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

RTK and TGF-β signaling pathways genes in the sea urchin genome

François Lapraz [a], Eric Röttinger [a], Véronique Duboc [a], Ryan Range [a], Louise Duloquin [a], Katherine Walton [b], Shu-Yu Wu [b], Cynthia Bradham [b], Mariano A. Loza [c], Taku Hibino [c], Karen Wilson [d], Albert Poustka [e], Dave McClay [b], Lynne Angerer [f], Christian Gache [a], Thierry Lepage [a,*]

[a] *UMR 7009 CNRS, Université Pierre et Marie Curie-Paris 6, Observatoire Oceanologique, 06230 Villefranche-sur-Mer, France*
[b] *Developmental, Molecular, and Cellular Biology Group, Duke University, Durham, NC, USA*
[c] *Sunnybrook and Women's Research Institute and Department of Medical Biophysics, University of Toronto. 2075 Bayview Ave., Toronto, Ontario, Canada M4N 3M5*
[d] *Kristineberg Marine Research Station 45034 Fiskebäckskil, Sweden*
[e] *Evolution and Development Group Max-Planck Institut fuer Molekulare Genetik Ihnestrasse 7314195, Berlin*
[f] *National Institute for Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD 20892, USA*

Received for publication 25 May 2006; revised 16 August 2006; accepted 16 August 2006

Abstract

The Receptor Tyrosine kinase (RTK) and TGF-β signaling pathways play essential roles during development in many organisms and regulate a plethora of cellular responses. From the genome sequence of *Strongylocentrotus purpuratus*, we have made an inventory of the genes encoding receptor tyrosine kinases and their ligands, and of the genes encoding cytokines of the TGF-β superfamily and their downstream components.

The sea urchin genome contains at least 20 genes coding for canonical receptor tyrosine kinases. Seventeen of the nineteen vertebrate RTK families are represented in the sea urchin. Fourteen of these RTK among which ALK, CCK4/PTK7, DDR, EGFR, EPH, LMR, MET/RON, MUSK, RET, ROR, ROS, RYK, TIE and TRK are present as single copy genes while pairs of related genes are present for VEGFR, FGFR and INSR.

Similarly, nearly all the subfamilies of TGF-β ligands identified in vertebrates are present in the sea urchin genome including the BMP, ADMP, GDF, Activin, Myostatin, Nodal and Lefty, as well as the TGF-β sensu stricto that had not been characterized in invertebrates so far. Expression analysis indicates that the early expression of *nodal*, *BMP2/4* and *lefty* is restricted to the oral ectoderm reflecting their role in providing positional information along the oral–aboral axis of the embryo. The coincidence between the emergence of TGF-β-related factors such as Nodal and Lefty and the emergence of the deuterostome lineage strongly suggests that the ancestral function of Nodal could have been related to the secondary opening of the mouth which characterizes this clade, a hypothesis supported by functional data in the extant species.

The sea urchin genome contains 6 genes encoding TGF-β receptors and 4 genes encoding prototypical Smad proteins. Furthermore, most of the transcriptional activators and repressors shown to interact with Smads in vertebrates have orthologues in echinoderms. Finally, the sea urchin genome contains an almost complete repertoire of genes encoding extracellular modulators of BMP signaling including Chordin, Noggin, Sclerotin, SFRP, Gremlin, DAN and Twisted gastrulation. Taken together, these findings indicate that the sea urchin complement of genes of the RTK and TGF-β signaling pathways is qualitatively very similar to the repertoire present in vertebrates, and that these genes are part of the common genetool kit for intercellular signaling of deuterostomes.

Keywords: Sea urchin; Genome; RTK; Receptor tyrosine kinase; TGF-β; SMAD; BMP; Activin; Nodal; Lefty; ADMP; GDF

Introduction

Cell interactions, which are critical both during embryonic development and adult life, are mediated by receptors that bind ligands and transduce signals to the cell machinery. The kinase receptors form a large group of membrane receptors that respond to ligand binding by modulating the catalytic activity of their intracellular kinase domain. These receptors form two families that differ by the substrate specificity of their kinase domain, their overall structure, their mechanism of action and their ligands. The first family includes the receptors that display

* Corresponding author.
E-mail address: lepage@obs-vlfr.fr (T. Lepage).

0012-1606/$ - see front matter © 2006 Published by Elsevier Inc.
doi:10.1016/j.ydbio.2006.08.048

a tyrosine kinase activity (RTK) and bind a variety of growth factors while the second comprises the receptors that phosphorylate serine or threonine residues and bind members of the TGF-β family (Hubbard and Till, 2000; Shi and Massague, 2003).

The kinase receptors are implicated in the control of a wide range of cellular processes, including cell cycle, metabolism, cell survival, specification of cell fate and differentiation. Alteration of their signaling ability is associated with many human diseases (Schlessinger, 2000; Robertson et al., 2000). The RTKs were among the first oncogenes discovered. Mutations in RTK genes are directly responsible for a variety of malignancies or are closely associated to these diseases (Schlessinger, 2000). Similarly, mutations in the TGF-β receptors or their downstream mediators, the Smads, cause various diseases, including cancers as well as vascular and bone disorders (Miyazono et al., 2001).

RTKs are major mediators of cell interactions that are essential in multicellular organisms. So far they have been identified only in metazoan and in their closest protozoan relatives, the choanoflagellates (King and Carroll, 2001) supporting the idea that RTK signaling may have played a role in the transition to muticellularity (Hunter and Cooper, 1985; King, 2004). The function of RTKs during development are extremely diverse and include determination of egg and embryonic polarity, formation of the germ layers, specification of particular cell types and regulation of cell migration (Shilo, 1992).

RTKs are generally big proteins (about 600 to 2400 amino acids, most of them between 800 and 1600 aa) that share a similar organization. All RTKs are single pass transmembrane proteins with an extremely conserved protein kinase domain in the intracellular C-terminal moiety. The extracellular N-terminal domain, which is responsible for the specificity of ligand binding, is highly variable and displays a modular architecture based on combinations of protein domains like Immunoglobulin, Fibronectin type III, Cadherin, Discoidin, Kringle, EGF, WIF or Plexin domains. RTKs can be subdivided into several families based on sequence similarity of the kinase domain, the composition and architecture of their extracellular domain and their exon/intron organization. About 60 RTK genes grouped in about 20 families have been identified in the human genome (Kostich et al., 2002; Manning et al., 2002).

RTKs bind a variety of growth factors including FGF, EGF, VEGF, TGF-α, Angiopoietin, Neurotrophins and Insulin. Upon ligand binding, monomeric RTKs dimerize and phosphorylate Tyr residues in their intracellular domains. These phosphorylated residues serve as docking sites for proteins that contains SH2 or PTB domains. Recruitment of these proteins leads to the downstream activation of a series of signaling molecules and ultimately to a change in cell state and gene expression. The different RTKs activate multiple downstream pathways like Ras/MAPK, JNK, PI3K/PKB, PI3K/Rac, PLCg/IP3 and STAT. Each pathway has many components, some of them being cytoplasmic Tyr-kinases or Ser/Thr-kinases. The signaling pathways activated by RTKs are linked to each other and cross talk with other transduction pathways. In addition, besides interactions with their cognate ligands, RTKs receive inputs relating to cell adhesion and to stress responses. Thus, RTKs and their ligands are essential components of a large signaling network (Schlessinger, 2000).

Another family of receptor kinases that play a cardinal role during development is the family of receptors that bind ligands of the TGF-β superfamily. The TGF-β superfamily, which comprises 45 members in humans, includes a large variety of cytokines with pleiotropic functions (Shi and Massague, 2003). Behind this apparent diversity, all members of the TGF-β superfamily are structurally related and are synthesized as precursors that are cleaved at the level of a RXXR site to release a 110–140 amino acid long peptide which is the mature form of the ligand. These C-terminal mature forms contain from 6 to 9 conserved cysteines, most of them being engaged in intramolecular disulfide bridges, and one of them being used for homo or heterodimerization. Structural studies revealed that all members of the TGF-β superfamily adopt a conserved three-dimensional structure, composed of two pairs of antiparallel β strands with a conserved pattern of disulfide bridges known as the "cysteine knot".

The BMP and Nodal subfamilies of TGF-β play pivotal roles in early development and regulate a number of essential developmental processes such as specification of the germ layers and body axes. Also, of particular interest for developmental biologists, some TGF-β members have been shown to act as morphogens, diffusing across fields of cells to specify a pattern of cell fates in a concentration-dependent manner (Chen and Schier, 2001; Dosch et al., 1997; Green and Smith, 1990; Lecuit et al., 1996; McDowell et al., 1997; Nellen et al., 1996; Wilson et al., 1997). Genes encoding cytokines of the TGF-β superfamily and their receptors are widespread in the animal kingdom and have been identified both in the Radiata (cnidarians, sponges) and Bilateria, probably reflecting an ancestral function in regulating cell proliferation and differentiation (Finnerty et al., 2004; Herpin et al., 2004; Suga et al., 1999). Since most members of the TGF-β superfamily are potent regulators of cell fate, cell proliferation and differentiation, fine regulation of their activity is essential during embryonic development (Khokha et al., 2005). This modulation is achieved in the extracellular space by secreted proteins such as Chordin and Noggin, that prevent ligand access to the signaling receptors (Balemans and Van Hul, 2002).

Despite the variety of cellular processes that they regulate, TGF-β ligands use a disarmingly simple set of receptors and transcription factors to mediate their effects. TGF-β ligands bind to transmembrane serine/threonine kinases receptors that share highly related sequences but that can be divided in two families based on their structure and their function (Derynck and Feng, 1997). The type II receptors are constitutively active and upon ligand binding, associate with and phosphorylate type I receptors resulting in activation of downstream transcription factors of the Smad family (Shi and Massague, 2003). Several structural features distinguish the type I and type II receptors. These features include the pattern of cysteines in the extracellular ligand binding domain (CCX4-5C for type I and CXCX4C for type II) and the presence in the type I receptors of a SGSGSG motif which defines the so-called GS box immediately before the kinase domain. Each family of receptors is further subdivided into 3 subfamilies, depending on the type

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

of ligand they preferentially bind i.e. the BMP, BMP/Activin or Nodal/Activin/TGF-β sensu stricto. Therefore, while there is a high level of structural and functional diversity within the TGF-β ligands, the assortment of receptors they bind to is much smaller. Despite the variety of cellular processes that they regulate and the large diversity of ligands present in some species, the TGF-β signal transduction pathway is surprisingly simple and relies on a handful of highly conserved transcription factors of the Smad family (Massague et al., 2005).

The sea urchin embryo, which has largely contributed to shape the concepts of embryonic induction and conditional specification, is an excellent model to unravel the gene networks and signaling networks that control cell interactions and development (Angerer and Angerer, 2003; Davidson et al., 2002). The assembly of the sea urchin (*Strongylocentrotus purpuratus*) genome provides an opportunity for a survey of RTK and TGF-β signaling pathway genes present in a basal invertebrate deuterostome genome.

t1.1 Table 1
t1.2 Identified RTK genes

t1.3	Provisional gene name	Official ID	Identified protein domains	Best blast hit (human)	Back blast	Tiling data	Human genes
t1.4	Sp-ALK	SPU-017036	/ TyrKin	AAB71619.1	<->	+	ALK, LTK
t1.5	Sp-CCK4/PTK7	SPU-010698	/ (Ig)5 / / TyrKin	NP-690620.1	<->	+	CCK4/PTK7
t1.6	Sp-DDR	SPU-026731	/ FA58C / TM / TyrKin	CAI17434.1	<->	+	DDR1, DDR2
t1.7	Sp-EGFR	SPU-008595	SP / rL / FU / rL / FU / FuR / TM / TyrKin	NP-005226.1	<->	+	EGFR, HER2, HER3, HER4
t1.8	Sp-EPH	SPU-027145	SP / EPH-lbd / EGF / (FN3)2 / TM / TyrKin/ SAM	NP-872272.1	<->	+	EphA1–8, 10, EphB1–4, 6
t1.9	Sp-FGFR 1	SPU-020677	SP / FN3 / (Ig)3 / TM / TyrKin	AAH15035	<->	+	FGFR 1–4
t1.10	Sp-FGFR 2	SPU-004746+ SPU-004747	(IG)3 / FN3 / TM / TyrKin	CAA40404.1	SPU-020677	+	FGFR 1–4
t1.11	Sp-ILGFR	SPU-002840	/ ANF / TM / TyrKin	AAB22215.1	SPU-003916	+	INSR, IRR, IGF1R
t1.12	Sp-INSR	SPU-003915+ SPU-003916	/ rL / (FN3)3 / TM / TyrKin	AAA59452.1	<->	+	INSR, IRR, IGF1R
t1.13	Sp-LMR	SPU-006026	SP / TM / TyrKin	NP-055731	<->	+	LMR 1–3
t1.14	Sp-MET/RON	SPU-013140	SP / SEMA / PSI / (TIG IPT) 3 / TM / TyrKin	CAA49634	<->	+	MET, RON
t1.15	Sp-MUSK	SPU-024610	/ (IG)2 / TM / TyrKin	AAB63044	<->	+	MUSK
t1.16	Sp-RET	SPU-016716	/ Cad / TM / TyrKin	NP-065681	<->	+	RET
t1.17	Sp-ROR	SPU-020646	SP / (Ig /Fz)2 / Kr / TM / TyrKin	NP-005003	<->	+	ROR1, ROR2
t1.18	Sp-ROS	SPU-007624+ SPU-028424	/ ((FN3)2 / (LY)2)2 / (FN3) / (LY)2 / (FN3)2 / (LY)2 / / TyrKin	NP-002935.2	<->	+	ROS
t1.19	Sp-RYK	SPU-010329	/WIF / TM / TyrKin	NP-001005861	<->	+	RYK
t1.20	Sp-TIE1/2	SPU-024044	/ IG / (EGF)3 / IG / (FN3)5 / TM / TyrKin	CAA43290	<->	+	TIE1, TIE2
t1.21	Sp-TRK	SPU-020803	/ IG / TM / TyrKin	AAC51371	<->	+	TRKA, TRKB, TRKC
t1.22	Sp-VEGFR-7	SPU-021021	/ (IG) 7 / TM / TyrKin	AAC16449	SPU-000310	+	VEGFR1, VEGFR2, VEGFR3
t1.23	Sp-VEGFR-10	SPU-000310	/ (IG) 10 / TM / TyrKin	AAC16449	<->	+	VEGFR1, VEGFR2, VEGFR3
t1.24	Sp-FGFR like 1	SPU-020680	SP / (IG)3 / TM /	AAK26742	<->	+	FGFR 5
t1.25	Sp-hypothetical 1	SPU-000667	SP / (FN3)2 / TM / TyrKin	NP-066124 (Ret)	SPU-016716		
t1.26	Sp-hypothetical 2	SPU-026272	FN3 / TM / TyrKin	NP-066124 (Ret)	SPU-016716		
t1.27	Sp-hypothetical 3	SPU-000806	(EGF)4 / TM / TyrKin	NP-075263 (FGFR2)	SPU-020677		
t1.28	Sp-hypothetical 4	SPU-009079	/ EGF / TM / TyrKin	NP-000133 (FGFR3)	SPU-020677		
t1.29	Sp-hypothetical 5	SPU-006004	(Hemi)7 / TM / TyrKin	NP-114141 (hemicentin) AAK51435 (FGFR4)[a]	SPU-011693 SPU-020677[a]		
t1.30	Sp-hypothetical 6	SPU-020532	SP / CCP / TM / TyrKin	NP-075263 (FGFR2)	SPU-020677		
t1.31	Sp-hypothetical 7	SPU-021843	/ CUB / (CCP)3 / TM / TyrKin/ (C2H2)7	NP-006725 (HIV-EBP) P35590 (Tie1)[a]	<-> SPU-024044[a]		

For each selected gene are indicated: Provisional gene name; SPU number; Domain organization of the predicted protein; Bidirectional blast analysis with the human genome: accession number (protein) for the best blast hit, <-> if best hits are reciprocal or Glean number if they are not; Tiling data: (+) indicates embryonic expression; Names of human genes of the same family. Protein domains: ANF, natriuretic peptide receptors; C2H2, zinc-finger; Cad, cadherin; CCP, CCP/sushi/SCR domain; CUB, CUB domain; EGF, EGF like domain; EPH-lbd, ephrin ligand binding domain; FA58C, coagulation factors 5/8 type C domain; FN3, fibronectin type III module; FU, furin; FuR, furin repeat; Fz, Frizzled cysteine-rich domain; Hemi, hemicentrin repeat; Ig, Ig like domain; LY, low density lipoprotein YWDT domain; PSI, Plexins, Semaphorins, Integrins domain; RL, Receptor L domain; SAM, Sterile Alpha Motif; SEMA, SEMA (semaphorin) domain; SP, signal peptide; TIG/IPT, Ig-like, Plexins, Transcription factor domain; TM, transmembrane domain; TyrKin, tyrosine kinase catalytic domain; WIF, Wnt inhibitory factor domain. Note that TMs
t1.32 were missing in three gene models (e.g. ALK, CCK4/PTK7 and ROS) and both the ECD and the TM are absent from one protein (ALK).
t1.33 [a] Blast done with the kinase domain alone.

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

The results of this survey indicate that most of the RTK and TGF-β signaling pathways genes are represented in the sea urchin suggesting that these genes are part of the common genetool kit for intercellular signaling of deuterostomes.

Results and discussion

A basic RTK gene set

The 28944 gene models predicted from the first draft of the sea urchin genome by the GLEAN program were surveyed for RTK genes using RTK sequences from deuterostome and protostome organisms. Twenty gene models (listed in the upper part of Table 1) can be confidently identified as RTK genes based on the following arguments: First, in all but a few cases, the predicted protein presents the general organization of RTKs: Extracellular domain (ECD)/Transmembrane domain (TM)/ Tyrosine Kinase domain (TyrK), with signal peptides (SP) sometimes detected. Second, BLAST analyses give the same hits with either the entire protein sequence or only the TyrK domain. Bidirectional best hit analysis carried out with the human and sea urchin genomes gave reciprocal hits in nearly all cases, or hits with closely related member of the same family in a few cases. Third, the domains identified in the ECD are those normally found in the family defined by the TyrK domain, although with some variation in the number and organization of the modules. Finally, in a phylogenetic tree of the TyrK domains, each sea urchin sequence clearly grouped with one known RTK family member (Fig. 1). This set of canonical RTKs includes two special cases. Identification of Sp-LMR does not rely on the structure of the ECD but on its absence, as paralogs found in vertebrates have only a vestigial extracellular domain reduced to a few amino acids. The prediction for the ALK receptor (Anaplastic Lymphoma Kinase) lacks both the ECD and the TM domains and thus resembles a cytoplasmic kinase. However, BLAST analysis and phylogeny consistently designate this kinase domain as closely related to ALK. Definitive assignment requires identification of the missing parts.

In the lower part of Table 1 are listed additional gene models that give BLAST hits with RTKs but that have been annotated as hypothetical RTK since they do not fulfill all the criteria described above. Among those putative RTK, seven models predict proteins containing TM and ECD upstream of Tyr kinase domains. However, BLAST analysis with human proteins does not produce reciprocal hits and when incorporated in the set of sequences used for a phylogenetic analysis, most of the kinase domains of these models failed to group with known RTK families (not shown). Exceptions are two models (SPU-000806 and SPU-020532), which cluster with the divergent Sp-FGFR2, and (SPU-000667), loosely connected to the RET family (bootstrap value below 50%, Fig. 1). Furthermore, several of these additional models display ECD components that do not correspond to those predicted from the similarity of their kinase domain. For example, SPU-000806 and SPU-020532, the 2 models that cluster with Sp-FGFR2, contain EGF (SPU-000806) or CCP (SPU-020532) domains, which have never been found associated with FGFRs so far. Similarly, FnIII domains are found associated with RET-related kinase domains in one model (SPU-000667). Finally, TyrKin domains were found associated with ECDs containing modules not previously found in any RTK such as hemicentin (SPU-020677) and the presence of 7 zinc-fingers in a long C-terminal domain downstream of the Tyr kinase domain of SPU-021843 appears unlikely. Although these predicted new architectures are potentially interesting, they need to be confirmed by further analysis of the genome and of the transcriptome.

Seventeen of the nineteen vertebrate RTK families are represented in the sea urchin

In vertebrates, 19 classes of RTK have been defined (Robertson et al., 2000; Kosticj et al., 2002; Manning et al., 2002), the size of which varies from a single member to 14 members for the Ephrin receptor family. The 20 identified sea urchin RTKs are distributed amongst 17 of the 19 vertebrate RTK families, as shown by the phylogenetic tree presented in Fig. 1. Most families have only one member. The INSR, FGFR and VEGFR families have two members, as in each case the 2 models identified seem to be too divergent to be haplotype pairs. This will have to be confirmed when a more advanced assembly of the sea urchin genome will be available. Only 2 families are not represented in the sea urchin genome, the ALX and PDGFR families. In human, the ALX family comprises 3 members: ALX, Tyro3 and Mer. These receptors are expressed in the immune, vascular and central nervous systems. No homolog have been identified in *Drosophila* or *C. elegans*, but Ci-TYRO3/AXL/MER was retrieved from the *Ciona* genome (Satou et al., 2003). Since neither the ALX receptor kinase nor its ligand Gas6 is represented in the sea urchin genome, it is likely that these genes appeared with the chordates.

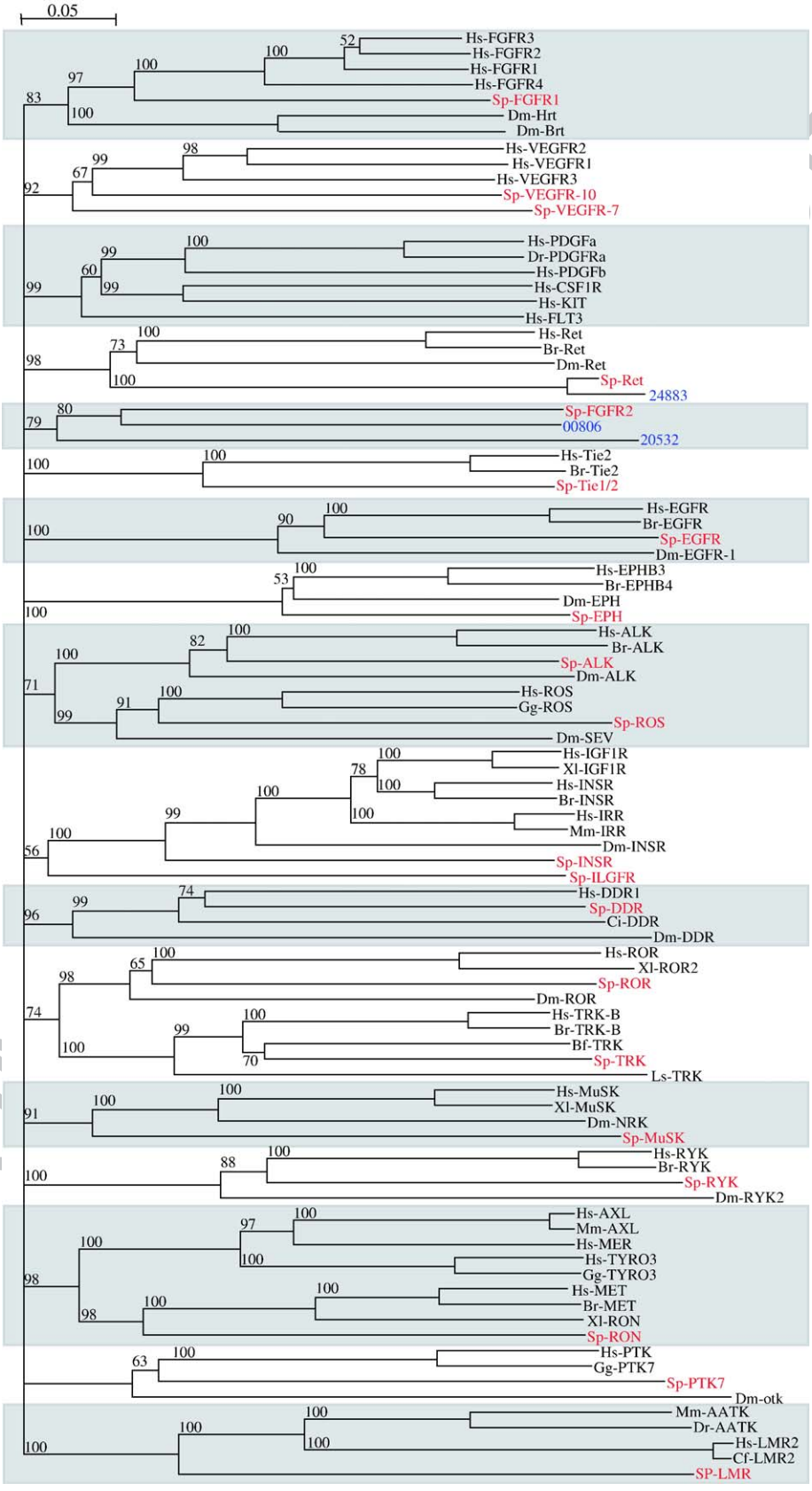
Fig. 1. Phylogenetic tree of the Tyr-kinase domain of the RTKs. Sequences from kinase domains were aligned with ClustalX and the tree was generated by the neighbor-joining method with 1000 bootstrap replications. Numbers indicate the percentage of times the corresponding node was supported in 1000 replications. Nodes that were insufficiently supported were collapsed. *S. purpuratus* sequence names are colored as follows: red, identified RTKs designated with their provisional name (Table 1); blue, Glean numbers of putative RTKs or isolated Tyr-kinase domains. Several predicted proteins (SPU-000667, 026272, 009079, 006004, 021843, 005055, 011509, 019799, 009842, 009990, 017493, 027311, 024883) that consist of isolated Tyr kinase domain or that display an unusual architecture do not appear in this tree. These proteins give non-reciprocal blast hits with RTKs. Furthermore, when incorporated in a phylogenetic analysis, most of these protein sequences failed to group with the classic RTK families (not shown). Therefore, these models cannot be confidently assigned as incomplete RTK gene models and will have to be reconsidered at a more advanced stage of assembly. Only SPU-024883 is closely related to Sp-Ret. Sp-Ret largely overlaps SPU-024883 on both sides. In the overlapping region, the nucleotide sequences are almost identical except for an insert in SPU-024883, which lies between 2 exons of Sp-RET. These 2 models resemble protein products from alternative splicing of the same gene. It is possible that they represent 2 different alleles.

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

The general picture that emerges is that the sea urchin genome contains a basic RTK gene set similar to that of vertebrates.

The PDGFR/VEGFR family

The PDGFR and VEGFR families are closely related. Their extracellular domains contain an array of Ig-like domains, 5 for



Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

PDGFR and 7 for VEGFR. In vertebrates, there are five PDGFR and three VEGFR paralogs. In contrast, *Drosophila* has only one receptor gene, PVR, that is related to both families, but possesses seven Ig domains and seems to be closer to VEGFR than to PDGFR. In the *Ciona* genome, a single gene similar to VEGFR was found but no orthologue of PDGFR. A careful phylogenetic study (Grassot et al., 2006) indicates that these two families evolved from a common ancestor which became duplicated after the protostome–deuterostome separation, the two genes having diverged before the appearance of urochordates. Other duplications occurred later during early evolution of the vertebrates to give the complete set of paralogs. In this hypothesis, the PDGFR gene would have been lost in ascidians. Apparently, the PDGFR gene is also lacking in the sea urchin genome. This is surprising since previous studies had strongly implicated the PDGF pathway in sea urchin development (Ramachandran et al., 1993, 1995, 1997). In contrast, two gene models for VEGFR have been found. Both proteins have a higher sequence similarity with VEGFR than with PDGFR, and their kinase domains group with those of the VEGFR (Fig. 1). One of these receptors displays the canonical seven Ig domains (Sp-VEGFR-7) and is likely the sea urchin orthologue to the vertebrate VEGFR. The other protein has a peculiar structure with 10 Ig domains (Sp-VEGFR-10). This structure was already known from cDNA cloning and sequencing in a closely related sea urchin species (C. Gache unpublished) and appears to be specific to the sea urchin. The presence of true VEGFR receptors in the sea urchin is also supported by the identification of several genes coding for their cognate ligands (Table 3). If PDGFR genes are absent in both echinoderms and ascidians, it is possible that a duplication from the common ancestor occurred later than expected. The origin of the atypical VEGFR receptor in the sea urchin is not understood.

INSR and ILGFR

Two gene models, SPU-002840 and SPU-003915, are related to the Insulin Receptor (INSR) family and were designated INSR and ILGRF based on BLAST hits. However, as shown in Fig. 1, their kinase domains do not group with any of the 3 vertebrate subfamilies INSR, IRR and IGF1R. Instead these genes branch at the base of the Insulin receptor sub tree.

FGFR

While two FGFRs (*breathless* and *heartless*) are present in *Drosophila*, only one FGFR (*egl5*) gene is found in *C. elegans* (DeVore et al., 1995) and in *Ciona* (Satou et al., 2003). The diversification leading to the 4 FGFR paralogs found in human is thought to have occurred through two large scale genome duplications during early vertebrate evolution (Itoh et al., 1995). It might thus be predicted that the sea urchin would have only one FGFR gene. However, several incomplete gene models give hits with known FGFRs, suggesting a moderate expansion of this family in Echinoderms. One of these incomplete gene model which encodes a kinase domain with reciprocal hits with FGFR (SPU-004747) is located downstream of a model predicted to contain 3 IG and 1 FnIII domains (SPU-004746), which are typically found in FGFRs. These two models are in fact parts of a single gene (termed FGFR2) since a cDNA clone from the Mediterranean sea urchin *Paracentrotus lividus* contains both the kinase domain and the IG and FnIII domains in a single molecule (T. Lepage, unpublished). The ECD of FGFR2 has the same composition as the FGF receptor (FGFR1) previously cloned but a different organization (McCoon et al., 1996). Its kinase domain is rather divergent and does not group with those from other FGFRs (Fig. 1). Its evolutionary relationship with other RTKs and FGFRs genes should be clarified using different phylogeny methods, focusing on RTKs containing Ig domains in their extracellular region.

We have included in Table 1 gene model SPU-20680, which lacks a catalytic domain. No exons coding for a kinase domain have been identified so far in the same genomic area. As the predicted protein shows strongest sequence similarities with vertebrate FGFRs that also lack kinase domains, it may belong to the same family of decoy proteins related to RTKs. Interestingly, the gene is located next to FGFR1 (SPU-020677) and in the opposite orientation, suggesting a common origin.

Other RTKs

For all other RTK families only one paralog was identified in the sea urchin genome. In vertebrates, MUSK, PTK7, RET, ROS and RYK are also present in the genome as “singletons”. In most cases, however, the vertebrate families consist of several paralogs and families that are implicated in highly specialized functions and organs like the nervous system are greatly expanded. This is clearly the case for the Ephrin receptors that increased during deuterostome evolution from one in sea urchin to 6 in ascidians and 14 in vertebrates.

Inactive RTKs

A number of RTKs are catalytically inactive due to amino acid changes in the kinase domain. The kinase domain has been divided in XI subdomains identified by consensus motifs harboring key amino acid residues (Hanks and Quinn, 1991; Hanks et al., 1988). Subdomain I contains the motif GXGXXGXV, which has a conformational role at the ATP binding site. In subdomain II, the lysine of the conserved VAVK motif interacts directly with the phosphate groups of ATP. The aspartic residue that is part of the motif HRDLAARN found in subdomain VIb is involved in catalysis while the aspartic residue within the DFG motif (subdomain VII) chelates the $Mg2+$ ions of ATP. Motifs that diverge from the consensus have been found in the sequence of the sea urchin RTKs. They are listed in Table 2, together with the sequences from their human homologs. The ROR, RYK and PTK7 kinases from human and other organisms are known to show divergence in these critical motifs. The sea urchin sequences have similar features. The changes in ROR are minor and Sp-ROR is probably active like its vertebrate homolog. In both human and sea urchin RYK, DNA replaces DFG. Some kinases displaying the DNA motif may be active but activity of human RYK was not demonstrated and RYK is generally considered to be inactive. In contrast, Sp-PTK7 lacks DFG and is probably inactive like other members of this family. Although these 2 kinases are catalytically inactive, they are

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

t2.1 Table 2
t2.2 Key residues of the Tyr-kinase catalytic domain

t2.3		GXGXXG	VAVK	HRDLXXXN	DFG	Activity
t2.4	Sp-PTK7Hs-PTK7	GHGAYGGKSEFG	VMVKVLVK	HGDLAARNHKDLAARN	TMSALG	Inactive
t2.5	Sp-RORHs-ROR	GTGTFGGEDRFG	IVIKVAIK	HRDLAARNHKDLATRN	DFGDLG	Active
t2.6	Sp-ROSHs-ROS	EHGSY_GSGAFG	LQLMVAVK	HRDLAARNHRDLAARN	DFGDFG	Active
t2.7	Sp-RYKHs-RYK	LEGTFGQEGTFG	VFIKAFVK	HKDLATRNHKDLATRN	DNADNA	Active ?

t2.8 Consensus motifs harboring key catalytic residues (underlined) of the Tyr-kinase catalytic domain are indicated. Subdomain I: GXGXXG; subdomain II: VAVK; subdomain VIb: HRDLXXXN; subdomain VII: DFG. Sequences from *S. purpuratus* proteins that do not fit with the consensus. Sequences from the human homologs are shown for comparison.

functional. In *Drosophila*, RYK is implicated in axon guidance and in vertebrates RYK is required for development of craniofacial structures probably by association with Ephrin receptors (Halford and Stacker, 2001). PTK7 is involved in the control of planar cell polarity in vertebrates (Lu et al., 2004).

The ROS case is puzzling. ROS is an active RTK but Sp-ROS lacks the VAVK motif and thus a critical K. The GXGXXG motif is also almost completely absent. At this stage, however, it would be premature to conclude that Sp-ROS is inactive. Sequencing errors or inaccuracy of the prediction should be carefully checked.

RTK ligands and docking proteins

As transducers of signals from outside to inside of the cell, RTKs interact with proteins on both sides of the membrane. In the extracellular space, they bind diffusible growth factors or proteins of the ECM. Inside the cell, they interact directly with membrane or cytoplasmic factors that are recruited upon RTK activation and set off the cascades of transduction events (Csiszar, 2006). These factors include enzymes (PLC γ, PI3Kinase p85) and adaptor proteins that interact with the RTKs and with each other through specific protein modules such as PH, PTB, SH2 and SH3 domains.

The cognate ligands that have been identified in the sea urchin genome are listed in Table 3 and some of the enzymes and adaptors that bind directly or are closely linked to the RTKs are listed in Table 4. This initial survey indicates that most of the key partners of the RTKs are indeed present in the sea urchin genome. The kinases that are important downstream components of the RTK transduction pathways are analyzed by Bradham et al. (this issue).

t3.1 Table 3
t3.2 Identified ligands for the RTKs

t3.3	RTK	Known ligand	Ligand name and official ID	Best blast hit (human)	Back blast
t3.4	ALK	orphan ? / pleiotrophin			
t3.5	CCK4/PTK7	(inactive kinase)	–		
t3.6	DDR	collagen	Numerous collagen fragments		
t3.7	EGFR	EGF, TGF-α	n.i.		
t3.8	EPH	ephrin	Sp-Eph, SPU-023757	NP-004084	<->
t3.9	FGFR	FGF	Sp-FGF 9/16/20, SPU-006242	NP-062825	<->
t3.10	ILGFR	insulin-like growth factor	Sp-IGF1, SPU-007203	NP-000609	<->
t3.11			Sp-IGF2, SPU-030139	Not significant	*
t3.12	INSR	insulin	n.i.		
t3.13	LMR	(vestigial ECD)	–		
t3.14	MET/RON	HGF (MSP)	Sp-HGF, SPU-017649	NP-001010933	SPU-000330
t3.15			Sp-HGF-like, SPU-000330	NP-000292	<->
t3.16	MUSK	agrin	SPU-002025+SPU-002467+ SPU-024494+SPU-022633+ SPU-022634		
t3.17	RET	GDNF (to coreceptor GFR)	n.i		
t3.18	ROR	orphan, WNT ?	11 WNT models		
t3.19	ROS	orphan ? BOSS ?	n.i.		
t3.20	RYK	wnt ?	11 WNT models		
t3.21	TIE1/2	angiopoietin	n.i.		
t3.22	TRK	NGF, BDNF, NT3, NT4	Sp-NT, SPU-030073	AAI07076	*
t3.23	VEGFR	VEGF	Sp-VEGF, SPU-014978	NP-004460	<->
t3.24			Sp-VEGF1, SPU-005737	NP-004460	SPU-014978, <->
t3.25			Sp-VEGF3, SPU-030148	NP-001020539	*
t3.26	AXL	Gas6	n.i.		
t3.27	PDGFR	PDGF, CSF1	n.i.		

t3.28 The cognate ligands for each RTK family have been searched amongst Glean models. Name and or SPU numbers are listed together with the results of reciprocal blast analyses except for agrin and collagen for which genes were not complete or not assembled, and for the Wnt (see article by Croce et al., this issue) that might be putative ligands for RYK and ROR. (n.i.), not identified; (<->), if best blast hits are reciprocal or SPU number if they are not; (*) not a Glean model, no back blast. Note that AXL and PDGFR genes have not been found (see Table 1).

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

Table 4
RTK intracellular ligands and close partners

Provisional gene name	Official ID	Best blast hit (human)	Back blast
Sp-Cbl	SPU-007862	NP-078063	<->
	SPU-007863		
Sp-Dok	SPU-021666	NP-003965	<->
Sp-GAB	SPU-007721	NP-536739	<->
Sp-GRB2	SPU-003586	NP-002077	<->
Sp-IRS	SPU-011063	NP-005535	<->
Sp-IRS	SPU-004492	NP-006331	<->
p53/58			
Sp-JAK	SPU-022023	NP-004963	SPU-006988
	SPU-020082		
Sp-JAK2	SPU-022495	NP-004963	SPU-006988
Sp-NCK	SPU-014752	NP-001004722	<->
Sp-PI3K-110	SPU-006197	NP-006209	<->
	SPU-027144		
	SPU-002836		
	SPU-022717		
Sp-PI3K-85	SPU-000206	NP-852556	<->
Sp-PLCg	SPU-027462	NP-002651	<->
Sp-SHC	SPU-008698	NP-079021	<->
Sp-SHP2	SPU-013810	NP-002822	<->
Sp-Src	SPU-004037	NP-0044374	SPU-022112
Sp-STAT	SPU-015108	NP-003143	<->

Gene numbers for proteins known for interacting with RTKs. Accession number of human proteins giving best blast hits; (<->), indicates that best blast hits are reciprocal, SPU number gives best back blast hits when not reciprocal.

Expression of RTK genes during sea urchin development.

As indicated by microarray expression data (Samanta et al., 2006, in this issue), most of the canonical RTKs identified in this study are expressed during early development. Some of these RTKs such as FGFR1 are expressed in surprisingly complex and dynamic pattern during development (Fig. 2) (McCoon et al., 1996, 1998). The complex expression pattern of FGFR1 in the sea urchin embryo is a good illustration of the repeated deployment of signaling pathways during embryogenesis and of their participation in different gene regulatory networks. FGFR1 is expressed ubiquitously during cleavage stages but begins to be expressed more strongly at the vegetal pole in the region where precursors of the skeletogenic mesenchyme (called PMCs) are located starting at the hatched blastula stage (Fig. 2B). Expression of FGFR1 transcripts intensifies in the PMCs at the time they start to ingress into the blastocel, giving the characteristic appearance of an open ring at the vegetal pole (Fig. 2D). Starting at the blastula stage, FGFR1 expression also becomes asymmetrical along the oral–aboral axis (Figs. 2C, D), with a stronger expression in the presumptive oral ectoderm. After ingression of the PMCs, two novel domains of expression appear at the animal pole and in a ring of cells at the vegetal pole that corresponds to the presumptive secondary mesenchyme cell territory (Figs. 2E, F). Cells within this territory will give rise to mesodermal derivatives such as pigment cells, muscle cells and blastocoelar cells. During gastrulation, restricted expression of FGFR1 persists at the animal pole (Fig. 2H) and in the oral ectoderm (Fig. 2I), but FGFR1 is now also transcribed actively in the presumptive endoderm and invaginated archenteron (Figs. 2H–J). Finally, at the prism stage, FGFR1 transcripts are confined to the tip of the archenteron where precursors of the coelomic pouches and pharyngeal muscles are located (McCoon et al., 1998).

These observations indicate that FGFR1 is expressed dynamically in all three germ layers and in several domains with sharp boundaries along both the animal and vegetal axis, which correspond to boundaries of cell fates and to regions undergoing morphogenesis.

In summary, the sea urchin genome harbors RTK gene orthologues that are expressed during development for almost every family found in vertebrates. The only absent families are AXL and PDGFR that might have appeared late during evolution, after the urochordate divergence. For most of the families that are multigenic in vertebrates, the sea urchin has a single paralog, except for two closely related Insulin-like receptors, two FGF receptors and an additional VEGFR receptor with a unique structure. The expansion of these families is known to have taken place during chordate or vertebrate evolution.

The repertoire of TGF-β ligands in the sea urchin genome

To identify the complement of TGF-β superfamily ligands, receptors, signal transducers as well as the transcription factors and regulators involved in TGF-β signaling, we searched the sea urchin genome database with individual vertebrate query sequences. This survey allowed us to identify 14 genes encoding TGF-β-related factors in the sea urchin genome. Phylogenetic analysis indicates that these sequences can be grouped into 11 distinct subfamilies (Fig. 3).

BMP2/4

Members of the BMP2/4 family, which includes the invertebrate gene *decapentaplegic* (*dpp*), are among the best known TGF-β, and have been characterized both in deuterostomes, protostomes and cnidarians such as hydra and *Nematostella* (Matus et al., 2006). Genetic analysis in *Drosophila* has demonstrated the crucial role played by *dpp* in dorsal–ventral patterning (Padgett et al., 1987). Members of the BMP2/4 family also play essential roles in patterning of the dorsal–ventral axis in vertebrates (De Robertis and Kuroda, 2004). The previously characterized *Sp-BMP2/4* gene clearly belongs to the BMP2/4 family as indicated by the phylogenetic and reciprocal best hit analyses (Fig. 3 and Table 5) (Angerer et al., 2000; Duboc et al., 2004). During sea urchin development, *BMP2/4* expression begins at the early blastula stage in the presumptive oral ectoderm and this restricted expression in the oral ectoderm persists during gastrulation (Figs. 4G–I). Intriguingly, at the end of embryogenesis, expression of *BMP2/4* switches from the ectoderm to the mesoderm and from the oral region to the aboral side (Fig. 4J). Functional analysis of BMP2/4 in *S. purpuratus* (Angerer et al., 2000) and in the Mediterranean species *P. lividus* (Duboc et al., 2004) indicates that the key role of this factor in dorsal ventral patterning in bilaterians is conserved in the sea urchin.

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

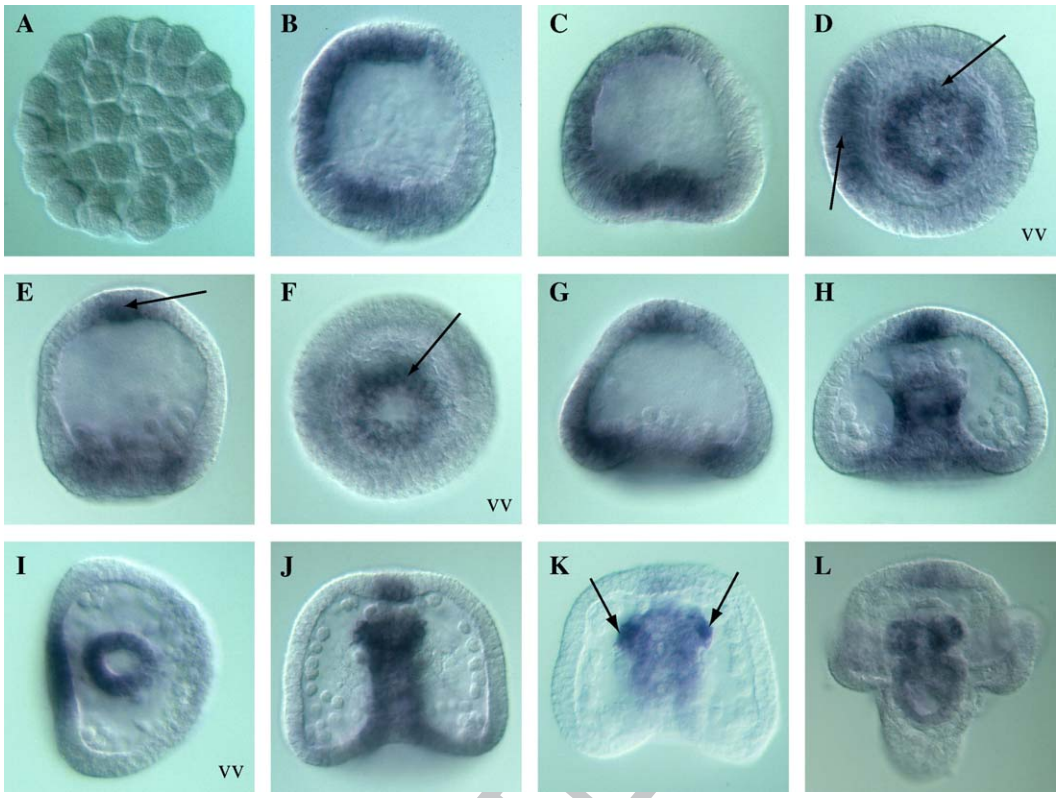


Fig. 2. Expression pattern of the FGFR1 during early development. Embryos of the Mediterranean sea urchin *Paracentrotus lividus* were fixed at the indicated stage and hybridized with sense (not shown) and antisense probes for FGFR1. All the embryos are oriented with the oral side on the left excepted in panels E, H, J, K and L which are viewed from the oral side. (A) 60-cell stage, (B) swimming blastula. (C, D) Early mesenchyme blastula, FGFR1 is expressed predominantly in the oral ectoderm and in the ring of precursors of the PMCs (arrows). (E, F) Mesenchyme blastula. The arrows in panels E and F point respectively to the animal pole region and to precursors of secondary mesenchyme cells. (G) Late mesenchyme blastula, (H, I) early gastrula, (J) late gastrula, (K) Prism stage (the arrows indicate the bilateral coelomic pouches), (L) early pluteus. (vv) Vegetal pole view.

Univin

The *univin* gene was the first TGF-β characterized in the sea urchin (Stenzel et al., 1994). Interestingly, the *univin* gene is located on the same scaffold as *BMP2/4* in the sea urchin genome, only 20 kilobases apart from *BMP2/4*. This close proximity suggests that the two genes originated by gene duplication. Indeed, sequence comparisons indicate that the mature form of Univin is highly related to BMP2/4 (60% identities); however, phylogenetic analysis indicates that this gene belongs to a distinct subfamily which includes GDF1 and GDF3. As shown previously (Stenzel et al., 1994), the *univin* gene is uniformly and strongly expressed maternally and during cleavage (Fig. 4A and data not shown see also Zito et al., 2003). Starting at the blastula stage, *univin* is expressed in a circumequatorial ring of ectodermal cells (Figs. 4B, C) and in the archenteron during gastrulation (Fig. 4D). At the end of embryogenesis, *univin* transcripts are confined to bilateral regions of the ectoderm between the arms of the young pluteus larva (Fig. 4E).

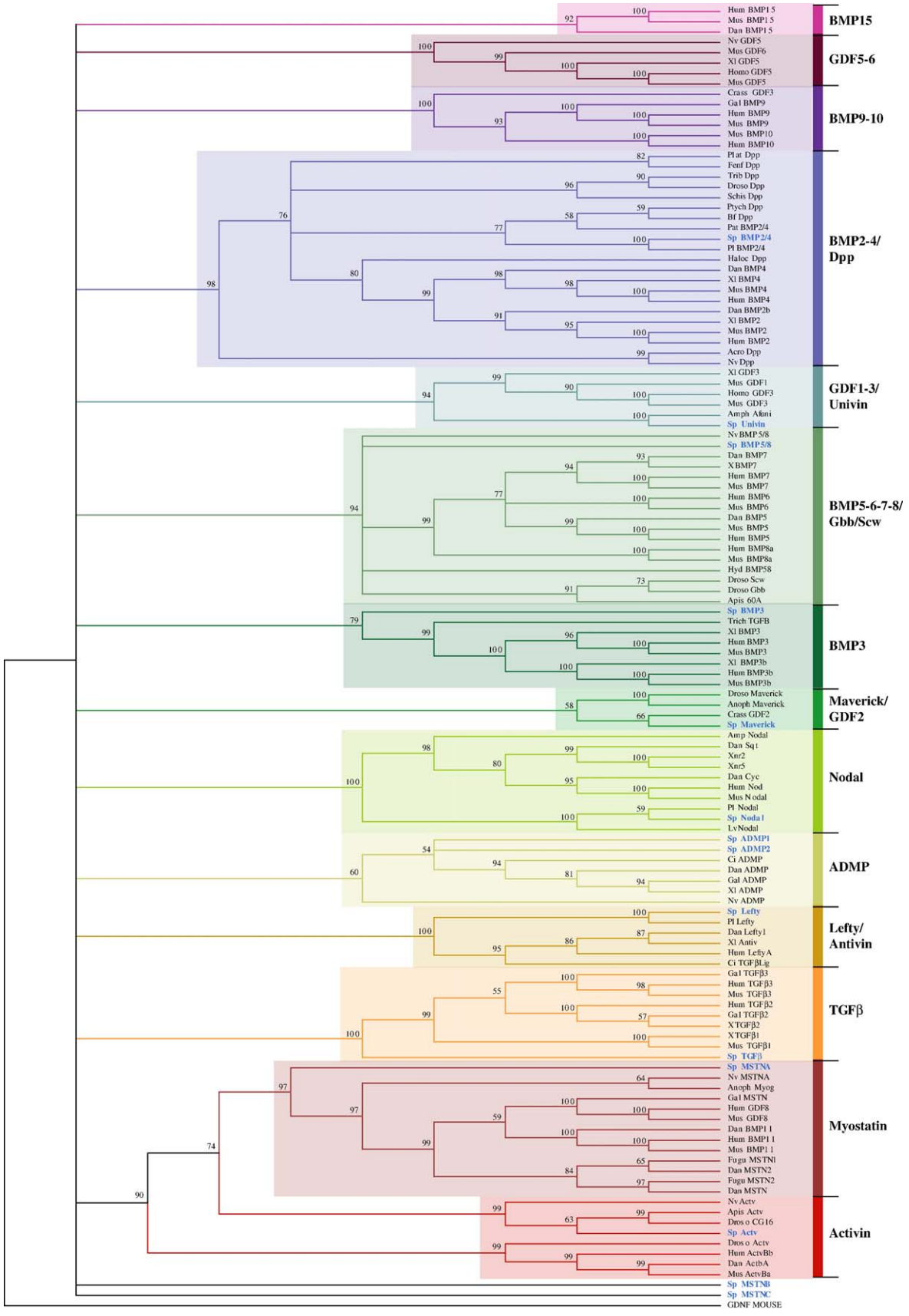
BMP5/6/7/8

The BMP5–8 group is another well-defined subgroup of BMP proteins that displays about 50% identity with BMP2/4. It includes 4 members in vertebrates, two members in *Drosophila*, called Glass bottom boat (Gbb) and Screw, and a single member in the cnidarian *Nematostella* (Matus et al., 2006). In *Drosophila*, Screw is required for patterning of the dorsal ventral axis through heterodimerization with Dpp (Shimmi et al., 2005) while Gbb is required for morphogenesis of the midgut and for growth and patterning of the imaginal discs (Wharton et al., 1999). In vertebrates, BMP5–8 members are required for kidney and eye development, but they do not appear critical for dorsal ventral patterning (Dudley et al., 1995; Luo et al., 1995). The sea urchin genome, like the ascidian genome, contains a single member of the BMP5–8 family (*Sp-BMP5–8*) that is equally related to *gbb* and *screw* (Fig. 3). The sequence of Sp-BMP5–8 was previously characterized by Ponce et al. (1999). The spatial expression pattern of BMP5–8 has not been reported, but microarray experiments indicate that this gene is expressed at a low level during sea urchin development.

BMP3

Members of the BMP3 family have only been described in deuterostomes so far. BMP3 is the most abundant Bone Morphogenetic protein present in demineralized bones but functional studies indicate that its biological activity is to antagonize bone formation (Daluiski et al., 2001). The sea urchin genome contains a single member of this family, whose sequence is about 40% identical with human BMP3 over the ligand region. Transcriptome analysis indicates that this BMP3-

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.



Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

like gene is expressed at very low levels during embryonic development (Samanta et al., 2006).

Maverick/GDF2

Maverick was identified in *Drosophila* (Nguyen et al., 2000) as a TGF-β that could not easily be assigned to previously defined families. The putative Maverick ligand domain contains 9 cysteines which are typically found in Activin and TGF-β sensu stricto factors as well as in a subgroup of BMP proteins that includes the vertebrate BMP/GDF8, BMP/GDF11 and BMP/GDF15. In our analysis, the *Sp-maverick* gene clusters with the fly and *Anopheles maverick* genes as well as with the recently characterized mollusk GDF2 (Herpin et al., 2004). Sp-Maverick shares 32% identical residues within the mature ligand domain with *Drosophila* Maverick. Phylogenetic analysis suggests that the sea urchin gene represents a deuterostome orthologue of the insect *maverick* (bootstrap value: 58%). In situ hybridization of *P. lividus* embryos (data not shown) and microarray array experiments (Samanta et al., 2006), both indicate that *maverick* is expressed at an extremely low level during embryogenesis.

ADMP

The founding member of this family, ADMP (antidorsalizing morphogenetic protein), was first described in *Xenopus* as a TGF-β related to BMP3, which, unlike other BMPs, was expressed exclusively on the dorsal side (Moos et al., 1995). Orthologues of ADMP have since been cloned in a number of vertebrate and chordate species (Hino et al., 2003; Lele et al., 2001; Willot et al., 2002). A single protostome sequence related to ADMP has been described so far (Matus et al., 2006). Therefore, it is not clear whether this gene is part of the ancestral complement of TGF-β in protostomes. Intriguingly, the sea urchin genome contains two distinct sequences that cluster with ADMP in our phylogenetic tree (Fig. 3), which we called *ADMP1* and *ADMP2*. Neither *ADMP1* nor *ADMP2* was accurately predicted by the prediction softwares. In the case of *ADMP1*, only the prodomain was predicted but tiling array data readily identified the missing exons in the adjacent sequence. No gene model was associated with *ADMP2*. The exons encoding the prodomain of this gene were accidentally fused to a gene encoding a transcription factor and the exons encoding the mature ligand were not predicted. RT-PCR analysis was therefore used to validate the structure and confirm the expression of these genes. Sea urchin *ADMP1* and *ADMP2* display about 40% identical residues in the mature ligand region and 26% in the prodomain and are equally similar to vertebrate ADMP (33% identical residues over the whole protein).

Nodal and Lefty

The sea urchin genome contains a single gene related to *nodal* and a single orthologue of *antivin*/*lefty*, which encodes a Nodal antagonist (Duboc et al., 2004; Thisse and Thisse, 1999). Nodal factors have not been described in protostomes so far suggesting that they arose independently in the deuterostome clade. In the sea urchin, Nodal is necessary for two important transitions during embryonic development: first, for the transition from radial to bilateral symmetry by establishing the oral–aboral (ventral–dorsal) axis of the embryo, then, for

Fig. 3. Phylogenetic tree of predicted *S. purpuratus* TGF-β ligands. The amino acid sequences of 16 GLEAN predictions were analyzed to build this tree. Careful examination of genomic sequences in the vicinity of some of these predictions allowed to add or to eliminate missing or incorrectly predicted exons and to detect three artifactual duplications. ADMP2 was not predicted by the GLEAN3 software but was found by TBLASTN analysis of the total genomic DNA (http://urchin.nidcr.nih.gov/blast/index.html). Abbreviations are: Acro: *Acropora milepora* (coral); Anoph: *Anopheles gambiae* (African malaria mosquito); Amph: *Amphiura filiformis* (brittle star); Amp: *Branchiostoma belcheri* (cephalochordate); Apis: *Apis mellifera* (honeybee); Bf *Branchiostoma floridae* (cephalochordate); Ci: *Ciona intestinalis* (ascidian); Crass: *Crassostrea gigas* (oyster); Dm: *Drosophila melanogaster*; Dan: *Danio rerio* (zebrafish); Ef: *Ephydatia fluviatilis* (sponge); fugu: *Takifugu rubripes* (fish); Gal: *Gallus gallus* (chicken); Haloc: *Halocynthia roretzi* (ascidian); Hum: *Homo sapiens*; Hyd: *Hydra littoralis*; Lv: *Lytechinus variegates* (green urchin, Atlantic ocean); Mus: *Mus musculus*; Nv: *Nematostella vectensis* (sea anemone); Pat, *Patella vulgata* (limpet); Pl: *Paracentrotus lividus* (Mediterranean urchin); Plat: *Platynereis dumerilii* (annelid); Pty: *Ptychodera flava* (hemichordate); Schis: *Schistocerca americana* (grasshopper); Sp: *Strongylocentrotus purpuratus* (purple urchin, Pacific ocean); Trib: *Tribolium castaneum* (red flour beetle); Trich: *Trichinella spiralis* (nematode); Xl: *Xenopus laevis*. The following sequences were used to construct the tree (accession number): Hum-BMP2A (P12643), Xl-BMP2A (P25703), Mus-BMP2A (P21274), Hum-BMP4 (P12644), Mus-BMP4 (P21275), Xl-BMP4 (P30885), Xl-ADMP (AAC59736), Dan-ADMP (NP-571951), Ci-ADMP (BAE06303), Gal-ADMP (NP-990153), Sp-Univin (P48970), Pl-BMP2/4 (DQ536194), Xl-BMP3b (Q7T2X6), Xl-BMP3 (Q7T2X7), Mus-BMP6 (P20722), Mus-BMP11 (Q9Z1W4), Hum-BMP11 (Q95390), Mus-BMP7 (P23359), Hum-BMP7 (P18075), Xl-BMP7 (AAT72008), Hyd-BMP58 (AAS01764), Dan-BMP5 (AAH54647), Hum-BMP5 (P22003), Mus-BMP15 (Q9Z0L4), Mus-BMP5 (NP-031581), Hum-BMP15 (NP-005439), Mus-BMP9 (Q9WV56), Hum-BMP9 (Q9UK05), Gal-BMP9 (P34822), Mus-GDF3 (NP-032134), Dan-BMP15 (NP-001018320), Dan-BMP11 (AAN03678), Mus-BMP3b (NP-665684), Hum-BMP6 (P22004), Dan-BMP4 (AAC60285), Dan-BMP2b (BAA24406), Mus-BMP3 (Q8BHE5), Hum-BMP3b (P55107), Hum-BMP3 (P12645), Mus-BMP8a (P34821), Hum-BMP8a (NP-861525), Droso-Gbb (P27091), Droso-Scw (P54631), Droso-Dpp (P07713), Sp-Actv (SPU-07004), Sp-MSTNA (SPU-17647/XP-789990), Sp-BMP3 (SPU-07822/XP-786367), Sp-Nodal (SPU-11064/XM-774841/XM-796712), Pl-Nodal (AAS00534), Mus-Nodal (P43021), Hum-Nodal (AAH33585), Lv-Nodal (AAY41193), Dan-Lefty1 (NP-571035), Xnr5 (BAB18971), Xnr2 (AAA97393), Dan-Cyc (AAC34361), Dan-Sqt (AAC34360), Ci-TGFbLig (BAE06534), Xl-Antiv (AAG35771), Hum-LeftyA (O00292), Droso-Actv (O61643), Mus-TGF-β1 (P04202), Xl-TGF-β2 (P17247), Hum-TGF-β2 (P61812), Xl-TGF-β1 (P16176), Hum-ActBb (P09529), Mus-ActBa (Q04998), Mus-TGF-β3 (P17125), Gal-TGF-β2 (P30371), Sp-TGF-β (SPU-03835/XP-793246), Sp-BMP24 (SPU-00669/XP-787248), Sp-BMP58 (SPU-12786/P48969), Sp-MSTNB (SPU-02795), Sp-MSTNC (SPU-22079/XP-788027), Fugu-MSTN1 (NP-001027843), Fugu-MSTN2 (NP-001027844), Dan-MSTN (O42222), Dan-MSTN2 (AAT95431), Gal-MSTN (O42220), Anoph-Myogl (AAT07311), Plat-Dpp (CAJ38807), Sp-Lefty (SPU-09911/XP-782698), Pl-Lefty (AAS00535), Mus-GDF8 (O08689), Hum-GDF8 (O14793), Dan-ActBa (AAH66402), Crass-GDF3 (CAD67715), Crass-GDF2 (CAD67714), Amp-Nodal (BAC82629), Pat-BMP24 (AAM33143), Nv-GDF5 (AAS77520), Nv-Dpp (AAR27580), Nv-BMP58 (ABC88372), Sp-Maverick (SPU-18248), Haloc-Dpp (BAA31132), Bf-Dpp (AAC97488), Ptych-Dpp (BAA89012), Trib-Dpp (Q26974), Schis-Dpp (AAA81169), Acro-Dpp (AAM54049), Droso-CG16987PA (AAF51204), Mus-GDF5 (P43027), Xl-GDF5 (AAT99303), Homo-GDF5 (P43026), Amph-Afuni (AAX54512), Trich-TGFB (AAQ72736), Apis-60A (XP-394252), Hum-GDF3 (Q9NR23), Xl-GDF3 (AAH73508), Mus-GDF1 (AAH79555), Mus-GDNF (P48540), Anoph-Mvrick (AAT07309), Droso-Mvrick (NP-524626), Sp-ADMP1 (SPU-21726), Mus-GDF6 (P43028), Hum-TGF-β3 (P10600), Gal-TGF-β3 (P16047), Hum-BMP10 (O95393), Mus-BMP10 (Q9R229). The following sequences were kindly provided by Mark Martindale: Nv Actv, Apis Actv, Nv ADMP, Nv MSTNA.

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

Table 5
Predicted TGF-β ligands

Provisional gene name	Official ID	NCBI corresponding accession numbers	Embryonic expression (Tiling data)	Best blast hit (human)	Back blast
Sp-Activin	SPU-007004	–	–	O95390 (GDF11)	SPU-017647 (007004 in 2nd)
Sp-ADMP	SPU-021726	–	+	P18075 (BMP7)	SPU-017647 (021726 in 13th)
Sp-ADMP2	No prediction	–	?	P18075 (BMP7)	–
Sp-BMP2/4	SPU-000669	XM-782155.1 (x)	+	P12644 (BMP4)	SPU-021497 (000669 in 2nd)
	SPU-021497 (x)	XM-785028.1	+		
Sp-BMP3	SPU-007822	XM-781274.1	–	P55107 (BMP3b)	SPU-007822
Sp-BMP5/8	SPU-012786	XM-777775.1	–	P18075 (BMP7)	SPU-012786 (002662 in 3rd)
	SPU-02662 (x)	NM-214655.1 (x)	–		
Sp-Lefty	SPU-009911	XM-777605.1	+	O75610 (LeftyB)	SPU-009911
Sp-Maverick	SPU-018248	–	–	O95390 (GDF11)	SPU-017647 (018248 in 3rd)
Sp-myostatinA	SPU-017647	XM-784897.1	–	O95390 (GDF11)	SPU-017647
Sp-myostatinB	SPU-002795	–	–	O14793 (GDF8)	SPU-017647 (002795 in 11th)
Sp-myostatinC	SPU-022079	XM-782934.1	–	O95390 (GDF11)	SPU-017647 (022079 in 13th)
Sp-Nodal	SPU-011064	XM-774841.1	+	Q96S42 (Nodal homolog)	SPU-011064
		XM-796712.1 (x)			
Sp-TGF-β	SPU-003835	XM-788153.1	–	P61812 (TGFb2)	SPU-003835 (022653 in 8th)
	SPU-022654 (x)	XM-789088.1	–		
Sp-Univin	SPU-000668	NM-214628.1	+	P12645 (BMP2)	SPU-021497 (000668 in 3rd)

The provisional gene name was chosen with respect to the phylogenic analysis and may differ from those of the corresponding Glean and NCBI predictions. SPU numbers are indicated for the predicted ligands. Three of these gene models, SPU-021497, SPU-002662 and SPU-022653 are most likely truncated, artificially duplicated or allelic versions of respectively SPU-000669, SPU-012786 and SPU-003835. These predictions were not incorporated into the phylogenic analysis. The accession numbers corresponding to the automated GNOMON gene predictions from NCBI are indicated when available. The star indicates that part of the NCBI prediction differs from the associated GLEAN3 prediction. Expression tilling data are derived from the hybridization embryonic array data in the genboree browser (www.genboree.org). (+) indicates a significant hybridization signal associated with the predicted exons. Best blast Human, indicates the accession numbers (Swissprot database) and the names of the human genes mostly related to the glean predictions using Blast analysis versus human proteins (www.ncbi.nlm.nih.gov/BLAST/). Back blast indicates the Glean numbers mostly related to the best blast human gene product.

the transition from bilateral to left–right asymmetry by restricting formation of the imaginal rudiment to the left side (Duboc et al., 2004, 2005). These two functions are highly homologous to the roles of Nodal during vertebrate embryogenesis where Nodal signals first specify the dorso-ventral polarity of the embryo and later direct establishment of left–right asymmetries by controlling asymmetrical positioning of various structures and organs. Starting at the 60-cell stage and during blastula and gastrula stages, *nodal* is expressed in the presumptive oral ectoderm territory (Figs. 4K–M, Duboc et al., 2004). At the end of gastrulation, the ectodermal expression of *nodal* is progressively shifted towards the right side of the larva and a novel domain of expression appears at the tip of the archenteron in a group of cells which correspond to the right coelomic pouch precursors (Figs. 4N, O, Duboc et al., 2005).

It is striking that the origin of *nodal* appears to coincide with the emergence of deuterostomes, which are defined by the secondary opening of the stomodeum. An interesting hypothesis is that the ancestral function of Nodal in deuterostomes could be in defining the region where the mouth opens (Chea et al., 2005; Duboc and Lepage, 2006). In sea urchins, which are basal deuterostomes, *nodal* is expressed precisely in the oral ectoderm and is essential for opening of the mouth. Embryos in which the function of Nodal is inhibited do not form a stomodeum. Reciprocally, overexpression of *nodal* results in a presumptive stomodeal region extending radially around the embryo. Furthermore, a random injection of *nodal* mRNA in a single blastomere in an embryo in which endogenous translation of *nodal* has been blocked is sufficient to fully rescue the formation of the mouth. These results are consistent with a function of Nodal in initiating a gene regulatory network that defines the stomodeal field and culminates with the fusion of the archenteron with the ectoderm and the opening of the larval mouth.

In conclusion, these findings indicate that the core of the Nodal signaling pathway was already present in the last common ancestor of chordates and echinoderms. They also suggest that an ancestral function of this pathway was the establishment of left–right asymmetry and perhaps the formation of the stomodeum.

Activin/Inhibins

In contrast to *nodal*, Activin/Inhibins related genes have been described in protostomes (Kutty et al., 1998) and are present in the genomes of organisms with mainly radial organization such as the cnidarian *Nematostella* (Matus et al., 2006). In vertebrates, Activins (which consist of dimers of Inhibin β subunits) are regulators of hormonal secretion and have been implicated in mesoderm formation but their exact function in more basal organisms is not known (Brummel et al., 1999). A single hit was obtained by searching the sea urchin genome against Activin sequences. The mature region of the Sp-Activin protein is about 35% identical to the human Activin and contains 9 cysteines typically found in Activin proteins. Tiling expression data indicate that Activin is expressed at an extremely low level during early development. In situ hybridizations performed on embryos of the Mediterranean sea urchin *P. lividus* indicate that this gene is expressed during late larval stages in the adult rudiment (data not shown).

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

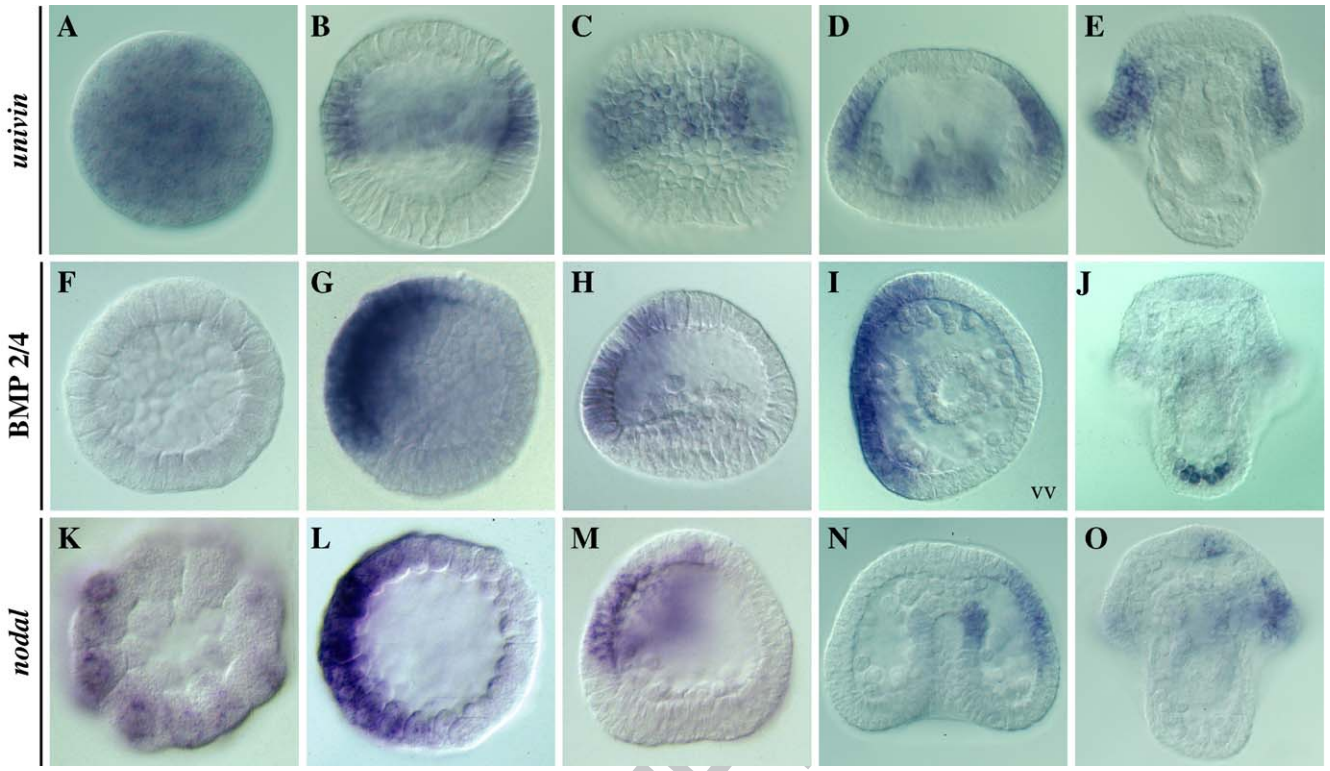


Fig. 4. Expression profiles of *BMP2/4*, *univin* and *nodal* during sea urchin development. (A–O) In situ hybridization of embryos fixed at different stages. All the embryos are oriented with the oral side on the left excepted in panels D, E and J, which are viewed from the oral side and panels N and O which are viewed from the aboral side. (A–E) *univin* probe. (A) Egg stage; (B, C) swimming blastula stage (side and surface views); (D) early gastrula stage; (E) prism stage. (F–J) *BMP2/4* probe, (F) early blastula stage; (G) swimming blastula stage; (H) mesenchyme blastula stage; (I) early gastrula stage animal view; (J) prism stage animal view. (K–L) *nodal* probe, (K) 60-cell stage; (L) early blastula stage; (M) mesenchyme blastula stage; (N) early gastrula stage; (O) prism stage animal view.

TGF-β sensu stricto

Members of the prototypic TGF-β subfamily were discovered as multifunctional cytokines that regulate proliferation, differentiation and inflammation during normal development and tissue repair. So far, clear orthologues of the original TGF-β have not been characterized in invertebrates. A sequence

Table 6
Extracellular modulators of TGF-β signaling and Proprotein convertases

Provisional gene name	Function	Official ID	Best blast hit (human)	Back blast
Sp-noggin	Antagonizes BMP signaling	SPU-024769	Q13253: Noggin	SPU-024769
Sp-chordin	Antagonizes BMP signaling	SPU-004983	Q9H2X0:Chordin	SPU-004983
Sp-follistatin	Antagonizes Activin and BMP signaling	SPU-024994	P19883: Follistatin	SPU-004994
Sp-Gremlin	BMP antagonist	SPU-020330	Q9H772: Gremlin-2	SPU-020330
Sp-Dan	May antagonize BMP signaling	SPU-019983	P41271: Neuroblastoma suppressor of tumorigenicity 1	SPU-019983
Sp-Sclerostin	May antagonize BMP signaling	Novel	NP-056279: Cystine knot-containing secreted protein	
Sp-SFRP	Antagonist of Wnt and BMP signaling	SPU-011271	Q5T4F7: Secreted frizzled-related protein 5	SPU-011271
Sp-tsg	Facilitates diffusion of TGF-β/Chordin complexes	SPU-009756	Q96K46: Twisted gastrulation	SPU-009756
Sp-BMP-1/tolloïd	Cleaves chordin/TGF-β complexes	SPU-007317	P13497: Bone morphogenetic protein 1	SPU-007317
		SPU-011551	Q9Y6L7: Tolloid-like protein 2	SPU-007317
		SPU-011552	Q9Y6L7: Tolloid-like protein 2	SPU-007317
Sp-LTBP	Forms complexes with TGF-β and ECM			
Sp-NOMO	Antagonizes Nodal signaling	SPU-014645	Q5JPE7: Nodal modulator 2	SPU-014645
		SPU-007315	Q5JPE7: Nodal modulator 2	SPU-014645
Sp-HtrA2	Antagonizes BMP/Actv/TGF-β signaling	SPU-012489	043464: Serine protease HTRA2	SPU-012489
Sp-Glypican3/5 Class	Antagonizes TGF-β signaling	SPU-013086	P78333: Glypican-5	SPU-013086
Sp-Furin	Processes TGF-β precursors	SPU-028030	P09958: Furin precursor	SPU-028030
		SPU-002615	Q92824: Proprotein convertase subtilisin/kexin type 5	SPU-002615
		SPU-010722	Q92824: Proprotein convertase subtilisin/kexin type 5	SPU-002615
Sp-Subtisilin	May process TGF-β precursors	SPU-026664	Q16549: Proprotein convertase subtilisin/kexin type 7	SPU-026664
		SPU-023813	P16519: Neuroendocrine convertase 2 precursor	SPU-023813

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

Table 7
TGF-β receptors and co-receptors

Provisional gene name	Official ID	Best blast hit (human)	Back blast
Type I receptors			
Sp-Alk2	SPU-016008	Q04771 (ACVR1)	SPU-016008
Sp-Alk3–6	SPU-016272	O00238 (BMR1B)	SPU-016272
Sp-Alk4–5–7	SPU-028066	P36897 (TGFR1)	SPU-028066
Type II receptors			
Sp-TGF-β receptor type II	SPU-017511	P37173 (TGFR2)	SPU-017511
Sp-BMP type II receptor	SPU-011711	Q13873 (BMPR2)	SPU-011711
Sp-ACVR2	SPU-024092	P27037 (AVR2A)	SPU-024092
Type III coreceptors			
Sp-Cryptic	SPU-000841	Q9GZR3 Cryptic)	SPU-000841
Sp-Tgfbr3	SPU-027380	Q03167 (TGBR3)	SPU-027380

strongly related to TGF-β sensu stricto (about 50% identical residues with the human TGF-β1 over the mature ligand domain) is present in the sea urchin genome (Table 5). This gene, called Sp-TGF-β, is the first TGF-β characterized in a non-chordate deuterostome (bootstrap value: 100%). Tiling array experiments (Samanta et al., 2006, in this issue), and RT-PCR analyses (data not shown) indicate that it is expressed at a low level during sea urchin early development.

Myostatins

Myostatins (GDF8), and the related TGF-β family protein BMP/GDF11, are potent negative regulators of skeletal muscle growth (McPherron et al., 1997). One gene highly related to myostatin has been characterized in *Drosophila* (Lo and Frasch, 1999) and in the sea anemone *Nematostella* (Matus et al., 2006). Intriguingly, searching the sea urchin genome against the vertebrate Myostatin protein yielded three different sequences highly related to Myostatin. As shown by the best hit analysis and the maximum likelihood analysis, one of them, Sp-myostatinA, is likely the orthologue of the vertebrate myostatin gene; however, it is important to note that the phylogenetic analysis failed to clearly assign Sp-myostatinB and Sp-myostatinC to any specific group and so, the phylogenetic relationships of these two TGF-β family proteins remain to be established.

In conclusion, the sea urchin genome contains at least 14 open reading frames encoding cytokines of the TGF-β superfamily. This number is significantly larger than the number of genes encoding TGF-β in *Nematostella* (6 genes), *C. elegans* (6 genes) or in *Drosophila* (9 genes) and even superior to the number of TGF-β identified in the ascidian genome (10 genes). Although comparisons between clades are difficult to make because some species are known to have undergone extensive secondary gene loss (Kortschak et al., 2003), the sea urchin family of TGF-β may provide a good example of the expansion of the gene tool kit that accompanied the emergence of the deuterostome lineage.

Extracellular modulators of TGF-β activity

We identified several genes encoding inhibitors of BMP signaling including Chordin, Noggin, SFRP (Secreted Frizzled related Proteins), Sclerostin and two members of the DAN/Cerberus family which contains five members in vertebrates (Table 6). Remarkably, several of these genes including *SFRP*, *Sclerostin* and *Dan* have not yet been described in protostomes. One possibility is that these genes emerged in the deuterostome lineage. Alternatively, the absence of these genes in the genomes of *Drosophila* or *C. elegans* may indicate that they have been lost during evolution of these phyla which are known to have undergone considerable secondary gene loss.

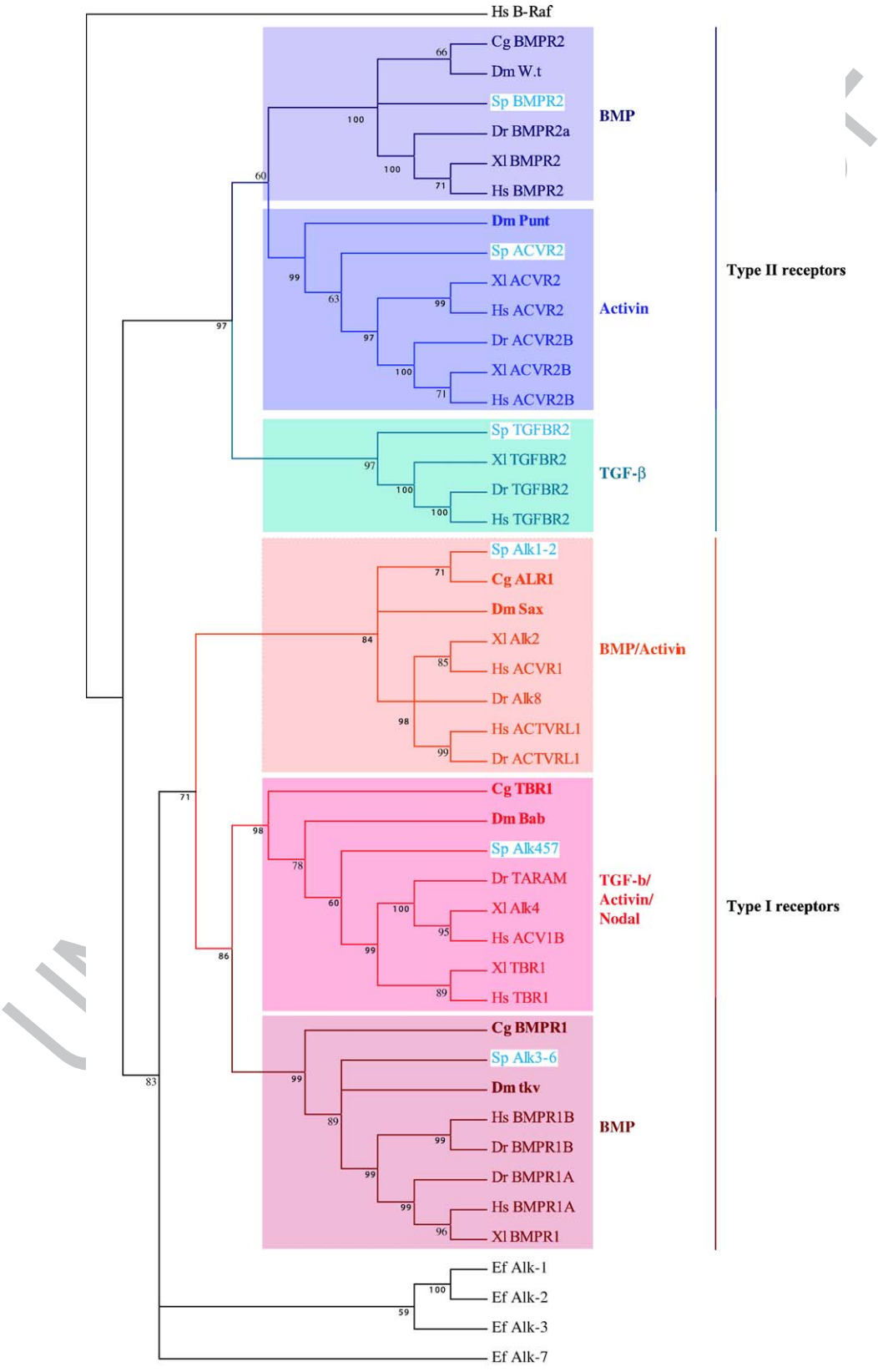
Follistatin is a secreted protein that contains cysteine rich domains also found in extracellular matrix proteins such as Agrin. Follistatin binds to Activin and prevents its binding to the receptor. In *Xenopus*, Follistatin was demonstrated to bind to and to inhibit BMPs (Fainsod et al., 1997; Hemmati-Brivanlou et al., 1994). We identified a gene likely encoding Follistatin in the sea urchin genome (Table 6). The corresponding protein shows a bidirectional best hit with the Human inhibitor of Activin and therefore likely corresponds to the orthologue of Follistatin.

The activity of TGF-β ligands is also regulated indirectly by metalloproteases of the BMP1/Tolloïd family that cleave Chordin complexed with BMP and Twisted gastrulation (De Robertis et al., 2000). In vertebrates, 3 *tolloid/BMP1-like* genes are known and two have been described in *Drosophila*. In the sea urchin, several gene models (SPU-007317, SPU-011551

Fig. 5. Phylogenetic relationships between TGF-β receptor superfamily members. This tree was generated by using an alignment made with ClustalW. From the alignment, a maximum likelihood based phylogenetic tree was constructed using PHYML with a substitution model WAG. Five hundred bootstraps were performed. Protostomes sequences are indicated in bold. Abbreviations are: Ef: *Ephydatia fluviatilis* (sponge), Cg: *Crassostrea gigas* (oyster), Dm: *Drosophila melanogaster*, Dr: *Danio rerio* (zebrafish), Hs: *Homo sapiens*, Sp: *Strongylocentrotus purpuratus*, Xl: *Xenopus laevis*, Xt: *Xenopus tropicalis*. The following sequences were used to construct the tree (accession number): Ef Alk-1 (BAA82601.1), Ef Alk-2 (BAA82602.1), Ef Alk-3 (BAA82603.1), Ef Alk-7 (BAA82607.1), Cg TBR1 (CAD66433.1), Cg BMPR1 (CAE11917.1), Cg ALR1 (CAC85263.1), Cg BMPR2 (CAD20574.1), Dm Tkv (AAA28996.1), Dm Sax (AAA18208.1), Dm Bab (NP-477000.1), Dm W.t (NP-524692.3), Dm Punt (AAC41566.1), Sp Alk1–2 (SPU-16008), Sp Alk3–6 (SPU-16272), Sp Alk457 (SPU-28066), Sp TGFBR2 (SPU-17511), Sp BMPR2 (SPU-11711), Sp ACVR2 (SPU-24092), Dr ACTVRL1 (AAI00044.1), Dr BMPR1A (NP-571696.1), Dr BMPRB (NP-571532.1), Dr TARAM (CAA63840.1), Dr Alk8 (AAG01346.1), Dr ACVR2 (Q56E96), Dr ACVR2B (Q9YGU4), Dr BMPR2a (Q288P3), Dr TGFBR2 (NP-878275.2), Xl Alk2 (AAB71328.1), Xl BMPR1 (AAA58707.1), Xl TBR1 (AAA84997.1), Xl Alk4 (AAB03621.1), Xl BMPR2 (P79954), Xl TGFBR2 (Q9DE31), Xl ACVR2 (P27039), Xt ACVR2B (Q6DEV8), Hs ACTVRL1 (P37023), Hs ACV1B(P36896), Hs TBR1 (P36897), Hs ACVR1 (Q04771), Hs BMPR1A (P36894), Hs BMPR1B (O00238), Hs TGFBR2 (P37173), Hs BMPR2 (Q13873), Hs ACVR2B (Q13705), Hs ACVR2 (NP-001607.1), Hs B-Raf (P15056).

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

and SPU-011552) encode proteins that are mostly similar to BMP1/Tolloïd. SPU-007317 encodes the uniformly expressed protein suBMP1 that has been previously cloned (Hwang et al., 1994) while SPU-011551 and SPU-011552 are probably parts of the same gene. Microarray data indicate that only SPU-007317 is expressed during development (Samanta et al., 2006). In addition to these genes, the sea urchin genome sequence contains a cluster of 5 genes encoding proteins mostly related to



Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

SPAN and BP10 proteins (Lepage et al., 1992; Reynolds et al., 1992), that are also related to Tolloïd (this feature is discussed in detail in the article by Angerer et al. in this issue). Although the function of these *tolloïd*-related genes is not known, the proteases they encode may potentially participate in the regulation of TGF-β activity in the extracellular space as suggested previously (Lepage et al., 1992; Reynolds et al., 1992).

In summary, an inventory of extracellular modulators of BMP signaling in the sea urchin genome indicates that Echinoderms have a large repertoire of such modulators. This repertoire is similar to that present in vertebrates suggesting that the expansion of the number of modulators accompanied the expansion of the number of TGF-β ligands.

TGF-β receptors

The sea urchin complement of TGF-β receptors is made of 3 type I and 3 type II receptors (Table 7). Sp-Alk1/2, Sp-Alk3/6 and Sp-Alk4/5/7 are the type I receptors while Sp-BMPR2, Sp-ACVR2 and Sp-TGFBR2 are the cognate type II receptors. Phylogenetic analysis and best-hit analysis unambiguously assigned each of these 6 receptors to one of the 6 known subfamilies of TGF-β receptors (Fig. 5). This complement of receptors is very similar to the complement of receptors found in *Drosophila*. In comparison, the vertebrate genome contains no less that 7 type I and 5 type II receptors, allowing potentially more than 30 combinations of homo and heterodimers. Therefore, the significant expansion of TGF-β ligands present in echinoderms was not accompanied by an increase in the repertoire of receptors raising the challenging question of how these different ligands use this limited set of receptors to mediate their effects.

In vertebrates, BMP signaling is negatively regulated by a pseudoreceptor called BAMBI (BMP and Activin Membrane Bound Inhibitor) in *Xenopus* or Nma in humans (Onichtchouk et al., 1999). The extracellular domain of BAMBI shows similarity to TGF-β receptors, but the protein lacks the intracellular kinase domain and behaves as a dominant negative receptor. We did not identify any orthologue of BAMBI in the current assembly of the sea urchin genome, suggesting that this gene emerged after the divergence of Echinoderms from the other deuterostome lineages or that it was lost in echinoderms. In contrast, we identified a member of the EGF-CFC family Oep/Crypto/FRL1 which in vertebrates is absolutely required for Nodal signaling and establishment of left right asymmetry (Gritsman et al., 1999).

Smads, Smad-interacting transcriptional regulators and Smad ubiquitin ligases

A survey of the Smad-related factors in the sea urchin revealed the classical triad of Receptor Regulated Smads, common Smads and Inhibitory Smads (see Howard et al. in this issue and Table 8). Two gene models (SPU-020722 and SPU-023107) are derived from the same gene and are homologous to Smad1, Smad5 and Smad8 which are recognized by BMP receptors (Massague, 1998; Miyazono et al., 2000). Sp-Smad2/3 is predicted by SPU-017642 and is homologous to the vertebrate Smad2 and Smad3 which mediate the effects of TGF-β sensu stricto, Nodal and Activin. Besides this pair of Receptor Regulated Smads, one gene encoding Sp-Smad4 is associated with two predictions (SPU-004287 and SPU-017971). Similarly, two gene models (SPU-001998 and SPU-018246) are predicted to encode an inhibitory Smad, Sp-Smad6/7 but are likely derived from the same gene. The sea urchin repertoire of Smads, which is made of 4 genes, is therefore very similar to the repertoire found in *Drosophila*. Intriguingly, one of the gene model predicted (SPU-000739) encodes a protein that contains a domain homologous to the MH2 region of Smads but which lacks a MH1 domain. The MH2 domain of SPU-000739 is preceded by a 180 amino acid region, which is not homologous to the SMADs and loosely homologous to various proteins. The absence of a MH1 domain linked to this MH2 region led us to provisionally exclude this sequence from the set of putative Smad factors.

The versatility of TGF-β factors and the large diversity of responses they can elicit result from the interaction of the Smads with a myriad of other protein partners (Massague et al., 2005). These protein partners regulate the interaction of the Smad complex with other transcriptional activators and repressors, accounting for the so-called "cellular context" that determines the transcriptional output of TGF-β signaling. Most of the transcription factors, coactivators and corepressors identified as Smad binding partners in vertebrates are present in the sea urchin genome (Table 9) including TGIF (Wotton et al., 1999), SIP1 (Verschueren et al., 1999), OAZ (Hata et al., 2000), Runx1 (Hanai et al., 1999), AP1 (Zhang et al., 1998), E2F (Chen et al., 2002), Sp1 (Pardali et al., 2000), Evi1 (Kurokawa et al., 1998) and FoxO (Seoane et al., 2004). A notable exception is the

Table 8
Smads and MH2 containing genes

Provisional gene name	Function	Official ID	Best blast hit (human)	Back blast
Sp-Smad1/5/8	Activated by BMPs	SPU-020722	Q99717 (SMAD5)	SPU-020722
		SPU-023107	Q99717 (SMAD5)	SPU-020722
Sp-Smad2/3	Activated by TGF-β and Activin	SPU-017642	P84022 (SMAD3)	SPU-017642
Sp-Smad4	Common mediator of TGF-βs (co-SMAD)	SPU-004287	Q13485 (SMAD4)	SPU-004287
		SPU-017971	Q13485 (SMAD4)	SPU-004287
Sp-Smad6	Antagonist of signaling by TGF-βs	SPU-001998	O43541 (SMAD6)	SPU-001998
		SPU-018246	O43541 (SMAD6)	SPU-001998
Sp-MH2		SPU-000739	Q99717 (SMAD5)	SPU-020722

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

t9.1 Table 9
t9.2 Smad interacting transcription factors

	Provisional gene name	Function	Official ID	Best blast hit (human)	References
t9.4	Sp-ATF2	ATF, CREB family, cooperates with Smad3	SPU-026905	NP-001871: activating transcription factor 2	(Sano et al., 1999)
t9.5	Sp-beta catenin	Functionally cooperates with Smad4	SPU-009155	P35222: CTNNB1 (β-catenin)	(Nishita et al., 2000)
t9.6	Sp-FAST	Fork head transcription factor cooperates with Smad2, 3	Not found		(Chen et al., 1997)
t9.7	Sp-GLI3	Zinc finger transcription factor	SPU-017627	NP-084657: GLI-Kruppel family member GLI2	(Liu et al., 1998)
t9.8	Sp-mix/mixer/milk	Paired-like homeodomain, cooperates with Smad2	SPU-004366	NP-114150: Mix-like homeobox protein 1	(Germain et al., 2000)
t9.9	Sp-Jun	AP-1 transcription factor complex, cooperates with Smads	SPU-003102	NP-002219: v-jun avian sarcoma virus 17 oncogene homolog	(Zhang et al., 1998)
t9.10	Sp-FoxG/BF-1	Transcriptional repressor	SPU-009771	Q14488: Forkhead box protein G1	(Rodriguez et al., 2001)
t9.11	Sp-Fos	AP-1 transcription factor complex, cooperates with Smads	SPU-021172 SPU-021174	NP-005244: FOS-like antigen 2	(Zhang et al., 1998)
t9.12	Sp-FoxO	Smad3 transcriptional partner for the activation of p21 cyclin -dependent inhibitors	SPU-009179	Q12778: Forkhead box protein O1A	(Seoane et al., 2004)
t9.13	Sp-E2F	Transcription activator	SPU-006753 SPU-028828	NP-001940: E2F transcription factor 3	(Chen et al., 2002)
t9.14	Sp-Evi-1	Zinc Finger transcription factor inhibits Smad3	SPU-018797	NP-955533: PR domain containing 16 isoform 2	(Kurokawa et al., 1998)
t9.15	Sp-Lef1	HMG box transcription repressor	SPU-009520	Q5VVR8: Transcription factor 7-like 2	(Nishita et al., 2000)
t9.16	Sp-NFKB	Functionally cooperates with Smad3	SPU-008177	P19838: NFKB1 (nuclear factor NF-kappa-B p105 subunit)	(Lopez-Rovira et al., 2000)
t9.17	Sp-p300CBP	Transcription coactivator, Histone deacetylase (HDAC)	SPU-019024	Q92793: CREBBP (CREB-binding protein)	(Feng et al., 1998; Janknecht et al., 1998; Nishihara et al., 1998; Pouponnot et al., 1998; Shen et al., 1998; Topper et al., 1998)
t9.18	Sp-P/CAF	Transcription coactivator, Histone deacetylase (HDAC)	SPU-000371	Q92830: GCNL2 (histone acetyltransferase GCN5)	(Itoh et al., 2000)
t9.19	Sp-Runx1	Runt domain protein, cooperates with Smads	SPU-006917 SPU-007853	Q01196: RUNX1 (Runt-related transcription factor 1)	(Hanai et al., 1999)
t9.20	Sp-SARA	Scaffold protein	SPU-014763	NP-004790: Zinc finger, FYVE domain containing	(Tsukazaki et al., 1998)
t9.21	Sp-SIP1	Zinc Finger Homeodomain transcriptional repressor	SPU-022242	NP-055610: zinc finger homeobox 1b	(Verschueren et al., 1999)
t9.22	Sp-Smicl	Cleavage and Polyadenylation Specificity Factor (CPSF)	SPU-022195 SPU-003053	Q8IXZ2: Zinc finger CCCH-type domain-containing protein 3	(Collart et al., 2005)
t9.23	Sp-SP1	Zinc finger transcription factor	SPU-024190	Q02446: SP4-HUMAN (Transcription factor Sp4)	(Pardali et al., 2000)
t9.24	Sp-Ski/Sno	Transcription co-repressor	SPU-010659 SPU-017676	NP-003027: v-ski sarcoma viral oncogene homolog NP-001032891: functional smad suppressing element	(Akiyoshi et al., 1999; Wang et al., 2000)
t9.25	Sp-Swift	BRCT domain containing protein cooperates with Smad2	SPU-027111	Q14676: Nuclear factor with BRCT domains 1	(Shimizu et al., 2001)
t9.26	Sp-TGIF	Transcription co-repressor	SPU-018126	NP-777480: TG-interacting factor isoform d	(Wotton et al., 1999)
t9.27	Sp-TFE3	HLH domain transcription factor	SPU-008175	P19532: Transcription factor E3	
t9.28	Sp-Tob/BTG	Negative regulator of BMP signaling	SPU-016792 SPU-021549	NP-005740: transducer of ERBB2, 1; NP-001722: B-cell translocation protein 1	(Yoshida et al., 2000)
t9.29	Sp-OAZ/EBF	Zinc finger transcription factor positive regulator of BMP signaling	SPU-004702	Q9H4W6: EBF3	(Hata et al., 2000)
t9.30	Other intracellular modulators				
t9.31	Sp-Smurf	Smad1 E3 ubiquitin ligases	SPU-025856	Q9HAU4: Smad ubiquitination regulatory factor 2	
t9.32	Sp-Dapper	promotes degradation of Nodal Receptor	not found		
t9.33	Sp-Ectodermin	Smad4 ubiquitin ligase	SPU-005708	Q13263: Ectodermin	
t9.34	Sp-FKBP12	Binds to the unphosphorylated GS box of the receptors	SPU-001569	P68106: FK506-binding protein 1B	
t9.35	Sp-LTBP				
t9.36	Sp-NOMO	Antagonizes Nodal signaling			
t9.37	Sp-HtrA1	Antagonizes TGF-β signaling			
t9.38	Sp-Glypican	Antagonizes TGF-β signaling	SPU-013086	P78333: Glypican-5	

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

Forkhead domain containing gene FoxH (FAST), which was the first transcription factor reported to interact with Smads and which mediates Nodal signaling in vertebrates. This gene appears to be absent from the sea urchin genome (see the article by Tu et al. in this issue). Finally, in addition to the highly conserved FKBP12 protein (Choi et al., 1996), several Smad cofactors such as Ski (Pardali et al., 2000), Tob (Yoshida et al., 2000), Smicl (Collart et al., 2005) and two genes encoding the Smad ubiquitin ligases Smurf and Ectodermin were identified (Table 9).

Conclusion

An in silico inventory of sea urchin genes belonging to two signaling pathways particularly important during embryonic development, the receptor tyrosine kinase and the TGF-β signaling pathways, indicates that an almost complete repertoire of these genes is represented in basal deuterostomes. Most of these genes are present as single copy in the sea urchin genome, and are expressed during early development with sometimes very complex and dynamic patterns suggesting their implication in different gene regulatory networks. Analysis of evolutionary relatedness shows that nearly all these genes are more related to vertebrate genes rather than to invertebrate sequences. Since echinoderms are basal deuterostomes, these genes can be considered as the part of the common genetic toolkit for intercellular signaling of deuterostomes. The next challenge will be to analyze the function of these factors during sea urchin development. With the apparent lack of gene redundancy and the availability of gene knockdown techniques by injection of antisense morpholino oligonucleotides, the sea urchin embryo, which has largely contributed for over a century to the study of the role of cell interactions during development, will undoubtedly continue to be a very attractive model to address these questions.

Materials and methods

The sea urchin genome database and GLEAN3 gene list (28944 predictions) were searched using TBLASTN and BLASTP (Altschul et al., 1997) using as queries a comprehensive set of individual vertebrates Receptor Tyrosine Kinases sequences as well as sequences belonging to the TGF-β, TGF-β receptors, Smads, transcription factors acting downstream of Smads, Smad cofactors and extracellular or intracellular modulators of this signaling pathway.

In the case of RTKs, either the entire RTK sequence or partial sequences corresponding to the kinase domain or interacting domains present in this class of proteins were used as query.

In the case of the TGF-β ligands, we also searched the Protein family (Pfam) database with PF00688, PF00019 which define the TGF-β propeptide and TGF-β mature ligand domains.

The predicted open reading frames were analyzed using the precomputed information available in the sea urchin annotation database and the GENBOREE viewer and the *S. purpuratus* genome research tools available at http://urchin.nidcr.nih.gov/blast/index.html. The domain organization of the putative proteins was deduced using algorithms from SMART (Simple Modular Architecture Research Tool) and InterproScan (http://www.ebi.ac.uk/InterProScan/).

The putative translated protein sequences were aligned with the protein sequences of known members from different species as well as with *P. lividus* sequences when available using ClustalW (Thompson et al., 1994). The global organization of the protein (length, nature, organization and number of domains, presence of a catalytic domain) was verified. When available, ESTs were used to validate the gene predictions. In most cases, the GLEAN3 program failed to predict accurately the 5′ end of the proteins and the signal peptides. The predicted exons/intron boundaries were checked against the tiling array expression data (Samanta et al., 2006).

Phylogenetic analysis

Predicted amino acid sequences corresponding to the catalytic domain of putative *S. purpuratus* Receptor Tyrosine kinases were selected using the SMART software. Sequences from kinase domains were aligned with ClustalX and the tree was generated by the neighbor-joining method with 1000 bootstrap replications.

For TGF-β and the TGF-β receptors, the sequences of the complete precursors (containing respectively the prodomains and mature ligands and the extracellular ligand binding domain and the kinase domain) were used in the alignments. Full-length sequences were aligned using ClustalW with default parameters (http://www.ebi.ac.uk/clustalw/), gap optimization and obvious alignment error corrections were made using Bioedit 7.0.5.3 (http://www.mbio.ncsu.edu/BioEdit/bioedit.html). The tree was calculated using the maximum likelihood method using the PHYML software (Guindon et al., 2005) with substitution model WAG (http://atgc.lirmm.fr/phyml/). A consensus tree with 50% cut off value was derived from 500 bootstrap analysis using Mega 3.1 (http://www.megasoftware.net/). Numbers above branches represent bootstrap values. The 113 additional taxons sequences were collected from diverse databases using the NCBI research tool (http://www.ncbi.nlm.nih.gov/).

In situ hybridization

In situ hybridization was performed following a protocol adapted from Harland (1991) with antisense RNA probes and staged embryos. A partial clone encoding the *P. lividus* FGFR1 cDNA (McCoon et al., 1996, 1998) was isolated in the course of an in situ hybridization screen (T. Lepage unpublished data). A full-length cDNA was subsequently isolated by library screening. The *P. lividus univin* cDNA was isolated using RT-PCR and library screening (T. Lepage unpublished). The *P.l BMP2/4* and *P.l nodal* clones were described previously (Duboc et al., 2004). All probes were synthesized from full-length cDNA clones in Bluescript after linearization with *Not*I and using T7 RNA polymerase.

The accession numbers for the *P. lividus* cDNA sequences described here have been submitted to Genebank: FGFR1: DQ536196, BMP2/4: DQ536194, Univin: DQ536195.

Acknowledgments

We thank Lydia Besnardeau for her help with the in situ hybridizations, Fabrice Girardot, Sebastien Fiorucci and Lucas Leclère for fruitful discussions and help with the bioinformatics, David Matus and Mark Martindale for sharing sequences alignments files, Evelyn Houliston for careful reading of the manuscript and our colleagues of the Marine station of Villefranche-sur-Mer for their support. The following people have contributed to the annotation of the genes listed in this paper: Dan Mellot (*Ephrin*), Andrew Cameron (*ectodermin*), Christine Byrum (*CCC4/PTK7*), Richard Hynes and Charlie Wittaker (*FGFR2, VEGFR*), Ronghui Xu (*RYK*), Mike Thorndyke (*IGF1*), Tony Robertson (*HGF-like*), Garry Wessel (*agrin*), Robert Burke (*agrin*), Christine Byrum (*VEGF*), Karl-Frederick Bergeron (*VEGF2*, β-Glycan), Cynthia Messier (*VEGF3*), Wendy Beane (*PI3Kp85, GRB2*), Bertrand Cosson (*IRSp53/58*), Kathy Foltz (*PLCγ*), Rebecca Thomason (*STAT*), Meredith Howard-Ashby (*Smad1*, *Smad3*, *Smad4*, *Smad5*, *TGIF, E2F*), Stephan C. Materna (*ATF2*, *Jun*, *NFKB*, *SIP1*, *Evi1*); James Coffman (*Runx1*); Jenifer Croce (*ATF1*, *p300CBP*, *HtrA2*), Paola Oliveri (*FoxO*, *FoxG*), Shuguang Liang (*OAZ*).

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

This work was supported by the Association pour la Recherche sur le Cancer (grant 4778 and 3287), the Ministère de la Recherche, the CNRS and the Université Pierre et Marie Curie-Paris 6. E.R was supported by a fellowship from the ARC foundation.

References

Akiyoshi, S., Inoue, H., Hanai, J., Kusanagi, K., Nemoto, N., Miyazono, K., Kawabata, M., 1999. c-Ski acts as a transcriptional co-repressor in transforming growth factor-beta signaling through interaction with smads. J. Biol. Chem. 274, 35269–35277.

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.

Angerer, L.M., Angerer, R.C., 2003. Patterning the sea urchin embryo: gene regulatory networks, signaling pathways, and cellular interactions. Curr. Top Dev. Biol. 53, 159–198.

Angerer, L.M., Oleksyn, D.W., Logan, C.Y., McClay, D.R., Dale, L., Angerer, R.C., 2000. A BMP pathway regulates cell fate allocation along the sea urchin animal–vegetal embryonic axis. Development 127, 1105–1114.

Balemans, W., Van Hul, W., 2002. Extracellular regulation of BMP signaling in vertebrates: a cocktail of modulators. Dev. Biol. 250, 231–250.

Brummel, T., Abdollah, S., Haerry, T.E., Shimell, M.J., Merriam, J., Raftery, L., Wrana, J.L., O'Connor, M.B., 1999. The *Drosophila* activin receptor baboon signals through dSmad2 and controls cell proliferation but not patterning during larval development. Genes Dev. 13, 98–111.

Chea, H.K., Wright, C.V., Swalla, B.J., 2005. Nodal signaling and the evolution of deuterostome gastrulation. Dev. Dyn. 234, 269–278.

Chen, X., Weisberg, E., Fridmacher, V., Watanabe, M., Naco, G., Whitman, M., 1997. Smad4 and FAST-1 in the assembly of activin-responsive factor. Nature 389, 85–89.

Chen, C.R., Kang, Y., Siegel, P.M., Massague, J., 2002. E2F4/5 and p107 as Smad cofactors linking the TGFbeta receptor to c-*myc* repression. Cell 110, 19–32.

Choi, J., Chen, J., Schreiber, S.L., Clardy, J., 1996. Structure of the FKBP12–rapamycin complex interacting with the binding domain of human FRAP. Science 273, 239–242.

Collart, C., Remacle, J.E., Barabino, S., van Grunsven, L.A., Nelles, L., Schellens, A., Van de Putte, T., Pype, S., Huylebroeck, D., Verschueren, K., 2005. Smicl is a novel Smad interacting protein and cleavage and polyadenylation specificity factor associated protein. Genes Cells 10, 897–906.

Csiszar, A., 2006. Structural and functional diversity of adaptor proteins involved in tyrosine kinase signalling. Bioessays 28, 465–479.

Daluiski, A., Engstrand, T., Bahamonde, M.E., Gamer, L.W., Agius, E., Stevenson, S.L., Cox, K., Rosen, V., Lyons, K.M., 2001. Bone morphogenetic protein-3 is a negative regulator of bone density. Nat. Genet. 27, 84–88.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z., Schilstra, M.J., Clarke, P.J., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., Bolouri, H., 2002. A genomic regulatory network for development. Science 295, 1669–1678.

De Robertis, E.M., Kuroda, H., 2004. Dorsal–ventral patterning and neural induction in *Xenopus* embryos. Annu. Rev. Cell Dev. Biol. 20, 285–308.

De Robertis, E.M., Larrain, J., Oelgeschlager, M., Wessely, O., 2000. The establishment of Spemann's organizer and patterning of the vertebrate embryo. Nat. Rev. Genet. 1, 171–181.

DeVore, D.L., Horvitz, H.R., Stern, M.J., 1995. An FGF receptor signaling pathway is required for the normal cell migrations of the sex myoblasts in *C. elegans* hermaphrodites. Cell 83, 611–620.

Derynck, R., Feng, X.H., 1997. TGF-beta receptor signaling. Biochim. Biophys. Acta 1333, F105–F150.

Dosch, R., Gawantka, V., Delius, H., Blumenstock, C., Niehrs, C., 1997. Bmp-4 acts as a morphogen in dorsoventral mesoderm patterning in *Xenopus*. Development 124, 2325–2334.

Duboc, V., Lepage, T., 2006. A conserved role for the nodal signaling pathway in the establishment of dorso-ventral and left–right axes in deuterostomes. J. Exp. Zoolog. B. Mol. Dev. Evol. 306, B:1–B:12.

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

Duboc, V., Rottinger, E., Besnardeau, L., Lapraz, F., Lepage, T., 2005. Left–right asymmetry in the sea urchin embryo is regulated by Nodal signalling on the right side. Developmental Cell 8, 1–12.

Dudley, A.T., Lyons, K.M., Robertson, E.J., 1995. A requirement for bone morphogenetic protein-7 during development of the mammalian kidney and eye. Genes Dev. 9, 2795–2807.

Fainsod, A., Deissler, K., Yelin, R., Marom, K., Epstein, M., Pillemer, G., Steinbeisser, H., Blum, M., 1997. The dorsalizing and neural inducing gene follistatin is an antagonist of BMP-4. Mech. Dev. 63, 39–50.

Feng, X.H., Zhang, Y., Wu, R.Y., Derynck, R., 1998. The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for smad3 in TGF-beta-induced transcriptional activation. Genes Dev. 12, 2153–2163.

Finnerty, J.R., Pang, K., Burton, P., Paulson, D., Martindale, M.Q., 2004. Origins of bilateral symmetry: Hox and dpp expression in a sea anemone. Science 304, 1335–1337.

Germain, S., Howell, M., Esslemont, G.M., Hill, C.S., 2000. Homeodomain and winged-helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif. Genes Dev. 14, 435–451.

Grassot, J., Gouy, M., Perriere, G., Mouchiroud, G., 2006. Origin and molecular evolution of receptor tyrosine kinases with immunoglobulin-like domains. Mol. Biol. Evol.

Green, B.A., Smith, J.C., 1990. Graded changes in dose of a *Xenopus* activin A homologue elicit stepwise transitions in embryonic cell fate. Nature 347, 391–394.

Gritsman, K., Zhang, J., Cheng, S., Heckscher, E., Talbot, W.S., Schier, A.F., 1999. The EGF–CFC protein one-eyed pinhead is essential for nodal signaling. Cell 97, 121–132.

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.

Halford, M.M., Stacker, S.A., 2001. Revelations of the RYK receptor. Bioessays 23, 34–45.

Hanai, J., Chen, L.F., Kanno, T., Ohtani-Fujita, N., Kim, W.Y., Guo, W.H., Imamura, T., Ishidou, Y., Fukuchi, M., Shi, M.J., Stavnezer, J., Kawabata, M., Miyazono, K., Ito, Y., 1999. Interaction and functional cooperation of PEBP2/CBF with Smads. Synergistic induction of the immunoglobulin germline Calpha promoter. J. Biol. Chem. 274, 31577–31582.

Hanks, S.K., Quinn, A.M., 1991. Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. Methods Enzymol. 200, 38–62.

Hanks, S.K., Quinn, A.M., Hunter, T., 1988. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241, 42–52.

Harland, R.M., 1991. In situ hybridization: an improved whole mount method for *Xenopus* embryos. In: Kay, B.K., Peng, H.J. (Eds.), Methods in Cell Biology, Vol. 36. Academic Press Inc, San Diego, CA, pp. 685–695.

Hata, A., Seoane, J., Lagna, G., Montalvo, E., Hemmati-Brivanlou, A., Massague, J., 2000. OAZ uses distinct DNA- and protein-binding zinc fingers in separate BMP-Smad and Olf signaling pathways. Cell 100, 229–240.

Hemmati-Brivanlou, A., Kelly, O.G., Melton, D.A., 1994. Follistatin, an antagonist of activin, is expressed in the Spemann organizer and displays direct neutralizing activity. Cell 77, 283–295.

Herpin, A., Lelong, C., Favrel, P., 2004. Transforming growth factor-beta-related proteins: an ancestral and widespread superfamily of cytokines in metazoans. Dev. Comp. Immunol. 28, 461–485.

Hino, K., Satou, Y., Yagi, K., Satoh, N., 2003. A genomewide survey of

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

developmentally relevant genes in *Ciona intestinalis*: VI. Genes for Wnt, TGFbeta, Hedgehog and JAK/STAT signaling pathways. Dev. Genes Evol. 213, 264–272.

Hubbard, S.R., Till, J.H., 2000. Protein tyrosine kinase structure and function. Annu. Rev. Biochem. 69, 373–398.

Hunter, T., Cooper, J.A., 1985. Protein–tyrosine kinases. Annu. Rev. Biochem. 54, 897–930.

Hwang, S.P., Partin, J.S., Lennarz, W.J., 1994. Characterization of a homolog of human bone morphogenetic protein 1 in the embryo of the sea urchin, *Strongylocentrotus purpuratus*. Development 120, 559–568.

Itoh, K., Tang, T., Neel, B., Sokol, S.Y., 1995. Specific modulation of ectodermal cell fates in *Xenopus* embryos by glycogen synthase kinase. Development 121, 3979–3988.

Itoh, S., Ericsson, J., Nishikawa, J., Heldin, C.H., ten Dijke, P., 2000. The transcriptional co-activator P/CAF potentiates TGF-beta/Smad signaling. Nucleic. Acids Res. 28, 4291–4298.

Janknecht, R., Wells, N.J., Hunter, T., 1998. TGF-beta-stimulated cooperation of smad proteins with the coactivators CBP/p300. Genes Dev. 12, 2114–2119.

Khokha, M.K., Yeh, J., Grammer, T.C., Harland, R.M., 2005. Depletion of three BMP antagonists from Spemann's organizer leads to a catastrophic loss of dorsal structures. Dev. Cell 8, 401–411.

King, N., 2004. The unicellular ancestry of animal development. Dev. Cell 7, 313–325.

King, N., Carroll, S.B., 2001. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. Proc. Natl. Acad. Sci. U.S.A. 98, 15032–15037.

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

Kostich, M., English, J., Madison, V., Gheyas, F., Wang, L., Qiu, P., Greene, J., Laz, T.M., 2002. Human members of the eukaryotic protein kinase family. Genome Biol. 3 (RESEARCH0043).

Kurokawa, M., Mitani, K., Irie, K., Matsuyama, T., Takahashi, T., Chiba, S., Yazaki, Y., Matsumoto, K., Hirai, H., 1998. The oncoprotein Evi-1 represses TGF-beta signalling by inhibiting Smad3. Nature 394, 92–96.

Kutty, G., Kutty, R.K., Samuel, W., Duncan, T., Jaworski, C., Wiggert, B., 1998. Identification of a new member of transforming growth factor-beta superfamily in *Drosophila*: the first invertebrate activin gene. Biochem. Biophys. Res. Commun. 246, 644–649.

Lecuit, T., Brook, W.J., Ng, M., Calleja, M., Sun, H., Cohen, S.M., 1996. Two distinct mechanisms for long-range patterning by Decapentaplegic in the *Drosophila* wing. Nature 381, 387–393.

Lele, Z., Nowak, M., Hammerschmidt, M., 2001. Zebrafish admp is required to restrict the size of the organizer and to promote posterior and ventral development. Dev. Dyn. 222, 681–687.

Lepage, T., Ghiglione, C., Gache, C., 1992. Spatial and temporal expression pattern during sea urchin embryogenesis of a gene coding for a protease homologous to the human protein BMP-1 and to the product of the *Drosophila* dorsal–ventral patterning gene tolloid. Development 114, 147–163.

Liu, F., Massague, J., Ruiz i Altaba, A., 1998. Carboxy-terminally truncated Gli3 proteins associate with Smads. Nat. Genet. 20, 325–326.

Lo, P.C., Frasch, M., 1999. Sequence and expression of myoglianin, a novel *Drosophila* gene of the TGF-beta superfamily. Mech. Dev. 86, 171–175.

Lopez-Rovira, T., Chalaux, E., Rosa, J.L., Bartrons, R., Ventura, F., 2000. Interaction and functional cooperation of NF-kappa B with Smads. Transcriptional regulation of the junB promoter. J. Biol. Chem. 275, 28937–28946.

Lu, X., Borchers, A.G., Jolicoeur, C., Rayburn, H., Baker, J.C., Tessier-Lavigne, M., 2004. PTK7/CCK-4 is a novel regulator of planar cell polarity in vertebrates. Nature 430, 93–98.

Luo, G., Hofmann, C., Bronckers, A.L., Sohocki, M., Bradley, A., Karsenty, G., 1995. BMP-7 is an inducer of nephrogenesis, and is also required for eye development and skeletal patterning. Genes Dev. 9, 2808–2820.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S., 2002. The protein kinase complement of the human genome. Science 298, 1912–1934.

Massague, J., 1998. TGF-beta signal transduction. Annu. Rev. Biochem. 67, 753–791.

Massague, J., Seoane, J., Wotton, D., 2005. Smad transcription factors. Genes Dev. 19, 2783–2810.

Matus, D.Q., Thomsen, G.H., Martindale, M.Q., 2006. Dorso/ventral genes are asymmetrically expressed and involved in germ-layer demarcation during cnidarian gastrulation. Curr. Biol. 16, 499–505.

McCoon, P.E., Angerer, R.C., Angerer, L.M., 1996. SpFGFR, a new member of the fibroblast growth factor receptor family, is developmentally regulated during early sea urchin development. J. Biol. Chem. 271, 20119–20125.

McCoon, P.E., Blackstone, E., Angerer, R.C., Angerer, L.M., 1998. Sea urchin FGFR muscle-specific expression: posttranscriptional regulation in embryos and adults. Dev. Biol. 200, 171–181.

McDowell, N., Zorn, A.M., Crease, D.J., Gurdon, J.B., 1997. Activin has direct long-range signalling activity and can form a concentration gradient by diffusion. Curr. Biol. 7, 671–681.

McPherron, A.C., Lawler, A.M., Lee, S.J., 1997. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. Nature 387, 83–90.

Miyazono, K., ten Dijke, P., Heldin, C.H., 2000. TGF-beta signaling by Smad proteins. Adv. Immunol. 75, 115–157.

Miyazono, K., Kusanagi, K., Inoue, H., 2001. Divergence and convergence of TGF-beta/BMP signaling. J. Cell Physiol. 187, 265–276.

Moos Jr., M., Wang, S., Krinks, M., 1995. Anti-dorsalizing morphogenetic protein is a novel TGF-beta homolog expressed in the Spemann organizer. Development 121, 4293–4301.

Nellen, D., Burke, R., Struhl, G., Basler, K., 1996. Direct and long-range action of a DPP morphogen gradient. Cell 85, 357–368.

Nguyen, M., Parker, L., Arora, K., 2000. Identification of maverick, a novel member of the TGF-beta superfamily in *Drosophila*. Mech. Dev. 95, 201–206.

Nishihara, A., Hanai, J.I., Okamoto, N., Yanagisawa, J., Kato, S., Miyazono, K., Kawabata, M., 1998. Role of p300, a transcriptional coactivator, in signalling of TGF-beta. Genes Cells 3, 613–623.

Nishita, M., Hashimoto, M.K., Ogata, S., Laurent, M.N., Ueno, N., Shibuya, H., Cho, K.W., 2000. Interaction between Wnt and TGF-beta signalling pathways during formation of Spemann's organizer. Nature 403, 781–785.

Onichtchouk, D., Chen, Y.G., Dosch, R., Gawantka, V., Delius, H., Massague, J., Niehrs, C., 1999. Silencing of TGF-beta signalling by the pseudoreceptor BAMBI. Nature 401, 480–485.

Padgett, R.W., St. Johnston, R.D., Gelbart, W.M., 1987. A transcript from *Drosophila* pattern gene predicts a protein homologous to the transforming growth factor-b family. Nature 325, 81–84.

Pardali, K., Kurisaki, A., Moren, A., ten Dijke, P., Kardassis, D., Moustakas, A., 2000. Role of Smad proteins and transcription factor Sp1 in p21(Waf1/Cip1) regulation by transforming growth factor-beta. J. Biol. Chem. 275, 29244–29256.

Ponce, M.R., Micol, J.L., Peterson, K.J., Davidson, E.H., 1999. Molecular characterization and phylogenetic analysis of SpBMP5–7, a new member of the TGF-beta superfamily expressed in sea urchin embryos. Mol. Biol. Evol. 16, 634–645.

Pouponnot, C., Jayaraman, L., Massague, J., 1998. Physical and functional interaction of SMADs and p300/CBP. J. Biol. Chem. 273, 22865–22868.

Ramachandran, R.K., Seid, C.A., Lee, H., Tomlinson, C.R., 1993. PDGF-BB and TGF-alpha rescue gastrulation, spiculogenesis, and LpS1 expression in collagen-disrupted embryos of the sea urchin genus Lytechinus. Mech. Dev. 44, 33–40.

Ramachandran, R.K., Govindarajan, V., Seid, C.A., Patil, S., Tomlinson, C.R., 1995. Role for platelet-derived growth factor-like and epidermal growth factor-like signaling pathways in gastrulation and spiculogenesis in the *Lytechinus* sea urchin embryo. Dev. Dyn. 204, 77–88.

Ramachandran, R.K., Wikramanayake, A.H., Uzman, J.A., Govindarajan, V., Tomlinson, C.R., 1997. Disruption of gastrulation and oral–aboral ectoderm differentiation in the *Lytechinus pictus* embryo by a dominant/negative PDGF receptor. Development 124, 2355–2364.

Reynolds, S.D., Angerer, L.M., Palis, J., Nasir, A., Angerer, R.C., 1992. Early mRNAs, spatially restricted along the animal–vegetal axis of sea urchin embryos, include one encoding a protein related to tolloid and BMP-1. Development 114, 769–786.

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-

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.

binding antagonist of transforming growth factor-beta signaling. J. Biol. Chem. 276, 30224–30230.
Sano, Y., Harada, J., Tashiro, S., Gotoh-Mandeville, R., Maekawa, T., Ishii, S., 1999. ATF-2 is a common nuclear target of Smad and TAK1 pathways in transforming growth factor-beta signaling. J. Biol. Chem. 274, 8949–8957.
Satou, Y., Sasakura, Y., Yamada, L., Imai, K.S., Satoh, N., Degnan, B., 2003. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*: V. Genes for receptor tyrosine kinase pathway and Notch signaling pathway. Dev. Genes Evol. 213, 254–263.
Seoane, J., Le, H.V., Shen, L., Anderson, S.A., Massague, J., 2004. Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. Cell 117, 211–223.
Shen, X., Hu, P.P., Liberati, N.T., Datto, M.B., Frederick, J.P., Wang, X.F., 1998. TGF-beta-induced phosphorylation of Smad3 regulates its interaction with coactivator p300/CREB-binding protein. Mol. Biol. Cell 9, 3309–3319.
Shi, Y., Massague, J., 2003. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell 113, 685–700.
Shilo, B.Z., 1992. Roles of receptor tyrosine kinases in *Drosophila* development. Faseb. J. 6, 2915–2922.
Shimizu, K., Bourillot, P.Y., Nielsen, S.J., Zorn, A.M., Gurdon, J.B., 2001. Swift is a novel BRCT domain coactivator of Smad2 in transforming growth factor beta signaling. Mol. Cell Biol. 21, 3901–3912.
Shimmi, O., Umulis, D., Othmer, H., O'Connor, M.B., 2005. Facilitated transport of a Dpp/Scw heterodimer by Sog/Tsg leads to robust patterning of the *Drosophila* blastoderm embryo. Cell 120, 873–886.
Stenzel, P., Angerer, L.M., Smith, B.J., Angerer, R.C., Vale, W.W., 1994. The univin gene encodes a member of the transforming growth factor-beta superfamily with restricted expression in the sea urchin embryo. Dev. Biol. 166, 149–158.
Suga, H., Ono, K., Miyata, T., 1999. Multiple TGF-beta receptor related genes in sponge and ancient gene duplications before the parazoan-eumetazoan split. FEBS Lett. 453, 346–350.
Thisse, C., Thisse, B., 1999. Antivin, a novel and divergent member of the TGFbeta superfamily, negatively regulates mesoderm induction. Development 126, 229–240.
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.
Topper, J.N., DiChiara, M.R., Brown, J.D., Williams, A.J., Falb, D., Collins, T., Gimbrone Jr., M.A., 1998. CREB binding protein is a required coactivator for Smad-dependent, transforming growth factor beta transcriptional responses in endothelial cells. Proc. Natl. Acad. Sci. U.S.A. 95, 9506–9511.
Tsukazaki, T., Chiang, T.A., Davison, A.F., Attisano, L., Wrana, J.L., 1998. SARA, a FYVE domain protein that recruits Smad2 to the TGFbeta receptor. Cell 95, 779–791.
Verschueren, K., Remacle, J.E., Collart, C., Kraft, H., Baker, B.S., Tylzanowski, P., Nelles, L., Wuytens, G., Su, M.T., Bodmer, R., Smith, J.C., Huylebroeck, D., 1999. SIP1, a novel zinc finger/homeodomain repressor, interacts with Smad proteins and binds to 5′-CACCT sequences in candidate target genes. J. Biol. Chem. 274, 20489–20498.
Wang, W., Mariani, F.V., Harland, R.M., Luo, K., 2000. Ski represses bone morphogenic protein signaling in *Xenopus* and mammalian cells. Proc. Natl. Acad. Sci. U.S.A. 97, 14394–14399.
Wharton, K.A., Cook, J.M., Torres-Schumann, S., de Castro, K., Borod, E., Phillips, D.A., 1999. Genetic analysis of the bone morphogenetic protein-related gene, gbb, identifies multiple requirements during *Drosophila* development. Genetics 152, 629–640.
Willot, V., Mathieu, J., Lu, Y., Schmid, B., Sidi, S., Yan, Y.L., Postlethwait, J.H., Mullins, M., Rosa, F., Peyrieras, N., 2002. Cooperative action of ADMP- and BMP-mediated pathways in regulating cell fates in the zebrafish gastrula. Dev. Biol. 241, 59–78.
Wilson, P.A., Lagna, G., Suzuki, A., Hemmati-Brivanlou, A., 1997. Concentration-dependent patterning of the Xenopus ectoderm by BMP4 and its signal transducer Smad1. Development 124, 3177–3184.
Wotton, D., Lo, R.S., Lee, S., Massague, J., 1999. A Smad transcriptional corepressor. Cell 97, 29–39.
Yoshida, Y., Tanaka, S., Umemori, H., Minowa, O., Usui, M., Ikematsu, N., Hosoda, E., Imamura, T., Kuno, J., Yamashita, T., Miyazono, K., Noda, M., Noda, T., Yamamoto, T., 2000. Negative regulation of BMP/Smad signaling by Tob in osteoblasts. Cell 103, 1085–1097.
Zhang, Y., Feng, X.H., Derynck, R., 1998. Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. Nature 394, 909–913.
Zito, F., Costa, C., Sciarrino, S., Poma, V., Russo, R., Angerer, L.M., Matranga, V., 2003. Expression of univin, a TGF-beta growth factor, requires ectoderm-ECM interaction and promotes skeletal growth in the sea urchin embryo. Dev. Biol. 264, 217–227.

Please cite this article as: François Lapraz et al., RTK and TGF-β signaling pathways genes in the sea urchin genome, Developmental Biology (2006), doi:10.1016/j.ydbio.2006.08.048.