EVOLUTION & DEVELOPMENT 8:2, 150–173 (2006)

Developmental expression of transcription factor genes in a demosponge: insights into the origin of metazoan multicellularity

Claire Larroux,[a] Bryony Fahey,[a] Danielle Liubicich,[a,1] Veronica F. Hinman,[a,2] Marie Gauthier,[a] Milena Gongora,[a] Kathryn Green,[a] Gert Wörheide,[a,3] Sally P. Leys,[a,b] and Bernard M. Degnan[a,*]

[a]School of Integrative Biology, University of Queensland, Brisbane, Qld 4072, Australia
[b]Department of Biological Sciences CW 405, University of Alberta, Edmonton, AB, Canada T6G 2E9
*Author for correspondence (email: b.degnan@uq.edu.au)
[1]Present address: Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, USA.
[2]Present address: Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA.
[3]Present address: Geoscience Centre Göttingen, Department of Geobiology, Goldschmidtstr.3, 37077 Göttingen, Germany.

SUMMARY Demosponges are considered part of the most basal evolutionary lineage in the animal kingdom. Although the sponge body plan fundamentally differs from that of other metazoans, their development includes many of the hallmarks of bilaterian and eumetazoan embryogenesis, namely fertilization followed by a period of cell division yielding distinct cell populations, which through a gastrulation-like process become allocated into different cell layers and patterned within these layers. These observations suggest that the last common ancestor (LCA) to all living animals was developmentally more sophisticated than is widely appreciated and used asymmetric cell division and morphogen gradients to establish localized populations of specified cells within the embryo. Here we demonstrate that members of a range of transcription factor gene classes, many of which appear to be metazoan-specific, are expressed during the development of the demosponge *Reniera*, including ANTP, *Pax*, *POU*, *LIM-HD*, *Sox*, *nuclear receptor*, *Fox* (*forkhead*), *T-box*, *Mef2*, and *Ets* genes. Phylogenetic analysis of these genes suggests that not only the origin but the diversification of some of the major developmental metazoan transcription factor classes took place before sponges diverged from the rest of the Metazoa. Their expression during demosponge development suggests that, as in today's sophisticated metazoans, these genes may have functioned in the regulatory network of the metazoan LCA to control cell specification and regionalized gene expression during embryogenesis.

INTRODUCTION

For most of the 20th Century, sponges—phylum Porifera—have been relegated to the realm of "almost animals"—the Parazoa. Their unusual adult body plans have been viewed simply as colonies of cells that lack many metazoan characters including true tissues, a true epithelium (i.e., with polarized cells, cell–cell junctions, and a clearly identifiable basement membrane), neurons and muscle cells. However, recent molecular phylogenies suggest something different, that sponges are basal metazoans (Aguinaldo et al. 1997; Collins 1998; Kruse et al. 1998; Muller et al. 1998; de Rosa et al. 1999; Medina et al. 2001). Although these phylogenies have yet to fully resolve the relationships amongst Porifera, Cnidaria, Ctenophora and Placozoa (e.g., see, Rodrigo et al. 1994; Manuel et al. 2003), they indicate that demosponges and hexactinellid sponges together form the most basal metazoan clade (Borchiellini et al. 2001; Medina et al. 2001; Cavalier-Smith and Chao 2003).

From the perspective of the adult body plan, sponges do indeed appear to have limited affinities with the rest of the animal kingdom. It is particularly striking then, that accumulating data tell a different story about sponge embryonic development and larval form—these may be much more similar to other animal phyla. Our recent analyses of early development and larval metamorphosis in *Reniera* sp., a haplosclerid demosponge, have revealed that it possesses many of the cardinal features of metazoan development (for details see Leys and Degnan 2001, 2002; Leys 2003a, 2004; Degnan et al. 2005; Leys et al. 2005, Fig. 1). After fertilization there is a period of cell division with little or no cell growth. Cleavage ends with an asymmetric cell division, which gives rise to two cell populations—micromeres and macromeres. This is followed by a period of cell sorting, clearly akin to gastrulation (Leys and Degnan 2002; Leys 2004; Leys et al. 2005) that produces an embryo consisting of inner macromeres and outer micromeres. Subpopulations of micromeres then directionally migrate to specific regions of the embryo along the

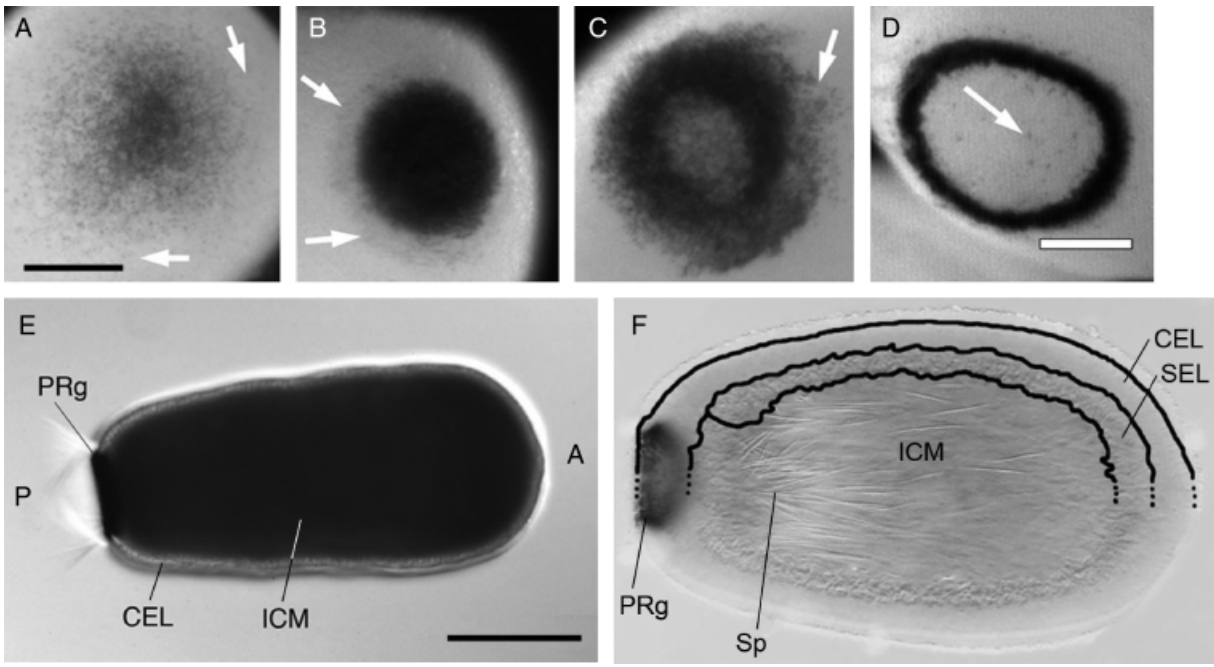

Fig. 1. *Reniera* embryos and larva. Development in *Reniera* occurs in brood chambers of 50–150 oocytes, embryos and larvae. At approximately the 128-cell stage, cleavage is unequal producing two populations of cells: micromeres and macromeres. Subpopulations of the micromeres begin differentiating to form either a single cilium or initiate spicule secretion; some also begin accumulating pigment. Micromeres and macromeres segregate to form a bi-layered embryo in which a layer of micromeres surrounds a solid inner mass of larger cells in a dense collagen matrix. Later a third cell layer will form between the outer micromere layer and the inner cell mass (Leys and Degnan 2001, 2002). (A) Anterior-posterior polarity of the embryo first becomes evident with the migration of pigmented ciliated micromeres and spicule-synthesizing sclerocytes to the posterior end. (B–D) The pigment cells coalesce to first form a dark spot and then eventually a dark ring around the posterior pole. (E) Swimming *Reniera* larva, anterior (A) to right. CEL, columnar epithelial layer; ICM, inner cell mass; PRg, pigment ring at the posterior (P). (F) Cleared mounted larva showing the three cell layers (demarcated by a black line in the upper half). SEL, sub-epithelial (middle) layer; Sp, spicules. Scale bars: A, 50 μm; D, 25 μm; E, 250 μm.

anterior–posterior (AP) axis (Degnan et al. 2005). At this stage, a middle cell layer also forms. The similarity between this layer and that seen in bilaterians and some cnidarians has been long recognized (Hyman 1940), although its relationship to the mesoderm in triploblastic animals is not yet understood. Embryogenesis culminates in a swimming larva composed of at least 11 cell types, each of which is allocated to a different cell layer and in some cases patterned along the AP axis, which is defined by the direction of larval swimming and can be readily identified by the posterior location of a pigment ring (Leys and Degnan 2001, 2002).

Similarities between the development of this demosponge and that of many eumetazoans lead us to infer that the last common ancestor (LCA) to all metazoans was likely to have required asymmetric cell division, organizers, morphogen gradients, and populations of cells with differing competencies to construct its body plan (Degnan et al. 2005). Here we explore the possibility that the metazoan LCA also employed highly conserved components of the regulatory networks known to underlie these developmental processes and features in extant eumetazoans (Davidson 2001; Carroll et al. 2005). We do this by determining if key regulatory transcription factor genes are expressed during *Reniera* development. Several studies already have identified the presence of metazoan developmental genes in demosponge genomes (e.g., Degnan et al. 1993, 1995; Seimiya et al. 1994, 1997; Hoshiyama et al. 1998; Adell et al. 2003; Coutinho et al. 2003; Perovic et al. 2003; Wiens et al. 2003a, b; Adell and Muller 2004) but none have yet addressed which of these are indeed expressed during embryogenesis, larval development and metamorphosis. Here, we demonstrate for the first time that an extensive range of metazoan transcription factor genes, including members of the ANTP class (outside *Hox*, *ParaHox*, and extended-*Hox* clades), *Pax*, *POU*, *LIM-HD*, *Sox*, *nuclear receptor* (*NR*), *Fox* (*forkhead*), T-*box*, *Mef2*, and *Ets* gene classes, are expressed during demosponge development, sometimes in a cell lineage restricted manner. These data combined with developmental gene expression patterns from other animals suggest that these genes may have been playing important regulatory roles in the embryos of the first metazoans.

MATERIALS AND METHODS

Sponge collection and RNA isolation

The demosponge *Reniera* sp. was collected from Heron Island Reef, Great Barrier Reef, and brood chambers and larvae were procured following procedures described in Leys and Degnan (2001, 2002). Total and poly(A) RNAs were isolated from embryos, larvae, and postlarvae as previously described (Hinman and Degnan 2000).

Isolation and identification of sponge transcription factor cDNAs

Transcription factor genes were amplified from cDNA synthesized from *Reniera* embryonic or larval total RNA as previously described (Hinman and Degnan 2000). Degenerate primers were designed to amplify sequences encoding the conserved DNA-binding domains of the following transcription factor gene classes: ANTP, *Pax*, *POU*, *LIM-HD*, *Sox*, *NR*, *Fox*, *T-box*, *Mef2*, and *Ets* (see Degnan et al. 1993, 1996; Degnan and Morse 1993; Finnerty and Martindale 1997; Martinez et al. 1997; Technau and Bode 1999; O'Brien and Degnan 2000; Devine et al. 2002; Spring et al. 2002). Sox forward (PanSox3A) and reverse (PanSox2B) primers were AAYGCNTTYATGRTNTGG and YGRTAYTTRTARTYNGGRT, respectively. The sequence of forward primer PanEn1 was AARMGNCCNMGNACNGCNTT. To search for *Hox* and *ParaHox* genes in *Reniera*, we also designed other Hox/ParaHox degenerate primers and digested PCR products amplified with Antp primers with the restriction enzyme Dde I to remove *Prox2*. All primer sequences are available upon request from the corresponding author.

RT-PCR products were size fractionated by gel electrophoresis and correct-sized products were cloned and sequenced as previously described (O'Brien and Degnan 2000). Nucleotide sequences belonging to the targeted class were identified using the BLASTx algorithm to search against the National Center for Biotechnology Information (NCBI) non-redundant protein database. 5′ and 3′ cDNA sequences were obtained by RACE PCR using the BD SMART RACE cDNA Amplification Kit (Clontech, Mountain View, CA, USA) according to the manufacturer's instructions. Putative RACE products were cloned, sequenced, and consensus sequences were created in Sequencher™ 3.1.1 (Gene Codes Corporation, Ann Arbor, MI, USA).

Phylogenetic analyses

Amino acid translations of *Reniera* transcription factor genes were aligned using Clustal X 1.64b (Thompson et al. 1994), to a selection of protein sequences that closely matched during the BLAST analysis and represented all the known families of the assigned transcription factor class. Alignments were edited visually in the Sequence Alignment Program Se-Al v1.d1 (Rambaut 1996) and regions of uncertain alignment were removed. Distance, Bayesian, and, in most cases (all but ANTP, *Fox*, and *T-box*), maximum likelihood (ML) analyses were performed on the aligned sequences. After a preliminary analysis, choice of genes was refined to represent more genes of the clade to which the sponge gene(s) convincingly belonged. With the exception of the ANTP and *NR* classes, all known families were represented in each analysis. After preliminary analyses, the *Reniera NR* genes were shown to belong to supergroup 2 (Auwerx et al. 1999) (data not shown) and, consequently, the final analyses focused on this clade, removing some genes belonging to other groups. Similarly, *Reniera* ANTP class genes did not belong to the *Hox*, *ParaHox*, and extended-*Hox* clades (Gauchat et al. 2000; Banerjee-Basu and Baxevanis 2001; Castro and Holland 2003) and the final analysis mostly excluded genes belonging to this clade.

Distance and ML were performed using the Unix PHYLIP v3.6a3 package (Felsenstein 2003). Distance neighbor joining (NJ) analyses with 1000 bootstraps were performed using Seqboot, Protdist, Neighbor, and Consense with default settings (using the Jones–Taylor–Thornton—JTT—matrix). ML analyses with 100 bootstraps were performed using Seqboot, Proml, and Consense with the likelihood model set to inv+gamma (invariant sites+gamma distribution) among-site rate variation. Estimates of gamma and the proportion of invariable sites were obtained using Codeml (set to the JTT amino acid substitution matrix) in the PAML package (Yang 1997) with an alignment and a preliminary Proml ML tree as input files. As there is a low effect of tree topology on parameter estimates (Sullivan et al. 2005), we considered these PAML estimates to be good approximations.

Bayesian analyses were performed using MrBayes v3 (Ronquist and Huelsenbeck 2003) with the amino acid substitution model prior set to the JTT fixed model and the likelihood model to the among-site rate variation invgamma (invariant sites+gamma distribution). A set of four independent simultaneous Metropolis-coupled Markov chains Monte Carlo was sampled every 100th generation. Initially, for five data sets, four separate analyses were conducted; three were run for one million generations and one for 10 million generations. Trees and parameters were compared between all four analyses. As there were no major differences, subsequent analyses (ANTP, *Fox*, *T-box*, and *Mef2*) were run twice, one for a million generations and another for 10 million generations. Runs were monitored for convergence and an adequate burn-in was removed. Bayesian posterior probabilities (PP) were used to assess the confidence value of each node.

Gene-specific cycle-restricted RT-PCR

RT-PCR was conducted with gene-specific oligonucleotide primers as described in O'Brien and Degnan (2000). RT-PCR was finally performed on 400 ng of total RNA from the different developmental stages using the optimized cycle conditions. RT-PCR for each gene was repeated three times on sets of RNA extracted from different cohorts of embryos, larvae, and postlarvae. Primer sequences are available upon request from the corresponding author.

In situ hybridization

Embryos and larvae were fixed in 4% paraformaldehyde, 0.05% glutaraldehyde, and 1 × MOPS buffer for 20 min on a nutator. They were stepped into and rinsed in 70% ethanol and stored at −20°C. Whole mount in situ hybridization (WMISH) was performed as described by Hinman and Degnan (2000) with minor modifications. Digoxigenin (DIG)-labeled antisense RNA probes were transcribed from liberalized pBSK+plasmids (Stratagene, La Jolla, CA, USA) following manufacturer instructions (Roche Applied Science, Indianapolis, IN, USA). Probes were transcribed

from cloned *Reniera ferritin*, *procollagen lysyl hydroxylase*, *galectin*, and *B-ZIP1* cDNAs, which were identified in an unpublished developmental EST set; probe lengths were 510, 215, 448, and 501 bp, respectively. The *RenHNF4* and *RenBsh* probes of 816 and 900 bp in length, respectively, which corresponded to the 5′ ends of the cDNAs, were synthesized and used in separate in situ hybridizations.

RESULTS

Normal development of *Reniera*

The haplosclerid demosponge *Reniera* broods its embryos throughout the year. From light and electron microscopic analysis of asynchronously developing embryos and larvae in the brood chamber, we have shown that this demosponge develops in a manner similar to that observed in disparate eumetazoans (for details see Leys and Degnan 2001, 2002; Leys 2003a, 2004; Degnan et al. 2005; Leys et al. 2005, Fig. 1). Cleavage is initiated after fertilization and establishes populations of micromeres and macromeres, which are initially mixed with each other. Differential cell movement—largely by multipolar ingression—produces an embryo consisting of inner macromeres and outer micromeres. This process is interpreted as gastrulation (Leys and Degnan 2002; Leys 2004; Leys et al. 2005). At this stage a number of differentiating cell types are evident in the micromere population, specifically a large group of uniciliated cells, and smaller groups of pigmented cells and spicule-producing sclerocytes. Once on the surface, pigment cells and some sclerocytes begin migrating to the future posterior pole of the larva in a predictable manner (Fig. 1, A and B). Through another series of cell sorting and morphogenetic events, the pigment cells form an external ring that surrounds the posterior pole as sclerocytes ingress into the inner cell mass (Fig. 1, C and D). At this stage the middle cell layer is formed; the origin of this important feature is still unclear. The cell patterning and differentiation events that occur during embryogenesis result in at least 11 morphologically discernable cell types whose spatial allocation confers larval functionality, namely phototaxis (Maldonado et al. 1997; Leys and Degnan 2001; Leys et al. 2002; Maldonado et al. 2003) and chemosensing (Jackson et al. 2002). In the obvious case of the pigment ring, cell patterning establishes a simple tissue-like structure whose organization confers a new capability beyond the ability of the individual cells that the tissue comprises (Leys and Degnan 2001, 2002; Degnan et al. 2005, Fig. 1). At metamorphosis the larva undergoes dramatic changes that include the loss of overt body axes, the migration and transdifferentiation of specific cell types and the formation of choanocyte chambers (Leys and Degnan 2002).

Identification of *Reniera* transcription factor genes

Using a variety of degenerate primer sets designed to amplify a range of transcription factor gene classes and families, we have identified 16 metazoan transcription factor genes belonging to 10 gene classes: ANTP (3 genes); *Pax*; *POU*; *LIM-HD*; *Sox* (3 genes); *NR*; *Fox* (2 genes); *T-box* (2 genes); *Mef2*; and *Ets*. In all cases, at least 20 clones were sequenced from a particular RT-PCR product and the expression of individual *Reniera* genes was confirmed by successful RT-PCR amplification of an independent set of developmental RNAs using gene-specific primers. 5′ and 3′ RACE products were obtained for these genes and contiguous cDNA sequences were constructed by aligning overlapping regions in the degenerate RT-PCR and RACE products. All *Reniera* genes have been given the prefix "*Ren*." We have annotated these genes based on alignments and phylogenetic analyses of sequences.

Putative orthology of *Reniera* genes was determined by NJ, ML, and Bayesian analyses (only Bayesian trees are shown), except in the cases of ANTP, *Fox*, and *T-box* genes. For each gene, tree topologies were generally congruent across the three independent analyses and consistently (1) generated the same monophyletic gene family clades, (2) grouped the *Reniera* gene with the sponge ortholog (when present), and (3) placed the sponge genes in a particular gene family. With the exception of *RenSoxC* and the *Sox C* family, these clades were strongly supported, with PP equal to or greater than 0.95 (or an NJ bootstrap greater than 90 in the case of *RenPouI* inside the *POU I* clade); critical Bayesian PP values (given for the 10 million generations run) varied by 1% at the most.

Putative orthology of *Reniera* ANTP, *Fox*, and *T-box* genes was determined only by NJ and Bayesian phylogenetic analyses. ML analyses, because of their computational need for a reduced number of taxa, produced trees with low resolution and statistical support for these classes. As above, tree topologies generally were consistent across both methods.

ANTP genes

Sequencing of clones derived from RT-PCRs using a variety of degenerate primers sets (Degnan et al. 1996) designed to amplify ANTP class homeoboxes yielded three consensus sequences. These sequences are most similar to *BarBsh* from the demosponge *Halichondria bowerbanki*, *HOXa1* from the demosponge *Suberites domuncula*, and *Prox1* from the demosponge *Ephydatia fluviatilis*, and have been named *RenBsh*, *RenProx2*, and *RenNK2*, respectively (Fig. 2). No *Hox*, *ParaHox*, extended-*Hox*, or *EHGbox* genes (Banerjee-Basu and Baxevanis 2001; Castro and Holland 2003; Garcia-Fernandez 2005) were obtained from *Reniera* using a combination of different degenerate primers and, thus far, none have been conclusively recovered from any other sponges. RACE produced 1471, 1368, and 1397 bp cDNA sequences for *RenBsh*, *RenProx2*, and *RenNK2*, respectively (Table 1). In addition to the homeodomain (HD), two amino acid positions were conserved just upstream of the HDs (Fig. 2B).

A distance tree depicting putative orthology of ANTP *Reniera* genes includes all recognized families outside the *Hox* and *ParaHox* subclasses of ANTP homeobox genes. *RenBsh* and *BarBsh* from *H. bowerbanki* are strongly supported sister taxa (NJ: 98, PP: 0.99) with an unresolved position within the *Bsh*+*BarH* clade; the NJ tree places them at the base of the *Bsh* family with low statistical support (NJ < 50) (Fig. 2C). The NCBI BLAST supports the assignment of *RenBsh* within the *Bsh* (vs. *BarH*) family (Fig. 2B). *RenProx2* and other demosponge genes, such as *Prox2* and *HOXa1*, form a well-supported monophyletic clade (NJ: 65, PP: 0.97; Fig. 2C) of unresolved position within the *Tlx*+*Lbx* clade; the NJ analysis places this group at the base of the *Tlx* family with low statistical support (NJ < 50; Fig. 2C). BLAST supports the assignment of these sponge genes inside the *Tlx* clade (Fig. 2B). *RenNK2* and *Prox1* are clear orthologs (NJ: 92, PP: 0.97) whereas *NKxD* from the calcareous sponge *Sycon raphanus* does not form a monophyletic clade with these demosponge genes with either method (Fig. 2C). Both analyses place all three genes within the *NK2* clade, albeit with low statistical support (NJ < 50, PP: 0.87; Fig. 2C), and the BLAST alignment warrants this grouping (Fig. 2B). In contrast, *Prox3* from the demosponge *E. fluviatilis* is confidently placed within the *Msh* family (NJ: 87, PP: 0.96) and most recognized families are well supported.

A

```
RenBsh      MKTGLLLHYNQPSASPSSKMMLATYPSPAHQQQQATSPPAFSAISQRTHPAAGGLLSTATGFIPTALPLAIAPSAVFGHHHHHHQPVALRGPVPAPQILQ  100
            HHFASVKVEEAKRRSESPRSRSPSSTPPAPQASKPQHTAFSIDAILSKKDEPKKHISSSSTSSEVSLAPHQTSLLASPPSVGHTPHLTRTTANGLFYIYT  200
            AASPTATGGHPTVVQAAQNPSPFSPFASLRGHHFDHHHHHHHELQRSPLGPLVLPPPEPSVLLSSKLKR**KRKLRTVFTEKQLEGLETKFSEKKYLSVPDR**  300
            **MELANRLELSETQVKTWFQNRRMKCKKQQ**QVEQGSSGEGCSEKDERYQKHFSAYIKEGLGPDDIEPMYKKAHEMIRADPVMQVKERKKMETQKRWTKQKL  400
            TLAQRKSKIKQKKASFLRQFKTDDDESED  429

RenProx2    MMESCCTALNQIQPQLQNSFRIDSILTGMGAHPPPPPQGYSVMHQHPSNAGGPGGVEAIKQFGNMFGKNEELNMQERKRESVSSADDDDDSYRK**KKKPRT**  100
            **AFSREQVSELEKKFTERKYLSSAERGELAEKLKLSDMQVKTWFQNRRMKFKRQN**EEAELEVKSPKFPYPPFVPYSSFYSYTMPHYKMMDQMQQQQQQQQQ  200
            SPSYTCTSPPSTSNSYQQSTPGPTPLVLGTNPAAHHISPPPHITSTYFNRAAIAHGVISNGSQSPCYHIGAGYLSDYNPGAPSGPIMSPPYCNSSGSSDW  300
            QRSPLPTPPAAP  312

RenNK2      MSSGKKKSFMICDILDDIYTSDSSRPSSTKLQVHPHIVHSSSVIYTASAAPSHIEGGLVGGEGDECSPGTDSGLGDCKPEAIRPESPTQHEPDSPSSSQS  100
            SEQGSTGAAKTR**KRRPRGLFSHAQIYELERRYALQKYLTAHEREQLANMLRLTETQVKIWFQNRRYKNKRQQ**LENARLSPKSAVAACSKTSDLFPSAIPP  200
            PPPLHSIAAAPSDIKLPSTTPAFPVNSPLSLNLTHAPPSALPIGVVATSLITPSGTQSPLYSITGPPPSLPPPSDYSFRYPSGHPPPIMNIKSTPPSLPK  300
            SMYYPAVAAAYGGSVTSVNATIPSGSYVSSICSCTTMPYQPLPRVPSPATSVTSIRSQY  359
```

B

```
                    10        20        30        40        50        60
                    .         .         .         .         .         .     %I   %P   Family
RenBsh      KRKRKLRTVFTEKQLEGLETKFSEKKYLSVPDRMELANRLELSETQVKTWFQNRRMKCKKQQ
BarBsh Hb   R..............D...S..AD.....................................     92   100  BarH/Bsh
Bsh Mm      C.R..A....SDS..S...KR.EIQR...T.E.V...TA.S.............H...L   66   79   Bsh
Barx1 Mm    .KG.RS......L..M...KR.EKQ....T...ID..ES.G..QL.....Y......W..IV   63   77   BarH
Ci-Mnx      N.S.RP..A..SQ..LE..RQ.KLN....R.K.F.V.TT.Q.T.....I........W.RSK   56   69   HLXB9

                    10        20        30        40        50        60
                    .         .         .         .         .         .     %I   %P   Family
RenProx2    RKKKKPRTAFSREQVSELEKKFTERKYLSSAERGELAEKLKLSDMQVKTWFQNRRMKFKRQN
Hoxa1 Sd    ..R..A.........A......Q.K.....N.............................S   89   95   Tlx
Prox2 Ef    TRR..A.........A......QNK.....T.........................Y...S   82   92   Tlx
Hox11 Mm    P.R.....S...V.IC....R.HRQ...A....AA..KS..MT.A..........T.WR..T   65   82   Tlx
Nkx2-3 Mm   .SRR...VL..QA..F...RR.KQQR...AP..EH..SS...TST...I......Y.C...R   56   73   NK2

                    10        20        30        40        50        60
                    .         .         .         .         .         .     %I   %P   Family
RenNK2      TRKRRPRGLFSHAQIYELERRYALQKYLTAHEREQLANMLRLTETQVKIWFQNRRYKNKRQQ
Prox1 Ef    HK.....A......V......F.V........QSK..TV.H.............S....   79   91   NK2
Nkx2-5 Ok   RQR.K..V...Q..V......FKQ.R..S.P.........K..S............C...R   72   87   NK2
bap Dm      S..K.S.AA.....VF.....F.Q.R..SGP..SEM.KS.................T..K.   67   83   NK2
Hex Mm      HKRKGGQVR..ND.TV...KKFET....SPP..KR..K..Q.S.R...T......A.WR.LK   46   69   Hex
```

Fig. 2. ANTP genes expressed during *Reniera* development. (A) Derived amino acid sequences of *RenBsh*, *RenProx2*, and *RenNK2*. The homeodomains (HD) are in bold. (B) Amino acid alignment of HDs (and two amino acids upstream) of RenBsh, RenProx2, RenNK2, their closest BLAST matches and selected representatives of other families. Dots represent identities. At the end of each aligned sequence, the percentage of identical amino acids (%I), conservative substitutions (%P), and the protein's family are shown. (C) Unrooted distance phylogenetic tree of *Antp*-class genes. Selected genes are for the most part outside the *Hox* or *ParaHox* subclasses. At key nodes, percentages of bootstrap support obtained with 1000 replicates are given. Bootstrap values above 50% are shown. An asterisk indicates a Bayesian posterior probability (PP) greater than or equal to 95%. Families are indicated on the right of the tree. *Reniera* genes are underlined and other sponge genes are indicated with a hash symbol. Ce, *Caenorhabditis elegans*, nematode; Ci, *Ciona intestinalis*, urochordate; Cv, *Chlorohydra viridissima*, hydrozoan cnidarian; Dm, *Drosophila melanogaster*, insect; Dt, *Discocelis tigrina*, platyhelminth; Ef, *Ephydatia fluviatilis*, demosponge; Em, *E. muelleri*, demosponge; Euf, *Eunapius fragilis*, demosponge; Hb, *Halichondria bowerbanki*, demosponge; Hv, *Hydra vulgaris*, hydrozoan cnidarian; Hys, *Hydractinia symbiolongicarpus*, hydrozoan cnidarian; Mm, *Mus musculus*, mammal; Nv, *Nematostella vectensis*, anthozoan cnidarian; Ok, *Octopus kaurna*, cephalopod; Pn, *Polycelis nigra*, platyhelminth; Sd, *Suberites domuncula*, demosponge; Sl, *Spongilla lacustris*, demosponge; Sp, *Strongylocentrotus purpuratus*, echinoderm; Sr, *Sycon raphanus*, calcareous sponge; Ta, *Tethya aurantia*, demosponge; Th, *Trochospongilla horrida*, demosponge.

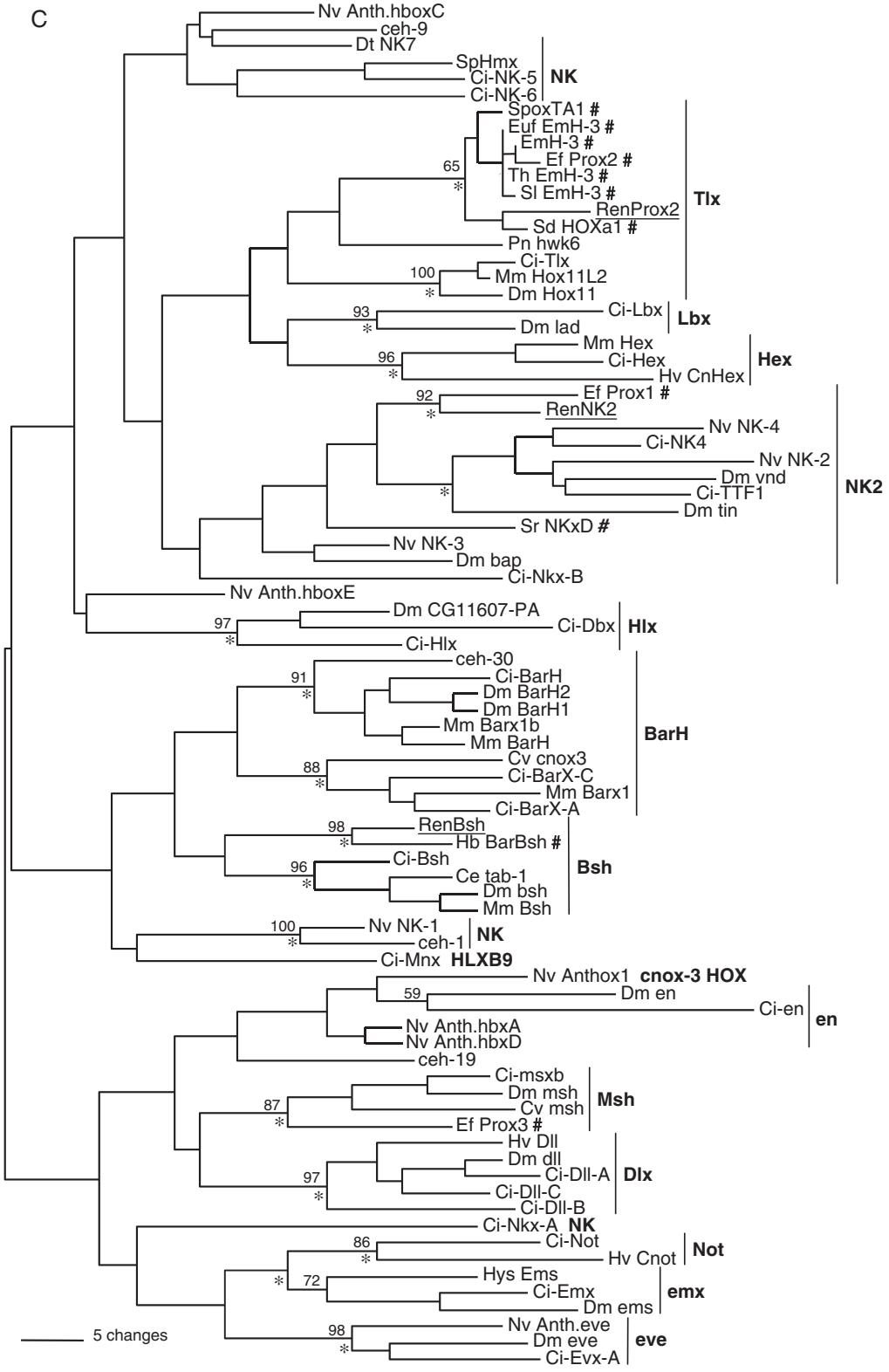

Fig. 2. Continued.

Table 1. Transcription factor genes expressed in *Reniera*

Gene name	cDNA length (bp)	5′ UTR (bp)	ORF (bp)	3′ UTR (bp)	Derived protein (aa)	Domains
RenBsh	1471	127	1287	57	429	Homeodomain
RenProx2	1368	150	936	282	312	Homeodomain
RenNK2	1397	282	1077	38	359	Homeodomain
RenPaxB	2117	174	1656	287	552	Paired and homeodomain
RenPouI	1360	45	972	343	324	POU and homeodomain
RenLim3	1634	252	1014	368	338	2 LIM and homeodomain
RenSoxB	1676	166	1311	199	437	Sox HMG domain
RenSoxC	937	166	771	—	257	Sox HMG domain
RenSoxF	782	36	672	74	224	Sox HMG domain
RenHNF4	2327	75	1908	344	636	NR DNA- & ligand-binding domains
RenFoxL1	1617	153	1464	—	488	Forkhead domain
RenFoxJ	144	—	150	—	50	Partial forkhead domain
RenTbxA	1650	102	1254	294	418	T-box domain
RenTbxB	1541	123	1350	68	450	T-box domain
RenMef2	1860	210	1629	21	543	MADS and Mef2 domains
RenEts	670	—	456	214	152	Partial ETS domain

NR, nuclear receptor.

Pax

A single *Pax* gene was obtained by RT-PCR. It is most similar to *sPax2/5/8* from the demosponge *E. fluviatilis* and has been named *RenPaxB*. RACE yielded a total transcript of 2117 bp and a predicted protein sequence of 552 amino acids (Table 1). A paired domain is upstream of a divergent HD (Fig. 3); no convincing hexapeptide (present in many *Pax* genes) can be detected between the two domains.

All three phylogenetic methods concur to group *RenPaxB* with *sPax-2/5/8* (NJ: <50, ML: 85, PP: 1) and the two sponge *Pax* genes are most closely related to the *Pax2/5/8* or *B* clade (NJ: <50, ML: 59, PP: 0.99) (Fig. 3C). In this analysis, *Pax* families and supergroups I and II (Sun et al. 1997) are also well supported. Pax3/7, Pax4/6, PaxC, and cnidarian PaxB proteins have full HDs, whereas Pax1/9, PaxA, and bilaterian PaxB proteins have truncated, degenerate, or no HDs (Miller et al. 2000; Fig. 3B). Both sponge PaxB proteins appear to have complete but slightly degenerate HDs. The *Reniera* PaxB seems more conserved at the C-terminal end of the HD than the *E. fluviatilis* PaxB (Fig. 3B).

POU

RT-PCR yielded a single *POU* gene with greatest similarity to the demosponge gene *spou1* from *E. fluviatilis* (Fig. 4B). It has been named *RenPouI*. With RACE-PCR, we obtained a full-length sequence of 1360 bp (Table 1). The resulting translated protein sequence of 324 amino acids contains a POU-specific domain upstream of a HD (Fig. 4). The linker region between the POU-specific domain and the HD, which is usually conserved within each family, as well as four amino acids just downstream of the HD, can be aligned with Pit1 proteins (Fig. 4B).

RenPouI and *spou1* form a monophyletic clade (NJ: 99, ML: 89, PP: 1), and these two sponge genes are most similar to vertebrate *Pit1* (*POU I*) genes (NJ: 91, ML: 58, PP: 0.84) (Fig. 4C). The other *POU* families are well supported in these analyses, but the position of the other sponge *POU* gene, *spou2*, is unresolved.

LIM-HD

A single consensus sequence was obtained by RT-PCR. It is most similar to *Lhx* from the demosponge *S. domuncula* (Fig. 5B) and has been named *RenLim3*. The full-length transcript obtained with RACE is 1634 bp (Table 1) and the predicted protein sequence of 338 amino acids contains the characteristic domains of LIM-HD proteins, two LIM domains upstream of a HD (Fig. 5). Four amino acids just upstream and two just downstream of the HD can also be aligned with other LIM-HD genes (Fig. 5B).

The two sponge LIM-HD genes, *RenLim3* and *Lhx* from *S. domuncula*, appear to be orthologs (NJ: 100, ML: 94, PP: 1) and are most closely related to the *Lim-3* family of *LIM-HD* genes (NJ: 100, ML: 88, PP: 1) (Fig. 5C). Deep nodes—family clades and relationships between families—are well supported. Previously defined families (Hobert and Westphal 2000) are supported and there appear to be two supergroups - [*Lim-3+lin-11+Lmx*] and [*Lhx6/7+apterous+islet*].

A

```
RenPaxB    MAKGXAGQGGVNQLGGVFVNGRPLPDIIRRKIVELAQNGVRPCDISRQLRVSHGCVSKILGRFYETGSVKPGIIGGSKPKVATPKVVNKIEDYKRENPSI  100
           FAWEIRDRLLTEGICNKVNVPSVSSINRIVRTRAQQRQKDLQEKAALGHFSILQTLPPPPHESYNGLIPSPYSHHTLSHFAGFPPRPGSGAMPPGCPGPW  200
           GAHPGLQEHPHQPPQHHHLDSGSSNGKLLTVFPAAATTTQQSATPSMPYFVYPTIPPTTGGGDSVSCHQEINSGYSQIQCTPSGASDIPNNPSPNMPSSN  300
           IDKAAIDHDMAGQSSYIRGAGANSPTTPSKSPVQGHSGVPNISGKDSQEQLIVKTEGTRKLTPFQIHQLELSYNTSHYPARPTVKQLASLLDLPESTIES  400
           WFAEHRQMSNNASSGVNVSSSWSHPPYSEPIQQPVDGLTHINYSLQTAAGFPHSSSRGSPVCVITPNVTIQTFHPMAAASTAASTGGHPHHVNQTNYPTP  500
           PSAAPPPPSMRPGTTNSPNQEPFGASLTPPPSNSSPPVATTAWNQTFPTYTC  552
```

B

```
                                                    Paired domain
                    10        20        30        40        50        60        70        80        90        100
                    .         .         .         .         .         .         .         .         .         .
RenPaxB             GQGGVNQLGGVFVNGRPLPDIIRRKIVELAQNGVRPCDISRQLRVSHGCVSKILGRFYETGSVKPGIIGGSKPKVATPKVVNKIEDYKRENPSIFAWEIRDRLLT
sPax2/5/8 Ef ..........L........ES........S.........................Y.....I...V..........S...L......Q..............Q
PaxB Pl      SH..................VV.QR..D..HS.........................Y.....IR..V.............S..A....Q..TM..........A
PaxB Am      .H.........L.........VV.SR..D...S.........................C.......I...V..........GP.....AE...N..TM..........S
Pax-2a Bf    .H.........Y........VV.HR.....HQ.......................R.Y.....I...V.............E..AE...Q..TM..........A
PaxC Am      SH..I.....P.........Y..HR..Q..AC.....E...R.L.....................IR..S............P.....VQ..QQ..T...........VE
Ci-Pax6      .HS.M.....M.........S..Q....F.H..A.......I.Q..N.......A.Y....TIR.RA......R....E.....AS....C.............N
amphipax1 Bl TF.E..............NA..LR......L.I.....................A.YN....IL..A......R.T..E..KA.KK..TLD.G...........A
Pax-37 Hr    ...R........I......NH..H....M.AH.I...V.................C.YQ....I...A........TNSEIES...Q..KDS..M.S.....Q.IK
```

```
                                                    Homeodomain
                    110       120                       10        20        30        40        50        60
                    .         .                         .         .         .         .         .         .
RenPaxB             EGICNKVNVPSVSSINRIVRTRA           TEGTRKLTPFQIHQLELSYNTSHYP-ARPTVKQLASLLDLPESTIESWFAEHRQMSNNAS
sPax2/5/8 Ef D.V.D.....................               GDN.PT.LDA.V.E..R.LGDCA..-DTA..QD..CR.G.T.GQ.Q..LKAR
PaxB Pl      ..V.E.D.............NK.                   KTNSGAFS.H..EAM.KAC.P.L..-DVYRKDIAQDGRGQVGMP.S.ASEY.HAHLEAQR
PaxB Am      ..V.STD.............N.I                   RRQRTTFSGE..E...KTFEKT...-DVF.REK..QDV..S.AR.QV..SNR.AKWRKQE
Pax-2a Bf    ....DNDT............NK.                   EQKRSTF..D.LEA..QAF.RG...TDPFNRDNMSNKV..SQTRVQD
PaxC Am      ..V.DRE.T.........L.NK.                   RRNRTTFS.E.LEM..KEFEK....-DVA.REE...KI.MS.ARVQV..SNR.AKWRRHQ
Ci-Pax6      .....ND.I........VL.NLD                   QRNRTSF.QI.VEA..KEFERT...-DVFARER..TKI....AR.QV..SNR.AKWRREE
amphipax1 Bl ..V.D.Y........S..L.NKI                   AQMGEGY.LNH.N.MQGQTDKVK..-DV
Pax-37 Hr    ..L.DRSSA.T..A.S..L.SKD                   RRSRTTFSAD.LEE..RCFERT...-DIY.REE..QRTR.T.ARVQV..SNR.ARWRKQM
```

```
                                                    %I   %P   Family
sPax2/5/8 Ef                                        89   97   2/5/8 or B
PaxB Pl                                             81   94   2/5/8 or B
PaxB Am                                             80   91   2/5/8 or B
Pax-2a Bf                                           80   91   2/5/8 or B
PaxC Am                                             75   88   pox neuro
Ci-Pax6                                             74   85   4/6
amphipax1 Bl                                        73   84   1/9
Pax-37 Hr                                           66   87   3/7
```

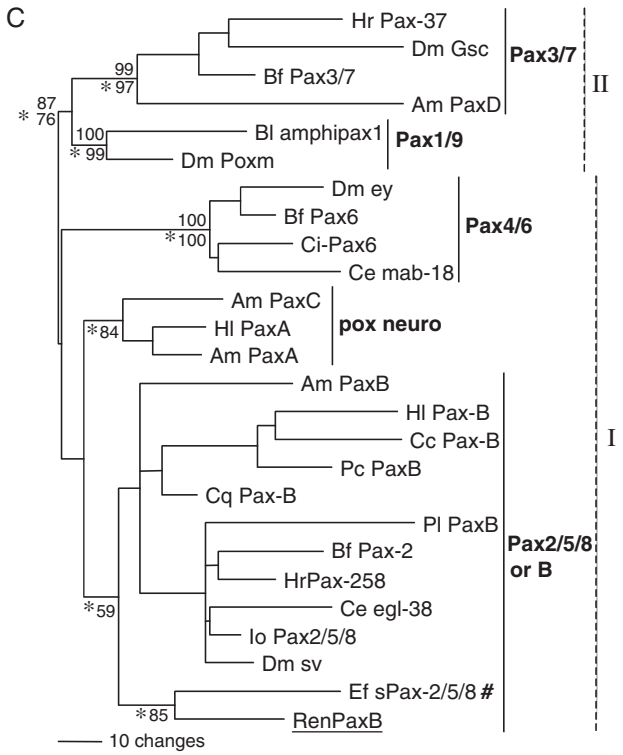

Fig. 3. *PaxB* gene expressed during *Reniera* development. (A) Derived amino acid sequence of *RenPaxB*. The paired domain is in bold letters and the homeodomain (HD) is underlined. (B) Amino acid alignment of conserved regions of RenPaxB, its closest BLAST matches, and selected representatives of other families. The percentages refer to the paired domain, as the HD is truncated and/or derived in many Pax proteins (it is absent in some). Unconserved sequence between the paired domain and the HD is not shown. (C) Unrooted Bayesian phylogenetic tree of the *Pax* class. At key nodes are given percentages of bootstrap support, obtained by distance (1000 replicates) above and maximum likelihood (100 replicates) below the branch. An asterisk indicates a Bayesian posterior probabilities greater than or equal to 95%. See Fig. 2 for further explanation of alignment and tree. In addition to previously described abbreviations: Am, *Acropora millepora*, anthozoan cnidarian; Bf, *Branchiostoma floridae*, cephalochordate; Bl, *Branchiostoma lanceolatum*, cephalochordate; Cc, *Cladonema californicum*, hydrozoan cnidarian; Cq, *Chrysaora quinquecirrha*, scyphozoan cnidarian; Hl, *Hydra littoralis*, hydrozoan cnidarian; Hr, *Halocynthia roretzi*, urochordate; Io, *Ilyanassa obsoleta*, mollusc; Pc, *Podocoryne carnea*, hydrozoan cnidarian; Pl, *Paracentrotus lividus*, echinoderm.

Sox

Three RT-PCR products were produced with degenerate primers. These yielded partial coding sequences of HMG domains that were most similar to sea urchin *soxB2* (one gene) and mouse *sox4* (two genes). These genes have been named *RenSoxB*, *RenSoxC*, and *RenSoxF* (Fig. 6). RACE on *RenSoxB*, *RenSoxC*, and *RenSoxF* yielded 1676, 937, and 782 bp sequences, respectively (Table 1).

A

```
RenPOUI   MDFEQYSNSAGQQQQQQGANGASNGTSQASEQGREGSPLRSPTVAMSMKSPRSQPASGPNDQQHKDLTAQVAAALQYSQYAAANPSQIPVHVIAAHQNLQ  100
          QAAGRIPVAVTTSFQMANGSPTLTPTLVHPNLTSLSSPQSFQPSGVIQAPGAAHPKNIPITAQVDPESPEVKNLEAFAVSFKSRRVKLGYTQTNVGQALA  200
          SVHGTNFSQTTICRFENLQLSYKNAQKLKPILEKWLEEAEKQGAIQHEEESMERHRKRRTTIGMSAKERLEQHFQVQPKPSSSDITKVADSLNLDKEVIR  300
          VWFCNRRQREKRVRASLGGSPEKE  324
```

B

```
                              POU - specific domain
                 10        20        30        40        50        60        70        80        90        100
                  .         .         .         .         .         .         .         .         .         .
RenPouI     AQVDPESPEVKNLEAFAVSFKSRRVKLGYTQTNVGQALASVHGTNFSQTTICRFENLQLSYKNAQKLKPILEKWLEEAEKQGAIQHEEESM---ERHRKRRTTIG
Spou1 Ef    Q.P..D.D...D..S..A......M.............RE....D.....................................PP....AS---PER.....S..
Pit1 Mm     EPI.MD...IRE..Q..NE..V..I.........E...A...SE...............F...C...A..S.......QV..LYN.KVGAN--..K.......S
Pou3f1 Mm   PHS.EDT.TSDD..Q..KQ..Q..I...F..AD..L..GTLY.NV..........A....F..MC....L.N......DSSSGSPTSIDKIAAQG.K..K..S.E

                 Homeodomain
                110       120       130       140       150
                  .         .         .         .         .         %I   %P   Family
RenPouI     MSAKERLEQHFQVQPKPSSSDITKVADSLNLDKEVIRVWFCNRRQREKRVRASL
Spou1 Ef    VG.Q.T..R..IN.........N.I....G.E...V..............PLRS   78   86   I
Pit1 Mm     VA..DA..R..GEHS....QE.MRM.EE...E...V..............KT..   67   83   I
Pou3f1 Mm   V.V.GA..S..LKC....AQE..SL....Q.E...V.........K...MTPPG   55   70   III
```

C

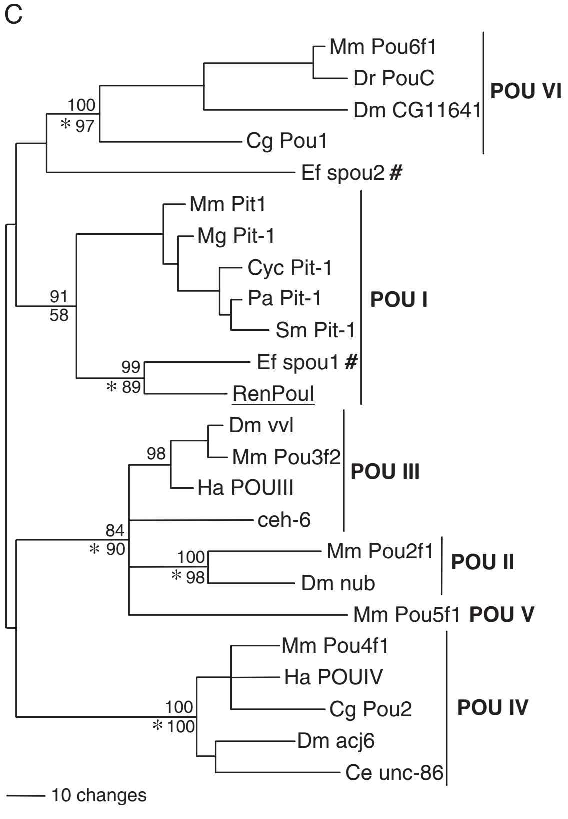

Fig. 4. *POUI* gene expressed during *Reniera* development. (A) Derived amino acid sequence of *RenPouI*. The POU-specific domain is in bold letters and the homeodomain is underlined. (B) Amino acid alignment of the conserved region of RenPouI, its closest BLAST matches, and a representative from the closest other family. (C) Unrooted Bayesian phylogenetic tree of the *POU* class. At key nodes are given percentages of bootstrap support, obtained by distance (1000 replicates) above and maximum likelihood (100 replicates) below the branch. An asterisk indicates a Bayesian posterior probabilities greater than or equal to 95%. See Fig. 2 for further explanation of alignment and tree. In addition to previously described abbreviations: Cg, *Condylactis gigantea*, anthozoan cnidarian; Cyc, *Cyprinus carpio*, fish; Dr, *Danio rerio*, fish; Ha, *Haliotis asinina*, mollusc; Mg, *Meleagris gallopavo*, bird; Pa, *Plecoglossus altivelis*, fish; Sm, *Scophthalmus maximus*, fish.

RenSoxB and *RenSoxF* clearly belong to the existing families *Sox B* (NJ: 98, ML: 81, PP: 1) and *Sox F* (NJ < 50, ML: 51, PP: 0.98), respectively (Fig. 6C). *RenSoxC* is placed within the *Sox C* family by all three methods, but with low statistical support. All existing families (Bowles et al. 2000), except *Sox C*, are well supported.

NR

A single gene, most similar to *RXR* from the demosponge *S. domuncula* (Fig. 7B) and named *RenHNF4* (based on preliminary phylogenetic analyses), was identified by RT-PCR. RACE produced a full-length sequence of 2327 bp encoding a 636 amino acid protein (Table 1). The translated protein sequence includes two conserved domains, a ligand binding domain downstream of a DNA binding domain (Fig. 7).

An ML analysis was undertaken with a lower number of taxa than in the NJ and Bayesian analyses, as reducing the number of taxa in the latter two yielded trees with lower statistical support; this reduction was necessary to do ML. The ML bootstrap values are indicated for the corresponding

A

```
RenLim3   MSYQDINTHQTTSSPVVDPFQVPPSLPSIQTPQPSGMAMLVQPQSAAPSSLSLAQTPTQEYHQNLSSPSNQPQQIQTSAPQAQDQNIAHVPFCAGCNTRI  100
          FDRFILRVQDKSWHAKCLRCSDCQCQLSDKCYSRSGQVYCKDDFSKRFGTRCAGCQQPIPPTQVVRRAQENVYHLQCFACFICQRQLSTGDEFYLMDDRK  200
          LVCKADYEAAKARADGSQKRPRTTISQKQLDLLKTAYCVSPKPSRHVRQELSDKTGLDMRVVQVWFQNKRAKDKRTKKDDGSDEGAGEGEGRGGDGDFSV  300
          STPTSATDDIQGQFILDPDIAQQGIDAPPPNPIGTTSA  338
```

B

```
                                        Lim domain 1                                         Lim domain 2
                    10         20         30         40         50         60         70         80         90        100
                     .          .          .          .          .          .          .          .          .          .
RenLim3     VPFCAGCNTRIFDRFILRVQDKSWHAKCLRCSDCQCQLSDKCYSRSGQVYCKDDFSKRFGTRCAGCQQPIPPTQVVRRAQENVYHLQCFACFICQRQLSTGDEFY
Lhx Sd      I.C....HHP.V.....K.L..P..S.....V..DML.T......D.E.F..A...R.........N..........................S..........
Lhx3 Mm     I.M....DQH.L.....KAL.RH..S...K....HVP.AER.F..RES.......F.....K..A..LG...........DF....H....VV.K...A......
Lhx5 Mm     MVH....ERP.L...L.N.L.RA..I..VQ.CE.KTN..E..F..E.KL...N..FR....K....A.G.S.SDL..K.RSK.F..N..T.MV.NK.....E.L.

                                                         Homeodomain
                   110        120                10         20         30         40         50         60
                     .          .                 .          .          .          .          .          .        %I   %P   Family
RenLim3     LMDDRKLVCKADYEAAKA  ADGSQKRPRTTISQKQLDLLKTAYCVSPKPSRHVRQELSDKTGLDMRVVQVWFQNKRAKDKRTKKD
Lhx Sd      ....K..........R.   -..N........TS....T..A..T..S....A..EQ.ANE...EV.........R.......T.G   78   88   Lim-3
Lhx3 Mm     ..E.SR........T..Q  .EATA.......TA...ET..S..NT....A....EQ..SE..............R...E..L...   69   84   Lim-3
Lhx5 Mm     VI.EN.F...D..LSSSS  SGTKRRG.....KA...ET..A.FAAT...T..I.EQ.AQE...N...I......R.S.ER.M.QL   52   77   lin-11
```

C

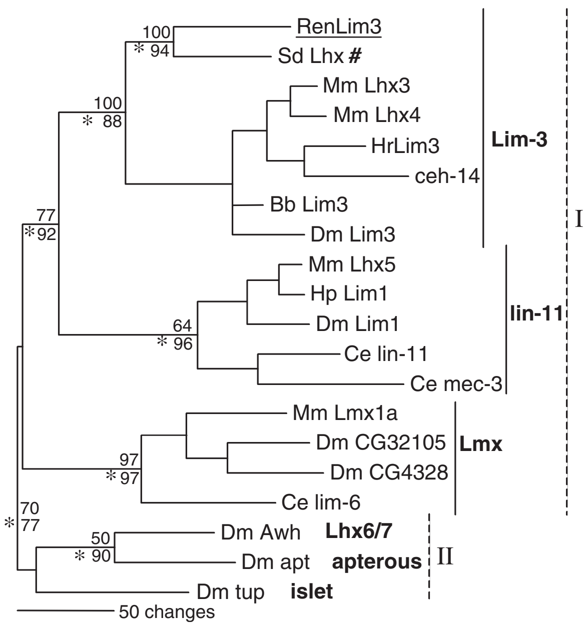

Fig. 5. *Lim3* gene expressed during *Reniera* development. (A) Derived amino acid sequence of *RenLim3*. The two LIM domains are underlined, Lim domain 1 fully and Lim domain 2 with a dotted line. The homeodomain (HD) is in bold letters. (B) Alignment of conserved regions of RenLim3, its closest BLAST matches, and a representative from the closest different family. Short regions upstream and downstream of the HDs are included. Unconserved sequence between Lim domain 2 and the HD is not shown. (C) Unrooted Bayesian phylogenetic tree of the *LIM-HD* class. At key nodes are given percentages of bootstrap support, obtained by distance (1000 replicates) above and ML (100 replicates) below the branch. An asterisk indicates a Bayesian posterior probabilities greater than or equal to 95%. See Fig. 2 for further explanation of alignment and tree. In addition to previously described abbreviations: Bb, *Branchiostoma belcheri*, cephalochordate; Hp, *Hemicentrotus pulcherrimus*, echinoderm.

major clades on the Bayesian tree (Fig. 7C). *RenHNF4* and *RXR* from *S. domuncula* are sister taxa (NJ: 100, ML: 100, PP: 1) and the close relationship of these two sponge *NR* genes to the *HNF4* clade is supported by all three methods (NJ < 50, ML < 50, PP: 0.99; Fig. 7C). Furthermore, supergroup II and each of its families (Auwerx et al. 1999) are clearly monophyletic groups. Relationships between supergroup II families are well supported, with the exception of the putative sister relationship between *HNF4* (*2A*) and *RXR* (*2B*), proposed by the three methods albeit with low statistical support.

Fox

Two sequences with significant similarity to *Fox* transcription factor genes were identified by sequencing cloned RT-PCR products (Fig. 8). The full-length coding sequence for one of these genes was obtained by RACE, resulting in a 1617 bp cDNA (Table 1). The forkhead domain of this ORF is most similar to those of vertebrate genes of the *Fox L1* class and this gene therefore has been named *RenFoxL1* (Fig. 8B). Partial sequence (150 bp, corresponding to 50 amino acids of the forkhead domain) obtained for the second gene is most similar to mouse *Foxj1* (Fig. 8B); this gene has been named *RenFoxJ*.

Both Bayesian and NJ analyses group *RenFoxL1* with another sponge *Fox* gene, *Sd-FoxF* from *S. domuncula*, with significant support (NJ: 77, PP: 1; Fig. 8C). However, whereas the NJ analysis places them at the base of the *Fox L1* family (not shown), the Bayesian tree places them at the base of a

A

```
RenSoxB   MLERESTMHIMSLPNMPGYPPHYAMPGPTGSTSIVPPQASELEHSVSPPLVGGQTGNGGPHGSNPSDMEED**DKVKRPMNAFMVWSRKMRKKIADENPKMH**  100
          **NSEISKRLGTQWKALSEEDKRPFIDEAKRLREAHMKKHPNYKYKPKRKKQ**TPTNTRIIPGIGSWPPYHQSRHIVHAGHIPSGGRWQYPQESGYYYTPGNG  200
          QHSYYTGYSGGYTRSPTVPSPSYHSWNNMTAPYPNTGQQHSPLPTDYATGSSVPVTMGPGACAQQSTTGPLQQSCIVNSYNDPLNTYSAAMSVRSNSSLA  300
          SVLTSPTSLVFSGMDSALGSPPKASSPVESLDSYSEVILSNCKVSDDCTSIHSNDSGAESDLRNMISTYLEESNSXPGPTETPPPSGSSRPPTAEFKLLN  400
          ASAQCTDFIASNSSNXTNSAESLLDGAGGTLPLQHLM  437

RenSoxC   MSAGDLQYTVILQQAPSPSCSSLSSPSSPYQSNLLDVCDVGTATHVHHNGGSRGGGNGGSAGKK**EHIKRPMNAFMVWAQLERRKMTTEFPDMHNAEISRR**  100
          **LGKLWRLLSDREKQPYIEESERLRIQHMKQYPDYKYRPRKKGG**KKPKPVSNTSYLGGNDSGSEEYYPTTMPTSSSNSCSCGAGIRRAPVPTCSIAVQCSM  200
          ELGEHVIEREPSSPKQTAEISIQVGNGSAHLQQNRRRSFSSFAGDKRPRDLSLSCPP  257

RenSoxF   MDQPVKEEDKIEKDISRDREGSEVGEEENETRNGFQEEEEVVDKQEEKESKGKKGKHGG**GRIKRPMNAFMVWSSLERKKLAEKEPNLHNTELSKRLGQMW**  100
          **KEMTEEDKTPYRQEATRLKDKLMEDHPEYKYKPRRRKD**LRHIQTSTGNIGLFSSHVIADRSYSAPPVPPILPSSNSLYHYNQDMIDGVGSGQTHYPYYHS  200
          TLTPHNKTANGRYQEKRITSSSLT  224
```

B

```
                  10        20        30        40        50        60        70
                  .         .         .         .         .         .         .          %I   %P   Family
RenSoxB           DKVKRPMNAFMVWSRKMRKKIADENPKMHNSEISKRLGTQWKALSEEDKRPFIDEAKRLREAHMKKHPNYKYKPKRKKQ
Sox2 Mm           .R.............GQ.R.M.Q..............AE..L...TE............AL...E..D...R.R..TK   76   89   B1
Sox21 Mm          .H.............GQ.R.M.Q..............AE..L.T.SE............AM...E..D...R.R..PK   75   87   B2
Sox17 Mm          SRIR.........AKDE..RL.QQ..DL..A.L..M..KS....TLAE....VE..E...VQ..QD......R.R.R..   58   84   F

                  10        20        30        40        50        60        70
                  .         .         .         .         .         .         .          %I   %P   Family
RenSoxC           EHIKRPMNAFMVWAQLERRKMTTEFPDMHNAEISRRLGKLWRLLSDREKQPYIEESERLRIQHMKQYPDYKYRPRKKGG
Sox4 Mm           G............S.I....IMEQS.........K....R.K..K.SD.I.F.Q.A....LK..AD...........VK   70   85   C
Sox14 Mm          D............SRGQ....AQ.N.K...S...K...AE.K...EA..R...D.AK...A....EH........R.PK   67   86   B2
Sox8 Mm           P.V............AA...LADQY.HL....L.KT.........ES..R.FV..A....V..K.DH.....Q..RRKS   63   85   E

                  10        20        30        40        50        60        70
                  .         .         .         .         .         .         .          %I   %P   Family
RenSoxF           GRIKRPMNAFMVWSSLERKKLAEKEPNLHNTELSKRLGQMWKEMTEEDKTPYRQEATRLKDKLMEDHPEYKYKPRRRKD
Sox17 Mm          S..R.........AKD...R..QQN.D...A....M..KS..AL.LAE.R.FVE..E..RVQH.Q...N...R.....Q   60   80   F
Sox4 Mm           .H............QI..R.IM.QS.DM..A.I.....KR..LLKDS..I.FI...E..RL.H.A.Y.D...R..KKVK   57   79   C
Sox9 Mm           PHV..........AQAA.R...DQY.H...A....T..KL.RLLN.SE.R.FVE..E..RVQHKK...D...Q.....S   54   77   E
```

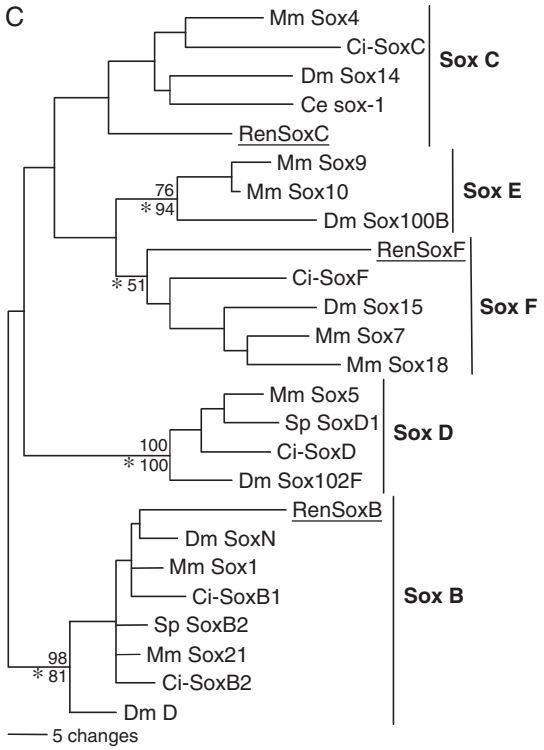

Fig. 6. *Sox* genes expressed during *Reniera* development. (A) Derived amino acid sequences of *RenSoxB*, *RenSoxC*, and *RenSoxF*. The HMG domains are in bold letters. (B) Alignment of the HMG domains of RenSoxB, RenSoxC, RenSoxF, their closest BLAST matches, and selected representatives of other families. (C) Unrooted Bayesian phylogenetic tree of the *Sox* class. At key nodes are given percentages of bootstrap support, obtained by distance (1000 replicates) above and maximum likelihood (100 replicates) below the branch. An asterisk indicates a Bayesian posterior probabilities greater than or equal to 95%. See Fig. 2 for further explanation of alignment and tree. Abbreviations as described in previous figures.

clade comprising both the *Fox L1* and *Fox C* families. Although the support values are low, a relationship between these genes and members of the *FoxL1* and *C* families is consistent with the BLAST results for RenFoxL1 (Fig. 8B). Both analyses group *RenFoxJ* within the *Fox J* family with a significant PP value (NJ < 50, PP: 0.96). Other sponge Fox genes, *Sd-FoxP*, *Sd-FoxD*, and *Sd-FoxL2*, are confidently placed within the *Fox P* (NJ: 100, PP: 0.99), *Fox D* (NJ: 72, PP: 1), and *Fox L2* (NJ: 99, PP: 1) families, respectively, whereas the position of *Sd-Fox1* is unresolved. Most recog-

A

```
RenHNF4   MCYKNLFLMEQQTSPIPNGTRPDPCYNELDLFDYSSTLPLSHGLELGDANTSPMSFASDCSPPPHPDFLDTSYGGVPRYPAGHYQKRVGYTSERDPFFSX  100
          QQPGMRGGMGHLSSSRAVYNGFANPPEYAYQQHQQHNRIPTHLPGYLQILQGFQDQYSPEDLDVKFAVPPSSLGSLTGPGSNXGSPPSTTIHPQSACKVC  200
          NDVASGNHFGVLSCEACKSFFRRSVRAGARYACRGTRNCSVEKHTRNRCQYCRLQKCLQTGMRKEAVQEERAPPVTRIQRAGTPTQLGYSASPAFSPTTF  300
          PNGPPLPPVPPMPSSFRYDGPVSPIMPSNFMTQSRLSGSLPNLKMGCGTGGYGDYDCPPMTPPIHMPPVEPYTGGSRPSTPSSINMTSGVPSSTVPPYSD  400
          SVSSMPEPTPSAVSPFVLVTADMQTESLPDNVTPSDQGIKLEDIFEGARQSLLRVIEWSKRIPAFTTLSLDDQVKLLKSCWCEHVLLKQATRIGPHSDTI  500
          LLSSGLTCRKDQIEDPEVRRIVERVSHEISYWFDVLHVDKVEMACLKGIILFNPDAKGLNPGTRKRVEIFQEQILQALETRCKTMYPVTPLRFSKLLLRL  600
          PPLRAAALESTQHMEVQRTLGNAKLDTIFGELLDFD  636
```

B

```
                                    DNA binding domain (DBD)
              10        20        30        40        50        60        70        80         10        20
               .         .         .         .         .         .         .         .          .         .
RenHNF4      ACKVCNDVASGNHFGVLSCEACKSFFRRSVRAGARYACRGTRNCSVEKHTRNRCQYCRLQKCLQTGMRKEAVQEERAPPV  NVTPSDQGIKLEDIFEGARQSLLRV
RXR Sd       I..I.G.T.......Q..............N.......S...AI................IAN...........T.QG  PIKTPASN.....L...LKT...K.
Rxrg Mm      I.AI.G.RS..K.Y..Y...G..G..K.TI.KDLI.T..DNKD.LID.RQ........Y....VM..KR.......QRSR  MNVENSTNDPVTN.CHA.DKQ.FTL
Hnf4 Mm      L.AI.G.R.T.K.Y.AS..DG..G.......KNHM.S..FS.Q.V.D.DK..Q.R....K..FRA..K.....N..DRIS  GDIRAKKIANIT.VC.SMKEQ..VL
Ci-COUP-TF   E.V..G.KS..K.Y.QYT..G.....K....RNLS.T...N...PIDQ.H..Q......N..VKI...R....RG.M..S  QCPVPN.IMGIDN.C.L.ARL.FSA

                                    Ligand binding domain (LBD)
          30        40        50        60        70        80        90       100       110       120       130
           .         .         .         .         .         .         .         .         .         .         .
RenHNF4      IEWSKRIPAFTTLSLDDQVKLLKSCWCEHVLLKQATRIGPHSDTILLSSGLTCRKDQIEDPEVRRIVERVSHEISYWFDVLHVDKVEMACLKGIILFNPDAKGLN
RXR Sd       ...A......VS............S....CT..L.AQN..KA..V..AN..S.NR..........VIN..FN.LA..L.Y.N..R..L......L.........S
Rxrg Mm      V..A....H.SD.T.E...I..RAG.N.LLIASFSH.SVSVQ.G...AT..HVHRSSAHSAG.GS.FD..LT.LVSKMKDMQM..S.LG..RA.V.........S
Hnf4 Mm      V..A.Y....CE.L.....A..RAHAG..L..GATK.SMVFK.VL..GNDYIVPRHCP.LA.MS.VSI.ILD.LVLP.QE.QI.DN.Y....A..F.D......S
Ci-COUP-TF   V..ARN..F.PE.QVT...AM..WV.S.LFV.NA.QSHM.LHVAP..AAAGLHTSMSADRVMTFMDHI.IFQ.QVERLKS....SA.YS...A.V..TA.SH..S

             140       150       160       170       180       190       200
               .         .         .         .         .         .         .       %I   %P   Family
RenHNF4      PGTRKRVEIFQEQILQALETRCKTMYPVTPLRFSKLLLRLPPLRAAALESTQHMEVQRTLGNAKLDTIFGELL
RXR Sd       .ASK............S....T......S.R..........A..K.IS..VM.......A...T.M.NL.....   72   85   2
Rxrg Mm      NPSE--..TLR.KVYAT..AYT.QK..EQ.G..A.......A..SIG.KCLE.LFFFKLI.DTPI.SFLM.M.   41   59   2B
Hnf4a Mm     DPGK--IKRLRS.VQVS..DYINDRQYDSRG..GE...L..T.QSITWQMIEQIQFIKLF.M..I.NLLQ.M.   40   58   2A
Ci-COUP-TF   DM.H--I.SV..KSQC...EYVRHQ..NQ.S..G.......S..TVSASAIEQLFFV.LV.KTPIE.LIRDM.   40   57   2F
```

C

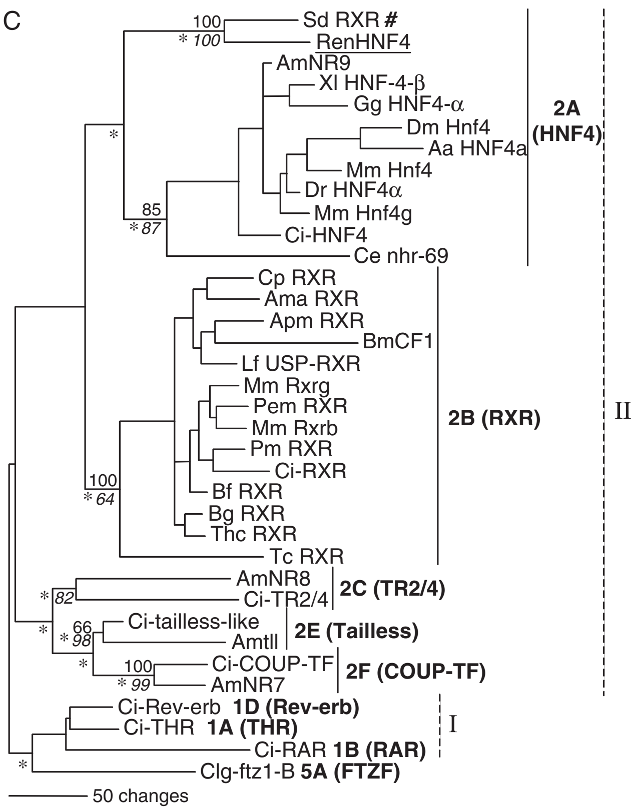

Fig. 7. *HNF4* gene expressed during *Reniera* development. (A) Derived amino acid sequence of *RenHNF4*. The DNA binding domain (DBD) is in bold letters and the ligand binding domain (LBD) is underlined. (B) Amino acid alignment of conserved regions of RenHNF4, its closest BLAST matches, and selected representatives of other closely related families. A short region downstream of the DNA binding domain is conserved. The divergent sequence between the DBD and the LBD is not shown. The LBD is displayed from the beginning of alpha helix 3. (C) Unrooted Bayesian phylogenetic tree of the *nuclear receptor* class. At key nodes, percentages of bootstrap support obtained by distance (1000 replicates) are given above the branch. Maximum likelihood (100 replicates) bootstrap values are given in italic below the branch, referring to an analysis undertaken with a subset of 25 genes; they are indicated for the corresponding major clades. An asterisk indicates a Bayesian posterior probabilities greater than or equal to 95%. See Fig. 2 for further explanation of alignment and tree. In addition to previously described abbreviations: Aa, *Aedis aegypti*, insect; Ama, *Amblyomma americanum*, chelicerate; Apm, *Apis mellifera*, insect; Bg, *Biomphalaria glabrata*, mollusk; Bm, *Bombyx mori*, insect; Clg, *Clarias gariepinus*, fish; Cp, *Celuca pugilator*, crustacean; Gg, *Gallus gallus*, bird; Lf, *Lithobius forficatus*, insect; Pem, *Petromyzon marinus*, fish; Pm, *Polyandrocarpa misakiensis*, urochordate; Tc, *Tripedalia cystophora*, cubozoan jellyfish; Thc, *Thais clavigera*, mollusk; Xl, *Xenopus laevis*, amphibian.

A

```
RenFoxL1   MIAATGPDQSMDQQRQRSYLPTYQYMKPSIAASSGAPVNQYSATTARPMYQLYDQSTQNGYMPLGSHTDYSLYSSYPYQSLSSRTAALCRAAGRTWPINQ  100
           AQETISAGSNTGLMAGGNATAMAMQGSSNRFQKPAYSYIALIAMSIECAPHKRATLSEICQFIRDRFPYYQQNCKQGWENSIRHNLSLNECFVKQPREQG  200
           RPGKGHYWTLDKNALKMFENGSFRRRKRRFKKGDVIGVEDHPESAGICSTMDALRTHGYIAGLAAGGSQGLPPGPHRGFAQPAGGLLQGEIISPCTPHYP  300
           AIRQPESAHGQHFVFPMGGSTVPGMSSALPSQHMPHGMQSHMLSMDQSGVTSSPLVGSMHGFNSQAPWMSGLMFPEVTSNSSIPDTNQTSSATTEKQIIY  400
           SPYDGHQGPQNNSPIHASAITSPLSTSQKQQQQQWSENSPLPHIPDIPSIPSCAESTGDNGQLSIGPSTTNGCGIGEITSIISHRCFQ  488

RenFoxJ    TTIIYLAIRSSKNDKVTLGEIYQWIKDHFMYYRVAEPTWQNSVRHNLSLN  50
```

B

```
                  10        20        30        40        50        60        70        80        90        100
                  .         .         .         .         .         .         .         .         .         .  %I  %P Family
RenFoxL1 KPAYSYIALIAMSIECAPHKRATLSEICQFIRDRFPYYQQNCKQGWENSIRHNLSLNECFVKQPREQGRPGKGHYWTLDKNALKMFENGSFRRRKRRFKK
FoxL1 Dr ..P.........A.KN..D......G.Y...M......HD.-....Q..........D..I.V...K......S.....TKC.D.....NY.....KCRT  73  83  L1
Foxl1 Mm ..P.........A.QD..EQ.V..NG.Y...M....F.HD.-R...Q..............V...K......S.....PRC.D.....NY.....KP.P  71  82  L1
Foxc2 Mm ..P.......T.A.QN..E.KI..NG.Y...M....F.RE.-....Q..............V..DDKK....S.....PDSYN.......L..R.....  70  83  C
Sd-FoxF  ..P.......TLA.MSKAERK...A....Y..ET.S..RE.................Q..Q.L.....K.......VI.PG.RH..DD..Y......YMR  66  83
Foxe1 Mm ..P.........A.AH..ER.L..GG.YK..TE...F.RD.-PKK.Q.......T..D..L.I...A......N..A..P..ED...S...L..RK...R  65  81  E

                    10        20        30        40        50
                    .         .         .         .         .  %I  %P  Family
RenFoxJ     TTIIYLAIRSSKNDKVTLGEIYQWIKDHFMYYRVAEPTWQNSVRHNLSLN
Foxj1 Mm    A.L.CM.MQA..AT.I..SA..K..T.N.C.F.H.D.....I.......  60  82  J
FhTF Ec     AQ..TQ...T.SAG.L..S...R..E.S.E...H.N.V.K..I.......  62  74
Hyfkh2 Hv   AAL.IM.MK.KVCG.M..S...K..G...PF.KY...S....I.......  58  78
FOXN2 Hs    SLL..M..EH.P.KCLPVK...S..L...P.FAT.PTG.K.........  54  66  N
```

C

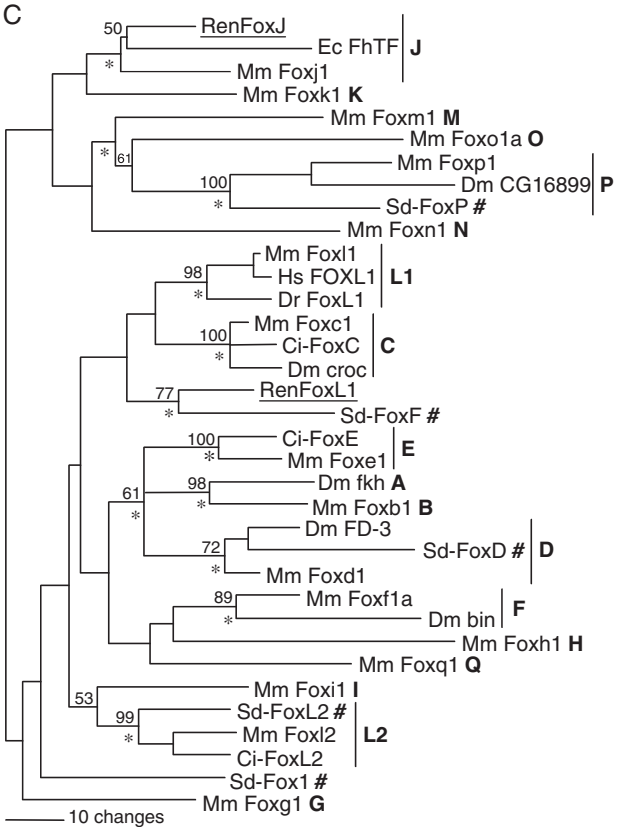

Fig. 8. *Fox* genes expressed during *Reniera* development. (A) Derived amino acid sequence of *RenFoxL1* and *RenFoxJ*. The forkhead domain (partial for RenFoxJ) is in bold letters. (B) Amino acid alignment of the forkhead domains of RenFoxL1, RenFoxJ, their closest BLAST matches, and selected representatives of closely related families. Note cysteine insertion specific to RenFoxL1 and Sd-FoxF at postion 42 of the Fox domain. (C) Unrooted Bayesian phylogenetic tree of the *Fox* class. At key nodes, percentages of bootstrap support obtained by distance (1000 replicates) are given above the branch. An asterisk indicates a Bayesian posterior probabilities greater than or equal to 95%. See Fig. 2 for further explanation of alignment and tree. In addition to previously described abbreviations: Ec, *Encephalitozoon cuniculi*, fungus; Hs, *Homo sapiens*, human.

nized families and some higher level groupings (Kaestner et al. 2000; Mazet et al. 2003) are supported by this analysis.

T-box

RT-PCR products encoding two distinct genes were obtained. Contiguous sequences of 1650 and 1541 bp were obtained by RACE for *RenTbxA* and *RenTbxB,* respectively (Table 1; Fig. 9). Sequence similarity is insufficient to assign either of the two genes to a particular *T-box* gene family (Fig. 9B), and they have therefore been named *RenTbxA* and *RenTbxB*.

Interestingly, the percent sequence identities within the T-box domain are higher for *RenTbxA* (46–50%) than for *RenTbxB* (37–42%), indicating that the latter is considerably

A

```
RenTbxA   MNTEVAPEEFEPEIKRGCWKEEPEDDELPLLNVPDNSRLGGVKMETADPSISLEICSKGGGSLLGQEELNPQSTTSTDSPEAEELEESPSPLLATPHSGI  100
          TLT LDNMDLWSKFSAVATEMIITKSGRRMFPTFSASICGLQMDQKYTMTFRCNPADNKRYKFVNTSWVVAGKGESHIDDLLAHVHPDSPALGKHWLHNKV  200
          HFKKVKLTNNKNTRKGQIALNSMHKYIPCVIVSKFITRKKSEVVHVQEFPEAYFYAVTAYQNDQVTQLKIDNNPFAKAFRDSPADSELQSGFNYHGNWFN  300
          NGISRWPVPSSPLLGPSSYPQNSPLLSPKWPPLSPLLSPYQGDFTQVTFSQVTPSAMPITGPSPLPQFGSVFGWDSSTGHFSLATSNMFRPSLPTTPTIP  400
          TLPPTTGGANIADATHDV  418

RenTbxB   MNSTATNIVVTAPSLSPVGTFPHLTFKYSPAPGTLHHSPIAAGYHFSPYSFPPIYLDTKPDSFSPVFLQQQPTNGGSSSDEEKQQNGSTPDLIATPPALA  100
          VGPAIYAPSSAPVPPPSATDKPAKHKEQLVPSAV LTENQLWKAFDTAGNEMIVTKPGRRLFPTILVTVSGLVPTIQYSIMVRFVSADKYRHKYITGQWTA  200
          VGESDIIHDESKMLFIHPSSPCYGEMWIKKPISFKSIKITHHSESQNGNILLHTMHKYRIEVVVQPHNDQLGEIVFPLAHSEFIAVTSYQSSAITKLKVD  300
          NNPFAKGF RDRETFISPYSQSYAPFLPPYTPFVHPLWANVTNTPSGMYPNPLAPVATVPPTPQIFSTSMTTTTVLTPTSSVQPLFPPSSNVFPVGASAQS  400
          SIDNGATVLEVDTPPSDLQSRSSSTSSKTEDEEEYEIIVKEDDDDIVDDD  450
```

B

```
                    10        20        30        40        50        60        70        80        90        100       110
                    .         .         .         .         .         .         .         .         .         .         .
RenTbxA   LDNMDLWSKFSAVATEMIITKSGRRMFPTFSASICGLQMDQKYTMTFRCNPADN--KRYKFVNTSWVVAGKGESHI-DDLLAHVHPDSPALGKHWLHNKVHFKKVKLTNNKNT
RenTbxB   .TENQ..KA.DTAGN...V..P...L...ILVTVS..VPTIQ.SIMV.FVS..K--Y.H.YITGQ.TAV.ESDIIHDESKMLFI..S..CY.EM.IKKPIS..SI.I.HHSES
org-1 Dm  .ETKA..DQ.H.QG.......T........QVR.G..DPHAT.ICMMDFV.M.DKRY..A.H.SC......ADPI--SPPRI........V.SN.MKQI.S.D.L.....QLD
Tbx18 Mm  .QGAE..KR.HEIG.......A......AMRVK.S..DPH.Q.YIAMDIV.V..KRY..VYHSSK.M...NAD.P--VPPRVYI......S.ET.MRQVIS.D.L.....ELD
Tbx20 Mm  .ETKE..D..HELG..............IRV.FS.VDPES..IVLMDIV.V..KRY..AYHRS..L....ADPP--LPARLY......FT.EQL.KQM.S.E.......ELD
Tbx2 Mm   .EAKE..DQ.HKLG...V..........P.KVRVS..DKKA..ILLMDIVA..D--C....H.SR.M....ADPE--MPKRMYI......T.EQ.MAKP.A.H.L.....ISD
Tbx5 Mm   .HERE..L..HE.G.......A......SYKVKVT..NPKT..ILLMDIV...D--H....ADNK.S.T..A.PA--MPGRLY......T.A..MRQL.S.Q.L.....HLD
Doc3 Dm   .E.NE..QQ.HSIG.......C......SMRV.LS..EEEAS.CVLLEMV.IGD--C....SGSQ..P..GA.PQ--SPQRMYL..E...T....QSQALL.S.......TLD
Pf-Tbrain .C.R...L..HQQN.......Q.......L.FRFT..DPSAH.NVFVDMVLS.P--NHW..QSGK..PC.QA.HVH-PGSNIYI.....NT.N..MKQE.V.S.L......GK
ApBra     .EDR...RR..KLTN...V..T......VL...VT..NPNAM.SILLDFT...E--H.W.Y..GE..PG..PD.P--PPST.YI.....NF.A..MKQC.S.S....S.-.LN

                    120       130                 140       150       160       170
                    .         .                   .         .         .         .
RenTbxA   RKGQIALNSMHKYIPCVIVSKFITRKKSE---------VVHVQEFPEAYFYAVTAYQNDQVTQLKIDNNPFAKAFRDS
RenTbxB   QN.N.L.HT....RIE.V.QPHNDQL------------GEI.FPLAHSE.I...S..SSAI.K..V.......G...R
org-1 Dm  EN.H.I.....R.Q.RFHLVYLPPKNA.LDENEHS--SHFRTFI...TS.T.......QR......SS.....G...D
Tbx18 Mm  DQ.H.I.H.....Q.R.H.IRKDCGDDLSPIKPVPSGEG.KAFS...TV.TT......Q.I.R....R.....G....
Tbx20 Mm  QH.H.I.......Q.R.HII.KKDHTA.LLNLKS---EEFRTFI...TV.T.......QLI.K....S.....G....
Tbx2 Mm   KH.FTI.......Q.RFHIVRANDIL.LPY-------STFRTYV...TD.I........KI............G...T
Tbx5 Mm   PF.H.I.......Q.RLHIV.ADENNGFGSKN-----TAFCTHV...TA.I...S...HKI.....E......G..G.
Doc3 Dm   NN.H.V.A.....Q.RLH.IRTSDLGQIPWA-------PQQAFV...TE.I........RI.K..........G..E.
Pf-Tbrain DN.H.V.......Q.RIH.IEVSPNRPPDQR-------TLQTHS...TQ.F.......TDI......H.....G...N
ApBra     GT...M...L...E.RIH.IRVGGPE.QR---------LIRSFS...TQ.I.......EDI.....KY.......L.I
```

	RenTbxA %I	RenTbxA %P	RenTbxB %I	RenTbxB %P	Family
RenTbxB	36	54			
org-1 Dm	49	58	38	55	Tbx1/10
Tbx18 Mm	46	59	42	59	Tbx15/18/22
Tbx20 Mm	50	60	41	61	Tbx20
Tbx2 Mm	48	61	41	63	Tbx2/3
Tbx5 Mm	49	59	38	62	Tbx4/5
Doc3 Dm	47	60	39	61	Tbx6
Pf-Tbrain	47	58	37	59	T-brain
ApBra	49	63	39	58	Brachyury

C

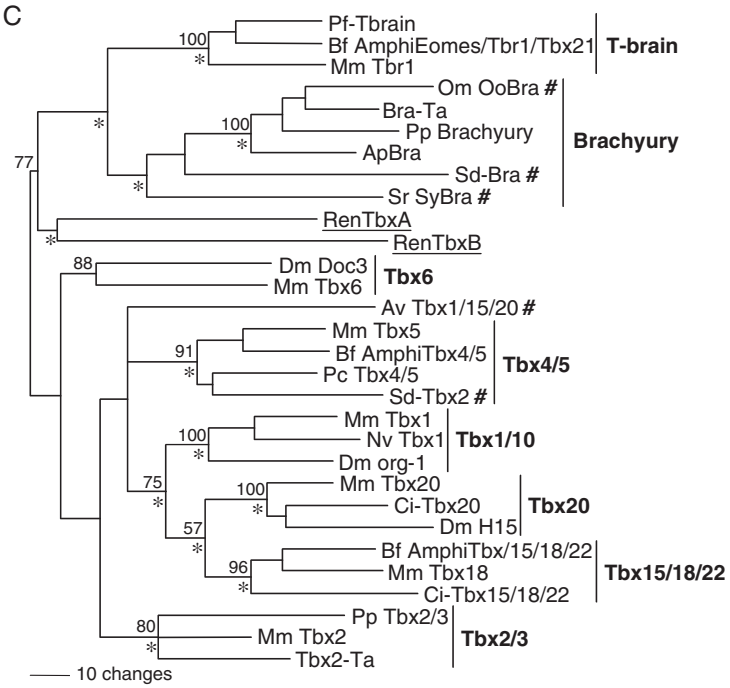

Fig. 9. *T-box* genes expressed during *Reniera* development. (A) Derived amino acid sequence of *RenTbxA* and *RenTbxB*. The T-box domain is in bold letters. (B) Alignment of the T-box domains of RenTbxA, RenTbxB, and selected representatives of all eight T-box families. (C) Unrooted Bayesian phylogenetic tree of the *T-box* class. At key nodes, percentages of bootstrap support obtained by distance (1000 replicates) are given above the branch. An asterisk indicates a Bayesian posterior probabilities greater than or equal to 95%. See Fig. 2 for further explanation of alignment and tree. In addition to previously described abbreviations: Ap, *Asterina pectinifera*, echinoderm; Av, *Axinella verrucosa*, demosponge; Pf, *Ptychodera flava*, hemichordate; Pp, *Pleurobrachia pileus*, ctenophore; Om, *Oopsacas minuta*, hexactinellid sponge; Ta, *Trichoplax adhaerens*, placozoan.

more divergent. Phylogenetic analyses are unable to resolve the position of either gene within the *T-box* class (Fig. 9C). The two proteins are grouped together in both trees with a significant Bayesian PP (0.98); it is possible that they represent a gene duplication event within the sponge lineage, but this cannot be determined from current data. In this analysis, *OmBra* from a hexactinellid sponge, *Sd-Bra* from a demosponge and *SyBra* from a calcareous sponge are confidently

placed within the *Brachyury* family (NJ <50, PP: 0.98). Demosponge *Sd-Tbx2* is placed within the *Tbx4/5* family with high support values (NJ: 91, PP: 1) whereas the position of demosponge *Tbx1/15/20* is unresolved. Recognized gene families (Papaioannou 2001; Takatori et al. 2004) and some higher level groupings are highly supported.

Mef2

The single *Mef2* consensus sequence obtained by RT-PCR is most similar to *Mef2* from *Podocoryne carnea* (Fig. 10) and has been named *RenMef2*. A 1860 bp transcript obtained by RACE (Table 1) encodes a 543 amino acid protein with the MADS/Mef2 domain spanning the N-terminal region from position 3 to position 86, as is the case for all other Mef2 proteins (Fig. 10).

RenMef2 most closely resembles other animal proteins from the Type II family of MADS box proteins (Fig. 10B), demonstrating sequence homology to Mef2 proteins both within and downstream from the Mef2 domain (not shown). Accordingly, phylogenetic analysis places *RenMef2* at the base of the metazoan *Mef2* clade with high statistical support (NJ: 83, ML: 59, PP: 0.97; Fig. 10C). This analysis also supports the Type I and II divisions (NJ: 90, ML <50, PP: <0.95) and lower level groupings.

Ets

RT-PCR and 3′ RACE sequence yielded a 670 bp partial cDNA encoding 152 amino acids of the protein (Table 1; sequence not shown). The sequence is most similar to two sponge genes, *Ets2* from *Tethya aurentia* and *Ets1* from *Haliclona* sp., and it contains an ETS domain. It appears to be part of the *Ets* family of the *Ets* class of transcription factor genes (Sharrocks 2001; data not shown) and has been named *RenEts*.

Temporal expression of *Reniera* transcription factors during development

In addition to determining the expression pattern of *Reniera* transcription factor genes throughout development, RT-PCR analyses served to confirm that the cDNAs we isolated did indeed originate from *Reniera*. Relative transcript levels of all of the transcription factor genes described above were assessed after cycle-restricted RT-PCR. Three replicates—from independent RNA preparations—displayed the same patterns, and controls for genomic DNA contamination were negative (data not shown). Comparison of expression patterns indicated that all staged RNAs were viable RT-PCR templates, as many different expression patterns were obtained (Fig. 11).

RT-PCR indicated that all of the transcription factors recovered from *Reniera* are expressed in embryos, newly emerged larvae, competent larvae, metamorphosing postlarvae, and juveniles (Fig. 11). Transcript levels increase in the postlarvae during early metamorphosis for most transcription factor genes, specifically *RenPaxB*, *RenBsh*, *RenProx2*, *RenPouI*, *RenLim3*, *RenSoxC*, *RenSoxF*, and *RenHNF4* (Fig. 11); larvae have barely detectable *RenPaxB* and *RenProx2* transcript levels. All of these genes except for *RenPouI*, *RenLim3*, and *RenSoxC* maintain high levels of expression in the juvenile. Transcript prevalence of *RenSoxB* and *RenFoxJ* appears to be greatest at the time larvae become competent to respond to inductive settlement cues. Only *RenSoxF* appears to have a marked increase in expression in larvae that have just emerged from the sponge. Both *RenTbx* gene transcripts appear at higher levels in early development (Fig. 11).

Localized expression of select ESTs in *Reniera* larvae

Four genes from the EST set (unpublished data)—encoding ferritin (Fig. 12, A–C), procollagen lysyl hydroxylase (Fig. 12, D–F), galectin (Fig. 12, G and H), and B-ZIP1 (Fig. 12, I and J)—were assessed by whole mount in situ hybridization (WMISH). We elected to use these arbitrarily selected ESTs as control probes to determine if gene-specific expression patterns could be obtained by WMISH, which had not been performed previously on demosponge embryos or larvae. We focused on the larval stage as this was when cell layers and localized cell types were most evident. Unique staining patterns were obtained for all four probes, with all of the genes restricted to either a specific cell layer or, in some cases, a subset of cells or a specific territory within a layer (Fig. 12). Together, these genes were expressed in all three layers.

Ferritin transcripts were localized to the outer columnar epithelial cells, with no staining detected in the inner cell mass, the middle subepithelial layer, or in a ring of columnar epithelium adjacent to the pigment ring (Fig. 12, A–C). Procollagen lysyl hydroxylase was expressed in distinct cell populations in all three cell layers (Fig. 12, D–F). In the outer epithelium, it was restricted to the ring of columnar cells that lack ferritin expression, adjacent to the pigment ring. These transcripts were also detected in a subset of cells interspersed throughout the subepithelial (middle) layer (Fig. 12F) and in a population of cells in the inner cell mass, which were predominantly localized in the centroanterior region of this layer. Galectin transcripts were restricted to inner portion of the inner cell mass, with no staining detected in the middle and outer layers (Fig. 12, G and H). Uncharacterized leucine zipper transcription factor gene B-ZIP1 was expressed in a subset of cells in the inner cell mass. These cells are chiefly localized to the outer territory of inner cell mass, just beneath the subepithelial layer, and in the center of the inner cell mass (Fig. 12, I and J).

A

```
RenMef2   MGRKKIQITRIQDERSRQITFSKRKAGLLKKAYELSILCDVEIAVIMIAHNSKLYQYASSNMNSILIRYTEFKDPDESKTNADILEDITKKERGRGDDLG  100
          DDDSIDASMLLAGDGGGGGSVPPRRVAVSVPPLSPNTQAQYQRIDEEFDHVLHTELQQAPPQPVPSPVPPSLVHVLPTKVPDAMPVSVPVKRLSSTGSS  200
          SLTDSPITPGSFPLKTIKSPLSVGSTGSMPPPAGTPTRLQTIRSQTLKLHLPGGGRNYIPLSYSGKSPLVDEGGGGGGEADGDNESTSNRPQLSVVIPGQ  300
          KGFMSRTIPSSPPPSAEGDNDHAPLSSSSDDSPSEQKPVPLTVATSHFHPPPVTVNTDSLTTPIMSLTTPSLFSSVQPGLSSFAADVPGLGDIGSASALV  400
          AFPVNIQPSSTGNPSASGPGGNIVIQAQNTGQPFLLTPKQTPGAVHTSGGLLSTPPNSNLRDLEQLKLQYDKIYQQISQTLQQHSHGGGGQGGAQNNEHK  500
          DSKPLTDSNGKLDSNRTEVIVPGPPPPLSRVTVSLGDSIRGRQ  543
```

B

```
                    MADS Domain                                   Mef2 Domain
                10        20        30        40        50        60        70        80
                 .         .         .         .         .         .         .         .          %I   %P   Type
RenMef2    MGRKKIQITRIQDERSRQITFSKRKAGLLKKAYELSILCDVEIAVIMIAHNSKLYQYASSNMNSILIRYT-EFKDPDESKTNADILE
Mef2 Pc    ........S..N...N..V..T...F..M...........C...L.IFNSGN..F....TD.DK..LK..-.YNE.H..R.....H.   67   87   II
Mef2d Mm   ........Q..T...N..V..T...F..M.......V...C...L.IFN.SN..F....TD.DKV.LK..-.YNE.H..R.....I.   66   89   II
mef-2 Ce   ..............N..V..T...F..M.......V...C...L.VFNSTN..F....TD.DKV.LK..-.YNE.H..R..N..M.   66   88   II
Rlm1p Sc   ...R..E.Q..S.D.N.AV..I......F...H...V..Q.D....ILGS.NTF.EFS.VDT.DLIYH.QNDKNLLH.V.DPS.YGD   46   67   II
Srf Mm     R..V..KMEF.DNKLR.YT......T.IM.......T.TGTQVLLLVASETGHV.TF.TR                            38   63   I
```

C

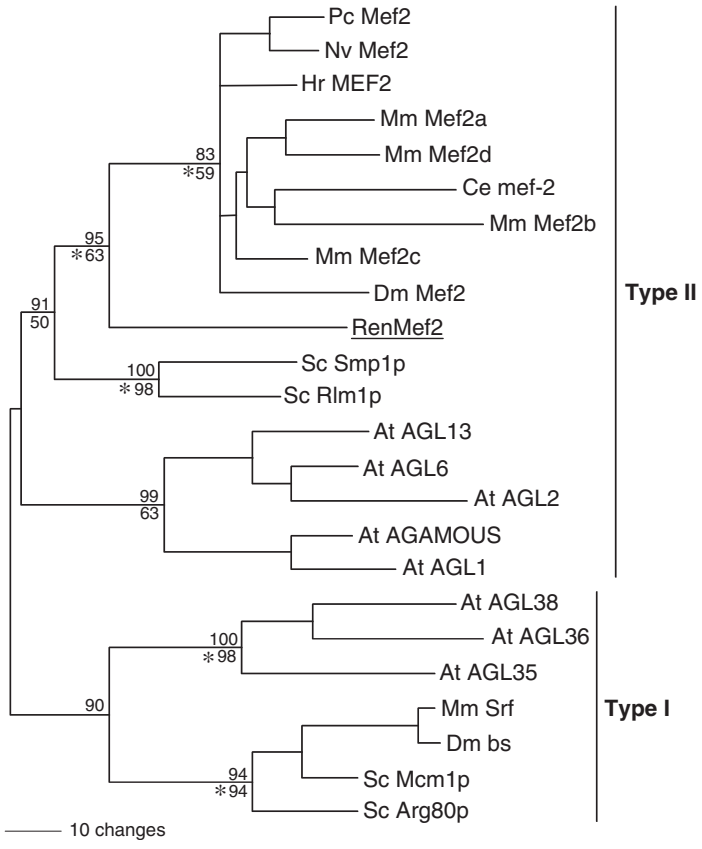

Fig. 10. *Mef2* gene expressed during *Reniera* development. (A) Derived amino acid sequence of *RenMef2*. The MADS domain is in bold letters and the Mef2 domain is underlined. (B) Amino acid alignment of the MADS and Mef2 domains of RenMef2 and selected representatives from the Mef2, yeast Type II, and animal Type I families of MADS box genes. Only the MADS domain of the mouse SRF protein is shown as it is the only region that aligns with the Mef2 proteins. Three amino acids upstream of the Mef2 domain are shown. (C) Unrooted Bayesian phylogenetic tree of MADS box genes. At key nodes are given percentages of bootstrap support, obtained by distance (1000 replicates) above and maximum likelihood (100 replicates) below the branch. An asterisk indicates a Bayesian posterior probabilities greater than or equal to 95%. MADS box type is indicated on the right of the tree. See Fig. 2 for further explanation of alignment and tree. In addition to previously described abbreviations: At, *Arabidopsis thaliana*, plant; Sc, *Saccharomyces cerevisiae*, fungus.

Localized expression of *RenHNF4* and *RenBsh*

RenHNF4 transcripts were detected by WMISH in a subset of cells of the outer epithelial layer of the larva (Fig. 12, A–C), similar to that observed for ferritin transcripts (Fig. 11, A–C). However unlike ferritin, only the small ciliated columnar cells, and not the larger epithelial cells, such as mucous and flask cells (Leys and Degnan 2001), appeared to be expressing *RenHNF4* (Fig. 12C). In contrast, *RenBsh* was expressed only in the inner cell mass in a pattern unlike any observed in the ESTs (Fig. 12, D and E), with unique staining of the inner cell mass beneath the pigment ring. Closer inspection revealed that *RenBsh* transcripts were localized to cells associated with spicules (Fig. 12E), which were dispersed throughout the inner cell mass but at highest density in the posterior region (Fig. 12F). Spicule-producing sclerocytes appear to begin differentiating from the larger micromere pool around onset of gastrulation (Leys and Degnan 2001, 2002; Leys 2003b). Later,

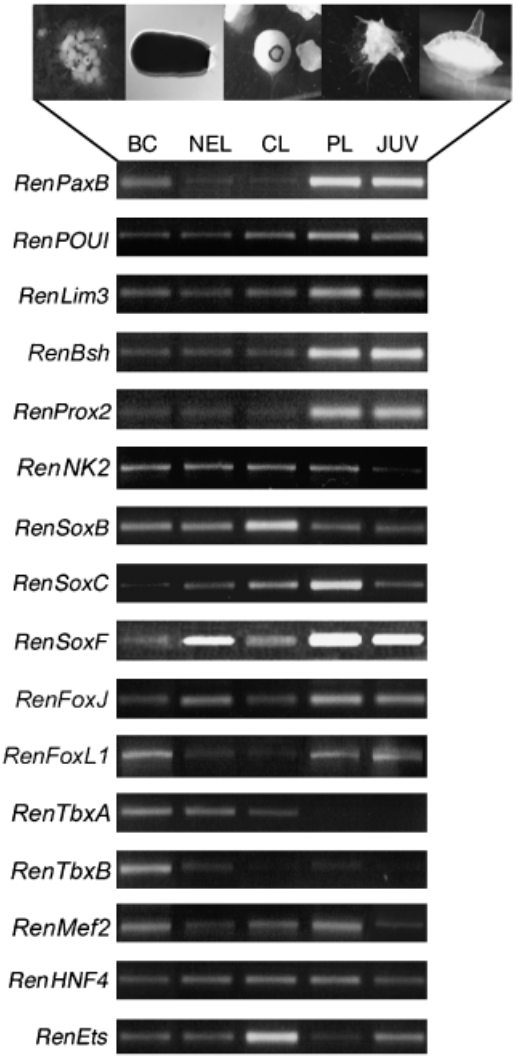

Fig. 11. Cycle-restricted gene-specific RT-PCR of *Reniera* transcription factor genes during a developmental time course. Columns correspond to stages shown in the pictures above, from left to right: BC, brood chamber composed of various staged embryos; NEL, newly released larvae; CL, competent larvae able to initiate metamorphosis; PL, postlarvae undergoing early metamorphosis; and JUV, 3-day-old juvenile.

some of these cells appear to migrate to the posterior pole alongside the pigmented cells and migrate into the inner cell mass. The sclerocytes in the larva become elongated and house an internal spicule. WMISH revealed that *RenBsh* was initially expressed in a small number of micromeres in the cleaving embryo, which were localized both internally and on the outer rim of the embryo (Fig. 14, A–D). These cells possessed a large vesicle suggesting that they were young sclerocytes. Later in development, at the stage pigment cells were coalescing into a spot (Fig. 14E) and then forming a ring (Fig. 14F), *RenBsh* expression was observed both in cells in the thick outer layer and the inner cell mass. In the larva, stained cells were associated with a spicule most of the time and elongated in the direction of the spicule (Figs 13, D, E and 14, G, H).

DISCUSSION

The basic repertoire of cell specification and morphogenetic processes employed during gastrulation and pattern formation in eumetazoans and bilaterians appears to be operational during *Reniera* embryogenesis. Early development in *Reniera* yields a three-layered embryo consisting of a number of patterned cell populations (Leys and Degnan 2002). Some of these differentiated cells (e.g., pigment cells) are arranged into localized functional units or simple tissue-like structures. Asymmetric cell division, organizers, and morphogen gradients appear to create and localize differentially competent populations of cells to specific territories within the demosponge embryo (Degnan et al. 2005). These developmental processes are operating probably during sponge metamorphosis—e.g., formation of choanocyte chambers—although the adult body plan axes and the demarcation of cell layers are often less obvious. Given the demosponge *Reniera* is likely to be a member of the most basal metazoan lineage (Borchiellini et al. 2001; Medina et al. 2001; Cavalier-Smith and Chao 2003), developmental mechanisms shared between this taxon and other living animals are inferred to have been present in the LCA to all extant metazoans.

Origin and diversification of transcription factors

Sequence alignments and phylogenetic analyses of the *Reniera* transcription factor genes have allowed for classification of these genes into specific gene classes and families, and in some cases the identification of *Reniera* orthologs of well-studied eumetazoan developmental genes. Despite extensive sequencing of developmental cDNAs derived from degenerate PCR, we identified only 1–3 members of any particular gene class. Although this may suggest that the repertoire of expressed transcription factor gene family members is markedly smaller in *Reniera* compared with eumetazoans, ongoing EST and genome projects need to be completed before such conclusions are made. Phylogenetic analyses of sponge developmental genes suggest that some gene duplication events occurred prior to metazoan cladogenesis (Hoshiyama et al. 1998; Miyata and Suga 2001) and that there exist transcription factor gene families in sponges additional to the ones detected in this survey.

Genes of the ANTP class have not been found outside of Metazoa. The apparent affinity of *RenBsh*, *RenProx2*, *RenNK2* (and their demosponge orthologs) to the *Bsh*, *Tlx*, and *NK2* families, respectively, plus the placement of *Prox3* within

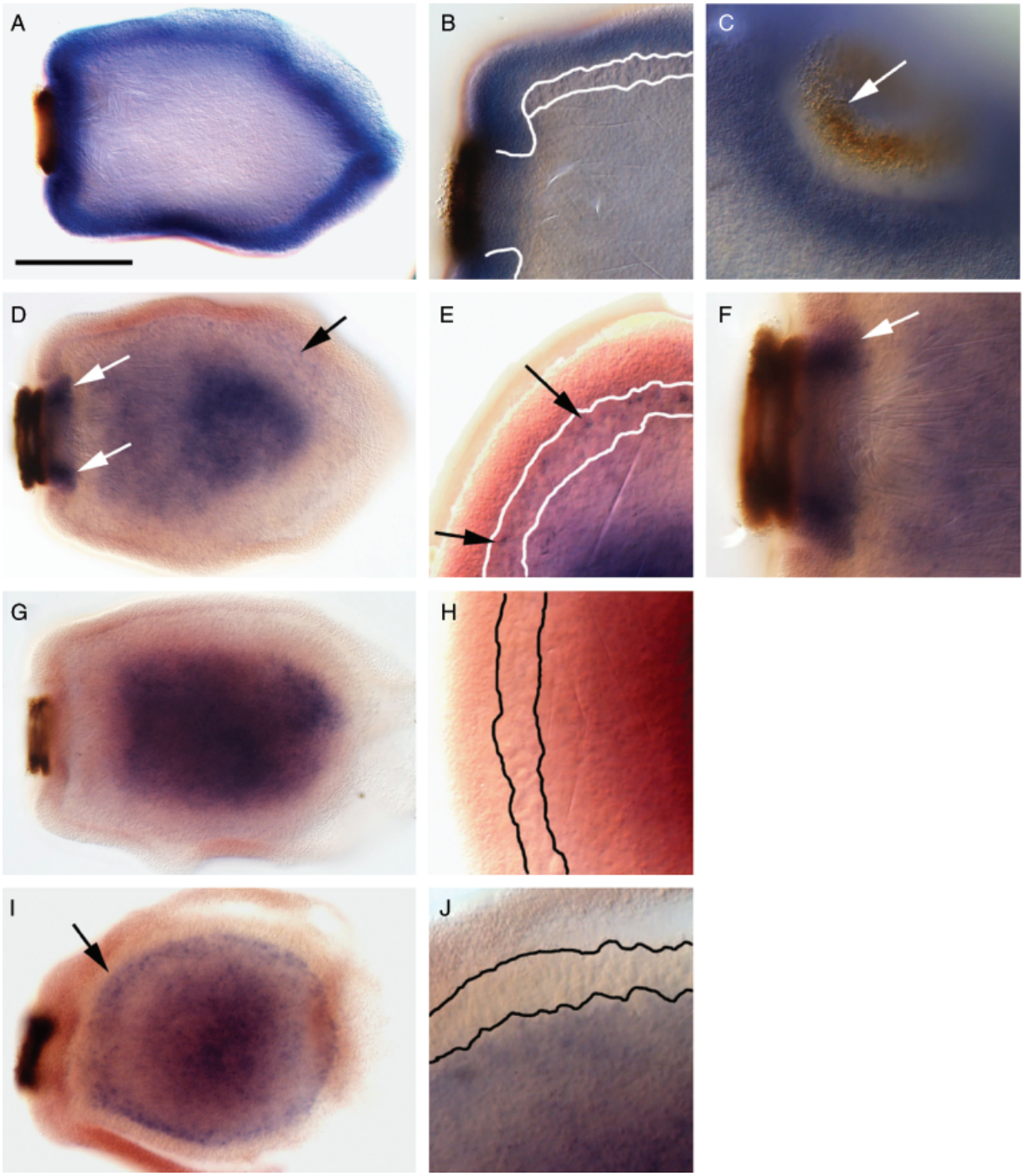

Fig. 12. In situ hybridization analysis of expression of select ESTs in *Reniera* larvae. Anterior to the right in plates A, D, G and I. (A-C) Ferritin transcripts are restricted to the outer columnar epithelial layer. (B) White lines demarcate the three larval cell layers at the posterior end of the larva (pigment ring evident) and show no ferritin expression in the inner cell mass and middle subepithelial cell layer. (C) A territory of outer columnar cells adjacent to the pigment ring (arrow) do not express ferritin. (D-F) Procollagen lysyl hydroxylase is expressed in subsets of cells in all three cell layers, being restricted to (1) the outer columnar epithelial cells adjacent to the pigment ring (white arrows in D and F), (2) a subset of cells interspersed throughout the middle subepithethial layer (black arrows in D and E) and (3) a subset of cells in the inner cell mass that are largely localized to the inner anterior half. White lines in E demarcate larval cell layers. (G, H) Galectin expression is restricted to the inner portion of the inner cell mass. Inner cell mass cells that border the subepithelial layer do not express galectin. Black lines in H demarcate larval cell layers. (I, J) B-ZIP1 is expressed in two populations of cells in the inner cell mass. One is localized predominantly to the boundary of the subepithelial layer and the other in the center of the cell mass. Black lines in J demarcate larval cell layers. Scale bar, 250 μm.

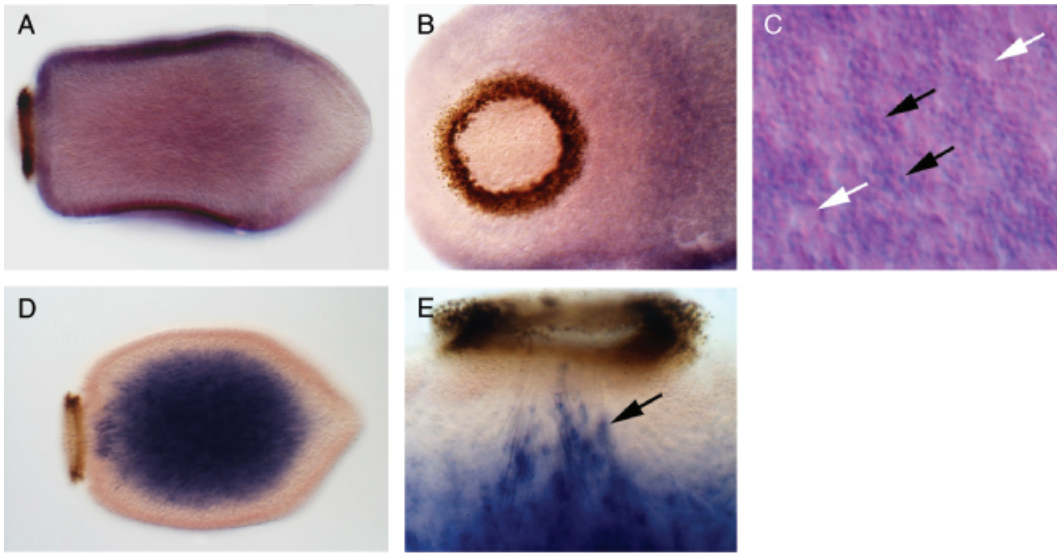

Fig. 13. Expression of *RenHNF4* and *RenBsh* in larvae. (A–C) *RenHNF4* expression is restricted to ciliated columnar cells in the outer epithelial layer. Transcripts are not detected (B) in the vicinity of the pigment ring and (C) in large cells in the epithelial layer (e.g., mucous and flask cells; white arrows), and are restricted to clusters of small ciliated columnar cells (black arrows). (D, E) *RenBsh* is expressed in the inner cell mass. The staining pattern co-localizes with spicules (arrow).

the *Msh* family, suggest that the metazoan LCA possessed at least four ANTP class genes, probably belonging to the *Bsh*/*BarH*, *Tlx*, *NK2*, and *Msh* families. The precise evolutionary origin of these families awaits a definitive answer on the presence or absence of these genes from whole genome sequencing of sponge and choanoflagellates. Present data indicate only that the ANTP class genes had already diversified prior to the sponge-eumetazoan split (Gauchat et al. 2000; Garcia-Fernandez 2005). Despite extensive efforts from our group and others, *Hox*, *ParaHox*, extended-*Hox* or *EHGbox* genes have not been identified in sponges, suggesting that a number of major eumetazoan ANTP subclasses and families may have evolved in the eumetazoan lineage after its separation from demosponges.

Current data confirm that *PaxB*, *POU I*, *Lim-3* (*LIM-HD*), and *HNF4* NR transcription factors, which belong to metazoan-specific gene classes, are present in sponges and therefore likely present in the metazoan LCA genome. The close relationship of these sponge genes to specific existing eumetazoan family clades suggests that genes belonging to other *Pax*, *POU*, *LIM-HD*, and *NR* families remain to be discovered in sponges and that some gene family diversification occurred prior to the metazoan radiation. Alternatively, it is possible that the clades comprising the sponge gene represents the ancestral form of the class, from which all other eumetazoan families evolved. In the *POU* class, the second sponge gene, *spou2*, could have arisen from a POU I-like ancestor. Sequencing of a sponge genome will enlighten us as to the extent of gene duplication and diversification occurred before demosponges split off from the main metazoan lineage.

RenPaxB appears to have a more conserved HD than the other sponge PaxB protein, sPax2/5/8. As most cnidarian PaxBs have a seemingly functional complete HD, whereas bilaterian PaxB HDs are truncated, degenerate or absent, it has been proposed that the ancestral PaxB had a complete and functional HD (Miller et al. 2000). The *Reniera PaxB* gene supports this proposal—its encoded HD is complete, although divergent in sequence.

POU I genes have thus far been identified only in vertebrates and demosponges. In our phylogenetic analyses, statistical support for the close relationship between *RenPouI*/*Ef spouI* and *Pit-1* genes is high. If the lack of recovery of *POU I* family genes from bilaterian invertebrates truly reflects the absence of these genes, then clearly they have been lost from multiple metazoan lineages. Recent comparison of coral ESTs with existing metazoan sequence databases revealed that lineage-specific gene loss probably has occurred extensively throughout metazoan evolution (Kortschak et al. 2003).

The HMG domain is found outside of Metazoa but Sox proteins form a well-supported monophyletic metazoan-specific clade. To date, they have been identified only in bilaterians. The three *Reniera Sox* genes identified here are the first found in a basal metazoan lineage, and indicate that the metazoan LCA genome included representatives of possibly three or more *Sox* families.

Transcription factors possessing the conserved forkhead DNA-binding domain have been found in both animals and fungi but not in plants, suggesting an origin of the class within the opisthokont lineage (Granadino et al. 2000; Carlsson and Mahlapuu 2002). However, a large number of Fox families,

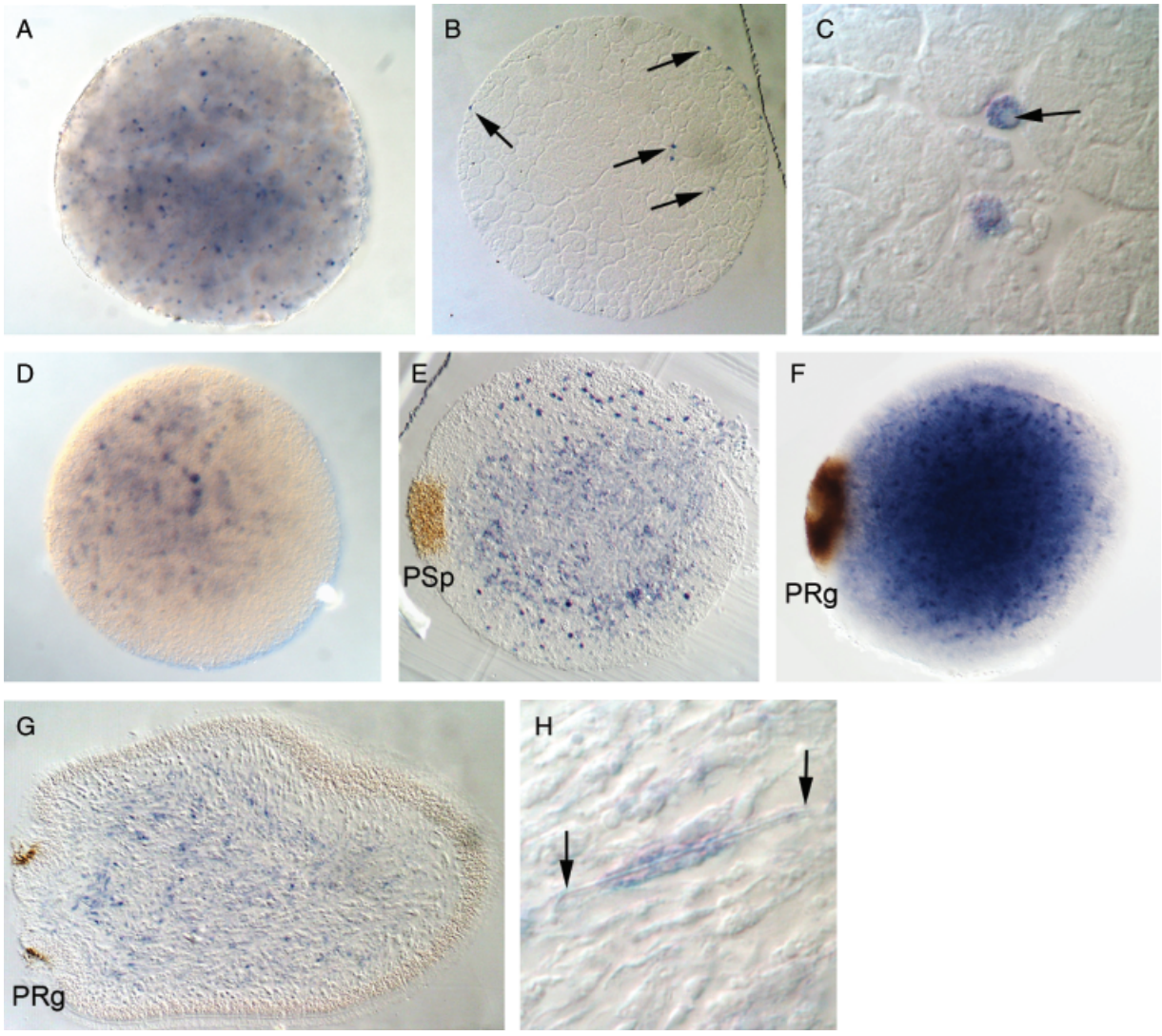

Fig. 14. Expression of *RenBsh* during *Reniera* development. A, D, and F, whole mount micrographs; B, C, E, G, and H, micrographs of sectioned embryos and larvae. (A–D) *RenBsh* is first detected in a subset of micromeres during cleavage. These cells are located both internally and on the surface (arrows in B). A large vesicle is present in many *RenBsh*-expressing cells (arrow in C) as observed in early sclerocytes (Leys and Degnan 2002). (E) During gastrulation [pigment spot (Psp) stage], *RenBsh* transcripts are localized to a subset of cells both in the micromere-rich outer layer and the forming inner cell mass. (F) Later in gastrulation when the pigment ring (PRg) is beginning to form, *RenBsh*-expressing cells are restricted to the inner cell mass, where they remain in the larva (G). (H) A larval sclerocyte expressing *RenBsh*. Arrows point to the end of the spicule.

many of which fall into a large monophyletic clade in this analysis (lower clade in Fig. 8C), appear to lack fungal orthologs (Granadino et al. 2000; Carlsson and Mahlapuu 2002). Of the isolated sponge *Fox* genes, *RenFoxJ* and *Sd-FoxP* seem to belong to the more basal lineages of the *Fox* class, whereas *RenFoxL1* and its apparent ortholog, *Sd-FoxF*, along with *Sd-FoxD* and *Sd-FoxL2*, fall within the putatively metazoan-specific clade. Another sponge gene, *Sd-Fox1*, is of uncertain affinity. Altogether representatives of at least five distinct *Fox* families (*J*, *D*, *L2*, *L1*, or *L1/C*) appear to have been present in the genome of the metazoan LCA. Of these, the *Fox D*, *L2*, *L1*, or *L1/C* may represent metazoan innovations, probably arising by gene duplication early in the metazoan lineage.

Phylogenetic analyses suggest that the *T-box* class, which appears to be metazoan specific, can be divided into eight families (Papaioannou 2001; Takatori et al. 2004; this study). Sponges convincingly possess representatives of two of the eight families. Our analyses provide additional support for the placement of four previously identified sponge *T-box* genes within the *Brachyury* and *Tbx 4/5* families and the presence of these two families in the metazoan LCA (Adell and Muller 2004; Manuel et al. 2004). By contrast, the relationships of the

Reniera T-box genes and of *Tbx1/15/20* from another demosponge to *T-box* genes from other animals are unclear. These may represent demosponge innovations arising from independent duplication and divergence events.

Mef2 genes are metazoan representatives of one of two major divisions (Type II) of MADS box transcription factors, found also in fungi and plants (Alvarez-Buylla et al. 2000; Nam et al. 2003). Although similar within the MADS box domain, the two Type II proteins from yeast do not appear to have a well conserved Mef2 domain (Alvarez-Buylla et al. 2000), suggesting that the defining features of the *Mef2* family evolved within the metazoan lineage. *RenMef2* represents the first *MADS box* gene to be isolated from sponges. The fact that *RenMef2* possesses a well-conserved Mef2 domain (and is convincingly placed at the base of the eumetazoan *Mef2* clade in phylogenetic analyses) indicates that this domain evolved prior to the divergence of poriferan clades and other metazoans.

Expression of *RenHNF4* and *RenBsh* during *Reniera* development

A range of metazoan transcription factors have been isolated previously from a number of demosponge species (Degnan et al. 1993, 1995; Coutinho et al. 1994, 2003; Seimiya et al. 1994, 1997; Hoshiyama et al. 1998; Richelle-Maurer et al. 1998; Adell et al. 2003; Lee et al. 2003; Perovic et al. 2003; Wiens et al. 2003a, b; Adell and Muller 2004). However, their roles in embryogenesis and metamorphosis are unknown. To date, our understanding of sponge gene expression is restricted chiefly to asexual reproductive processes, such as gemmulation, cell aggregation, and primorph formation (e.g., Seimiya et al. 1997; Richelle-Maurer et al. 1998; Adell et al. 2003; Perovic et al. 2003; Wiens et al. 2003a; Adell and Muller 2004; Leys et al. 2005), which have unclear affinities with eumetazoan developmental events. Here, we demonstrate that underlying *Reniera* development is the expression of a suite of transcription factors whose orthologs are involved in a wide range of eumetazoan developmental phenomena.

WMISH analysis of a number of ESTs indicates that underlying the external characteristics of the *Reniera* larva are complex spatial patterns of gene expression. The four ESTs used in this study—encoding ferritin, procollagen lysyl hydroxylase, galectin and B-ZIP1, an uncharacterized leucine zipper transcription factor—display unique patterns of expression, indicating that WMISH is a valid approach to studying the developmental expression of *Reniera* genes. Their expression in subsets of cells within each of the layers suggests that there exists multiple cell types within these layers and that these can be defined initially based on gene expression patterns.

In this study, we have elected to further investigate the expression of two transcription factor genes—*RenHNF4* and *RenBsh*. *HNF4* and its close relative *RXR* have a wide range of developmental and physiological functions in eumetazoans, and their role as transcriptional regulators is contingent upon the binding of a specific lipophilic signal to their ligand binding domains (Auwerx et al. 1999). The restricted expression of *RenHNF4* to ciliated epithelial cells of the larva and not other cells of the outer layer suggests that these cells are competent to respond to an unidentified lipophilic signal. The transdifferentiation of these cells at metamorphosis into choanocytes and possibly other cell types (Leys and Degnan 2002) may be contingent upon *RenHNF4* expression and the presence of this signal.

RenBsh appears to be the ortholog of *BarBsh-Hb*, which is expressed in the demosponge *Halichondria* sp. during cell reaggregation (Hill et al. 2004), and is a member of the *Bar/Bsh* family of ANTP genes. Although bilaterian members of this family largely are involved in nervous system development (e.g., Jones and McGinnis 1993; Patterson et al. 2000; Saba et al. 2003), *Barx1* plays a role in mammalian skeletal patterning (Tucker et al. 1998; Pacifici et al. 2000). *RenBsh* expression in the inner cell mass of the *Reniera* larva suggests that it might have a role in sclerocyte development and spiculogenesis, neither of which have clear homologues in eumetazoans. Following the expression of this gene through development lends support for a role for *RenBsh* in sclerocyte specification. WMISH staining indicates that *RenBsh* transcripts are restricted to a subset of the micromeres located primarily on the outer rim in the early embryo concurs with observations that sclerocytes are one of the first cells to differentiate in *Reniera* (Leys and Degnan 2002; Leys 2003a). However, unlike ultrastructural observations (Leys 2003a), we find that not all *RenBsh*-expressing cells migrate with the pigment cells. This may reflect a continued replenishment of this cell population throughout embryogenesis and larval development.

The differential expression of the other transcription factor genes during *Reniera* embryogenesis and metamorphosis suggests that they are playing different and specific developmental roles, as is the case in other metazoans. Indeed, a number of the *Reniera* transcription factor genes display lineage-restricted patterns of expression in a manner akin to *RenBsh*, albeit in different cells (Larroux and Degnan, unpublished data). It appears that *RenBsh* expression is initiated prior to the start of the differentiation of sclerocytes and other micromere cell lineages (Leys and Degnan 2002), supporting the supposition that micromere specification and differentiation precedes morphogenesis in *Reniera*. Given that the formation and maintenance of the sponge body plan is largely reliant upon cell-level morphogenetic events (Degnan et al. 2005), we propose that the primary function of *RenBsh* and other transcription factor genes is in the specification of the different cell lineages, although some of these genes are likely to be involved also in giving cells positional identities along the embryonic AP axis.

Origin of the metazoan regulatory network

Cell behaviors during demosponge embryogenesis show remarkable similarity to those occurring in eumetazoan development. Analysis of gene expression and cell movements during sponge development suggests that the LCA to all metazoans had the ability to specify and determine the fates of multiple cell types, establish fixed body axes, allocate different cell types along these axes, and form integrated multicellular structures (i.e., simple tissues) using a developmental repertoire remarkably similar to that used by extant eumetazoans.

The presence of a large number of demosponge transcription factor genes suggests that major genome innovation and gene duplication events took place prior to metazoan cladogenesis. Metazoan-specific innovations appear to include the emergence, duplication, and divergence of a number of transcription factor families and classes. A similar situation appears to have occurred with genes encoding signaling pathway components (reviewed in Miyata and Suga 2001). In some gene classes, these early duplication events may have outnumbered those that occurred later around the time of the Cambrian explosion (Miyata and Suga 2001). The metazoan LCA apparently possessed at least four ANTP, one *Pax*, one *POU*, one *LIM-HD*, three *Sox*, one *NR*, five *Fox*, two *T-box*, one *Mef2*, and one *Ets* gene. All but one of these gene classes, *Fox*, appear to have originated in the period between the fungi-metazoan-choanoflagellate LCA and the sponge-eumetazoa LCA around 1 Ba (Nikoh et al. 1997). These genes may have evolved after metazoan and choanoflagellate lineages split; full genome sequencing will help resolve this uncertainty.

Reniera transcription factors are likely to contribute to establishing regionalized patterns of gene expression during development, similar to the roles of their orthologs in other metazoans. From these observations, we infer that underlying embryogenesis in the metazoan LCA was a relatively sophisticated regulatory network that included a range of metazoan-specific transcription factors. We postulate that these genes play instrumental roles in cell specification and morphogenetic processes in sponge development in a homologous manner to their orthologs in bilaterians and eumetazoans (Davidson 2001; Carroll et al. 2005), and, as such, were developmentally important in the metazoan LCA.

The development of all metazoans is reliant upon an inheritable program of gene network interactions in which transcription factors can act as crucial regulatory nodes (Davidson 2001; Carroll et al. 2005). Although multicellularity, per se, is not contingent upon a robust network, developmental programs appear to be. At the origin of animal multicellularity there appears to have been an expansion and diversification of signaling pathways (Miyata and Suga 2001), allowing cells to differentially respond to extrinsic signals, many of which may have been originally derived from an unpredictable external environment (Wolpert 1994). Along this evolutionary lineage eventually evolved metazoan gametogenesis and embryogenesis, which entrained intercellular interactions into a robust and encoded network. Our data from *Reniera* suggest that the origin of the metazoan body plan—with its requisite embryogenesis and metamorphosis—included the establishment of a developmental regulatory repertoire that consisted of many of the same signalling and transcription factor components employed by all living animals. The origin of new transcription factor families and their co-option into a multicellular system where cells already responded differentially to external signals may have paved the way for the evolution of robust networks that appear to underlie the predictable cell decisions and behaviors that occur not only during embryogenesis but throughout the metazoan life cycle.

Acknowledgments

We thank R. Raff, S. Degnan, and an anonymous reviewer for critical comments, which greatly improved the original manuscript. This research was supported by Australian Research Council grants to BMD and SPL. D. L. was supported by AAEF and the Fulbright Association. Part of the phylogenetic analysis was undertaken during C. L.'s attendance at the Workshop on Molecular Evolution at the MBL Woods Hole.

REFERENCES

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

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

Aguinaldo, A. M. A., et al. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387: 489–493.

Alvarez-Buylla, E. R., et al. 2000. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. USA* 97: 5328–5333.

Auwerx, J., et al. 1999. A unified nomenclature system for the nuclear receptor superfamily. *Cell* 97: 161–163.

Banerjee-Basu, S., and Baxevanis, A. D. 2001. Molecular evolution of the homeodoamin family of transcription factors. *Nucleic Acids Res.* 29: 3258–3269.

Borchiellini, C., Manuel, M., Alivon, E., Boury-Esnault, N., Vacelet, J., and Le Parco, Y. 2001. Sponge paraphyly and the origin of Metazoa. *J. Evol. Biol.* 14: 171–179.

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

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

Carroll, S. B., Grenier, J. K., and Weatherbee, S. D. 2005. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. 2nd Ed. Blackwell Science, Malden.

Castro, L. F. C., and Holland, P. W. H. 2003. Chromosomal mapping of ANTP class homeobox genes in amphioxus: piecing together ancestral genomes. *Evol. Dev.* 5: 459–465.

Cavalier-Smith, T., and Chao, E. E. Y. 2003. Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote megaevolution. *J. Mol. Evol.* 56: 540–563.

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

Coutinho, C. C., Fonseca, R. N., Mansure, J. J. C., and Borojevic, R. 2003. Early steps in the evolution of multicellularity: deep structural and functional homologies among homeobox genes in sponges and higher metazoans. *Mech. Dev.* 120: 429–440.

Coutinho, C. C., Vissers, S., and Van de Vyver, G. 1994. Evidence of homeobox gene in the freshwater sponge *Ephydatia fluviatilis*. In R. W. M. Van Soest, T. M. G. Van Kempen, and J. C. Braekman (eds.). *Sponges in Time and Space*. Balkema Press, Rotterdam, pp. 385–388.

Davidson, E. H. 2001. *Genomic Regulatory Systems. Development and Evolution*. Academic Press, San Diego.

Degnan, B. M., Degnan, S. M., Giusti, A. F., and Morse, D. E. 1995. A *hom/hox* gene in sponges. *Gene* 155: 175–177.

Degnan, B. M., Degnan, S. M., Naganuma, T., and Morse, D. E. 1993. The *ets* multigene family is conserved throughout the Metazoa. *Nucleic Acids Res.* 21: 3479–3484.

Degnan, B. M., Leys, S. P., and Larroux, C. 2005. Sponge development and antiquity of animal pattern formation. *Integr. Comp. Biol.* 45: 335–341.

Degnan, B. M., and Morse, D. E. 1993. Identification of eight homeobox-containing transcripts expressed during larval development and at metamorphosis in the gastropod mollusc *Haliotis rufescens*. *Mol. Mar. Biol. Biotechnol.* 2: 1–9.

Degnan, B. M., Rohde, P. R., and Lavin, M. F. 1996. Normal development and embryonic gene activity of the ascidian *Herdmania momus*. *Mar. Freshw. Res.* 47: 543–551.

Devine, C., Hinman, V. F., and Degnan, B. M. 2002. Evolution and developmental expression of nuclear receptor genes in the ascidian *Herdmania curvata*. *Int. J. Dev. Biol.* 46: 687–692.

Felsenstein, J. 2003. *PHYLIP (Phylogeny Inference Package)*. Distributed by the author, Department of Genetics, University of Washington, Seattle.

Finnerty, J. R., and Martindale, M. Q. 1997. Homeoboxes in sea anemones (Cnidaria; Anthozoa): a PCR-based survey of *Nematostella vectensis* and *Metridium senile*. *Biol. Bull.* 193: 62–76.

Garcia-Fernandez, J. 2005. Hox, parahox, protohox: facts and guesses. *Heredity* 94: 145–152.

Gauchat, D., et al. 2000. Evolution of Antp-class genes and differential expression of *Hydra Hox/parahox* genes in anterior patterning. *Proc. Natl. Acad. Sci. USA* 97: 4493–4498.

Granadino, B., Perez-Sanchez, C., and Rey-Campos, J. 2000. Forkhead transcription factors. *Curr. Genomics* 1: 353–382.

Hill, A., Tetrault, J., and Hill, M. 2004. Isolation and expression analysis of a poriferan Antp-class Bar-/Bsh-like homeobox gene. *Dev. Genes Evol.* 214: 515–523.

Hinman, V. F., and Degnan, B. M. 2000. Retinoic acid perturbs *Otx* gene expression in the ascidian pharynx. *Dev. Genes Evol.* 210: 129–139.

Hobert, O., and Westphal, H. 2000. Functions of LIM-homeobox genes. *Trends Genet.* 16: 75–83.

Hoshiyama, D., et al. 1998. Sponge *Pax* cdna related to *Pax-2/5/8* and ancient gene duplications in the *Pax* family. *J. Mol. Evol.* 47: 640–648.

Hyman, L. H. 1940. *The Invertebrates*. McGraw-Hill, New York.

Jackson, D., Leys, S. P., Hinman, V. F., Woods, R., Lavin, M. F., and Degnan, B. M. 2002. Ecological regulation of development: induction of marine invertebrate metamorphosis. *Int. J. Dev. Biol.* 46: 679–686.

Jones, B., and McGinnis, W. 1993. A new Drosophila homeobox gene, *bsh*, is expressed in a subset of brain cells during embryogenesis. *Development* 117: 793–806.

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

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

Kruse, M., Leys, S. P., Muller, I. M., and Muller, W. E. G. 1998. Phylogenetic position of the Hexactinellida within the phylum Porifera based on the amino acid sequence of the protein kinase C from *Rhabdocalyptus dawsoni*. *J. Mol. Evol.* 46: 721–728.

Lee, S. E., Gates, R. D., and Jacobs, D. K. 2003. Gene fishing: the use of a simple protocol to isolate multiple homeodomain classes from diverse invertebrate taxa. *J. Mol. Evol.* 56: 509–516.

Leys, S. P. 2004. Gastrulation in sponges. In C. Stern (ed.). *Gastrulation: from Cells to Embryos*. Cold Spring Harbor Laboratory Press, New York, pp. 23–31.

Leys, S. P. 2003a. Comparative study of spiculogenesis in demosponge and hexactinellid larvae. *Microsc. Res. Tech.* 62: 300–311.

Leys, S. P. 2003b. The significance of syncytial tissues for the position of the Hexactinellida in the Metazoa. *Integr. Comp. Biol.* 43: 19.

Leys, S. P., Cronin, T. W., Degnan, B. M., and Marshall, J. N. 2002. Spectral sensitivity in a sponge larva. *J. Comp. Physiol. A Sens. Neural Behav. Physiol.* 188: 199–202.

Leys, S. P., and Degnan, B. M. 2001. Cytological basis of photoresponsive behavior in a sponge larva. *Biol. Bull.* 201: 323–338.

Leys, S. P., and Degnan, B. M. 2002. Embryogenesis and metamorphosis in a haplosclerid demosponge: gastrulation and transdifferentiation of larval ciliated cells to choanocytes. *Invertebr. Biol.* 121: 171–189.

Leys, S. P., Rokhsar, D. S., and Degnan, B. M. 2005. Quick guide—sponges. *Curr. Biol.* 15: R114–R115.

Maldonado, M., Durfort, M., Mc Carthy, D. A., and Young, C. M. 2003. The cellular basis of photobehavior in the tufted parenchymella larva of demosponges. *Mar. Biol.* 143: 427–441.

Maldonado, M., George, S. B., Young, C. M., and Vaquerizo, I. 1997. Depth regulation in parenchymella larvae of a demosponge: relative roles of skeletogenesis, biochemical changes and behavior. *Mar. Ecol. Prog. Ser.* 148: 115–124.

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

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

Martinez, D. E., Dirksen, M. L., Bode, P. M., Jamrich, M., Steele, R. E., and Bode, H. R. 1997. Budhead, a fork head HNF-3 homologue, is expressed during axis formation and head specification in hydra. *Dev. Biol.* 192: 523–536.

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

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

Miller, D. J., et al. 2000. Pax gene diversity in the basal cnidarian *Acropora millepora* (Cnidaria, Anthozoa): implications for the evolution of the Pax gene family. *Proc. Natl. Acad. Sci. USA* 97: 4475–4480.

Miyata, T., and Suga, H. 2001. Divergence pattern of animal gene families and relationship with the Cambrian explosion. *BioEssays* 23: 1018–1027.

Muller, W. E. G., Muller, I. M., Lukic, L., Cetkovic, H., and Gamulin, V. 1998. Monophyly of metazoa: sponges as living fossils. *Period Biol.* 100: 373–381.

Nam, J., de Pamphilis, C. W., Ma, H., and Nei, M. 2003. Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol. Biol. Evol.* 20: 1435–1447.

Nikoh, N., et al. 1997. An estimate of divergence time of parazoa and eumetazoa and that of cephalochordata and vertebrata by aldolase and triose phosphate isomerase clocks. *J. Mol. Evol.* 45: 97–106.

O'Brien, E. K., and Degnan, B. M. 2000. Expression of POU, Sox, and Pax genes in the brain ganglia of the tropical abalone *Haliotis asinina*. *Mar. Biotechnol.* 2: 545–557.

Pacifici, M., Koyama, E., Iwamoto, M., and Gentili, C. 2000. Development of articular cartilage: what do we know about it and how may it occur? *Connect. Tissue Res.* 41: 175–184.

Papaioannou, V. E. 2001. T-box genes in development: from hydra to humans. *Int. Rev. Cytol.* 207: 1–70.

Patterson, K. D., Cleaver, O., Gerber, W. V., White, F. G., and Krieg, P. A. 2000. Distinct expression patterns for two Xenopus Bar homeobox genes. *Dev. Genes Evol.* 210: 140–144.

Perovic, S., et al. 2003. Expression of one sponge Iroquois homeobox gene in primmorphs from *Suberites domuncula* during canal formation. *Evol. Dev.* 5: 240–250.

Rambaut, A. 1996. Se–Al: Sequence Alignment Editor. Available at http://evolve.zoo.ox.ac.uk

Richelle-Maurer, E., Van de Vyver, G., Vissers, S., and Coutinho, C. C. 1998. Homeobox-containing genes in freshwater sponges: characterization, expression, and phylogeny. In W. E. G. Muller (ed.). *Progress in Molecular Subcell Biology*. Springer-Verlag, Berlin, Heidelberg, pp. 158–175.

Rodrigo, A. G., Bergquist, P. R., and Bergquist, P. L. 1994. Inadequate support for an evolutionary link between the Metazoa and the Fungi. *Syst. Biol.* 43: 578–584.

Ronquist, F., and Huelsenbeck, J. P. 2003. Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

de Rosa, R., et al. 1999. Hox genes in brachiopods and priapulids and protostome evolution. *Nature* 399: 772–776.

Saba, R., Nakatsuji, N., and Saito, T. 2003. Mammalian barh1 confers commissural neuron identity on dorsal cells in the spinal cord. *J. Neurosci.* 23: 1987–1991.

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

Seimiya, M., Watanabe, Y., and Kurosawa, Y. 1997. Identification of POU-class homeobox genes in a freshwater sponge and the specific expression of these genes during differentiation. *Eur. J. Biochem.* 243: 27–31.

Sharrocks, A. D. 2001. The Ets-domain transcription factor family. *Nat. Rev. Mol. Cell Biol.* 2: 827–837.

Spring, J., Yanze, N., Josch, C., Middel, A. M., Winninger, B., and Schmid, V. 2002. Conservation of *Brachyury*, *Mef2*, and *Snail* in the myogenic lineage of jellyfish: a connection to the mesoderm of bilateria. *Dev. Biol.* 244: 372–384.

Sullivan, J., Abdo, Z., Joyce, P., and Swofford, D. L. 2005. Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation. *Mol. Biol. Evol.* 22: 1386–1392.

Sun, H. M., et al. 1997. Evolution of paired domains: isolation and sequencing of jellyfish and hydra Pax genes related to Pax-5 and Pax-6. *Proc. Natl. Acad. Sci. USA* 94: 5156–5161.

Takatori, N., et al. 2004. T-box genes in the ascidian *Ciona intestinalis*: characterization of cDNAs and spatial expression. *Dev. Dyn.* 230: 743–753.

Technau, U., and Bode, H. R. 1999. Hybra1, a Brachyury homologue, acts during head formation in Hydra. *Development* 126: 999–1010.

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

Tucker, A. S., Matthews, K. L., and Sharpe, P. T. 1998. Transformation of tooth type induced by inhibition of BMP signaling. *Science* 282: 1136–1138.

Wiens, M., Batel, R., Korzhev, M., and Muller, W. E. G. 2003a. Retinoid X receptor and retinoic acid response in the marine sponge *Suberites domuncula*. *J. Exp. Biol.* 206: 3261–3271.

Wiens, M., et al. 2003b. The molecular basis for the evolution of the metazoan bodyplan: extracellular matrix-mediated morphogenesis in marine demosponges. *J. Mol. Evol.* 57: S60–S75.

Wolpert, L. 1994. The evolutionary origin of development—cycles, patterning, privilege and continuity. *Development* (suppl. S): 79–84.

Yang, Z. 1997. Paml: a program for package for phylogenetic analysis by maximum likelihood. *CABIOS* 15: 555–556.