbers of alternative splice forms, but there is a known exception: the zinc finger genes (Kopelman et al., 2005). The mysteries surrounding this prominent, rapidly evolving, and ubiquitous gene family are likely to resolve only when at last we have learned something of the functional roles in development of the majority of C_2H_2 zinc finger genes.

Acknowledgments

The authors would like to thank C.T. Brown for help with computation, T. Minokawa for initial help with WMISH, and J. Nam for critical support of phylogenetic analysis. We are grateful for helpful comments and suggestions on the manuscript by V. Hinman and J. Nam. This research was supported by NIH grant HD37105 and by the Office of Science (BER), US Department of Energy, grant DE-FG02-03ER63584.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ydbio.2006.08.032.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H.C., Blazej, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Miklos, G.L.G., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, A., de Pablos, B., Delcher, A., Deng, Z.M., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F.C., Gorrell, J.H., Gu, Z.P., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, D.A., Howland, T.J., Wei, M.H., Ibegwam, C., et al., 2000. The genome sequence of *Drosophila melanogaster*. Science 287, 2185–2195.

Andreazzoli, M., Delucchini, S., Costa, M., Barsacchi, G., 1993. RNA-binding properties and evolutionary conservation of the *Xenopus* multifinger protein Xfin. Nucleic Acids Res. 21, 4218–4225.

Aruga, J., 2004. The role of Zic genes in neural development. Mol. Cell. Neurosci. 26, 205–221.

Choo, Y., Castellanos, A., GarciaHernandez, B., SanchezGarcia, I., Klug, A., 1997. Promoter-specific activation of gene expression directed by bacteriophage-selected zinc fingers. J. Mol. Biol. 273, 525–532.

Chung, H.R., Schäfer, U., Jäckle, H., Böhm, S., 2002. Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in *Drosophila*. EMBO Rep. 3, 1158–1162.

Corbo, J.C., Erives, A., DiGregorio, A., Chang, A., Levine, M., 1997. Dorsoventral patterning of the vertebrate neural tube is conserved in a protochordate. Development 124, 2335–2344.

Croker, J.A., Ziegenhorn, S.L., Holmgren, R.A., 1998. regulation of the Drosophila transcription factor, *Cubitus interruptus*, by two conserved domains. Dev. Biol. 291, 368–381.

Dai, X., Schonbaum, C., Degenstein, L., Bai, W.Y., Mahowald, A., Fuchs, E., 1998. The ovo gene required for cuticle formation and oogenesis in flies is involved in hair formation and spermatogenesis in mice. Genes Dev. 12, 3452–3463.

Davidson, E.H., 1986. Gene Activity in Early Development, 3rd ed. Academic Press, Orlando, FL.

Davidson, E.H., Erwin, D.H., 2006. Gene regulatory networks and the evolution of body plans. Science 311, 796–800.

Filippova, G.N., Qi, C.F., Ulmer, J.E., Moore, J.M., Ward, M.D., Hu, Y.J., Loukinov, D.I., Pugacheva, E.M., Klenova, E.M., Grundy, P.E., Feinberg, A.P., Cleton-Jansen, A.M., Moerland, E.W., Cornelisse, C.J., Suzuki, H., Komiya, A., Lindblom, A., Dorion-Bonnet, F., Neiman, P.E., Morse, H.C., Collins, S.J., Lobanenkov, V.V., 2002. Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter its DNA-binding specificity. Cancer Res. 62, 48–52.

Glusman, G., Yanai, I., Rubin, I., Lancet, D., 2001. The complete human olfactory subgenome. Genome Res. 11, 685–702.

Herdegen, T., Leah, J.D., 1998. Inducible and constitutive transcription factors in the mammalian nervous system: control of gene expression by Jun, Fos and Krox, and CREB/ATF proteins. Brains Res. Rev. 28, 370–490.

Howard-Ashby, M., Materna, S.C., Brown, C.T., Chen, L., Cameron, R.A., Davidson, E.H., 2006a. The identification and characterization of transcription factor families in early *Strongylocentrotus* development. Dev. Biol. (this issue).

Howard-Ashby, M., Materna, S.C., Brown, C.T., Chen, L., Davidson, E.H., 2006b. 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., Tu, Q., Oliveri, C.T., Davidson, P., submitted. The usage of transcription factor families in sea urchin development: implications for development and evolution of the Bilateria. Science.

Howard, E.W., Newman, L.A., Oleksyn, D.W., Angerer, R.C., Angerer, L.M., 2001. SpKrl: a direct target of β-catenin regulation required for endoderm differentiation in sea urchin embryos. Development 128, 365–375.

Kaczynski, J., Cook, T., Urrutia, R., 2003. Sp1- and Krüppel-like transcription factors. Genome Biol. 4.

Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511–518.

Knight, R.D., Shimeld, S.M., 2001. Identification of conserved C2H2 zinc-finger gene families in the Bilateria. Genome Biol. 2 research0016.1-16-8.

Kopelman, N.M., Lancet, D., Yanai, I., 2005. Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. Nat. Genet. 37, 588–589.

Krishna, S.S., Majumdar, I., Grishin, N.V., 2003. Structural classification of zinc fingers. Nucleic Acids Res. 31, 532–550.

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

Laity, J.H., Lee, B.M., Wright, P.E., 2001. Zinc finger proteins: new insights into structural and functional diversity. Curr. Opin. Struct. Biol. 11, 39–46.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M.,

Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Livi, C.B., Davidson, E.H., 2006. Expression and function of blimp1/krox, an alternatively transcribed regulatory gene of the sea urchin endomesoderm network. Dev. Biol. 293, 512–525.

Lomberk, G., Urrutia, R., 2005. The family feud: turning off Sp1 by Sp1-like KLF proteins. Biochem. J. 392, 1–11.

Lu, D., Searles, M.A., Klug, A., 2003. Crystal structure of a zinc-finger–RNA complex reveals two modes of molecular recognition. Nature 426, 96–100.

Manzanares, M., Locascio, A., Nieto, M.A., 2001. The increasing complexity of the Snail gene superfamily in metazoan evolution. Trends Genet. 17, 178–181.

Manzanares, M., Blanco, M.J., Nieto, M.A., 2004. Snail3 orthologues in vertebrates: divergent members of the Snail zinc-finger gene family. Dev. Genes Evol. 214, 47–53.

Materna, S.C., Berney, K., Cameron, R.A., 2006. The *Strongylocentrotuspur-puratus* genome: A comparative perspective. Dev. Biol 300, 485–495.

Mayor, R., Aybar, M.J., 2001. Induction and development of neural crest in *Xenopus laevis*. Cell Tissue Res. 305, 203–209.

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

Miwata, K., Chiba, T., Horii, R., Yamada, L., Kubo, A., Miyamura, D., Satoh, N., Satou, Y., 2006. Systematic analysis of embryonic expression profiles of zinc finger genes in *Ciona intestinalis*. Dev. Biol. 292, 545–546.

Morgenstern, B., Atchley, W.R., 1999. Evolution of bHLH transcription factors: modular evolution by domain shuffling? Mol. Biol. Evol. 16, 1654–1663.

Nemer, M., Rondinelli, E., Infante, D., Infante, A.A., 1991. Polyubiquitin RNA characteristics and conditional induction in sea-urchin embryos. Dev. Biol. 145, 255–265.

Reece-Hoyes, J.S., Deplancke, B., Shingles, J., Grove, C.A., Hope, I.A., Walhout, A.J., 2005. A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. Genome Biol. 6, R110.1–R110.12.

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.

Rozen, S., Skaletzky, H.J., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., Misener, S. (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365–386.

Rubin, G.M., Yandell, M.D., Wortman, J.R., Miklos, G.L.G., Nelson, C.R., Hariharan, I.K., Fortini, M.E., Li, P.W., Apweiler, R., Fleischmann, W., Cherry, J.M., Henikoff, S., Skupski, M.P., Misra, S., Ashburner, M., Birney, E., Boguski, M.S., Brody, T., Brokstein, P., Celniker, S.E., Chervitz, S.A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R.F., Gelbart, W.M., George, R.A., Goldstein, L.S.B., Gong, F.C., Guan, P., Harris, N.L., Hay, B.A., Hoskins, R.A., Li, J.Y., Li, Z.Y., Hynes, R.O., Jones, S.J.M., Kuehl, P.M., Lemaitre, B., Littleton, J.T., Morrison, D.K., Mungall, C., O'Farrell, P.H., Pickeral, O.K., Shue, C., Vosshall, L.B., Zhang, J., Zhao, Q., Zheng, X.Q.H., Zhong, F., Zhong, W.Y., Gibbs, R., Venter, J.C., Adams, M.D., Lewis, S., 2000. Comparative genomics of the eukaryotes. Science 287, 2204–2215.

Ruvkun, G., Hobert, O., 1998. The taxonomy of developmental control in *Caenorhabditis elegans*. Science 282, 2033–2041.

Samanta, M.P., Tongprasit, W., Istrail, S., Cameron, A., Tu, Q., Davidson, E. H., submitted. A high-resolution transcriptome map of the sea urchin embryo. Science.

Sea Urchin Sequencing Consortium, submitted. The sequence of the sea urchin genome. Science.

Shannon, M., Hamilton, A.T., Gordon, L., Branscomb, E., Stubbs, L., 2003. Differential expansion of zinc-finger transcription factor loci in homologous human and mouse gene clusters. Genome Res. 13, 1097–1110.

Sweetman, D., Münsterberg, A., 2006. The vertebrate spalt genes in development and disease. Dev. Biol. 293, 285–293.

Tu, Q., Brown, C.T., Davidson, E.H., Oliveri, P., 2006. Sea urchin forkhead gene family: phylogeny and embryonic expression. Dev. Biol. 300, 49–62.

Tupler, R., Perini, G., Green, M.R., 2001. Expressing the human genome. Nature 409, 832–833.

Urrutia, R., 2003. KRAB-containing zinc-finger repressor proteins. Genome Biol. 4, 231.

Wang, D.G.W., Kirchhamer, C.V., Britten, R.J., Davidson, E.H., 1995. SpZ12-1, a negative regulator required for spatial control of the territory-specific cyiiia gene in the sea-urchin embryo. Development 121, 1111–1122.

Wang, Q.R., Lan, Y., Cho, E.S., Maltby, K.M., Jiang, R.L., 2005. Odd-skipped related 1 (Odd1) is an essential regulator of heart and urogenital development. Dev. Biol. 288, 582–594.

Wheelan, S.J., Church, D.M., Ostell, J.M., 2001. Spidey: a tool for mRNA-to-genomic alignments. Genome Res. 11, 1952–1957.

Williams, A.J., Blacklow, S.C., Collins, T., 1999. The zinc finger-associated SCAN box is a conserved oligomerization domain. Mol. Cell. Biol. 19, 8526–8535.

Wong, M.L., Medrano, J.F., 2005. Real-time PCR for mRNA quantitation. Biotechniques 39, 75–85.

Yamada, L., Kobayashi, K., Degnan, B., Satoh, N., Satou, Y., 2003. A genome-wide survey of developmentally relevant genes in *Ciona intestinalis*: IV. Genes for HMG transcriptional regulators, bZip and GATA/Gli/Zic/Snail. Dev. Genes Evol. 213, 245–253.

Zhang, Z.L., Gerstein, M., 2004. Large-scale analysis of pseudogenes in the human genome. Curr. Opin. Genet. Dev. 14, 328–335.