

Genesis and Expansion of Metazoan Transcription Factor Gene Classes

Claire Larroux,* Graham N. Luke,† Peter Koopman,‡ Daniel S. Rokhsar,§|| Sebastian M. Shimeld,¶ and Bernard M. Degnan*

*School of Integrative Biology, The University of Queensland, Brisbane, Queensland, Australia; †School of Biological Sciences, The University of Reading, Whiteknights, Reading, United Kingdom; ‡Institute of Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia; §US Department of Energy Joint Genome Institute, Walnut Creek, CA; ||Department of Molecular and Cell Biology, Center for Integrative Genomics, University of California, Berkeley; and ¶Department of Zoology, University of Oxford, Oxford, UK

We know little about the genomic events that led to the advent of a multicellular grade of organization in animals, one of the most dramatic transitions in evolution. Metazoan multicellularity is correlated with the evolution of embryogenesis, which presumably was underpinned by a gene regulatory network reliant on the differential activation of signaling pathways and transcription factors. Many transcription factor genes that play critical roles in bilaterian development largely appear to have evolved before the divergence of cnidarian and bilaterian lineages. In contrast, sponges seem to have a more limited suite of transcription factors, suggesting that the developmental regulatory gene repertoire changed markedly during early metazoan evolution. Using whole-genome information from the sponge *Amphimedon queenslandica*, a range of eumetazoans, and the choanoflagellate *Monosiga brevicollis*, we investigate the genesis and expansion of homeobox, Sox, T-box, and Fox transcription factor genes. Comparative analyses reveal that novel transcription factor domains (such as Paired, POU, and T-box) arose very early in metazoan evolution, prior to the separation of extant metazoan phyla but after the divergence of choanoflagellate and metazoan lineages. Phylogenetic analyses indicate that transcription factor classes then gradually expanded at the base of Metazoa before the bilaterian radiation, with each class following a different evolutionary trajectory. Based on the limited number of transcription factors in the *Amphimedon* genome, we infer that the genome of the metazoan last common ancestor included fewer gene members in each class than are present in extant eumetazoans. Transcription factor orthologues present in sponge, cnidarian, and bilaterian genomes may represent part of the core metazoan regulatory network underlying the origin of animal development and multicellularity.

Introduction

Within the Opisthokonta are 2 true multicellular lineages—the Metazoa and the Fungi (e.g., Cavalier-Smith and Chao 2003; Steenkamp et al. 2006). Metazoans form a clade within the Opisthokonta—the Holozoa—with their apparent sister group, the choanoflagellates, which include many colonial species, and a number of other unicellular lineages (e.g., Cavalier-Smith et al. 1996; King and Carroll 2001; Snell et al. 2001; Lang et al. 2002; Burger et al. 2003; Cavalier-Smith and Chao 2003; King et al. 2003; Steenkamp et al. 2006). Although these phylogenies suggest metazoans evolved from a unicellular ancestor, we know little about the genomic events that led to the evolution of animal multicellularity. Choanoflagellates express homologues of genes involved in cell communication and adhesion in animals, indicating that some of the molecular prerequisites for multicellularity predate the origin of the animal kingdom (King and Carroll 2001; King et al. 2003; King 2004). Within the Metazoa, there are 2 ancient lineages of extant animals (e.g., Cavalier-Smith et al. 1996; Borchiellini et al. 2001; Medina et al. 2001; Collins 2002; Wallberg et al. 2004). One lineage consists of a tremendously diverse range of the body plans—the Eumetazoa (ctenophores, cnidarians, and bilaterian phyla)—and the other a single, simple aquiferous-like body plan—phylum Porifera (sponges) (Brusca RC and Brusca GJ 2003). Unlike most eumetazoans, the sponge body plan appears to have remained relatively unchanged since well before

the Cambrian (Li et al. 1998), with extant species lacking true muscle and nerve cells, integrated tissue, and organ systems (Simpson 1984).

The disparate evolutionary outcomes of choanoflagellate, sponge, and eumetazoan lineages imply that there exist inherent and long-standing genetic differences to which extant forms are contingent upon. With the recent sequencing of the genomes of representatives of these clades—the choanoflagellate *Monosiga brevicollis*, the sponge *Amphimedon queenslandica*, and the cnidarians *Nematostella vectensis* (Putnam et al. 2007) and *Hydra magnipapillata*—we have begun to gain insights into these differences and the early history of the metazoan genome. Embryonic development is the metazoan synapomorphy underlying the primary generation of differentiated cell types and their patterning into larval and adult body plans (Wolpert 1994; Degnan et al. 2005; Degnan SM and Degnan BM 2006). Underpinning embryonic cell behaviors and specification events is a gene regulatory network consisting of conserved signaling pathways and transcription factors (Davidson and Erwin 2006). Although the causal link between the evolution of a developmental regulatory network and the evolution of metazoan embryogenesis is currently a mystery, comparison of choanoflagellate and early-branching metazoan genomes can provide a general indicator of the relative size and complexity of these networks in key ancestors. Transcription factors act as regulatory nodes in developmental specification and patterning events and their expansion in very early metazoan evolution may have provided the necessary preadaptive conditions for the evolution of metazoan embryogenesis and multicellularity (Larroux et al. 2006, 2007).

Cnidarians appear to be the sister group to the Bilateria (e.g., Cavalier-Smith et al. 1996; Borchiellini et al. 2001; Medina et al. 2001; Collins 2002; Wallberg et al. 2004).

Key words: developmental genes, *Amphimedon queenslandica*, sponge, homeodomain, Sox, Fox, T-box.

E-mail: b.degnan@uq.edu.au.

Mol. Biol. Evol. 25(5):980–996, 2008

doi:10.1093/molbev/msn047

Advance Access publication February 21, 2008

Although they were traditionally thought to have a radial diploblastic grade of body plan organization, anatomical and gene expression data suggest that they may instead be descended from a bilateral triploblastic ancestor (reviewed in Martindale 2005). Comparative analyses of cnidarian and bilaterian genomes have revealed that transcription factor and signaling ligand gene classes were generally of a similar size in the cnidarian–bilaterian and the protostome–deuterostome last common ancestors (LCAs) (Kusserow et al. 2005; Magie et al. 2005; Miller et al. 2005; Chourrout et al. 2006; Kamm et al. 2006; Ryan et al. 2006; Putnam et al. 2007; Yamada et al. 2007), indicating that much of the gene repertoire necessary for complex bilaterian development had evolved before the divergence of cnidarian and bilaterian lineages. Limited evidence to date from whole-genome information suggests that sponges may have a narrower suite of transcription factors (Larroux et al. 2007; Simionato et al. 2007).

To gain further understanding of the role of transcription factor gene evolution in the evolution of metazoan development, we analyzed a number of major transcription factor gene groups that have essential developmental roles in bilaterians, specifically homeobox, Sox, T-box, and Fox/forkhead genes (POU reviewed in Ryan and Rosenfeld 1997; LIM-HD reviewed in Hobert and Westphal 2000; homeobox genes reviewed in Banerjee-Basu and Baxevasis 2001; T-box reviewed in Papaioannou 2001; Fox reviewed in Carlsson and Mahlapuu 2002; Pax reviewed in Chi and Epstein 2002; and Sox reviewed in Schepers et al. 2002). Through comparative genomic and phylogenetic analyses, we trace the genesis and expansion of these genes using publicly available whole-genome information from the demosponge *Amphimedon*, a range of eumetazoans, the choanoflagellate *Monosiga*, and fungi. As sponges are considered one of, if not the earliest, branching lineages of modern metazoans (e.g., Cavalier-Smith et al. 1996; Medina et al. 2001; Wallberg et al. 2004), genes that are present in the *Amphimedon*, cnidarian, and bilaterian genomes are likely to have been present in the LCA to all metazoans. Comparison of the constituency of the reconstructed metazoan LCA genome with that of fungi and choanoflagellate genomes allows us to place the origin of metazoan innovations (Ruiz-Trillo et al. 2007). This knowledge can be used to identify the canonical features of the metazoan genome and their role in the evolution of multicellularity and development.

Materials and Methods

TBlastN using conserved domains from each known family within each studied class was performed on *A. queenslandica* genome traces (available at ftp://ftp.ncbi.nih.gov/pub/TraceDB/reniera_sp__jgi-2005/) and developmentally expressed sequence tags (ESTs, sequenced at the Joint Genome Institute). Sponge genes were classified, and selected traces were assembled using an in-house assembly pipeline as described in Larroux et al. (2007). A primary genome assembly from the Joint Genome Institute was later consulted. Putative genomic structure of genes for which no cDNA sequences were available was determined with

GeneScan or by visually searching for open reading frames and aligning translations of conserved regions with related genes from other taxa. Similarly, sequences were obtained from genome traces of *N. vectensis* (available at http://genome.jgi-psf.org/Nemve1/), *Lottia gigantea* (available at ftp://ftp.ncbi.nih.gov/pub/TraceDB/lottia_gigantea), and *M. brevicollis* (available at ftp://ftp.ncbi.nih.gov/pub/TraceDB/monosiga_brevicollis). We did not aim to obtain the full complement of cnidarian genes but, if no cnidarian or poriferan gene was identified from a specific family, a thorough search for a *Nematostella* representative was undertaken. We searched for *Lottia* (and *Drosophila melanogaster*) genes when the presence or absence of a specific family in this taxon could shed light on the origin of the family.

Bayesian, distance, and maximum likelihood (ML) analyses were undertaken on conserved regions using representatives from all families of each class. Bayesian analyses were performed using MrBayes 3.1 (Ronquist and Huelsenbeck 2003) with the likelihood model set to invgamma (invariant sites + gamma distribution) among-site rate variation and the amino acid substitution model prior set to the Jones-Taylor-Thornton (JTT) fixed model. A set of 4 independent simultaneous Metropolis-coupled Markov Chains Monte Carlo was sampled every 100th generation, and a burn-in of 1,300 trees was removed (except for the Fox class where the burn-in was 8,000 trees). Convergence of each run was assessed by plotting the log likelihood against the number of generations. Two Bayesian analyses were run for each data set, 1 for 1 million generations and another for 10 million generations, except the Fox class for which analyses were run twice for 20 million generations. Convergence between the 2 separate analyses was assessed by comparing tree topology and posterior probabilities (PPs). Distance Neighbor-Joining (NJ) analyses with 1,000 bootstraps were performed using the Unix PHYLIP 3.6 package (Felsenstein 2003). For ML analyses, the JTT amino acid substitution matrix was used and, for each data set, the among-site rate variation likelihood model was determined using ProtTest 1.4 (Abascal et al. 2005) selecting between 1) no among-site rate variation, 2) invariant sites, 3) gamma distribution, and 4) invariant sites and gamma distribution. For analyses comprising less than 50 taxa (Pax, POU, LIM-HD, Six, TALE, reduced Sox data set), proml with 100 bootstraps from the Unix PHYLIP 3.6 package (Felsenstein 2003) with the selected likelihood model was used. Estimates of the parameters (proportion of invariant sites and/or gamma) were initially obtained with Codeml (set to the JTT amino acid substitution matrix) from the PAML package (Yang 1997) using an alignment and a proml tree as input files. These estimates were considered good approximations as there is little effect of tree topology on parameter estimation (Sullivan et al. 2005). For analyses comprising 50 or more taxa (paired [prd]-like, large Sox data set, T-box, and Fox), ML analyses with 100 bootstraps were undertaken using the PHYML online Web server (Guindon et al. 2005) with the likelihood model selected in ProtTest. With the exception of the prd-like and TALE trees, trees were not rooted as preliminary analyses suggested that the possible outgroups were too divergent.

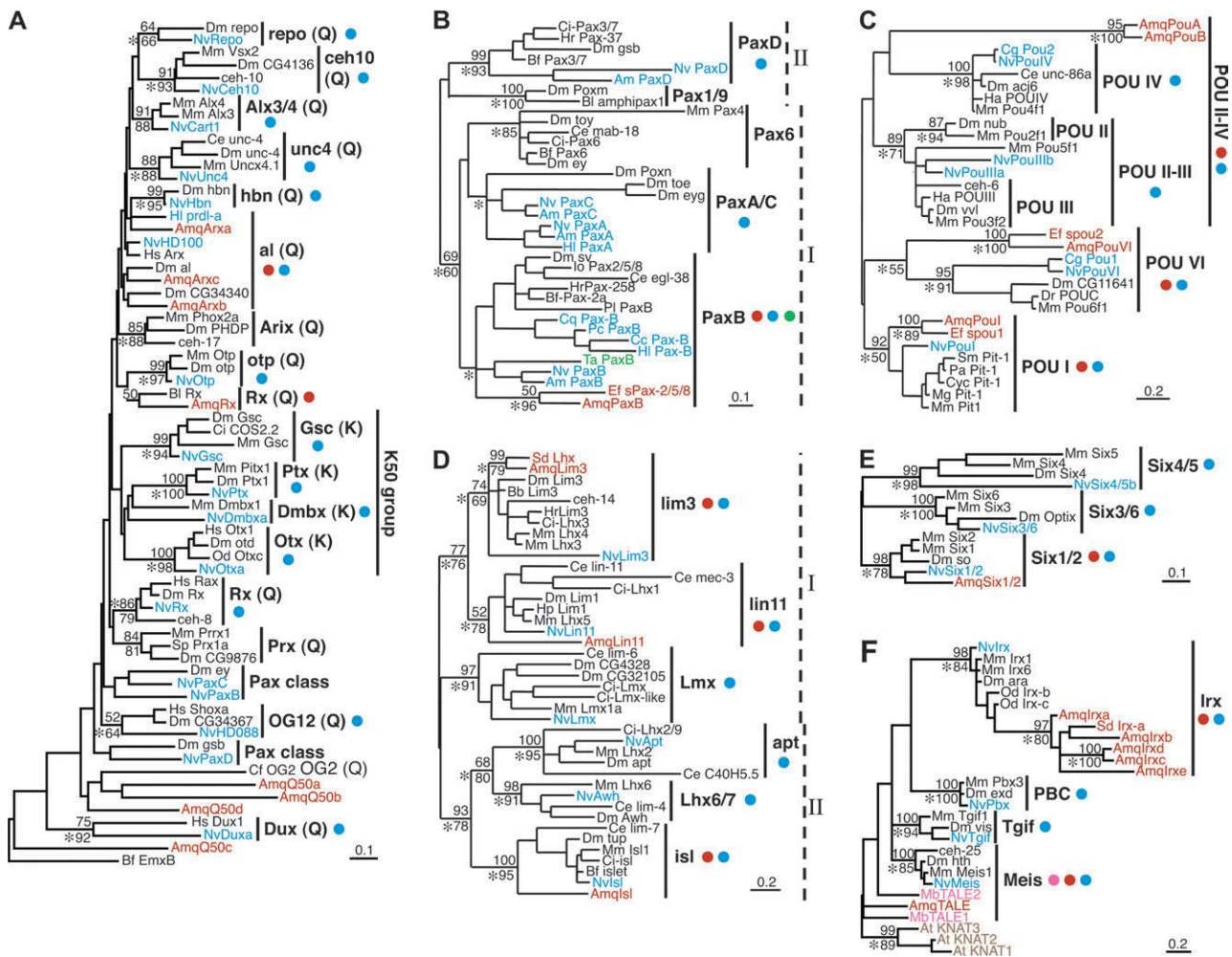


FIG. 2.—Phylogenetic trees of (A) prd-like, (B) Pax, (C) POU, (D) LIM-HD, (E) Six, and (F) TALE homeobox gene classes. Bayesian, distance NJ, and ML analyses were undertaken. Bayesian trees are shown for all but the prd-like class where the NJ tree is shown. At key nodes are given percentages of bootstrap support obtained by distance (NJ; 1 000 replicates) above the branch and by ML (100 replicates) below the branch. Bootstrap values above 50% are shown. An asterisk indicates a Bayesian PP greater than or equal to 95%. Pax, POU, LIM-HD, and Six trees are unrooted, whereas the prd-like tree is rooted with an ANTP gene, *EmxB*, and the TALE tree is rooted with plant *KNAT* genes. Families and higher level groupings are shown on the right of the tree. Sponge genes are in red, cnidarian genes in blue, and placozoan genes in green. Plant TALE genes are in brown. Clades containing representatives of these taxa are indicated by a circle of the corresponding color. Abbreviations: Am: *Acropora millepora*, anthozoan cnidarian; Amq: *Amphimedon queenslandica*, demosponge; At: *Arabidopsis thaliana*, plant; Bb: *Branchiostoma belcheri*, cephalochordate; Bf: *Branchiostoma floridae*, cephalochordate; Bl: *Branchiostoma lanceolatum*, cephalochordate; Cc: *Cladonema californicum*, hydrozoan cnidarian; Ce: *Caenorhabditis elegans*, nematode; Cf: *Canis familiaris*, vertebrate; Cg: *Condylactis gigantea*, anthozoan cnidarian; Ci: *Ciona intestinalis*, urochordate; Cq: *Chrysaora quinquecirrha*, scyphozoan cnidarian; Cyc: *Cyprinus carpio*, vertebrate; Dm: *Drosophila melanogaster*, insect; Dr: *Danio rerio*, vertebrate; Ef: *Ephydatia fluviatilis*, demosponge; Ha: *Haliotis asinina*, mollusk; Hl: *Hydra littoralis*, hydrozoan cnidarian; Hp: *Hemicentrotus pulcherrimus*, echinoderm; Hr: *Halocynthia roretzi*, urochordate; Hs: *Homo sapiens*, human; Io: *Ilyanassa obsoleta*, mollusk; Mb: *Monosiga brevicollis*, choanoflagellate protist; Mg: *Meleagris gallopavo*, vertebrate; Mm: *Mus musculus*, vertebrate; Nv: *Nematostella vectensis*, anthozoan cnidarian; Od: *Oikopleura dioica*, urochordate; Pa: *Plecoglossus altivelis*, vertebrate; Pc: *Podocoryne carnea*, hydrozoan cnidarian; Pl: *Paracentrotus lividus*, echinoderm; Sd: *Suberites domuncula*, demosponge; Sm: *Scophthalmus maximus*, vertebrate; Sp: *Strongylocentrotus purpuratus*, echinoderm; and Ta: *Trichoplax adhaerens*, placozoan.

domains, with Bayesian, distance NJ, and ML phylogenetic analyses (fig. 2).

The Prd-Like Genes

The Prd-like genes possess a prd-like HD but not a paired domain, found in Pax genes, and are classified into 2 groups based on whether they have a glutamine (Q50 group) or lysine (K50 group) at position 50 of their HD (Galliot et al. 1999). There is no evidence of this class

of homeobox genes existing outside the Metazoa (Burglin 2005; this study). *Amphimedon queenslandica* has 8 Q50 prd-like genes and no K50 genes (fig. 1; table 1). Six of these genes encode at least 5 of the 6 diagnostic amino acids in the prd HD (P26, D27, E32, R44, Q46, and A54) as in the case with most prd proteins; the 2 other genes only have 2–4 conserved codons (fig. 1; Galliot et al. 1999). Additionally, all *Amphimedon* prd-like genes have a conserved intron between codons 46 and 47 in their homeobox (fig. 1), as observed in many other prd genes (Miller et al. 2000; Burglin

Table 1
Transcription Factor Numbers in *Amphimedon* and Inferred Minimal Transcription Factor Content in the Demosponge–Eumetazoan LCA, the Cnidarian–Bilaterian LCA, and the PDA

Gene Number	<i>Amphimedon queenslandica</i>	Demosponge–Eumetazoan LCA	Cnidarian–Bilaterian LCA	PDA
Homeobox	31	17–20	61–62	82
ANTP ^a	8	6–7	26	36
prd-like	8	2–3	14	16
Pax	1	1	2–3	5
POU	4	2–3	4	5
LIM-HD	3	3	6	6
Cut	0	0	1 ^b	3 ^c
Pros	0	0	0 ^b	1 ^c
ZF-HD	0	0	0 ^b	2 ^c
HNF	0	0	1 ^b	1 ^c
Six	1	1	3	3
TALE ^d	6	2	4	4
SOX	4	2–4	3–5	6
T-box	7	3	6	8
Fox ^d	16	10–11	17–18	19
Fox group I	7	3–4	9–10	11
Fox group II ^d	9	7	8	8

^a From Larroux et al. (2007).

^b From Ryan et al. (2006).

^c From Burglin (2005).

^d Found in fungi and choanoflagellates.

2005; Ryan et al. 2006). Amino acid alignments suggest half of the *Amphimedon* prd-like genes are affiliated with specific eumetazoan families (supplementary fig. 1, Supplementary Material online). However, phylogenetic analyses cannot resolve their relationships clearly. The Bayesian tree results in polytomy at the node between most families, so the distance NJ tree is shown in figure 2A. It is rooted with an ANTP gene and an array of Pax genes are included in the analysis.

Three *Amphimedon* prd-like genes are more similar to eumetazoan *aristalless* (*al*) genes than to other prd-like genes and may be paralogues (supplementary fig. 1, Supplementary Material online). These genes—named *AmqArxa*, *AmqArxb*, and *AmqArxc*—show high sequence similarity to one another but are more similar to bilaterian *al* genes. In the Bayesian tree, these genes as well as each eumetazoan *al* gene are in unresolved positions. However, 2 of these genes are grouped with some eumetazoan *al* genes in the NJ tree, whereas all *al* and *Arix* genes form a monophyletic clade in the ML tree, suggesting that the 3 sponge genes are indeed *al* genes (fig. 2A). Another *Amphimedon* prd-like gene, *AmqRx*, shows higher sequence similarity to members of the Rx family than to other prd-like families (supplementary fig. 1, Supplementary Material online). *AmqRx* is a poorly supported sister taxon to a lancelet *Rx* in phylogenetic analyses, whereas other eumetazoan *Rx* genes belong to a separate clade (fig. 2A). There are 4 prd-like genes of uncertain relatedness during sequence alignment and named *AmqQ50a* to *d* (supplementary fig. 1, Supplementary Material online). *AmqQ50a* and *c* have fewer prd diagnostic codon positions than other prd genes, suggesting that they are relatively divergent. *AmqQ50d* is marginally more similar to genes belonging to the *al* family and *AmqQ50b* to genes belonging to the OG12 and OG2 families. In the ML and NJ trees, *AmqQ50c*, *Dux* genes and a clade of *OG2*, *AmqQ50a*, *b*,

and *d* are in a basal position to the rest of the prd genes, whereas in the Bayesian tree, *AmqQ50c* and *Dux* genes are in such a position and *AmqQ50a*, *b*, and *d* are grouped with *OG2* but elsewhere in the tree (fig. 2A). These 4 *Amphimedon* genes may represent descendants of an early metazoan *OG2*-like gene. However, all these genes have long branches and may be clustered together and alongside the outgroup due to long-branch attraction (Bergsten 2005).

All *Amphimedon* prd-like genes have a glutamine at position 50 of their HD, suggesting that the proto-prd-like gene was a Q50 gene (Galliot et al. 1999). Additionally, K50 genes form a monophyletic clade nested among Q50 genes in all 3 analyses implying that the proto-K50 gene arose after the divergence of sponge and eumetazoan lineages (fig. 2A). Although they are not monophyletic in the NJ or ML trees, genes belonging to the Pax class are monophyletic in the Bayesian tree (PP: 0.95). Analysis of the *Amphimedon* genome suggests that the metazoan LCA may have had at least 3 ancestral prd-like genes, *al*-like, *Rx*-like, and *OG2*-like (table 1; fig. 3A). In this analysis, 14 families containing *Nematostella* representatives are generally well supported (1 extra family to that found in Ryan et al. 2006) and 2 families, Prx and Arix, are only present in bilaterians (fig. 2A). Hence, it appears that the cnidarian–bilaterian ancestor had a minimum of 14 prd-like genes and the PDA a minimum of 16 genes, representing extensive subsequent diversification (table 1; fig. 3A).

Pax Genes

A single *PaxB* gene containing a prd domain upstream of a homeobox, renamed here *AmqPaxB* (genes were formerly prefixed with *Ren* based on the previous name of this sponge, *Reniera* sp.), has been previously characterized by reverse transcriptase–polymerase chain reaction (RT-PCR) (Larroux et al. 2006). It appears to be the only Pax gene

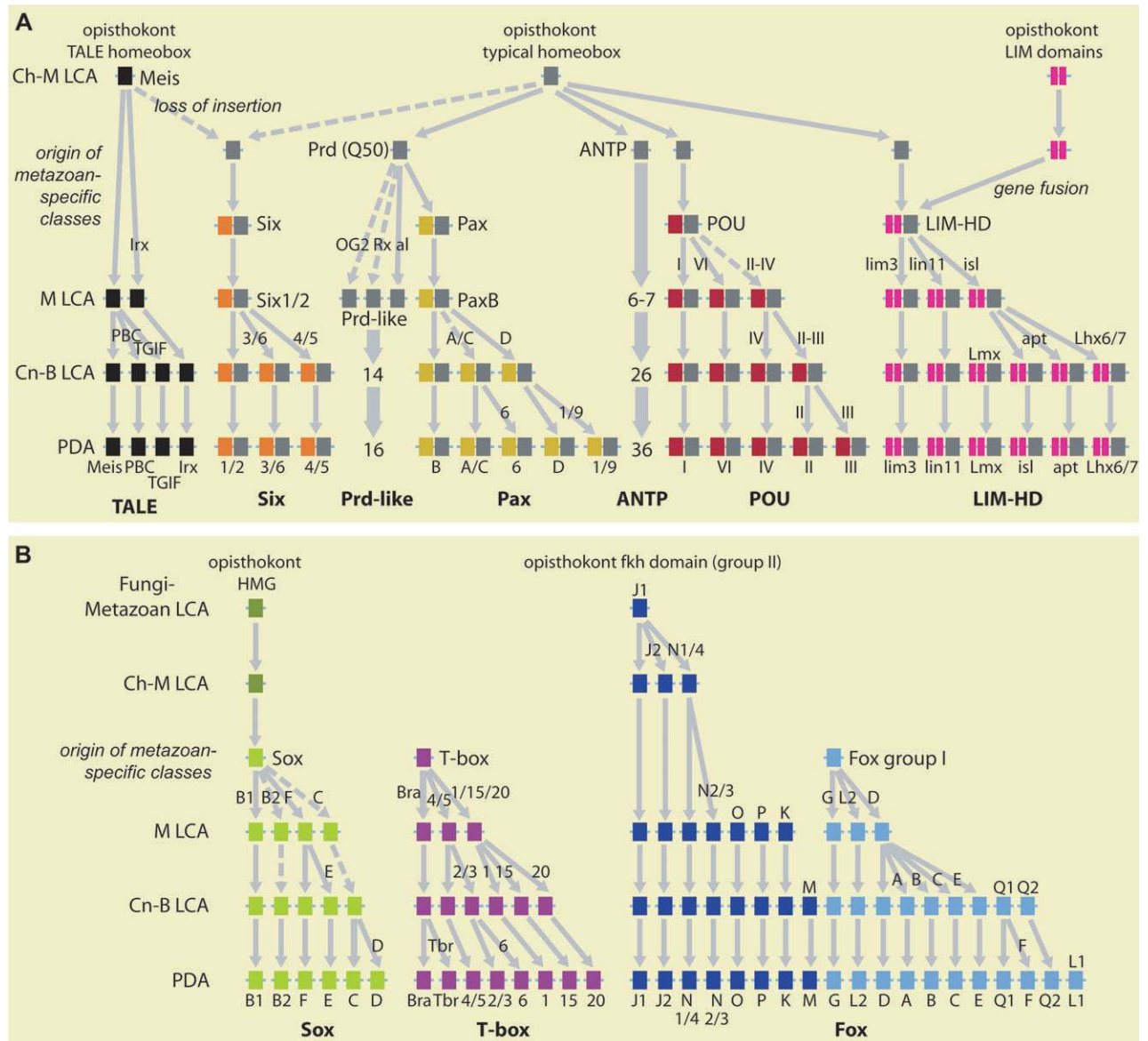


FIG. 3.—Reconstruction of the early evolution of metazoan transcription factor classes. (A) Homeobox gene classes. See table 1 for Cut, Pros, ZF-HD, and HNF classes. (B) Sox, T-box, and Fox gene classes. Dashed lines indicate poorly supported inferences. B, bilaterian; Ch, choanoflagellate; Cn, cnidarian; LCA, last common ancestor; M, metazoan; and PDA, protostome–deuterostome ancestor.

present in the *Amphimedon* genome (fig. 1; table 1). No prd domains have been found in nonmetazoan genomes (Burglin 2005; this study). The *AmqPaxB* HD is relatively divergent; it has a Glu instead of the generally invariant Asp at position 51 and has only 1 of the 6 conserved prd HD amino acid positions (fig. 1). *AmqPaxB* has an Ala at position 50, unlike most Pax HDs, which have a Ser (Galliot et al. 1999). *AmqPaxB* has the conserved prd HD intron (fig. 1; supplementary fig. 2 [Supplementary Material online]). The presence of this conserved intron in all *Amphimedon* prd HDs and in 31 of 33 *Nematostella* prd HDs (Ryan et al. 2006) suggests that the first prd gene possessed an intron in this position.

All 3 phylogenetic methods group the 2 orthologous demosponge Pax genes, *AmqPaxB* and *sPax-2/5/8* (Hoshiyama et al. 1998), within the PaxB clade (NJ < 50, ML < 50, PP:

0.98) (fig. 2B). The placozoan Pax gene and many cnidarian genes also fall within the PaxB family. The NJ tree sees *PoxN* and *PaxA* genes outside of a clade (NJ: 79) comprising *Pax6* genes on one hand and {*PaxC* + *toe* + *eyg*} on the other. However, a *PaxA/C/PoxN* clade comprising cnidarian *PaxA* and *PaxC* genes along with the 3 *Drosophila* genes, *toe*, *eyg*, and *PoxN*, is supported in the Bayesian and ML analyses (PP: 0.99; ML: 48). As there are also non-chordate deuterostome *pox neuro* (*PoxN*) genes, which were not included in this analysis, this suggests that a *PaxA/C/PoxN* gene in the cnidarian–bilaterian LCA gave rise to 2 genes in the PDA, *Pax6* and *PaxA/C/PoxN*. Moreover, there are clear cnidarian representatives in the PaxD family but not in the Pax1/9 family. These analyses suggest that a *PaxB*-like gene founded the Pax class before the divergence of sponge and eumetazoan lineages (fig. 3A).

PaxD and *PaxA/C/PoxN* genes appear to have arisen in the lineage leading to the cnidarian–bilaterian LCA and *Pax6* and *Pax1/9* genes in the period preceding the PDA (table 1; fig. 3A; Matus et al. 2007).

POU Genes

The *Amphimedon* genome includes 4 POU genes, which contain a POU-specific domain upstream of a HD (fig. 1; table 1). POU domains have not been detected outside the Metazoa (Burglin 2005; this study). *AmqPouI* was previously characterized (*RenPouI* in Larroux et al. 2006). The other 3 genes—*AmqPouVI*, *AmqPouA*, and *AmqPouB*—were detected in the genome or ESTs (supplementary fig. 3 and table 1, Supplementary Material online). The full-length amino acid sequences of the latter 2 genes are very similar, suggesting that they are the result of a recent lineage-specific duplication (supplementary fig. 3, Supplementary Material online). We identified the POU-specific domains of the 5 *Nematostella* POU genes, whose homeoboxes were previously described (Ryan et al. 2006), in order to include them in phylogenetic analyses (supplementary fig. 11 and table 3, Supplementary Material online).

In phylogenetic analyses, 2 *Amphimedon* POU genes are orthologous to other demosponge genes (Seimiya et al. 1997) and clearly affiliated with existing bilaterian families I and VI (POU I: NJ: 92, ML: 50, PP: 0.97 and POU VI: NJ: 69, ML: 55, PP: 0.99) (fig. 2C). The paralogous relationship between *AmqPouA* and *AmqPouB* is supported by phylogenetic analyses with these genes not clearly belonging to any specific family. They are inside a poorly supported clade of POU II–IV families in the ML and Bayesian trees (ML < 50, PP: 0.91). The NJ tree places *AmqPouA* and *AmqPouB* inside the POU VI clade alongside the demosponge *POU VI* genes. This may be an artifact due to long-branch attraction between the sponge genes (Bergsten 2005), and this pair of paralogous sponge genes may be descended from an ancestral gene that gave rise to *POU II, III*, and *IV* genes. Additionally, cnidarian genes convincingly belong to the POU I, IV, and VI families and to a well-supported clade of POU II and III families. In the NJ and ML trees, *NvPouIIa* and *NvPouIIb* are placed within a clade of *POU III* genes (ML < 50, NJ: 65).

It thus appears that the POU gene class arose at the dawn of the Metazoa and expanded to at least 2 genes before the separation of extant phyla (table 1; fig. 3A). The cnidarian–bilaterian LCA appears to have had a minimum of 4 POU genes (I, II–III, IV, and VI), and another POU gene appears to have arisen in the period preceding the PDA yielding the 5 POU families shared by most bilaterians (as found in Ryan et al. 2006). The POU V family seems to be a vertebrate innovation; no representative was found in the genomes of *Ciona intestinalis* (urochordate) (Wada et al. 2003), *Drosophila* (ectdysozoan), or *Lottia* (lophotrochozoan). Additionally, *POU I* genes have thus far been identified only in vertebrates, cnidarians, and demosponges. They are absent from completed *Ciona* (Wada et al. 2003), *Strongylocentrotus purpuratus* (echinoderm), *Drosophila*, and *Lottia* genomes, suggesting that they have been lost from multiple metazoan lineages (Larroux et al. 2006).

LIM-HD Genes

In addition to a previously characterized *Lim3* gene (named here *AmqLim3*; Larroux et al. 2006), 2 LIM-HD genes—*AmqLin11* and *AmqIsl*—were identified in the *Amphimedon* genome or ESTs (fig. 1; table 1; supplementary table 1 [Supplementary Material online]). Their predicted protein sequences contain the characteristic 2 LIM domains upstream of a HD (supplementary fig. 4, Supplementary Material online); this domain configuration is not observed in other eukaryotes (Burglin 2005; this study). Phylogenetic analyses provide high support for relationships between LIM-HD families and classification of sponge and cnidarian genes within existing bilaterian families (fig. 2D), allowing us to confidently reconstruct the early evolution of these genes. All sponge genes are in well-supported clades: the *lim3* (NJ: 74, ML: 69, PP: 0.99), *lin11* (NJ: 52, ML: 78, PP: 1), and *islet* (NJ: 100, ML: 95, PP: 1) families. We also identified the LIM domains of 6 *Nematostella* LIM-HD genes, whose homeoboxes were previously described (Ryan et al. 2006; supplementary fig. 11 and table 3 [Supplementary Material online]). They conclusively belong to each of the 6 previously defined (Hobert and Westphal 2000) bilaterian families of LIM-HD genes. These analyses suggest that the proto-LIM-HD originated early in metazoan evolution and duplicated to give rise to at least 3 genes in the metazoan LCA and 6 genes in the cnidarian–bilaterian LCA (table 1; fig. 3A).

CUT, HNF, Pros, and ZF-HD Genes

No genes belonging to the small Cut, HNF, Prospero (Pros), or ZF-HD homeobox classes were identified in the *Amphimedon* and nonmetazoan opisthokont genomes (Burglin and Cassata 2002; Burglin 2005; this study; table 1). TblastN with cut and CMP domains of Cut genes did not produce any matches. Pros and ZF-HD homeobox genes apparently are also absent from the *Nematostella* genome, although a Cut and an HNF genes are present (Ryan et al. 2006). These data suggest that the ancestral Cut and HNF genes arose in the eumetazoan lineage after it had diverged from the sponge lineage and that the other 2 homeobox classes are bilaterian specific, arising after cnidarian and bilaterian lineages split.

Six Genes

A single Six homeobox gene with a Six/sine oculis domain directly upstream of the HD—*AmqSix1/2*—seems to be present in *Amphimedon* (fig. 1; table 1). There is no evidence of Six domains in nonmetazoan genomes (Burglin 2005; this study). This class of non-TALE/typical homeobox genes appears to be closely related to TALE/atypical homeobox genes, which encode larger HDs (Derelle et al. 2007). *AmqSix1/2* seems to be orthologous to partially characterized genes from the 3 sponge classes (Bebenek et al. 2004; supplementary fig. 5 [Supplementary Material online]) and convincingly belongs to the Six1/2 family in phylogenetic analyses (NJ: 98, ML: 78, PP: 1) (fig. 2E). We have identified the Six domains of 3 *Nematostella* Six genes

whose homeoboxes were characterized in Ryan et al. (2006) (supplementary table 3, Supplementary Material online). These genes clearly belong to each of the 3 previously defined bilaterian families of Six genes (fig. 2E; Dozier et al. 2001; as found in Ryan et al. 2006 and Hoshiyama et al. 2007). These data suggest that the ancestral Six gene emerged prior to metazoan cladogenesis and resembled extant *Six1/2* genes and that subsequent gene duplication and divergence events early in the eumetazoan lineage gave rise to the 2 other Six gene families (table 1; fig. 3A).

TALE Genes

Of the 6 TALE genes identified in the *Amphimedon* genome, 5 are *Iroquois* (*Irx*) genes—*AmqIrx**a*, *b*, *c*, *d*, and *e*—and 1 gene—*AmqTALE*—is of uncertain affinity (fig. 1; table 1; supplementary fig. 6 [Supplementary Material online]). Genes with an atypical TALE homeobox likely arose early in eukaryote evolution (Derelle et al. 2007). After preliminary phylogenetic analyses including representatives of all fungi and plant TALE families (according to Burglin 2005), only plant *KNAT* genes were kept for subsequent analyses as other fungi and plant genes were divergent and their positions were unresolved. Two TALE genes are present in the *Monosiga* genome (supplementary table 3, Supplementary Material online) and were included in phylogenetic analyses. Trees were rooted with the plant *KNAT* genes as it is likely that the plant–fungi–metazoan ancestor had only 1 TALE gene (Burglin 2005; Derelle et al. 2007).

Amphimedon Irx genes conclusively belong to the *Irx* family (NJ: 98, ML: 84, PP: 1) and form a monophyletic clade, together with another demosponge gene (Perovic et al. 2003) (fig. 2F). Along with sequence similarity, intron position, and/or genomic linkage (supplementary fig. 6, Supplementary Material online), this suggests that *Amphimedon Irx* genes are the result of lineage-specific duplications. The *Amphimedon* genome also includes *Irx* pseudogenes (supplementary fig. 6, Supplementary Material online). In the 3 trees, the 2 choanoflagellate genes and *AmqTALE* are outside of a clade comprising all other metazoan TALE genes, with the eumetazoan *Meis* genes separating first in the NJ tree (fig. 2F). These data suggest that the 3 genes are descendants of a *Meis*-like TALE ancestor that predates the divergence of choanoflagellate and metazoan lineages. Alignment of the full-length sequence of *AmqTALE* with plant and metazoan TALE genes (supplementary fig. 6, Supplementary Material online) as well as Blast sequence similarity of the *Monosiga* genes supports their affiliation with *Meis* genes. The 2 *Monosiga* TALE genes may be the result of a lineage-specific duplication as they form a poorly supported clade in the NJ analysis. Genes belonging to each of the 4 bilaterian TALE families are present in the *Nematostella* genome (fig. 2F; Ryan et al. 2006). Thus, it appears that 1) the ancestral TALE gene, which predates the origin of the Metazoa, was *Meis*-like; 2) an ancestral *Irx* gene arose prior to metazoan cladogenesis; and 3) the ancestors of the 2 other metazoan TALE families—PBC and TGIF—arose early in the eumetazoan lineage (table 1; fig. 3A).

Sox Genes

Three Sox genes, named here *AmqSoxB1*, *AmqSoxC*, and *AmqSoxF*, were previously characterized by RT-PCR (Larroux et al. 2006; *AmqSoxB1* was previously named *RenSoxB*), and a fourth Sox gene, named *AmqSoxB2* based on sequence similarity, was detected in the genome survey and in developmental ESTs (table 1; supplementary fig. 7 and table 1 [Supplementary Material online]). Whereas non-Sox HMG domains are present outside the Metazoa, Sox HMG domains are not present in choanoflagellate or fungal genomes and appear to have arisen early in metazoan evolution (Soullier et al. 1999; this study). In contrast to the 4 Sox genes in *Amphimedon*, there are 14 Sox genes in the *Nematostella* genome (Magie et al. 2005). As only 6 families of Sox genes are present in bilaterians (Bowles et al. 2000), at least 8 of the 14 *Nematostella* Sox genes are likely to be the result of lineage-specific duplications. For this reason, cnidarian sequences were not included in a first phylogenetic analysis (fig. 4A; ML with PHYMLIP). Sponge, cnidarian, and ctenophore Sox genes characterized by Jager et al. (2006) were not included in our analyses as they are partial HMG domains (the analysis resulted in bootstrap values generally under 50%).

AmqSoxB1 and *AmqSoxB2* clearly belong to the SoxB clade (NJ: 98, ML: 70, PP: 1). Within SoxB, 2 families are recognized in bilaterians, B1 and B2 (Bowles et al. 2000). As *AmqSoxB1* is most similar to *SoxB1* genes (Larroux et al. 2006) and *AmqSoxB2* to *SoxB2* genes (supplementary fig. 7, Supplementary Material online), the 2 sponge genes may correspond to each of those subclasses but evidence is not conclusive. In the Bayesian tree, a clade comprising bilaterian *SoxB1* genes and *AmqSoxB1* is well supported (PP: 0.96), but *AmqSoxB2* and bilaterian *SoxB2* genes are in a paraphyletic position at the base of the B clade. Similarly, the ML tree places *AmqSoxB1* in the B1 clade (ML: 56), with *Drosophila Dichaete* (*D*) (B2), and *AmqSoxB2* outside of the B1 clade and a partial B2 clade. The NJ tree places both sponge genes outside of 2 poorly supported B1 and B2 clades. *AmqSoxF* is placed inside the SoxF family in all 3 trees; the Bayesian analysis gives high support for this grouping (NJ < 50, ML: 57, PP: 0.98). *AmqSoxC* is most similar to *SoxC* genes (Larroux et al. 2006), but it belongs to the SoxC clade only in the ML analysis (ML < 50). It is at the base of the SoxE clade in the NJ tree and in an unresolved position between SoxC, SoxD, and SoxE + F clades in the Bayesian tree.

Selected cnidarian sequences and more bilaterian genes were included in a second set of phylogenetic analyses (fig. 4B; ML with PHYML). Unlike other conserved transcription factor domains spanning a similar number of amino acid residues, *Nematostella* Sox genes do not clearly fall into previously defined bilaterian families, suggesting that they have diverged markedly. In all 3 trees, the 2 sponge *SoxB* genes and 6 *Nematostella* genes are included in a SoxB clade albeit with low support (NJ < 50, ML: 51, PP: 0.87), and *NvSoxB1* is inside a B1 clade. This is in contrast to the high support for the SoxB clade in the previous analysis. *NvSoxA* seems to be a misnomer as *SoxA* (*sry*) genes probably arose from a *SoxB* gene in the mammalian lineage (Bowles et al. 2000; Koopman et al. 2004).

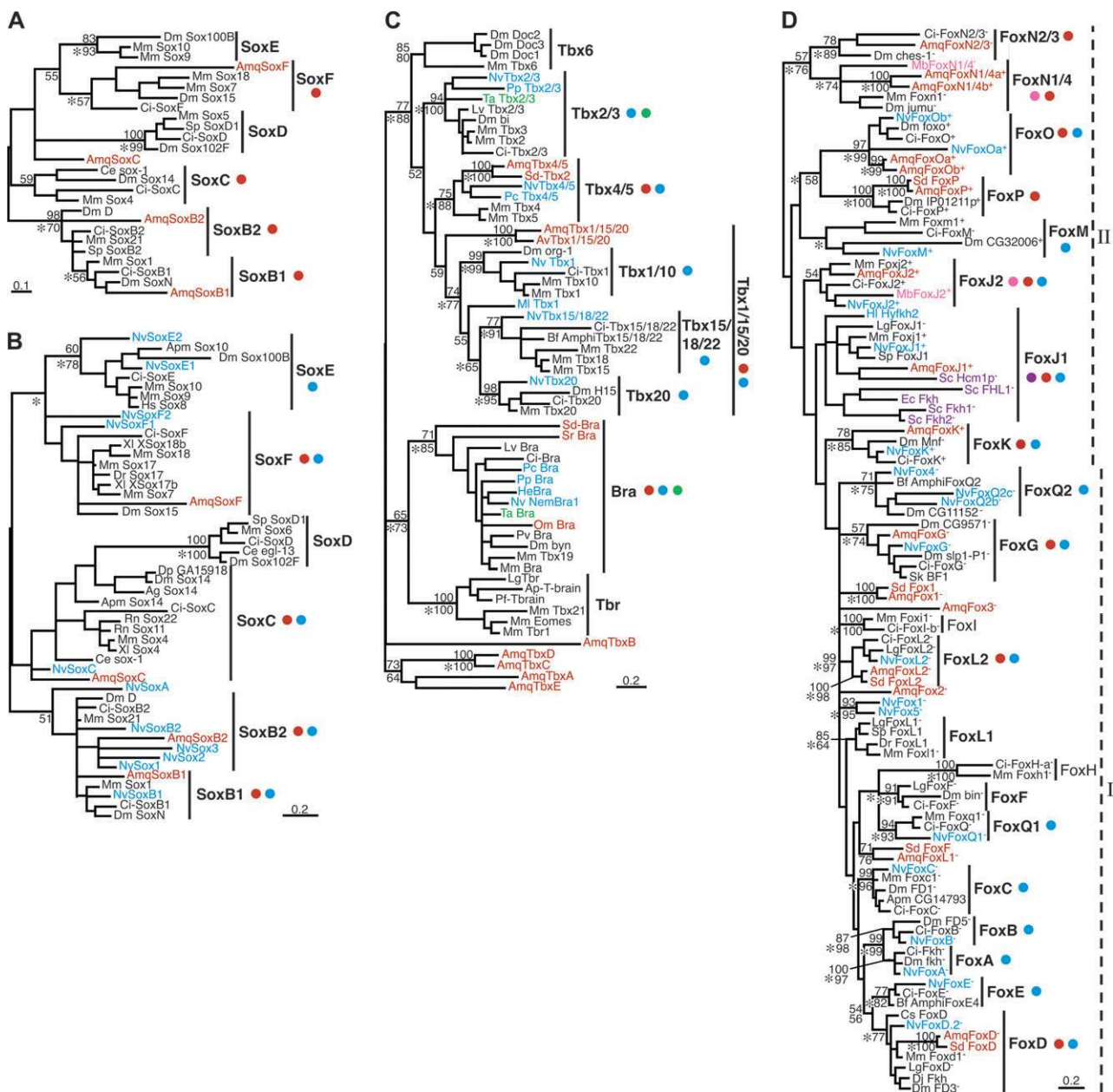


FIG. 4.—Phylogenetic trees of (A, B) Sox, (C) T-box, and (D) Fox gene classes. Bayesian, distance NJ, and ML analyses were undertaken; unrooted Bayesian trees are shown. At key nodes are given percentages of bootstrap support obtained by distance (NJ; 1 000 replicates) above the branch and by ML (100 replicates) below the branch. Bootstrap values above 50% are shown. An asterisk indicates a Bayesian PP greater than or equal to 95%. Families and higher level groupings are shown on the right of the tree. Sponge genes are in red, cnidarian and ctenophore genes in blue, and placozoan genes in green. Fungal and choanoflagellate Fox genes are indicated in purple and pink, respectively. Clades containing representatives of these taxa are indicated by a circle of the corresponding color. For the Sox class, an analysis excluding cnidarian genes (A) provides higher support for family clades than an analysis including cnidarian genes (B). In the Fox tree, presence and absence of the conserved fkh domain intron is indicated by a plus or minus sign in taxa where it is known. In addition to abbreviations in figure 2. Ag: *Anopheles gambiae*, insect; Ap: *Asterina pectinifera*, echinoderm; Apm: *Apis mellifera*, insect; Av: *Axinella verrucosa*, demosponge; Cs: *Ciona savignyi*, urochordate; Dj: *Dugesia japonica*, flatworm; Dp: *Drosophila pseudoobscura*, insect; Ec: *Encephalitozoon culiculi*, fungus; He: *Hydractinia echinata*, hydrozoan cnidarian; Lg: *Lottia gigantea*, mollusk; Lv: *Lytechinus variegatus*, echinoderm; Ml: *Mnemiopsis leidyi*, ctenophore; Om: *Opsacas minuta*, hexactinellid sponge; Pf: *Ptychodera flava*, hemichordate; Pp: *Pleurobrachia pileus*, ctenophore; Pv: *Patella vulgaris*, mollusk; Rn: *Rattus norvegicus*, vertebrate; Sc: *Saccharomyces cerevisiae*, fungus; Sk: *Saccoglossus kowalevskii*, hemichordate; Sr: *Sycon raphanus*, calcareous sponge; and Xl: *Xenopus laevis*, vertebrate.

A SoxE + F clade comprising *AmqSoxF* and 4 cnidarian genes is recovered in all analyses and well supported in the Bayesian tree (NJ < 50, ML < 50, PP: 0.98). Whereas *AmqSoxF* is within a SoxF clade in the Bayesian tree, it is outside a clade of SoxE and F families in the NJ and ML

trees. A SoxE group including 2 cnidarian sequences is well supported, and *NvSoxF1* is inside the SoxF clade in the 3 analyses. *SoxD* genes are embedded within *SoxC* genes in the Bayesian and ML trees, but the NJ analysis supports 2 separate clades, *SoxC* and *SoxD*. Although *AmqSoxC* and

NvSoxC are placed at the base of a group of *SoxC* and *SoxD* genes in all analyses, this is poorly supported (NJ < 50, ML < 50, PP: 0.65).

From these analyses, it appears that the metazoan LCA had at least a proto-*SoxB* and proto-*SoxF* (or *SoxE/F* progenitor) gene as well as possibly a second *SoxB*-like gene and a *SoxC*-like gene (table 1; fig. 3B). A *SoxE* gene seems to have arisen in the lineage leading to the cnidarian–bilaterian LCA. Gene duplication appears to have yielded at least a *SoxD* gene in the first bilaterians, resulting in the 6 bilaterian Sox families (B1, B2, C, D, E, and F; Bowles et al. 2000).

T-box Genes

Seven T-box genes were identified in the *Amphimedon* genome (table 1; supplementary fig. 8 [Supplementary Material online]). Of these, 2 were previously characterized by RT-PCR (named here *AmqTbxA* and *AmqTbxB*; Larroux et al. 2006). Two others, *AmqTbx1/15/20* and *AmqTbxE*, were uncovered in the developmental ESTs (supplementary table 1, Supplementary Material online). The remaining 3 T-box genes, *AmqTbxC*, *AmqTbxD*, and *AmqTbx4/5*, were detected in the genome. A representative subset of the 13 T-box genes present in the *Nematostella* genome (Yamada et al. 2007) was included in phylogenetic analyses. The T-box class is divided into 8 families (Papaioannou 2001; Takatori et al. 2004), which are well supported in this analysis (fig. 4C). Aside from the relationships within the Tbx1/15/20 group, all 3 analyses yield exactly the same relationships between families. Genomic linkage, genomic structure, and/or sequence similarity indicate that *AmqTbxA* and *AmqTbxE*, as well as *AmqTbxC* and *AmqTbxD*, are probably the result of lineage-specific duplications (supplementary fig. 8, Supplementary Material online). The paralogous relationship between *AmqTbxC* and *AmqTbxD* is confirmed by phylogenetic analyses (fig. 4C). In all 3 trees, these 2 genes along with *AmqTbxA* and *AmqTbxE* form a monophyletic clade (NJ: 73, ML: 64, PP: 0.68). *AmqTbxB* forms a monophyletic clade with these genes in the NJ analysis (NJ < 50), whereas it does not in the Bayesian and ML trees but is in a similar position. These 5 sponge genes are between well-supported clades—{Bra + Tbr} (NJ: 65, ML: 73, PP: 1) and {Tbx1 + Tbx15 + Tbx20 + Tbx4/5 + Tbx2/3 + Tbx6} (NJ: 77, ML: 88, PP: 0.99)—and may have evolved through a set of lineage-specific duplications. In contrast, *AmqTbx4/5* and its demosponge orthologue *Sd-Tbx2* (Adell et al. 2003) are confidently placed within a Tbx4/5 clade alongside cnidarian sequences (NJ: 75, ML: 88, PP: 1). *AmqTbx1/15/20* and its orthologue from the demosponge *Axinella verrucosa* (Martinelli and Spring 2005) are at the base of a well-supported clade comprising Tbx1/10, Tbx15/18/22, and Tbx20 families (NJ: 74, ML: 77, PP: 1) in all analyses but their position lacks support (NJ < 50, ML: 59, PP: 0.86). In the Bayesian and ML trees, *MITbx1* from a ctenophore is in an intermediate position between Tbx1/10 and {Tbx15/18/22 + Tbx20} clades, whereas in the NJ tree, it belongs to the Tbx1/10 family (NJ: 53).

Convincing *Brachyury* (*Bra*) genes (NJ: 71, ML: 85, PP: 1) have been isolated from the 3 different classes of

sponges (Adell et al. 2003; Manuel et al. 2004), numerous diploblasts, and the placozoan *Trichoplax adhaerens* (Martinelli and Spring 2003) but appear to be absent from *Amphimedon*. It may be due to gene loss or the ancestral *Bra* may have duplicated and diverged to give rise to the group of 5 divergent T-box genes in the *Amphimedon* lineage, leading to some loss of phylogenetic signal. Additionally, there clearly are diploblast representatives in the Tbx2/3, Tbx1/10, Tbx15/18/22, and Tbx20 families and a placozoan *Tbx2/3* gene (as found in Yamada et al. 2007). However, well-supported Tbx6 and T-brain (*Tbr*) families lack any genes from basal metazoans. Although it is apparently absent from *Drosophila*, a *Tbr* gene was characterized from *Lottia* (supplementary fig. 11, Supplementary Material online), implying that a *Tbr* gene was present in the PDA.

T-box genes have not been found outside of Metazoa (Papaioannou 2001; this study), suggesting that the T-box domain is a metazoan innovation. Of the 8 T-box families that seem to have been present in the PDA, 6 were probably already present in the cnidarian–bilaterian ancestor—Tbx2/3, Tbx15/18/22, Tbx20, Tbx1/10, Tbx4/5, and Brachyury—with Tbx6 and T-brain arising later in the period leading up to the PDA (table 1; fig. 3B; Papaioannou 2001; Takatori et al. 2004; Yamada et al. 2007; this study). Altogether, T-box genes in sponges suggest that 3 T-box genes were present in the metazoan LCA, a *Bra*, a *Tbx4/5*, and a *Tbx1/15/20* gene. The *Tbx1/15/20* gene would have given rise to Tbx1/10, Tbx15/18/22, and Tbx20 eumetazoan families.

Fox Genes

A total of 16 Fox genes were characterized in the *Amphimedon* genome (table 1; supplementary fig. 9 [Supplementary Material online]). The first Fox gene probably arose in a stem opisthokont as this gene class is also present in fungi (Kaestner et al. 2000). The full cDNA sequence of *AmqFoxL1* and partial cDNA sequence of *AmqFoxJ1* (previously named *RenFoxJ*) were characterized in Larroux et al. (2006). *AmqFoxD*, *AmqFoxG*, *AmqFoxN2/3*, and *AmqFoxP* are present in the EST data set along with partial sequences of *AmqFoxL2* (supplementary table 1, Supplementary Material online). In addition, *AmqFoxJ2*, *AmqFoxK*, *AmqFoxN1/4a*, *AmqFoxN1/4b*, *AmqFoxOa*, *AmqFoxOb*, *AmqFox1*, *AmqFox2*, and *AmqFox3* were found in the genome. Sequence similarity, intron position, and/or genomic linkage suggest that the pairs of *FoxO* and *FoxN1/4* genes are the result of lineage-specific duplications (supplementary fig. 9, Supplementary Material online); this is confirmed by phylogenetic analyses (fig. 4D). In addition to the Fox genes published in Magie et al. (2005), 10 forkhead (fkh) domains were identified in the *Nematostella* genome (supplementary fig. 11 and table 3, Supplementary Material online). A total of 7 Fox genes were detected in the *Monosiga* genome, and 5 Fox genes were identified in the *Lottia* genome (supplementary fig. 11 and table 3, Supplementary Material online).

After preliminary phylogenetic analyses, all choanoflagellate genes except *MbFoxN1/4* and *MbFoxJ2* were removed from the data set as they were divergent and their

positions were unresolved. In the Bayesian, NJ, and ML analyses, *MbFoxN1/4* and *MbFoxJ2* conclusively belong to N1/4 (NJ < 50, ML: 74, PP: 1) and J2 clades (NJ: 54, ML < 50, PP: 0.94), respectively (fig. 4D). These analyses suggest that *AmqFoxD* (NJ: not supported [at the base of D + E], ML: 77, PP: 1), *AmqFoxG* (NJ: 97 [without *CG9571*], ML: 74, PP: 1), *AmqFoxL2* (NJ: 99, ML: 97, PP: 1), *AmqFoxK* (NJ: 78, ML: 85, PP: 1), *AmqFoxJ2* (NJ: 54 [63 without *MbFoxJ2*], ML < 50, PP: 0.94), *Amq-FoxN2/3* (NJ: 78, ML: 89, PP: 1), the 2 *Amphimedon FoxN1/4* genes (NJ < 50 [75 in an analysis excluding *Monosiga* genes], ML: 74, PP: 1), the 2 *Amphimedon FoxO* genes (NJ: 97, ML: 99, PP: 1), and *AmqFoxP* (NJ: 100, ML: 100, PP: 1) convincingly belong to the families they were assigned to during alignments. A FoxJ1 clade comprising *AmqFoxJ1*, *NvFoxJ1*, bilaterian *FoxJ1* genes, and fungal genes is weakly supported in all 3 trees (NJ < 50, ML < 50, PP: 0.56); the NJ and ML clades do not include all fungal genes. As in the alignments (supplementary fig. 9, Supplementary Material online), *AmqFoxL1* and *AmqFox3* are grouped with *FoxL1* and *FoxI* genes, respectively, in the NJ tree (the *FoxI* clade is also present in the ML tree), but both genes are in an unresolved position in the Bayesian analysis. In all analyses, the affinities of *AmqFox1* and *AmqFox2* are unresolved. Five of the *Amphimedon* genes seem to be orthologous to Fox genes isolated from the demersal sponge *Suberites domuncula* (Adell and Muller 2004).

The grouping of *Nematostella* Fox genes within existing families A, B, C, D, E, G, L2, Q1, Q2, J2, K, M, and O (Kaestner et al. 2000; Mazet et al. 2003) is supported by all 3 analyses with generally high statistical support; a partial fkh domain sequence of a *FoxN* gene (Magie et al. 2005) was not included in the analysis. Both FoxN families form a well-supported clade, as do FoxA + B and FoxO + P families. The Bayesian analysis supports a clade of *FoxF*, *Q1*, and *H* genes (only *FoxF* and *Q1* in the NJ and ML trees), of which a cnidarian representative is found in the FoxQ1 clade. Families apparently lacking sponge or cnidarian representatives are FoxF, H, L1, and I (traces were probed with fkh domain from each of these families); these clades are well supported. Of these, *FoxH* genes were not detected in *Strongylocentrotus*, *Lottia*, or *Drosophila*; they have only been found in *Ciona* and vertebrates so far and may be chordate innovations. *FoxI* genes are absent from the 2 protostome genomes but present in *Ciona* and *Strongylocentrotus*; they may have arisen early in the deuterostome lineage. The other families—FoxF and L1—have protostome and deuterostome representatives and were thus present in the PDA. Although *FoxJ2* is apparently absent from *Drosophila*, a partial fkh sequence was recovered from *Lottia*, which is clearly affiliated with the FoxJ2 family (C Larroux and BM Degnan unpublished data). In contrast, *FoxE* and *FoxQ1* genes are present in *Nematostella*, *Ciona*, and vertebrates but apparently absent from *Drosophila* and *Lottia*; they may have been lost early on in the protostome lineage.

There clearly is a group of *Amphimedon* Fox genes—comprising *AmqFoxD*, *AmqFoxG*, *AmqFoxL1*, and *Amq-FoxL2*, as well as probably *AmqFox1*, *AmqFox2*, and *AmqFox3*—with very few introns, either none at all or only introns in the 5' untranslated regions (supplementary fig. 10,

Supplementary Material online). In contrast, another group of Fox genes, comprising *AmqFoxK*, *AmqFoxJ1*, *Amq-FoxJ2*, *AmqFoxN1/4a-b*, *AmqFoxN2/3*, *AmqFoxOa-b*, and *AmqFoxP*, have numerous introns in their open reading frames. These 2 groups fall into 2 separate clades during the Bayesian analysis; the former clade is named here group I and the latter group II (fig. 4D). In the NJ and ML trees, these groupings are recovered except that *FoxH* genes fall within group II (alongside fungal *FHL1*) rather than group I. *FHL1* and *FoxH* genes have long branches and may be grouped together due to long-branch attraction. Interestingly, *Nematostella* Fox genes characterized by rapid amplification of cDNA ends (*NvFoxA*, *NvFoxB*, *NvFoxC*, *NvFoxD.1*, *NvFoxE*, and *NvFox1*; Magie et al. 2005), all belong to group I and also do not have any introns in their open reading frames. However, of the 8 *Drosophila* group I genes shown in the tree, 2 have introns in their open reading frame (*FD3* and *bin*), suggesting that this may be an initial functional constraint that was subsequently lost in some lineages. Among the *Amphimedon* group II Fox genes, 1 intron position is conserved between all of them except *AmqFoxN2/3*, corresponding to between amino acid positions 48 and 49 of a typical fkh domain (WX-N). The presence and absence of this intron position is indicated for each taxon for which it is known in the phylogenetic tree (fig. 4D). Within the {A-I + L1-2 + Q1-2} group (bottom of tree, group I), none of the genes have this intron. In the {K + M-P + J} group (top of tree, group II), *FoxO*, *P*, and *J2* genes possess this intron, whereas it is present in some of the genes of the FoxK, M, and J1 families. Among *FoxN* genes, only *AmqFoxN1/4a* and *b* have this intron, possibly reflecting an ancestral state with intron loss occurring later in the lineage.

All choanoflagellate and fungal genes included in this analysis belong to group II in the Bayesian, NJ, and ML trees (fig. 4D). Bayesian analyses including the other 5 *Monosiga* genes result in polytomy at the node separating the 2 groups but, during Blast, these genes are most similar to *FoxJ1* or *FoxJ2* genes, suggesting that they are also affiliated with group II. Regarding the diagnostic intron position, *MbFoxJ2*, 1 of the 2 *FoxJ2*-like genes and 2 of the 3 *FoxJ1*-like genes have the conserved group II intron position, supporting their affiliation with group II. As with all *FoxN1/4* genes except the *Amphimedon* gene, this intron position is not present in the choanoflagellate single *FoxN1/4* gene. No Fox genes from the fungus *Saccharomyces cerevisiae* possess the conserved intron position. However, only 14% of introns that were probably in the plant–fungi–metazoan LCA are present in the yeast genome, suggesting that massive intron loss occurred in this lineage (Roy and Gilbert 2006). Thus, the lack of intron in the fungus probably reflects this trend rather than a phylogenetic signal. Generally, intron loss seems to occur more frequently than previously thought (Roy and Gilbert 2006), and the absence of the characteristic intron position in some group II genes probably reflects intron loss. The presence and absence of this intron supports the 2 large groups, which in turn support a metazoan-specific diversification in group I. More generally, with all classes presented here, particular trends can be observed regarding intron composition, possibly reflecting deep structural and functional

evolutionary constraints specific to each class (supplementary fig. 10, Supplementary Material online).

Based on fungal, *Monosiga*, and *Amphimedon* genomes, the fungi–choanoflagellate–metazoan LCA probably had at least a *FoxJ1* gene (fig. 3B). *FoxJ2* and *N1/4* genes seem to have arisen in the choanoflagellate–metazoan LCA and *FoxK*, *N2/3*, *O*, and *P* genes in the period preceding the metazoan LCA. Prior to demosponges branching off the main metazoan lineage, a new type of Fox genes (group I) with novel intron composition appears to have evolved. This gene seems to have duplicated resulting in 3 or 4 clade I genes in the metazoan LCA (D, G, L2, and possibly L1) (table 1; fig. 3B). Later, in the lineage leading up to the cnidarian–bilaterian LCA, 6 new gene families appear to have evolved in the group I lineage (A, B, C, E, Q1, and Q2) and 1 gene family in group II (M). At least 1 new Fox gene (F) seems to have arisen in the period preceding the PDA. Thus, diversification of the Fox class principally occurred in 2 steps, with duplications mainly happening within group I in the period preceding the cnidarian–bilaterian LCA and resulting in almost the full complement of 19 Fox genes found in most bilaterians.

Discussion

In terms of transcription factor gene composition, *Amphimedon* has much closer affinity to other metazoan genomes than to any of the other opisthokont genomes, including the choanoflagellate *Monosiga* (Larroux et al. 2007; Simionato et al. 2007; this study). The *Amphimedon* genome has representatives of a large majority of transcription factor gene families and classes that have previously been found in eumetazoan genomes, including homeobox genes belonging to ANTP, prd-like, Pax, POU, LIM-HD, Six, and TALE classes, as well as basis helix-loop helix (bHLH), Sox, T-box, and Fox genes (Larroux et al. 2006, 2007; Simionato et al. 2007; this study).

Genesis of Transcription Factor Classes in the First Metazoans

As seems to have been the case in the ancestor of all eukaryotes, the common ancestor of fungi and animals probably had 1 typical (non-TALE) homeobox gene and 1 TALE homeobox gene (Burglin 2005; Derelle et al. 2007); TALE homeoboxes have a 3 codon insertion, resulting in additional amino acids between the first and second helices of their HDs. Of the classes present in *Amphimedon*, Fox genes are also found in fungi (Kaestner et al. 2000). Only TALE homeobox and Fox genes seem to be present in the *Monosiga* genome, suggesting that a typical homeobox gene was lost in the *Monosiga* lineage. ANTP, prd-like, Pax, POU, LIM-HD, Six, Sox, and T-box transcription factor classes appear to have evolved early in metazoan evolution, between choanoflagellate–metazoan and demosponge–eumetazoan divergences (fig. 3). Thus, a large suite of genomic innovations appears to have occurred at the dawn of the Metazoa, prior to the divergence of all major extant animal lineages. Early metazoan transcription

factor gene evolution included the de novo evolution of new regulatory domains, the diversification of existing domains, and the shuffling of domains to yield novel combinations.

The ancestral opisthokont typical homeobox gene most likely duplicated to yield progenitors of most metazoan homeobox classes (fig. 3A; Larroux et al. 2007); the progenitor of Six genes may have evolved from a TALE gene having lost the HD insertion (Derelle et al. 2007). Prior to metazoan cladogenesis, a Prd, a POU-specific, and a Six domain seem to have evolved alongside a homeobox to generate the first Pax, POU, and Six genes, respectively. Additionally, the more ancient LIM domains became linked to a HD to give rise to another novel metazoan homeobox class, LIM-HD.

Analyses of *Amphimedon* prd-like genes support the hypothesis that a Q50 prd-like gene founded the Prd superclass in early metazoans (Galliot et al. 1999) and suggest a proto-K50 gene arose from a Q50 gene after the sponge–eumetazoan split. The apparent diversification among prd-like genes and presence of only 1 Pax gene in this sponge support the idea that prd-like genes are more ancient than Pax genes and that Pax genes arose from a prd-like gene. Based on the presence of genes possessing a Prd domain but no homeobox in cnidarians, it was proposed that the fusion of 2 genes —1 containing a Prd domain and 1 containing a prd-like homeobox—led to the first Pax gene possessing both domains (Galliot et al. 1999). However, the absence of Prd genes without a homeobox in the *Amphimedon* genome suggests that a Prd domain arose alongside a prd-like homeobox and not independently with subsequent gene fusion (fig. 3A).

Similarly, it appears that a *mef2* domain arose alongside a MADS domain during early metazoan evolution as *Amphimedon* has a *mef2* gene (Larroux et al. 2006) but *Monosiga* does not. Unlike non-Sox HMG proteins that are probably involved in structural regulation of transcription (Grosschedl et al. 1994), metazoan Sox proteins bind specifically to target sequences and have complex developmental expression patterns and various key developmental roles (Schepers et al. 2002). As this particular type of HMG domain and the T-box domain have not been found outside of Metazoa (Soullier et al. 1999; Papaioannou 2001; this study) and are present in the genome of this demosponge, they appear to have arisen early in the metazoan lineage (fig. 3B). Although Fox genes probably evolved early in opisthokont evolution (Kaestner et al. 2000), a novel type of Fox genes (group I) with initial functional constraints regarding intron composition seems to have arisen in the first metazoans.

Gradual Expansion of Transcription Factor Classes in Early Metazoans

Reconstruction of the ancestor from which all living metazoans arose and of the early evolutionary history of metazoans is contingent upon both a clear understanding of the evolutionary relationship of basal metazoan taxa and robust gene phylogenies. Contentious questions regarding the phylogeny of early metazoans include whether

sponges are monophyletic or paraphyletic (e.g., Cavalier-Smith et al. 1996; Kruse et al. 1998; Borchiellini et al. 2001) and whether placozoans are the earliest branching metazoans or derived eumetazoans (e.g., Collins 1998; Dellaporta et al. 2006). While we and others have recognized the caveats associated with these various scenarios when reconstructing the state and evolution of the early metazoan genome (Larroux et al. 2007; Simionato et al. 2007), we take the traditional view here that the Porifera is a monophyletic group that has stemmed from the ancestor to all living metazoans. In molecular phylogenetic analyses supporting sponge paraphyly, Calcarea are more closely related to Eumetazoa than the 2 siliceous sponge classes, Demospongia and Hexactinellida (e.g., Kruse et al. 1998; Borchiellini et al. 2001; Medina et al. 2001; Wallberg et al. 2004). As *Amphimedon* is a demosponge and thus belongs to the more ancient sponge lineage in both monophyly and paraphyly hypotheses, our inferences remain valid whichever scenario is correct.

Superimposing the evolution of specific gene classes and families on the organismal tree is restricted by the confidence in specific nodes within a given gene tree, which is reliant on the phylogenetic signal inherent in the sequence. This signal varies among the classes of transcription factors analyzed here and previously (Larroux et al. 2007; Simionato et al. 2007). For example, the 60 amino acid HD has a limited phylogenetic signal, making evolutionary reconstructions difficult, particularly in the case of prd-like and TALE genes (see also Ryan et al. 2006; Larroux et al. 2007). In contrast, the inclusion of additional conserved Pax, POU, LIM-HD, and Six domains with HDs in phylogenetic analyses results in greater support for tree topologies. In most cases, our evolutionary reconstructions in figure 3 are conservative—based on high support values and/or congruence between the 3 phylogenetic methods.

Our findings show that a wide range of transcription factor gene classes in *Amphimedon* are significantly smaller than their eumetazoan counterparts. There are 2 alternative explanations for this observation: 1) the sponge–eumetazoan LCA represented an intermediate condition in the evolution of these transcription factor gene classes and extensive gene duplications occurred in the eumetazoan lineage after it had diverged from the sponge lineage or 2) most transcription factor gene families had already diversified in the sponge–eumetazoan LCA and extensive gene loss took place in the sponge lineage. We assume that there has been some gene loss in the sponge lineage, as has been observed in all eukaryotic genomes to date. Nonetheless, the position of sponges at the base of Metazoa is compatible with gradual diversification of transcription factor classes. By studying a large range of classes, we show that the limited number of transcription factor genes in sponges is likely to be a genome-wide phenomenon and not related to specific cases of gene loss.

There is some evidence for gene loss specifically in the lineage leading to *A. queenslandica* with this sponge clearly lacking a gene found in other sponges: the *Bra* T-box gene (Adell et al. 2003). However, we have not detected any other genes previously identified in other sponges (e.g., Seimiya et al. 1997; Hoshiyama et al. 1998; Perovic et al. 2003; Wiens et al. 2003; Adell and Muller 2004;

Bebenek et al. 2004) without an orthologue in *Amphimedon*, suggesting that distribution of *Amphimedon* genes in molecular phylogenies closely reflects the ancestral sponge condition.

The amount of gene loss in sponges may be partly determined by the phylogenetic position of placozoans. If they are at the base of the metazoan tree (as hypothesized in Dellaporta et al. 2006), it is likely that the genome of the LCA to placozoans, sponges, and eumetazoans was markedly more complex than observed in either extant sponge or placozoan genomes, given the lack of overlap in the constituencies of specific gene families in these animals (Peterson and Sperling 2007 and references therein). However, although this phylogenetic position is supported by analysis of the placozoan mitochondrial genome (Dellaporta et al. 2006), it is not supported by 18S ribosomal RNA (rRNA) phylogenies, which place this group within the Eumetazoa (e.g., Collins 1998; Cavalier-Smith and Chao 2003). A recent phylogenetic analysis of the large subunit as well as the small subunit rRNA resulted in conflicting relationships depending on the data set or method used but the position of Placozoa as a sister group to the rest of Metazoa received low support (da Silva et al. 2007). Recent sequencing of a placozoan genome should provide the opportunity to finally clarify the phylogenetic position of this small enigmatic phylum.

One way to infer whether there has been gene loss may be to root the gene trees (e.g., Hoshiyama et al. 2007; Peterson and Sperling 2007). However, we have chosen not to root most of our gene trees as outgroups that are too distant can often lead to an erroneous tree topology, mainly through long-branch attraction (Bergsten 2005). In the case of the ANTP class of homeobox genes (Larroux et al. 2007), we had initially analyzed our data set using 3 homeobox genes from 3 other classes and these genes were incorrectly distributed in different areas of the tree. Hence, we had decided to use an unrooted tree. In this study, it is likely that the root position of the prd-like class is incorrect and due to long-branch attraction of prd-like genes with the outgroup (fig. 2A). Even with closely related Pax and prd-like genes, NJ and ML trees do not result in a monophyletic Pax class (as occurred in Galliot et al. 1999).

In our analyses, the only clear occurrence of a sponge gene at the base of a clade of multiple families—apparently representing a descendant of an ancestral gene with a sequence equally similar to the daughter families—are the *Tbx1/15/20* and possibly *POU II-IV* genes, the ancestral genes having given rise to 3 families (figs. 2C and 3C). In many cases, sponge genes can instead be confidently classified into specific families within a transcription factor class, whereas sister families have no sponge representatives (figs. 2 and 4; Larroux et al. 2007; Simionato et al. 2007). Such a tree topology can be interpreted as evidence that an ancestral gene has been lost in the sponge lineage (cf., Peterson and Sperling 2007). For example, in the well-resolved LIM-HD tree, the nesting of *Amphimedon* genes within 3 of the 6 families within the larger groups I and II (fig. 2D) may suggest that the ancestor to sponges and eumetazoans had a more complex LIM-HD gene repertoire than observed in *Amphimedon*, comprised of 5 or 6 genes. However, an alternate explanation is that, after gene duplication, 1 duplicated gene remains similar to the ancestral

form while the other diverges. This differential rate of evolution may reflect functional constraints on the protein's original role, allowing only 1 copy to evolve and acquire a new function. Indeed, there is some evidence that, following duplication, the 2 daughter genes often have asymmetric evolutionary rates—with 1 duplicate evolving faster than the other (reviewed in Taylor and Raes 2004). Until genomes of organisms spanning the breadth of basal metazoan phyla are sequenced and we have a clear understanding of the relationship of these taxa, we cannot know with certainty the extent to which we are underestimating the number of genes present in key ancestors. However, with the 3 genomes analyzed here, we can propose a scenario based on hard evidence of presence and absence of genes, which may be revised in the future.

The demosponge *A. queenslandica* has 31 homeobox genes, with representatives of both typical (non-TALE) and TALE genes. Although the affinities of a number of *Amphimedon* genes are poorly resolved, phylogenetic analyses suggest that the metazoan LCA had 17–20 homeobox genes in contrast to the 62 genes probably present in the cnidarian–bilaterian LCA and 82 genes in the PDA (table 1; this study; Burglin 2005; Ryan et al. 2006; Larroux et al. 2007). This approximate 3-fold increase in homeobox gene number between sponge and cnidarian divergences is similar to that observed for the bHLH class (Simionato et al. 2007).

Detailed analysis reveals differential gene expansion in the various transcription factor groups during early metazoan evolution. For example, although the progenitors of the Six and Pax classes did not seem to duplicate in the first metazoans, a 3-fold increase in gene numbers likely occurred in the prd-like, POU, LIM-HD, T-box, and Fox I groups, after the origin of the first representative, during this period (table 1; fig. 3). In the subsequent period, between metazoan and cnidarian–bilaterian LCAs, it seems that Six, prd-like, Pax, ANTP, and Fox I groups expanded 3- to 4.3-fold, whereas TALE, POU, LIM-HD, Sox, T-box, and Fox II group increased in size 1.1–2 times. In the period leading to the PDA, TALE, Six, LIM-HD, and Fox II genes did not appear duplicate, whereas prd-like, Pax, ANTP, POU, Sox, T-box, and Fox I groups seem to have expanded only slightly (1.2–1.7 times). The evolution of these transcription factor classes by gene duplication and divergence hence appears to have principally taken place in the periods both preceding and following the demosponge–eumetazoan divergence. However, there are marked differences in the extent and timing of gene duplication events in each of the classes. This suggests that classes followed different evolutionary trajectories and that duplications of transcription factor genes were not part of whole-genome duplications in any particular period but independent events. Evidence that 5 ANTP-class genes were the result of *cis* duplications is still present in the *Amphimedon* genome (Larroux et al. 2007). In contrast, aside from those apparently issued from recent lineage-specific duplications, other genes do not seem to be clustered in the genome (supplementary table 2, Supplementary Material online). Unlike the *Nematostella* genome (Magie et al. 2005; Chourrout et al. 2006; Ryan et al. 2006; Putnam et al. 2007; Simionato et al. 2007; Yamada et al. 2007), there are few cases of apparent lineage-specific duplications in *Amphimedon*; these

are in the TALE, prd-like, POU, T-box, and Fox lineages (figs. 2 and 4).

The Ancestral Metazoan Developmental Program

A period of genome innovation and gene duplication, prior to the divergence of all the major extant metazoan phyla, appears to have led to the genesis of many of the regulatory components found in the modern metazoan developmental program. The emergence of novel transcription factor genes prior to the separation of modern animal lineages is compatible with the supposition that early innovations in the ancestral genome provided the regulatory foundation for the evolution of multicellularity and embryogenesis. These novel transcription factors, with new DNA-binding specificities, would have extended the regulatory capacity of the genome, with combinatorial interactions between transcription factors further expanding regulatory complexity (Phillips and Luisi 2000; Wilson and Koopman 2002). In modern metazoans, members of these transcription factor gene classes act as critical and often conserved regulatory nodes in developmental genetic networks. Their presence and developmental expression in the sponge *Amphimedon* (this study; Larroux et al. 2006; C Larroux and BM Degnan unpublished data) suggests that the ancestral developmental network was populated by many of the same regulatory components that are operating in modern complex metazoans. This conservation extends to cell–cell signaling by hedgehog-like, Wnt, and TGF- β ligands (Nichols et al. 2006), all of which are developmentally expressed in *Amphimedon* (Adamska, Degnan, et al. 2007; Adamska, Matus, et al. 2007).

Early in eumetazoan genome evolution, after the divergence of sponges and eumetazoans, a second period of expansion gave rise to almost the full diversity of genes in these transcription factor classes, as evidenced in the *Nematostella* genome. The duplication and divergence of transcription factor genes in the eumetazoan lineage allowed their co-option into new roles, which may have been the first step toward the evolution of complex eumetazoan body plans and life cycles. Sponges appear to represent an intermediate phase in the evolution of the metazoan genome, with a limited suite of developmental transcription factors correlating with a simpler body plan that has not changed since well before the Cambrian. From these comparative analyses, we infer that the developmental network of the metazoan LCA must have been smaller than that of the ancestor that gave rise to cnidarians and bilaterians. This core developmental network may have been sufficient for the evolution of metazoan multicellularity and development.

Supplementary Material

Genes presented will be provided with accession numbers. Supplementary tables 1–3 and figures 1–11 and are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We gratefully acknowledge the significant contribution and support of The US Department of Energy Joint

Genome Institute in the production of *Amphimedon* (*Reniera*) genomic and EST sequences used in this study through the Community Sequencing Program. The research was supported by grants from the Australian Research Council to B.M.D. G.N.L and S.M.S were supported by the Biotechnology and Biological Sciences Research Council.

Literature Cited

- Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. 21: 2104–2105.
- Adamska M, Degnan SM, Green KM, Adamski M, Craigie A, Larroux C, Degnan BM. 2007. Wnt and TGF- β expression in the sponge *Amphimedon queenslandica* and the origin of metazoan embryonic patterning. *PLoS ONE*. 2:e1031.
- Adamska M, Matus DQ, Adamski M, Green KM, Martindale MQ, Degnan BM. 2007. The evolutionary origin of hedgehog proteins. *Curr Biol*. 17:R836–R837.
- Adell T, Grebenjuk VA, Wiens M, Muller WEG. 2003. Isolation and characterization of two T-box genes from sponges, the phylogenetically oldest metazoan taxon. *Dev Genes Evol*. 213:421–434.
- Adell T, Muller WEG. 2004. Isolation and characterization of five Fox (Forkhead) genes from the sponge *Suberites domuncula*. *Gene*. 334:35–46.
- Banerjee-Basu S, Baxevanis AD. 2001. Molecular evolution of the homeodomain family of transcription factors. *Nucleic Acids Res*. 29:3258–3269.
- Bebenek IG, Gates RD, Morris J, Hartenstein V, Jacobs DK. 2004. *Sine oculis* in basal Metazoa. *Dev Genes Evol*. 214:342–351.
- Bergsten J. 2005. A review of long-branch attraction. *Cladistics*. 21:163–193.
- Borchellini C, Manuel M, Alivon E, Boury-Esnault N, Vacelet J, Le Parco Y. 2001. Sponge paralogy and the origin of Metazoa. *J Evol Biol*. 14:171–179.
- Bowles J, Schepers G, Koopman P. 2000. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev Biol*. 227: 239–255.
- Brusca RC, Brusca GJ. 2003. *The invertebrates*. Sunderland (MA): Sinauer Associates.
- Burger G, Forget L, Zhu Y, Gray MW, Lang BF. 2003. Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc Natl Acad Sci USA*. 100:892–897.
- Burglin TR. 2005. Homeodomain proteins. In: Meyers RA, editor. *Encyclopedia of molecular cell biology and molecular medicine*. Weinheim (Germany): Wiley. p. 179–222.
- Burglin TR, Cassata G. 2002. Loss and gain of domains during evolution of cut superclass homeobox genes. *Int J Dev Biol*. 46:115–123.
- Carlsson P, Mahlapuu M. 2002. Forkhead transcription factors: key players in development and metabolism. *Dev Biol*. 250:1–23.
- Cavalier-Smith T, Chao EE. 2003. Phylogeny of choanozoa, apusozoa, and other protozoa and early eukaryote mega-evolution. *J Mol Evol*. 56:540–563.
- Cavalier-Smith T, Chao EE, Boury-Esnault N, Vacelet J. 1996. Sponge phylogeny, animal monophyly, and the origin of the nervous system: 18S rRNA evidence. *Can J Zool*. 74:2031–2045.
- Chi N, Epstein JA. 2002. Getting your Pax straight: pax proteins in development and disease. *Trends Genet*. 18:41–47.
- Chourrout D, Delsuc F, Chourrout P, et al. (11 co-authors). 2006. Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature*. 442:684–687.
- Collins AG. 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.
- Collins AG. 2002. Phylogeny of Medusozoa and the evolution of cnidarian life cycles. *J Evol Biol*. 15:418–432.
- da Silva FB, Muschner VC, Bonatto SL. 2007. Phylogenetic position of Placozoa based on large subunit (LSU) and small subunit (SSU) rRNA genes. *Genet Mol Biol*. 30: 127–132.
- Davidson EH, Erwin DH. 2006. Gene regulatory networks and the evolution of animal body plans. *Science*. 311:796–800.
- Degnan BM, Leys SP, Larroux C. 2005. Sponge development and antiquity of animal pattern formation. *Integr Comp Biol*. 45:335–341.
- Degnan SM, Degnan BM. 2006. The origin of the pelagobenthic metazoan life cycle: what's sex got to do with it? *Integr Comp Biol*. 46:683–690.
- Dellaporta S, Xu A, Sagasser S, Jakob W, Moreno MA, Buss LW, Schierwater B. 2006. Mitochondrial genome of *Trichoplax adhaerens* supports Placozoa as the basal lower metazoan phylum. *Proc Natl Acad Sci USA*. 103:8751–8756.
- Derelle R, Lopez P, Le Guyader H, Manuel M. 2007. Homeodomain proteins belong to the ancestral molecular toolkit of eukaryotes. *Evol Dev*. 9:212–219.
- Dozier C, Kagoshima H, Niklaus G, Cassata G, Burglin TR. 2001. The *Caenorhabditis elegans* Six/sine oculis homeobox gene *ceh-32* is required for head morphogenesis. *Dev Biol*. 236:289–303.
- Felsenstein J. 2003. PHYLIP (phylogeny inference package). Seattle (WA): Department of Genome Sciences, University of Washington. Distributed by the author.
- Galliot B, de Vargas C, Miller D. 1999. Evolution of homeobox genes: q(50) paired-like genes founded the paired class. *Dev Genes Evol*. 209:186–197.
- Grosschedl R, Giese K, Pagel J. 1994. HMG domain proteins—architectural elements in the assembly of nucleoprotein structures. *Trends Genet*. 10:94–100.
- Guindon S, Lethiec F, Duroux P, Gascuel O. 2005. PHYML online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res*. 33:W557–W559.
- Hobert O, Westphal H. 2000. Functions of LIM-homeobox genes. *Trends Genet*. 16:75–83.
- Hoshiyama D, Iwabe N, Miyata T. 2007. Evolution of the gene families forming the Pax/Six regulatory network: isolation of genes from primitive animals and molecular phylogenetic analyses. *FEBS Lett*. 581:1639–1643.
- Hoshiyama D, Suga H, Iwabe N, Koyanagi M, Nikoh N, Kuma K, Matsuda F, Honjo T, Miyata T. 1998. Sponge Pax cDNA related to Pax-2/5/8 and ancient gene duplications in the Pax family. *J Mol Evol*. 47:640–648.
- Jager M, Queinnee E, Houliston E, Manuel M. 2006. Expansion of the SOX gene family predated the emergence of the Bilateria. *Mol Phylogenet Evol*. 39:468–477.
- Kaestner KH, Knochel W, Martinez DE. 2000. Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev*. 14:142–146.
- Kamm K, Schierwater B, Jakob W, Dellaporta SL, Miller DJ. 2006. Axial patterning and diversification in the Cnidaria predate the Hox system. *Curr Biol*. 16:1–7.
- King N. 2004. The unicellular ancestry of animal development. *Dev Cell*. 7:313–325.
- King N, Carroll SB. 2001. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. *Proc Natl Acad Sci USA*. 98:15032–15037.

- King N, Hittinger CT, Carroll SB. 2003. Evolution of key cell signaling and adhesion protein families predates animal origins. *Science*. 301:361–363.
- Koopman P, Schepers G, Brenner S, Venkatesh B. 2004. Origin and diversity of the Sox transcription factor gene family: genome-wide analysis in *Fugu rubripes*. *Gene*. 328:177–186.
- Kruse M, Leys SP, Muller IM, Muller WEG. 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.
- Kusserow A, Pang K, Sturm C, et al. (11 co-authors). 2005. Unexpected complexity of the Wnt gene family in a sea anemone. *Nature*. 433:156–160.
- Lang BF, O'Kelly C, Nerad T, Gray MW, Burger G. 2002. The closest unicellular relatives of animals. *Curr Biol*. 12:1773–1778.
- Larroux C, Fahey B, Degnan SM, Adamski M, Rokhsar DS, Degnan BM. 2007. The NK homeobox gene cluster predates the origin of Hox genes. *Curr Biol*. 17:706–710.
- Larroux C, Fahey B, Liubicich D, Hinman VF, Gauthier M, Gongora M, Green K, Worheide G, Leys SP, Degnan BM. 2006. Developmental expression of transcription factor genes in a demosponge: insights into the origin of metazoan multicellularity. *Evol Dev*. 8:150–173.
- Li CW, Chen JY, Hua TE. 1998. Precambrian sponges with cellular structures. *Science*. 279:879–882.
- Magie CR, Pang K, Martindale MQ. 2005. Genomic inventory and expression of Sox and Fox genes in the cnidarian *Nematostella vectensis*. *Dev Genes Evol*. 215:618–630.
- Manuel M, Le Parco Y, 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.
- Martindale MQ. 2005. The evolution of metazoan axial properties. *Nat Rev Genet*. 6:917–927.
- Martinelli C, Spring J. 2003. Distinct expression patterns of the two T-box homologues Brachyury and Tbx2/3 in the placozoan *Trichoplax adhaerens*. *Dev Genes Evol*. 213:492–499.
- Martinelli C, Spring J. 2005. T-box and homeobox genes from the ctenophore *Pleurobrachia pileus*: comparison of Brachyury, Tbx2/3 and Tlx in basal metazoans and bilaterians. *FEBS Lett*. 579:5024–5028.
- Matus DQ, Pang K, Daly M, Martindale MQ. 2007. Expression of Pax gene family members in the anthozoan cnidarian, *Nematostella vectensis*. *Evol Dev*. 9:25–38.
- Mazet F, Yu JK, Liberles DA, Holland LZ, Shimeld SM. 2003. Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. *Gene*. 316:79–89.
- Medina M, Collins AG, Silberman JD, Sogin ML. 2001. Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. *Proc Natl Acad Sci USA*. 98:9707–9712.
- Miller DJ, Ball EE, Technau U. 2005. Cnidarians and ancestral genetic complexity in the animal kingdom. *Trends Genet*. 21:536–539.
- Miller DJ, Hayward DC, Reece-Hoyes JS, Scholten I, Catmull J, Gehring WJ, Callaerts P, Larsen JE, Ball EE. 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.
- Nichols SA, Dirks W, Pearse JS, King N. 2006. Early evolution of animal cell signaling and adhesion genes. *Proc Natl Acad Sci USA*. 103:12451–12456.
- Papaioannou VE. 2001. T-box genes in development: from hydra to humans. *Int Rev Cytol*. 207:1–70.
- Perovic S, Schroder HC, Sudek S, Grebenjuk VA, Batel R, Stifanic M, Muller IM, Muller WEG, Nicholas KB, Nicholas HB. 2003. Expression of one sponge Iroquois homeobox gene in primmorphs from *Suberites domuncula* during canal formation. *Evol Dev*. 5:240–250.
- Peterson KJ, Sperling EA. 2007. Poriferan ANTP genes: primitively simple or secondarily reduced? *Evol Dev*. 9:405–408.
- Phillips K, Luisi B. 2000. The virtuoso of versatility: pOU proteins that flex to fit. *J Mol Biol*. 302:1023–1039.
- Putnam N, Srivastava M, Hellsten U, et al. (19 co-authors). 2007. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science*. 317:86–94.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19:1572–1574.
- Roy SW, Gilbert W. 2006. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat Rev Genet*. 7:211–221.
- Ruiz-Trillo I, Burger G, Holland PWH, King N, Lang BF, Roger AJ, Gray MW. 2007. The origins of multicellularity: a multi-taxon genome initiative. *Trends Genet*. 23:113–118.
- Ryan AK, Rosenfeld MG. 1997. POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev*. 11:1207–1225.
- Ryan JF, Burton PM, Mazza ME, Kwong GK, Mullikin JC, Finnerty JR. 2006. The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes. Evidence from the starlet sea anemone, *Nematostella vectensis*. *Genome Biol*. 7:R64.
- Schepers GE, Teasdale RD, Koopman P. 2002. Twenty pairs of Sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev Cell*. 3:167–170.
- Seimiya M, Watanabe Y, 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.
- Simionato E, Ledent V, Richards G, Thomas-Chollier M, Kerner P, Coornaert D, Degnan BM, Vervoort M. 2007. Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. *BMC Evol Biol*. 7:33.
- Simpson TL. 1984. The cell biology of sponges. New York: Springer.
- Snell EA, Furlong RF, Holland PWH. 2001. Hsp70 sequences indicate that choanoflagellates are closely related to animals. *Curr Biol*. 11:967–970.
- Soullier S, Jay P, Poulat F, Vanacker JM, Berta P, Laudet V. 1999. Diversification pattern of the HMG and SOX family members during evolution. *J Mol Evol*. 48:517–527.
- Steenkamp ET, Wright J, Baldauf SL. 2006. The protistan origins of animals and fungi. *Mol Biol Evol*. 23:93–106.
- Sullivan J, Abdo Z, Joyce P, Swofford DL. 2005. Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation. *Mol Biol Evol*. 22:1386–1392.
- Takatori N, Hotta K, Mochizuki Y, Satoh G, Mitani Y, Satoh N, Satou Y, Takahashi H. 2004. T-box genes in the ascidian *Ciona intestinalis*: characterization of cDNAs and spatial expression. *Dev Dyn*. 230:743–753.
- Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. *Annu Rev Genet*. 38:615–643.
- Wada S, Tokuoka M, Shoguchi E, et al. (13 co-authors). 2003. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. II. Genes for homeobox transcription factors. *Dev Genes Evol*. 213:222–234.

- Wallberg A, Tholleson M, Farris JS, Jondelius U. 2004. The phylogenetic position of the comb jellies (Ctenophora) and the importance of taxonomic sampling. *Cladistics*. 20: 558–578.
- Wiens M, Mangoni A, D'Esposito M, Fattorusso E, Korchagina N, Schroder HC, Grebenjuk VA, Krasko A, Batel R, Muller IM, Muller WEG. 2003. The molecular basis for the evolution of the metazoan bodyplan: extracellular matrix-mediated morphogenesis in marine demosponges. *J Mol Evol*. 57:S60–S75.
- Wilson M, Koopman P. 2002. Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. *Curr Opin Genet Dev*. 12:441–446.
- Wolpert L. 1994. The evolutionary origin of development—cycles, patterning, privilege and continuity. *Development*. (Suppl):79–84.
- Yamada A, Pang K, Martindale MQ, Tochinai S. 2007. Surprisingly complex T-box gene complement in diploblastic metazoans. *Evol Dev*. 9:220–230.
- Yang Z. 1997. PAML: a program for package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 15:555–556.

Billie Swalla, Associate Editor

Accepted February 12, 2008