The Early Expansion and Evolutionary Dynamics of POU Class Genes

David A. Gold,[‡,1] Ruth D. Gates,[§,1] and David K. Jacobs[*,1]

[1]Department of Ecology and Evolution, University of California, Los Angeles

[‡]Present address: Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA.

[§]Present address: Hawaii Institute of Marine Biology, University of Hawaii–Manoa, Kaneohe, HI

Corresponding author: E-mail: djacobs@ucla.edu.

Associate editor: Stuart Newfeld

Abstract

The POU genes represent a diverse class of animal-specific transcription factors that play important roles in neurogenesis, pluripotency, and cell-type specification. Although previous attempts have been made to reconstruct the evolution of the POU class, these studies have been limited by a small number of representative taxa, and a lack of sequences from basally branching organisms. In this study, we performed comparative analyses on available genomes and sequences recovered through "gene fishing" to better resolve the topology of the POU gene tree. We then used ancestral state reconstruction to map the most likely changes in amino acid evolution for the conserved domains. Our work suggests that four of the six POU families evolved before the last common ancestor of living animals—doubling previous estimates—and were followed by extensive clade-specific gene loss. Amino acid changes are distributed unequally across the gene tree, consistent with a neofunctionalization model of protein evolution. We consider our results in the context of early animal evolution, and the role of *POU5* genes in maintaining stem cell pluripotency.

Key words: POU, Metazoa, homeobox, EvoDevo, stem cells, gene duplication.

Introduction

The POU genes represent a large class of DNA-binding transcription factors known for their roles in cell-type specification and developmental regulation (Ryan and Rosenfeld 1997; Phillips and Luisi 2000). The POU homolog *Oct-4* has been extensively studied, as it is the most critical of the four "Yamanaka factors" used to induce pluripotent stem cells in mammals (Niwa et al. 2000; Takahashi and Yamanaka 2006; Ng and Surani 2011). The POU name is an acronym derived from the mammalian genes *Pit-1*, *Oct-1*, and *Oct-2*, as well as the *Caenorhabditis elegans* gene *unc-86*, which all share a 150-amino acid region of high sequence similarity (Herr et al. 1988). Although POU genes have been identified in animals as diverse as sponges and humans, there exists strong conservation within the major domains (fig. 1). POU genes feature a modular, tripartite structure, consisting of an N-terminal POU-specific domain (POU_S), a C-terminal homeodomain (POU_{HD}), and a linker region of varying length connecting the two. The secondary structure of both POU_S and POU_{HD} domains consists of a series of α-helices, which make multiple contacts with DNA through hydrogen bonding with the phosphate backbone or directly to nucleotides (Jacobson et al. 1997; Reményi et al. 2001; Jauch et al. 2010; Esch et al. 2013). In both domains, the third helix serves as the recognition helix, binding to the major groove of DNA and making the majority of direct contacts with nucleotides (Assa-Munt et al. 1993; Dekker et al. 1993; Jacobson

et al. 1997). As figure 1 suggests, these contact regions are often, though not always, the most invariant sites within the POU class.

Despite this significant conservation, POU proteins are capable of generating high levels of conformational diversity through complex interactions with DNA and other transcription factors. POU genes form a variety of heterodimers and homodimers that can bind to noncontiguous DNA strands (Voss et al. 1991; Jacobson et al. 1997; Scully et al. 2000; Reményi et al. 2001; Rodda et al. 2005). It is common for POU paralogs to share partially overlapping functions (Erkman et al. 1996; Tichy et al. 2008), and certain POU knockouts can be rescued by a paralog that is not normally expressed in the region (Friedrich et al. 2005). Some POU genes take on multiple isoforms, which oppose each other in regulation, or work together to bind multiple trans factors (Konzak and Moore 1992; Lee and Salvaterra 2002; Theodorou et al. 2009). Even changes in the spacing between the two DNA binding domains can allow the same transcript to act as an activator in one scenario and a repressor in another (Scully et al. 2000). The last two amino acids of the homeodomain may be particularly important in driving dimerization (Reményi et al. 2001), which could be as important for the recognition of *cis*-regulatory modules as the DNA-binding interface (Jauch et al. 2010). Interestingly, this final dipeptide appears to be one of the most variable positions and may be a site involved in functional protein evolution (fig. 1).
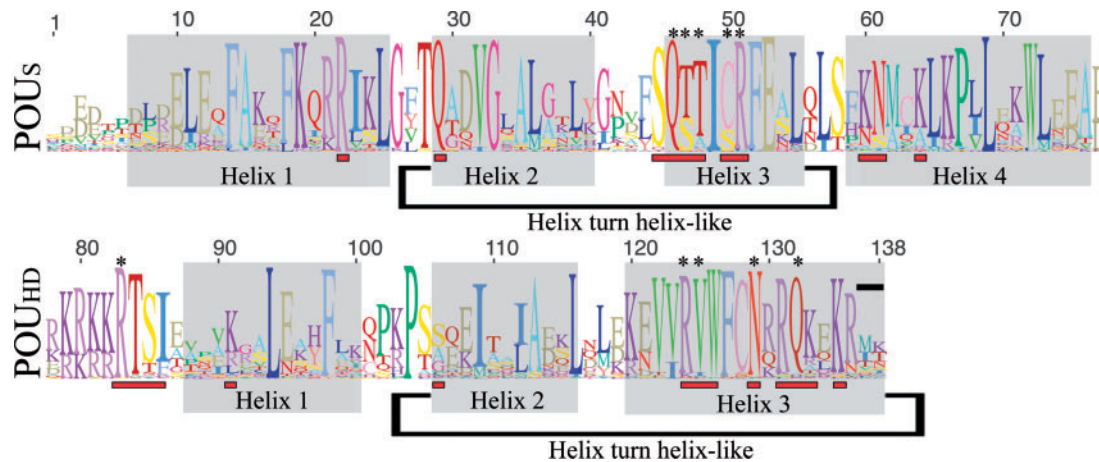
Open Access

Article

Fig. 1. Structure and variation within the POU$_S$ and POU$_{HD}$ domains. Probable α-helix subdomains are shaded in gray, amino acids known to make contact with DNA in at least one POU class are marked with a red bar, and the subset which makes direct contacts with nucleotides is marked with an asterisk (based on Klemm et al. 1994; Jacobson et al. 1997; Reményi et al. 2001, 2003; Jauch et al. 2010; Esch et al. 2013). The combined height of the amino acids at each position indicates the degree of sequence conservation, whereas the height of each individual amino acid indicates its relative frequency. This figure is based on the alignment we used for our phylogenetic analyses (see Materials and Methods and supplementary file S1, Supplementary Material online), and was created using the Sequence Logo function in Geneious.

Table 1. Division of POU Homologs into the Six Major Classes, Including Common Names.

	Mammalian Homologs	Drosophila Homologs	Caenorhabditis Homologs
POU1	POU1F1 (Pit-1)	None	None
POU2	POU2F1 (Oct-1), POU2F2 (Oct-2), POU2F3 (Oct-11)	pdm-1 (nubbin; dPOU-19; twain; dOct1) pdm-2 (miti-mere; dOct-2)	Ceh-18
POU3	POU3F1 (Oct-6; SCIP), POU3F2 (Oct-7; Brn-2), POU3F3 (Oct-8; Brn-1), POU3F4 (Oct-9; Brn-4; DFN3)	vvl (cf1a; drifter)	Ceh-6
POU4	POU4F1 (Brn-3a; RDC-1; Oct-T1), POU4F2 (Brn-3b; Brn-3.2), POU4F3 (Brn-3c; Brn-3.1; DFNA15)	acj6 (Ipou)	Unc-86
POU5	POU5F1(Oct-3; Oct-4), POU5F2 (SPRM-1), Pou2/V	None	None
POU6	POU6f1 (Brn-5; mPOU), POU6f2 (Emb; RPF-1)	pdm-3	None

Since their initial discovery, more than 1,000 POU sequences have been recovered from across the Metazoa. These are generally organized into six families (POU1–POU6). Multiple POU families have been described in every annotated animal genome, and many lineages, particularly vertebrates, have multiple paralogs in multiple families. The resultant extensive nomenclature is summarized in table 1. POU genes have only been recovered from Metazoa, suggesting that the POU$_S$ domain represents an animal novelty that was incorporated into a more ancient homeodomain containing gene during the early evolution of animals (Degnan et al. 2009). However, the presence of multiple, and often nonoverlapping, POU families in early-branching animal lineages makes rooting the POU gene tree difficult, and has led to conflicting topologies in gene tree reconstruction (Kamm and Schierwater 2007; Larroux et al. 2008; Ryan et al. 2010).

To better understand the diversity and evolution of POU genes, we adopted a comparative genomic approach to reconstruct the class topology. The results of this study were corroborated with gene fishing, using degenerate polymerase chain reaction (PCR) primers to capture novel POU homologs from a variety of understudied animal clades.

Finally, we used ancestral state reconstruction to track the most likely trajectory of POU sequence evolution. Taken together, our results suggest that four of the six major families of POU genes (POU6, POU1, POU3, and POU4) were present before the last common ancestor of all living animals, which is double the previous estimate. The POU families appear to have evolved primarily through gene duplication followed by neofunctionalization (Lynch and Conery 2000; Innan and Kondrashov 2010), where one paralog retains the ancestral amino acid sequence (and presumably aspects of the ancestral function), whereas the other duplicate incorporates significantly more nonsynonymous mutations.

Results

A Comparative Genomics Approach Resolves Many Aspects of the POU Gene Tree Topology

We began by surveying available animal genomes for POU-domain sequences. For phylogenetic analyses, we ultimately chose a subsample of taxa that included model laboratory animals as well as representatives of major clades from across the animal tree (discussed in detail in the Materials and Methods section). We employed both maximum likelihood
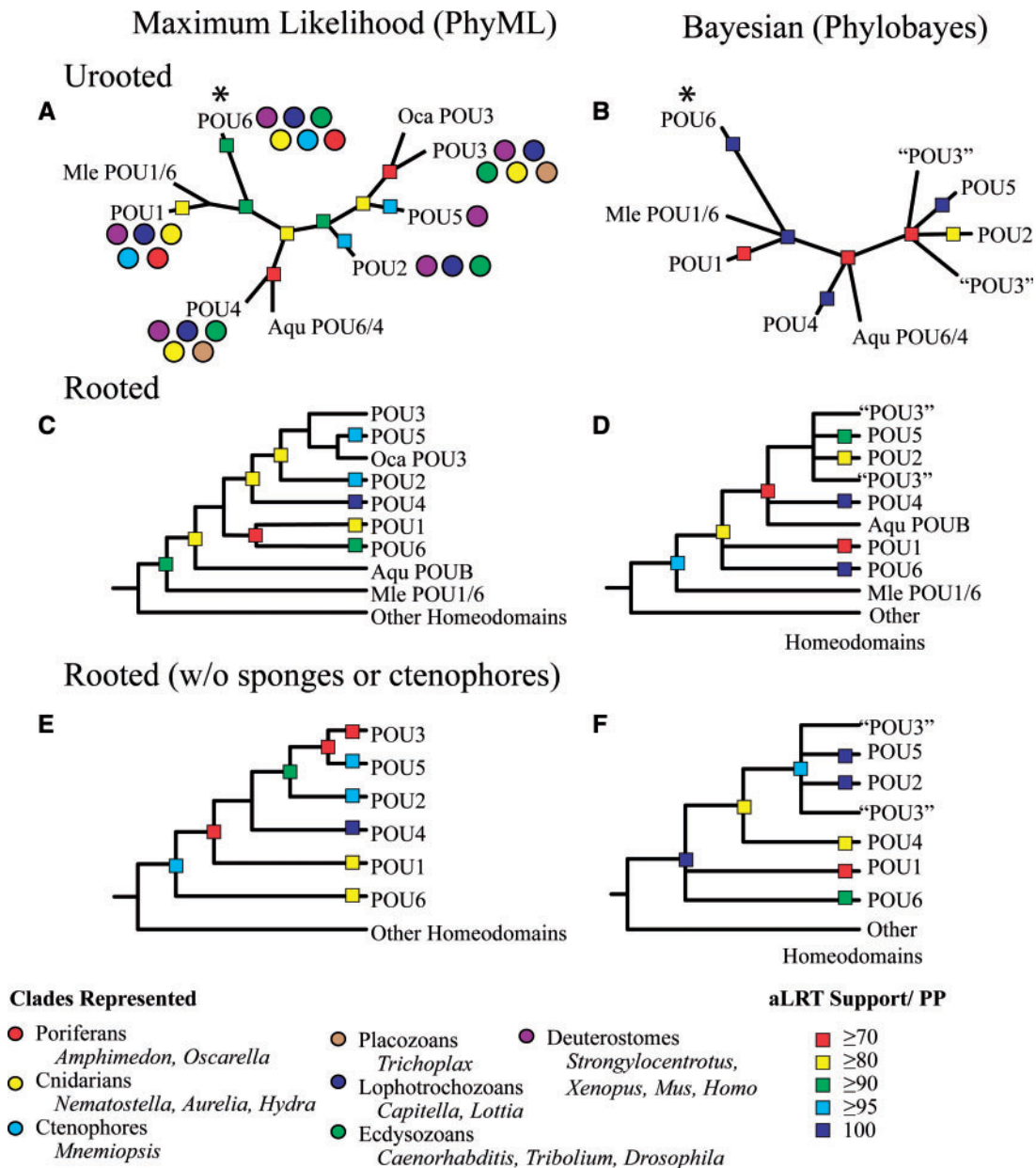
FIG. 2. Summary of ML and Bayesian reconstructions of our POU data set. See the Materials and Methods section and supplementary material, Supplementary Material online, for more complete information on taxon sampling and support values for all nodes. Genes with uncertain phylogenetic position from *Amphimedon* (Aqu), *Oscarella* (Oca), and *Mnemiopsis* (Mle) are singled out. (A, B) Unrooted topologies. The location of the midpoint root is marked with an asterisk. (C, D) Topologies that have been rooted by the inclusion of additional homeodomains. (E, F) Rooted topologies with all sequences from *Amphimedon*, *Oscarella*, and *Mnemiopsis* removed.

(ML) and Bayesian approaches to tree building. To account for the low phylogenetic support of sequences from the sponges *Amphimedon queenslandica* and *Oscarella carmela*, as well as the ctenophore *Mnemiopsis leidyi*, we also attempted tree reconstruction excluding these taxa. The results of these analyses are summarized in figure 2 (see supplementary figs. S1–S6, Supplementary Material online, for full trees).

Although we were unable to generate a single topology across all analyses, we were able to resolve some areas of uncertainty regarding the relationships between POU families. Previous studies rooting the POU class with homeodomains have recovered *POU6* as sister to all other POU families.

However, there has been disagreement whether the next family to diverge was *POU4* (e.g., Ryan et al. 2010) or *POU1* (e.g., Larroux et al. 2008; Millane et al. 2011); different rooting methods have also produced alternate topologies for the same data set (Kamm and Schierwater 2007). In contrast, all of our analyses identified *POU1* as the closest paralog to *POU6*, although the nature of that relationship varied, with *POU1* and *POU6* occasionally forming a sister clade or a polytomy (fig. 2C, D, and F). Still, there are several reasons to prefer *POU6* as the outgroup to the other extant POU families. Topologies illustrated in figure 2C and D include highly divergent and poorly supported sequences from *Mnemiopsis* and *Amphimedon*, which increases the

probability of long-branch attraction artifacts. When these sequences are removed, ML strongly supports POU6 as the outgroup (fig. 2E), whereas the Bayesian phylogeny generates a polytomy between POU6 and POU1 (fig. 2F). Given the small number of phylogenetically informative sites in our alignment, it is likely that the Bayesian approach lacks sufficient information to produce a topology with strong posterior support. Midpoint rooting on the unrooted topologies places the root within the POU6 family (fig. 2A and B), and an additional rooting process used during ancestral state reconstruction (discussed in detail in the Materials and Methods section) also supports POU6 as the outgroup.

Following POU6 and POU1, all of our analyses support POU4 as the next paralog to diverge. One gene from the sponge Amphimedon, which has previously been described as POUB (Larroux et al. 2008), has an affinity with POU4 in some of our analyses, and POU6 in others. This was followed by either a split between POU2 and POU3/5 (maximum-likelihood analyses; fig. 2A, C, and E) or a polytomous POU3 "bush," which includes monophyletic POU2 and POU5 classes (Bayesian analyses; fig. 2B, D, and F). As figure 2A illustrates, POU3 includes representatives from a number of basally branching animal taxa, including cnidarians, the placozoan Trichoplax adherans, and possibly the sponge Oscarella. POU2 was only recovered from bilaterian animals, whereas POU5 appears restricted to vertebrates, which supports the hypothesis that these families are products of more recent, clade-specific duplications.

Gene Fishing Recovers Putative POU3 and POU4 Classes in Sponges

By using a diverse selection of taxa, we uncovered several unanticipated results regarding the distribution of POU genes across the animals. First, our analyses provide good support for a POU3 homolog in Oscarella, as well as moderate support for a POU4 homolog in Amphimedon. This potentially doubles the number of POU families identified in the sponges, as previous analyses have only recognized POU6 and POU1 homologs in the Porifera (Larroux et al. 2008). A second surprise comes from the taxon distribution of the POU1 family. POU1 is present in early-branching animals, such as cnidarians, ctenophores, and sponges, as well as vertebrates and the chordate amphioxus (Jacobs and Gates 2003; Candiani et al. 2008). Our analyses suggest that POU1 is also present in the annelid Capitella teleta, but absent from all other sampled protostomes. The identification of POU1 in annelids is not new, as it has previously been described in the polychaete worm Platynereis dumerilii (Raible et al. 2005), but the hypothesis that the annelids are the only protostomes to retain this homolog has not been formalized. Indeed, although we were also able to recover a candidate POU1 from the genome of the leech Helobdella robusta (supplementary fig. S7, Supplementary Material online), we found no other protostome POU1 candidates in the NCBI database, or in any additional publically available protostome genomes.

To corroborate these results, we performed a gene fishing experiment, using degenerate PCR primers to amplify POU genes from a variety of understudied animal lineages (summarized in table 2). Family designations for the recovered genes were determined using Basic Local Alignment Search Tool (BLAST), alignments of the linker regions (supplementary fig. S7, Supplementary Material online), and phylogenetic analysis (supplementary fig. S8, Supplementary Material online). In our phylogenetic analyses, the linker was discarded, for although the region is

Table 2. Results of Gene Fishing Experiments.

Species Name	Phylum	Class	POU Genes Recovered
Acarnus erithacus	**Porifera**	**Demospongiae**	*POU1, POU4* (2)
Tethya aurantia	**Porifera**	**Demospongiae**	*POU4*
Spongilla sp.	**Porifera**	**Demospongiae**	*POU4*
Haliclona sp.	**Porifera**	**Demospongiae**	*POU1*
Rhabdocalyptus dawsoni	**Porifera**	**Hexactinellida**	*POU1* (2), *POU3*
Pleurobrachia sp.	**Ctenophora**	**Tentaculata**	*POU1*
Agaricia sp.	**Cnidaria**	**Anthozoa**	*POU1, POU3*
Anthopleura elegantissima	**Cnidaria**	**Anthozoa**	*POU3*
Fungia sp.	**Cnidaria**	**Anthozoa**	*POU1*
Pelagia colorata	**Cnidaria**	**Scyphozoa**	*POU4*
Convolutriloba sp.	**Acoelomorpha**	**Acoela**	*POU3, POU4*
Notoplana acticola	**Platyhelminthes**	**Turbellaria**	*POU3, POU4*
Stylochus tripartitus	**Platyhelminthes**	**Turbellaria**	*POU3, POU4* (3)
Alitta virens (formally *Nereis virens*)	**Annelida**	**Polychaeta**	*POU3, POU4*
Phragmatopoma californica	**Annelida**	**Polychaeta**	*POU4*
Hydroides sp.	**Annelida**	**Polychaeta**	*POU4*
Acanthina sp.	**Mollusca**	**Gastropoda**	*POU4*
Kelletia kelletii	**Mollusca**	**Gastropoda**	*POU3* (2), *POU4*
Transennella sp.	**Mollusca**	**Bivalvia**	*POU4*
Crassostrea gigas	**Mollusca**	**Bivalvia**	*POU3*

often conserved within POU families, it is difficult to homologize between them. However, this also makes the linker a good candidate for supporting family affinity, as it reduces the probability that our phylogenetic results are caused by convergent evolution within the otherwise largely invariant POU$_S$ and POU$_{HD}$ domains.

Our gene fishing results are consistent with comparative genomic inferences regarding clade-specific gene gain and loss of POU families. The recovery of a *POU3* homolog in the hexactinellid *Rhabdocalyptus dawsoni* and *POU4* homologs from the demosponges *Acarnus erithacus*, *Tethya aurantia*, and *Spongilla* sp. strongly supports our interpretation of the *Amphimedon* and *Oscarella* data, and collectively doubles the number of POU families known from the sponges. Because sponges commonly house a variety of symbiotic and commensal organisms (Brusca RC and Brusca GJ 2003), contamination is a concern. However, a number of observations argue against contamination. First, we obtained different *POU* genes from different sponge clades, with *POU4* being exclusive to demosponges (including *Amphimedon*), whereas *POU3* was recovered in the hexactenellid *Rhabdocalyptus* and the homoscleromorph *Oscarella*. Second, we obtained *POU4* genes from both the saltwater demosponges *Acarnus* and *Thethya*, as well as the freshwater sponge *Spongilla*. Third, although the sponge *POU* genes do not form monophyletic clades in phylogenetic analyses (supplementary fig. S8, Supplementary Material online), they also show no consistent affinity to any other animal phyla across NCBI BLAST searches. Consequently, we infer that *POU1*, *POU3*, *POU4*, and *POU6* were all present in the last common ancestor of sponges.

As with any gene fishing expedition, one must be cautious about making hard conclusions regarding gene absence. For example, we did not recover any *POU2* or *POU6* genes, even though *POU6* homologs have been identified in every annotated metazoan genome (excluding nematodes) and *POU2* in every annotated bilaterian genome. Therefore, it is unclear how we should interpret our failure to recover *POU1* genes from the annelids *Alitta*, *Phragmatopoma*, or *Hydroides*, despite their presence in the three annelid genomes. A previous gene fishing study also failed to recover *POU1* in the earthworm *Lumbricus terrestris* (Shah et al. 2000), which suggests that *POU1* has either been lost in many annelid lineages, or that annelid *POU1* genes are difficult to amplify with degenerate primers.

Still, the distribution of gene absences might provide some information. For example, our inability to find *POU3* or *POU4* homologs in *Pleurobrachia* supports the hypothesis that these families are absent from the two major ctenophore lineages (the Tentaculata and Nuda), and thus missing from the phylum altogether. Similarly, although *POU1* genes were recovered from cnidarians, sponges, and a ctenophore, none was recovered from flatworms, acoels, or molluscs, which is consistent with their absence in publically available genomes.

Ancestral State Reconstruction Supports a Pattern of Gene Duplication Followed by Protein Neofunctionalization

Resolving the topology and affinity of metazoan POU homologs allowed us to study the directionality of evolution within the POU$_S$ and POU$_{HD}$ domains. We used maximum-likelihood-based ancestral state reconstruction on a species-tree-corrected gene tree to track all amino acid changes that occurred at each node up to the common ancestor of the extant POU classes (fig. 3). Out of 173 amino acid changes, 117 occurred within an α-helix domain, and 95 changes were "significant," which we define as a shift from one major type of amino acid to another (i.e., positively charged [K, R, H], negatively charged [D, E], hydrophilic [S, T, N, Q, C, G, P], and hydrophobic [A, I, L, M, F, W, V, Y]). Our results suggest that mutations are not distributed evenly across the tree; after most bifurcations, one lineage appears to accumulate more amino acid changes than the other. To verify this pattern, we used the DIVERGE (v3.0) software package to perform pairwise comparisons between gene families (Gu et al. 2013; see Materials and Methods for more information). In our tests for differences in significant amino acid substitution rates, we determined that *POU4* was significantly different from *POU2*, *POU3*, or *POU5*, and that *POU5* was significantly different from *POU2*. Taken collectively, our analyses suggest three major times of significant increase in amino acid substitutions 1) When *POU6* split from the last common ancestor of all other POU families, 2) when *POU4* split from the last common ancestor of *POU2/3/5*, and 3) when *POU5* and *POU2* split from *POU3*. These results appear consistent with neofunctionalization models of gene duplication, which predict purifying selection on one gene duplicate, and a release of purifying selection combined with the evolution of a new adaptive function in the other duplicate (Innan and Kondrashov 2010). This would be in contrast to subfunctionalization models that predict relaxed purifying selection on both gene duplicates (Innan and Kondrashov 2010). These results also lead to some unintuitive conclusions regarding the similarity between extant POU families and their ancestral nodes. For example, although *POU6* appears to be the earliest branching family, *POU1* has accumulated far fewer significant amino acid substitutions during its evolution, presumably as a function of purifying selection and functional continuity. Similarly, although *POU2* appears to be sister to a *POU3/POU5* clade, *POU3* has accumulated far fewer significant substitutions since splitting from the common ancestor than either *POU2* or *POU5*. Similar to our presence–absence data described earlier, these results are consistent with the hypothesis that the last common ancestor of the *POU2/3/5* superfamily was *POU3*-like, and that *POU2* and *POU5* represent clade-specific duplications in bilaterians and vertebrates, respectively.

Given the modular nature of POU genes, we were curious whether there was any evidence of some modules evolving at different rates than others. Mutations appear to be fairly evenly distributed between POU$_S$ and POU$_{HD}$ domains at every node, but more mutations occur in α-helices in
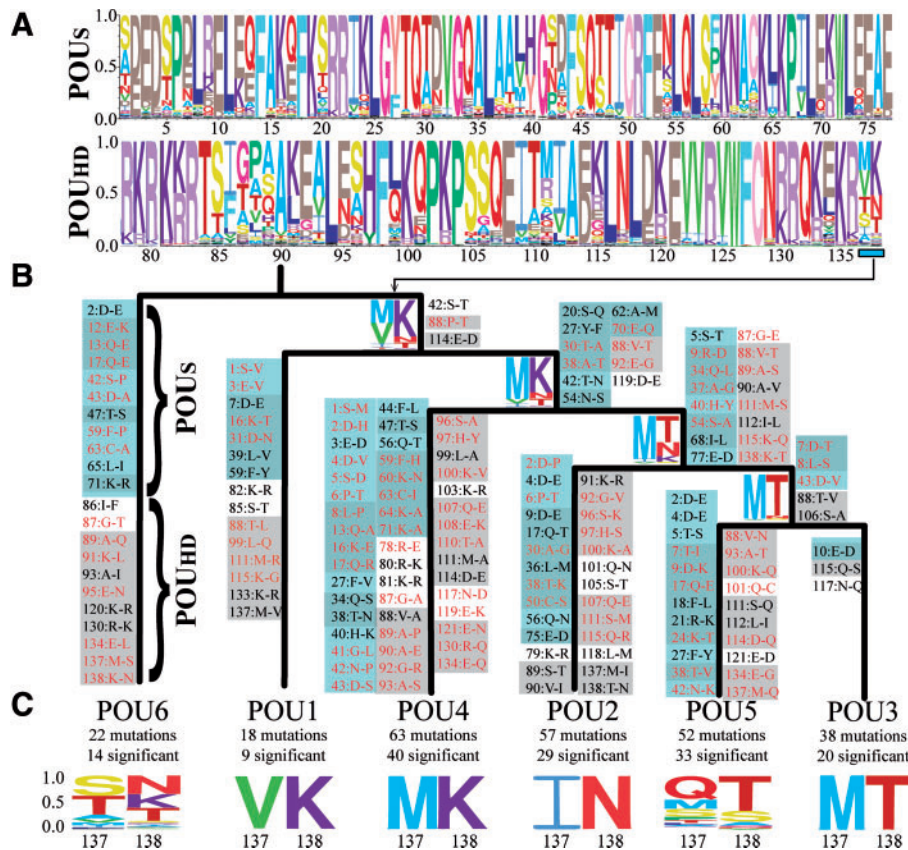
Fig. 3. Ancestral sequence and evolutionary trajectory of the POU_S and POU_{HD} domains. (A) Ancestral state reconstruction of the original POU_S and POU_{HD} domains. The probability of an amino acid being the ancestral state at each site is represented by the height of the letter, with the most probable peptide at the top. (B) Amino acid substitutions that occurred at each node, based on the most likely peptide at each node versus the ancestral node. Significant amino acid substitutions (i.e., moving between amino acids with positively charge, negatively charge, polar uncharged, or hydrophobic side chains) are colored in red. Mutations in the POU_S domain are highlighted in blue, and mutations occurring within an α-helix subdomain are highlighted in gray. (C) Total number of mutations that occurred between the common ancestor of each POU class and the ancestral POU sequence. The probability of the final dipeptide for the ancestor of each POU class is visualized at the bottom of the figure, and at the bifurcation of each ancestral node.

the POU_{HD} domain (66 of 87 amino acid changes for POU_{HD} vs. 51 of 86 changes for POU_S), even though the α-helix portion of POU_{HD} is smaller than in POU_S. As mentioned earlier, most amino acids known to play a direct role in DNA binding are largely invariant across the gene family. However, there are several significant amino acid substitutions in *POU4* (60K→N 64K→A) and *POU6* (91K→L) at positions involved in DNA binding in other POU classes. The consequences of these substitutions are unclear; the crystal structure has not been studied in *POU4* or *POU6*, so the impact that these substitutions have on DNA binding/bending is unknown. The protein folding prediction software I-TASSER (Roy et al. 2010) suggests that these substitutions have a minor impact on the shape of α-helices (fig. 4).

The last two amino acids of the POU_{HD} domain are distinct at each family-level bifurcation, and in four of the six cases there is a conserved combination of an aliphatic residue followed by a charged residue. This supports the hypothesis that this dipeptide is important in driving functional differentiation between the classes (Jauch et al. 2010). In *POU1*, amino acids 135–138 sit in an extended conformation beyond the terminus of the alpha helix (Jacobson et al. 1997), which is likely critical in driving dimerization in the

final dipeptide. Thus, there might be an implicit loss of dimerization specification in *POU6* and *POU5*, the two families that have lost this aliphatic/charged motif in the final dipeptide. Position 134, 2 bp upstream of this final dipeptide, also exhibits an interesting evolutionary pattern; at each bifurcation, the ancestral peptide (glutamic acid) is retained in one lineage, whereas the other lineage exhibits a significant substitution (*POU6* E→L, *POU4* E→Q, *POU5* E→G). Protein folding predictions of the POU_{HD} domain suggest that these substitutions have impacted the conformation of the recognition helix C-terminus (fig. 5); *POU6*, *POU1*, and *POU3* have retained the structure of the ancestral POU_{HD} protein (see fig. 4B), whereas *POU2*, *POU4*, and *POU5* exhibit an unwinding of the final dipeptide.

Discussion

The results of this study are summarized in figure 6. The POU gene tree is marked by an early diversification, followed by significant gene loss in multiple clades. Our study suggests that two major events shaped the evolution of POU genes at the family-level: The radiation of POU paralogs prior to the evolution of the Bilateria and the evolution of *POU5* in the vertebrates.
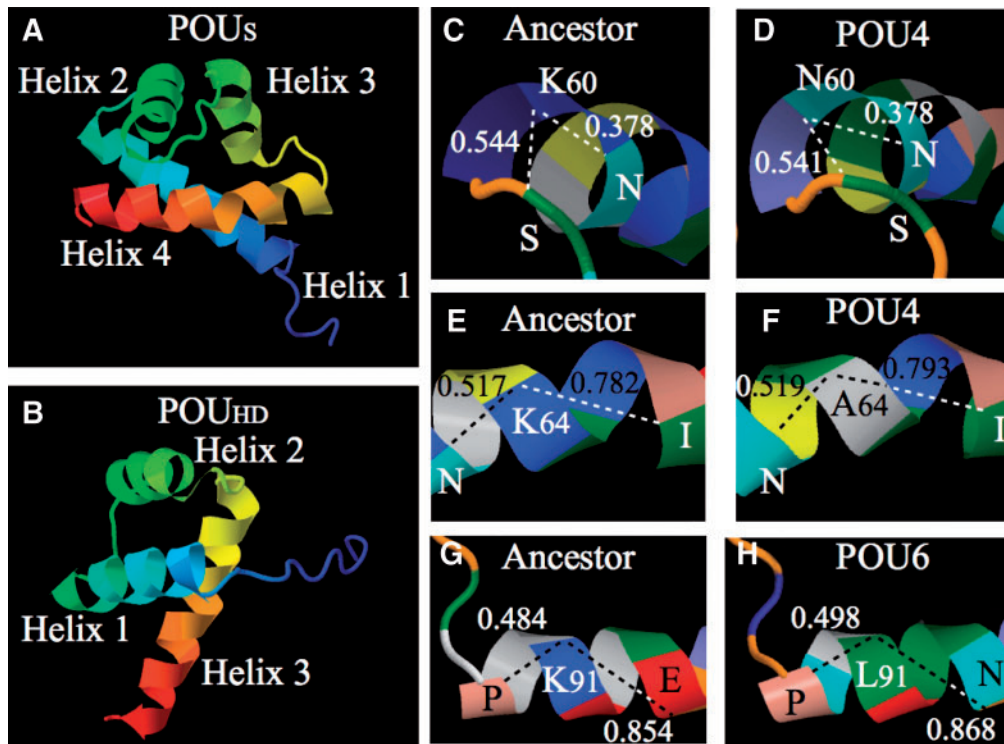
FIG. 4. Predicted structure of the ancestral POU$_S$ and POU$_{HD}$ domains, and the effects of significant amino acid substitutions on protein folding. Amino acid sequences were taken from the ancestral state reconstruction analysis, and folding was predicted using the I-TASSER server. The protein models were manipulated in Jmol. Structure of the ancestral (A) POU$_S$ and (B) POU$_{HD}$ domains. (C–H) Comparisons of protein folding in the ancestral sequences versus ancestral POU family members for significant amino acid substitutions to known DNA-binding sites. All measurements are in nanometers. (C) Ancestral condition of position 60. (D) Derived condition of position 60 in the last common ancestor of POU4. (E) Ancestral condition of position 64. (F) Derived condition of position 64 in the last common ancestor of *POU4*. (G) Ancestral condition of position 91. (H) Derived condition of position 91 in the last common ancestor of *POU6*.
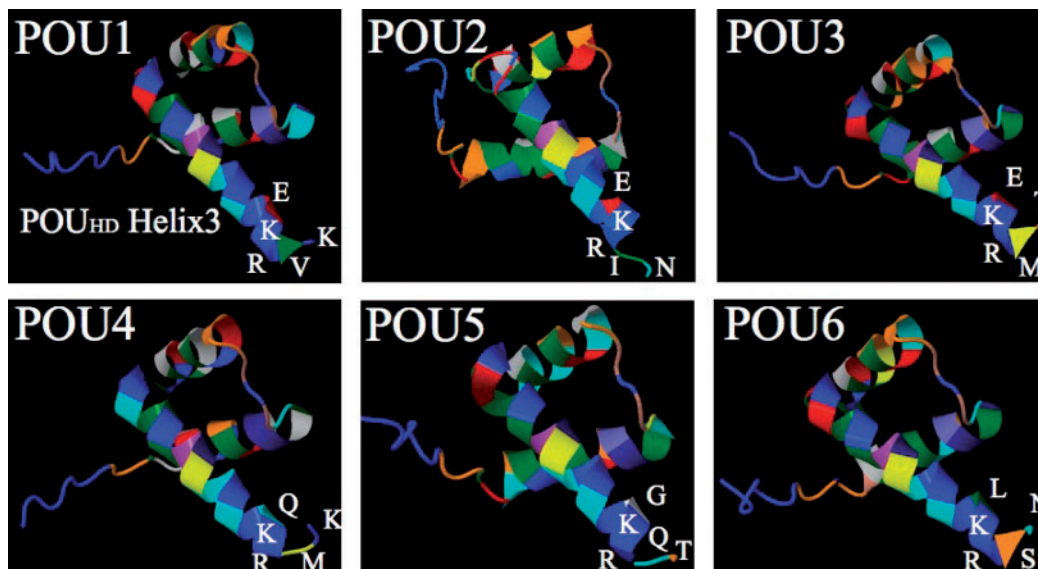

FIG. 5. Predicted folding of POU$_{HD}$ domains in the last common ancestor of each POU family, with a focus on the C-terminus. The last common ancestors of *POU1*, *POU3*, and *POU6* exhibit C-termini that are similar to the ancestral POU protein (see fig. 4B). In contrast, the last common ancestors of *POU2*, *POU4*, and *POU5* display an unwinding of the final dipeptide from the recognition α-helix.

If we accept sponges as the earliest-branching animal clade, then four of the six POU families had evolved before the last common ancestor of living animals (*POU6*, *POU1*, *POU3*, and *POU4*), with a fifth family (*POU2*) evolving in the stem lineage leading to modern bilaterian animals. However, several recent phylogenomic studies have suggested that ctenophores represent the oldest living animals (Ryan et al. 2013; Moroz et al. 2014). Our gene phylogeny is
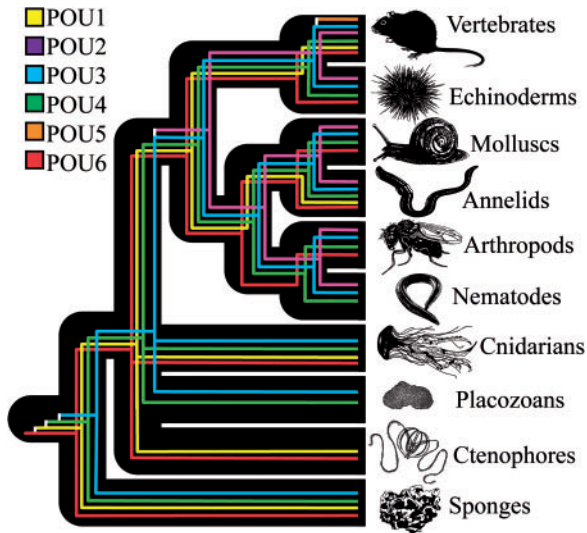
FIG. 6. Reconciliation of the POU gene tree and our animal phylogeny. This figure summarizes our hypothesis regarding how major POU families were gained and lost across the major animal phyla. Presence/absence results as they retain to each phylum were verified using BLAST searches on NCBI and through publically available genome data sets. Some of the animal images in this figure were modified under the creative commons agreement from the OpenLearn Tree of Life project (http://www.open.edu/openlearn/nature-environment/natural-history/tree-life, last accessed August 24, 2014). The base of the animal tree, particularly the placement of ctenophores and placozoans, is an active area of research. Opposing animal phylogenies to the one we present here have the potential to alter how rapid this initial expansion of POU classes was, but all currently debated animal topologies would still require the divergence of the first four POU genes prior to the evolution of the *Eumetozoa*.

consistent with this evolutionary scenario, as ctenophores have *POU6* and *POU1* homologs—the two original POU families determined in our analyses. In either evolutionary scenario, the lineage leading up to the sponges exhibited a dramatic increase in POU paralogs, which is in marked contrast to their simple bodyplans. This discrepancy could be explained by a secondary reduction of the sponge bodyplan from a more complex ancestor, or by the hypothesis that various eumetazoan organs regulated by POU genes share a deep common ancestry among the small number of cell types that sponges possess (reviewed in Jacobs and Gates 2001; Jacobs et al. 2007, 2010). Supporting the later hypothesis, reverse transcription PCR in *Ephydatia* shows that multiple POU genes are expressed during the period of canal formation, suggesting that they might serve overlapping roles in choanocyte formation (Seimiya et al. 1997). Additional work on resolving the base of the animal tree, alongside an increased study of POU genes in early-branching animals, should help resolve these competing hypotheses (e.g., see Nakanishi et al. 2010; Hroudova et al. 2012 for examples of POU-class gene expression in complex cnidarians).

Our results also constrain the evolution of the *POU5* family to the vertebrates, which is particularly germane to understanding of the evolutionary role of *POU5f1/Oct-4* in maintaining stem cell pluripotency. Frankenberg and Renfree

(2013) previously suggested that *POU5* is a gnathostome novelty, as the gene is missing from the lamprey, lancelet, and tunicate genomes. However, significant lineage-specific gene loss has occurred following a pregnathostome duplication event; the *POU5* paralog *Pou2* (not to be confused with the *POU2* family; see table 1) was lost in eutherian mammals and squamates, whereas the paralog *POU5f1/Oct-4* was lost in teleost fish, archosaurs, and anurans (Niwa et al. 2008; Frankenberg et al. 2010; Frankenberg and Renfree 2013). It is uncertain whether all *POU5*-family genes play a functional role in stem cell pluripotency, or whether that function is restricted to *POU5f1/Oct-4*. Morrison and Brickman (2006) found that zebrafish *Pou2* has little to no ability to support the self-renewal of mouse embryonic stem cells, but *Pou2* from other taxa—including axolotl, medaka (Tapia et al. 2012), and to a lesser extent chicken (Lavial et al. 2007)—can regulate pluripotency in mammalian stem cells. This suggests that all vertebrate *POU5* paralogs can play a role in regulating pluripotency (Onichtchouk 2012; Tapia et al. 2012). However, major differences in mammalian and nonmammalian stem cell dynamics exist (Fernandez-Tresguerres et al. 2010; Esch et al. 2013), and the extent that *Pou2* can co-opt the role of *POU5f1* remains unclear. For example, the *Pou2* homolog from the medaka fish can reprogram human fibroblasts into pluripotent stem cells (Tapia et al. 2012), but replacing the linker region of human *POU5f1* with the medaka linker abolishes the gene's ability to induce pluripotency (Esch et al. 2013).

Assuming all vertebrate *POU5* genes play some collective role in cell pluripotency, is it possible that they inherited this function from an ancestral invertebrate POU paralog? POU genes are involved in stem cell dynamics of the planarian worm *Schmidtea mediterranea* (Onal et al. 2012) as well as the cnidarian *Hydractinia echinata* (Millane et al. 2011). Although these taxa were not included in our analyses, the results of our study would suggest that these invertebrate genes do not represent genuine *POU5* orthologs. Adding these proteins to our phylogenetic alignment suggests that candidate *Schmidtea* and *Hydractinia* genes represent *POU4* and *POU3* paralogs, respectively (supplementary fig. S9, Supplementary Material online). Additionally, these invertebrate POU sequences lack the α-helix domain that exists in the linker of amniote *POU5* peptides (supplementary fig. S10, Supplementary Material online), which is necessary for inducing pluripotency in mammalian cells according to Esch et al. (2013). This could be interpreted as further evidence for the independent evolution of invertebrate and mammalian stem cells (Gold and Jacobs 2013), although additional regulatory and epigenetic similarities between planarian and mammalian stem cells suggest that there might still be deep underlying conservation of the pluripotency network, even if disparate POU paralogs are ultimately utilized in different animal lineages (Onal et al. 2012). Such uncertainty only reinforces the point that we are just beginning to appreciate how dynamically evolving protein families become integrated into ancestral and novel genetic networks.

In an era of comparative and functional genomics, the elucidation of gene trees will prove just as important as

the resolution of species trees. Our results suggest that POU genes have undergone a complex series of lineage-specific duplication and loss, which will only be fully clarified by using an extensive and diverse sampling of animals. Greater study of POU genes in animal clades such as sponges, cnidarians, ctenophores, and annelids should help elucidate the functional evolution of the POU class, and will be critical to determining cellular homologies between the invertebrates and vertebrates. This will likely prove important for establishing invertebrate model systems for a variety of developmental phenomena, including neurogenesis and stem cell dynamics.

Materials and Methods

Data Collection and Alignment

For our phylogenetic analysis, we searched for POU sequences from the publically available genomes of *A. queenslandica* (demosponge), *O. carmella* (homoscleromorph sponge), *Hydra magnipapillata* (cnidarian), *Nematostella vectensis* (cnidarian), *M. leidiy* (ctenophore), *T. adherans* (placozoan), *Cap. teleta* (annelid), *Lottia gigantea* (mollusc), *Cae. elegans* (nematode), *Tribolium castaneum* (arthropod), *Drosophila melanogaster* (arthropod), *Strongylocentrotus purpuratus* (echinoderm), *Xenopus tropicalis* (vertebrate), *Mus musculus* (vertebrate), and *Homo sapiens* (vertebrate). We also included sequences based on our unpublished transcriptomic data for *Aurelia* sp.1 (cnidarian). Databases were queried using the Human *Pit-1* POU$_S$ domain: DSPEIRELEKFANEFKVRRIKL-GYTQTNVGEALAAVHGSEFSQTTICRFENLQLSFKNACKLK-AILSKWL. Sequences from *Hydra*, *Nematostella*, *Lottia*, *Caenorhabditis*, *Tribolium*, *Drosophila*, *Strongylocentrotus*, *Xenopus*, *Mus*, and *Homo* were collected from Metazome (http://www.metazome.net/, last accessed August 24, 2014) using BLASTP against the predicted proteomes. For *A. queenslandica*, we used TBLASTN against the Spongezome Metazome database (http://spongezome.metazome.net, last accessed August 24, 2014). Sequences from *Capitella* and *Trichoplax* were collected from the Joint Genome Institute using BLASTP. *Oscarella* sequences were obtained from the predicted protein models (OCAR G-PEP) available on the Compagen website (Hemmrich and Bosch 2008). *Mnemiopsis* proteins were recovered using BLASTP against the protein models (v2.2) available at the NIH *Mnemiopsis* Genome Project Portal (http://research.nhgri.nih.gov/mnemiopsis/blast/, last accessed August 24, 2014). The proteins we recovered for *Amphimedon* and *Mnemiopsis* are not identical to those that have been previously published (Larroux et al. 2008; Ryan et al. 2010); we interpreted this as resulting from improvements in the respective genome/proteome assemblies, and chose to work with the POU proteins we recovered. For the *Capitella POU1* gene, we recovered an alternative transcript using TBLASTN against the genome, which contained part of the POU$_S$ domain missing from the predicted peptide; this longer sequence was used for subsequent analyses. Accession numbers for all genes are included in the alignment, available as supplementary file S1, Supplementary Material online.

Sequences were aligned using the MUSCLE algorithm (Edgar 2004) in Geneious (v.5.4.6., created by Biomatters and available from http://www.geneious.com/, last accessed August 24, 2014). The alignment was edited by hand and restricted to the POU$_S$ and POU$_{HD}$ domains. Redundant sequences, unalignable sequences, and uninformative (unique) insertions were manually removed. The final alignment is available as supplementary file S1, Supplementary Material online.

Phylogenic Analyses

We used ProtTest3 (Darriba et al. 2011) to determine the best-fitting model of amino acid evolution for our alignments. The program strongly preferred the LG model in conjunction with a gamma distribution and four substitution rate categories. We used PhyML (Guindon et al. 2010) to perform maximum-likelihood estimates; node values were determined using approximate likelihood ratio tests (aLRT) with Shimodaira-Hasegawa (SH)-like support. We used PhyloBayes 3.3 (Lartillot et al. 2009) for our Bayesian analyses. PhyloBayes was ran with the commands "pb -d {Alignment} -lg -nchain 2 100 0.3 100 {Output}," which means that the program ran two chains in parallel, checking every 100 cycles to see whether all discrepancies between the two chains were less than or equal to 0.3, and that all effective sizes were larger than 100. The runs were automatically stopped once these conditions were met.

Gene Fishing

Animals were starved for at least 48 h prior to sampling. Genomic DNA was extracted using either a classic C-Tab protocol (Bebenek et al. 2004) or the DNeasy Kit (Qiagen). Degenerate PCR primers were designed to capture conserved regions of the POU$_S$ and POU$_{HD}$ domains (F1: CAA GCA GMG RMG VAT MAA RYT RGG; F2: CTB ACB YTB TCV CAY AAC AAC ATG; R1: CKY TTY TCN GGH GCV GCR ATR S; R2: RTT RCA RAA CCA SAC BCK MAC MAC). For each gene recovered, we used BLAST as well as phylogenetic analysis (supplementary fig. S8, Supplementary Material online) to assign a family identity to each gene. These family identities were supported with MUSCLE-based alignments of the linker regions, performed in Geneious (supplementary fig. S7, Supplementary Material online).

Ancestral State Reconstruction

Accurate ancestral state reconstruction requires a gene tree that is consistent with the species tree, which is not generally the result of a standard ML or Bayesian analysis. To generate a gene tree informed by the species tree, we created an additional topology using TreeBeST (Vilella et al. 2008). Because of uncertainties in the topology at the base of the animal tree, we removed *Oscarella*, *Mnemiopsis*, and *Trichoplax* from our ancestral state reconstruction. We invoked the commands "treebest best -f {Input tree} -o {Output tree} {Alignment}," which resulted in a gene tree that was reconciled with the species tree, rooted by minimizing the number of duplications and losses, and bootstrapped 100 times.

The output of TreeBeST did a good job at creating a gene tree that was consistent with the species tree, with one exception. It produced a topology in the POU6 family where all bilaterian invertebrate POU6 genes were derived from one of the two vertebrate homologs (data not shown). This scenario would require a duplication of POU6 at the base of the bilaterians, with the same paralog being lost in every invertebrate clade. A more likely scenario is that there was a single POU6 gene in invertebrate bilaterians, and this gene duplicated in the vertebrates; a scenario that occurred in POU2, POU3, and POU4 families. To modify the TreeBeST topology and get adjusted initial branch lengths, we ran the original POU alignment through BEUTi/BEAST (Drummond et al. 2012) for 500,000 generations, constraining every node as a prior to reflect the TreeBeST topology with our modification. For this analysis, we ultimately decided to exclude Amphimedon POUB and a Nematostella POU3 paralog, as both sequences were highly derived, and we wished to avoid biasing our ancestral states with these sequences. However, it is worth noting that when Amphimedon POUB was included in the TreeBest analysis, it grouped with POU4. The final tree used for ancestral state reconstruction is available as supplementary file S2, Supplementary Material online.

The modified consensus tree and the relevant protein alignment were imported into the FastML server (Ashkenazy et al. 2012), using the LG substitution model, optimization of branch lengths, and gamma distribution options. The probabilities of the ancestral POU sequence were graphically exported using the WebLogo (Crooks et al. 2004) function in FastML and recolored in Adobe Illustrator to be consistent with MacClade-style amino acid coloration (as seen in fig. 1). The most probable ancestral state at each relevant node was exported from the FastML output, and amino acid substitutions were determined manually.

Tests of Asymmetric Functional Divergence

We tested for functional divergence following gene duplication using the DIVERGE (v3.0) package (Gu et al. 2013). The tree used for ancestral state reconstruction (supplementary file S2, Supplementary Material online) and the relevant sequences were imported into DIVERGE to calculate the coefficient of functional divergence (or $\ominus$) for each pairwise comparison between POU families. We performed tests for type-I functional divergence (differences in amino acid variability between POU families) and type-II functional divergence (differences in significant amino acid substitutions between families, using the "significance" criteria described earlier). Z values were calculated by dividing $\ominus$ by the standard error, and P values were determined using a two-tailed Z-score test (normal distribution test). The results of all tests are available in supplementary figure S11, Supplementary Material online.

Supplementary Material

Supplementary figures S1–S11, file S1, and file S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

References

Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi G, Zomer O, Pupko T. 2012. FastML: a web server for probabilistic reconstruction of ancestral sequences. Nucleic Acids Res. 40: W580–W584.

Assa-Munt N, Mortishire-Smith RJ, Aurora R, Herr W, Wright PE. 1993. The solution structure of the Oct-1 POU-specific domain reveals a striking similarity to the bacteriophage λ repressor DNA-binding domain. Cell 73:193–205.

Bebenek IG, Gates RD, Morris J, Hartenstein V, Jacobs DK. 2004. sine oculis in basal Metazoa. Dev Genes Evol. 214:342–351.

Brusca RC, Brusca GJ 2003. Invertebrates. 2nd ed. Sunderland (MA): Sinauer Associates.

Candiani S, Holland ND, Oliveri D, Parodi M, Pestarino M. 2008. Expression of the amphioxus Pit-1 gene (AmphiPOU1F1/Pit-1) exclusively in the developing preoral organ, a putative homolog of the vertebrate adenohypophysis. Brain Res Bull. 75:324–330.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res. 14:1188–1190.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27: 1164–1165.

Degnan BM, Vervoort M, Larroux C, Richards GS. 2009. Early evolution of metazoan transcription factors. Curr Opin Genet Dev. 19:591–599.

Dekker N, Cox M, Boelens R, Verrijzer CP, van der Vliet PC, Kaptein R. 1993. Solution structure of the POU-specific DNA-binding domain of Oct-1. Nature 362:852–855.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 29: 1969–1973.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Erkman L, McEvilly RJ, Luo L, Ryan AK, Hooshmand F, O'Connell SM, Keithley EM, Rapaport DH, Ryan AF, Rosenfeld MG. 1996. Role of transcription factors a Brn-3.1 and Brn-3.2 in auditory and visual system development. Nature 381:603–606.

Esch D, Vahokoski J, Groves MR, Pogenberg V, Cojocaru V, Vom Bruch H, Han D, Drexler HC, Araúzo-Bravo MJ, Ng CK, et al. 2013. A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat Cell Biol. 15:295–301.

Fernandez-Tresguerres B, Cañon S, Rayon T, Pernaute B, Crespo M, Torroja C, Manzanares M. 2010. Evolution of the mammalian embryonic pluripotency gene regulatory network. Proc Natl Acad Sci U S A. 107:19955–19960.

Frankenberg S, Pask A, Renfree MB. 2010. The evolution of class V POU domain transcription factors in vertebrates and their characterisation in a marsupial. Dev Biol. 337:162–170.

Frankenberg S, Renfree MB. 2013. On the origin of POU5F1. BMC Biol. 11:56.

Friedrich RP, Schlierf B, Tamm ER, Bösl MR, Wegner M. 2005. The class III POU domain protein Brn-1 can fully replace the related Oct-6 during schwann cell development and myelination. Mol Cell Biol. 25:1821–1829.

Gold DA, Jacobs DK. 2013. Stem cell dynamics in Cnidaria: are there unifying principles? Dev Genes Evol. 223:53–66.

Gu X, Zou Y, Su Z, Huang W, Zhou Z, Arendsee Z, Zeng Y. 2013. An update of DIVERGE software for functional divergence analysis of protein family. *Mol Biol Evol.* 30:1713–1719.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.

Hemmrich G, Bosch TCG. 2008. Compagen, a comparative genomics platform for early branching metazoan animals, reveals early origins of genes regulating stem-cell differentiation. *Bioessays* 30:1010–1018.

Herr W, Sturm RA, Clerc RG, Corcoran LM, Baltimore D, Sharp PA, Ingraham HA, Rosenfeld MG, Finney M, Ruvkun G, et al. 1988. The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and *Caenorhabditis elegans* unc-86 gene products. *Genes Dev.* 2:1513–1516.

Hroudova M, Vojta P, Strnad H, Krejcik Z, Ridl J, Paces J, Vlcek C, Paces V. 2012. Diversity, phylogeny and expression patterns of Pou and Six homeodomain transcription factors in hydrozoan jellyfish *Craspedacusta sowerbyi*. *PLoS One* 7:e36420.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet.* 11:4–108.

Jacobs DK, Gates RD. 2001. Evolution of POU/homeodomains in basal metazoa: implications for the evolution of sensory systems and the pituitary. *Dev Biol.* 235:241.

Jacobs DK, Gates RD. 2003. Developmental genes and the reconstruction of metazoan evolution—implications of evolutionary loss, limits on inference of ancestry and type 2 errors. *Integr Comp Biol.* 43:11–18.

Jacobs D, Nakanishi N, Yuan D. 2007. Evolution of sensory structures in basal metazoa. *Integr Comp Biol.* 47:712–723.

Jacobs DK, Gold DA, Nakanishi N, Yuan D, Camara A, Nichols SA, Hartenstein V. 2010. Basal metazoan sensory evolution. In: DeSalle R, Schierwater B, editors. Key transitions in animal evolution. Enfield (NH): Science Publishers. p. 175–196.

Jacobson EM, Li P, Leon-del-Rio A, Rosenfeld MG, Aggarwal AK. 1997. Structure of Pit-1 POU domain bound to DNA as a dimer: unexpected arrangement and flexibility. *Genes Dev.* 11:198–212.

Jauch R, Choo SH, Ng CKL, Kolatkar PR. 2010. Crystal structure of the dimeric Oct6 (POU3f1) POU domain bound to palindromic MORE DNA. *Proteins* 79:674–677.

Kamm K, Schierwater B. 2007. Ancient linkage of a POU class 6 and an anterior Hox-like gene in Cnidaria: implications for the evolution of homeobox genes. *J Exp Zool B Mol Dev Evol.* 308:777–784.

Klemm JD, Rould MA, Aurora R, Herr W, Pabo CO. 1994. Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell* 77:21–32.

Konzak KE, Moore DD. 1992. Functional isoforms of Pit-1 generated by alternative messenger RNA splicing. *Mol Endocrinol.* 6:241–247.

Larroux C, Luke GN, Koopman P, Rokhsar DS, Shimeld SM, Degnan BM. 2008. Genesis and expansion of metazoan transcription factor gene classes. *Mol Biol Evol.* 25:980–996.

Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288.

Lavial F, Acloque H, Bertocchini F, Macleod DJ, Boast S, Bachelard E, Montillet G, Thenot S, Sang HM, Stern CD, et al. 2007. The Oct4 homologue PouV and Nanog regulate pluripotency in chicken embryonic stem cells. *Development* 134:3549–3563.

Lee M-H, Salvaterra PM. 2002. Abnormal chemosensory jump 6 is a positive transcriptional regulator of the cholinergic gene locus in *Drosophila* olfactory neurons. *J Neurosci.* 22:5291–5299.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.

Millane RC, Kanska J, Duffy DJ, Seoighe C, Cunningham S, Plickert G, Frank U. 2011. Induced stem cell neoplasia in a cnidarian by ectopic expression of a POU domain transcription factor. *Development* 138:2429–2439.

Moroz LL, Kocot KM, Citarella MR, Dosung S, Norekian TP, Povolotskaya IS, Grigorenko AP, Dailey C, Berezikov E, Buckley KM, et al. 2014. The ctenophore genome and the evolutionary origins of neural systems. *Nature* 510:109–114.

Morrison GM, Brickman JM. 2006. Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development. *Development* 133:2011–2022.

Nakanishi N, Yuan D, Hartenstein V, Jacobs DK. 2010. Evolutionary origin of rhopalia: insights from cellular-level analyses of Otx and POU expression patterns in the developing rhopalial nervous system. *Evol Dev.* 12:404–415.

Ng H-H, Surani MA. 2011. The transcriptional and signalling networks of pluripotency. *Nature* 13:490–496.

Niwa H, Sekita Y, Tsend-Ayush E, Grützner F. 2008. Platypus Pou5f1 reveals the first steps in the evolution of trophectoderm differentiation and pluripotency in mammals. *Evol Dev.* 10:671–682.

Niwa H, Smith AG, Miyazaki J-I. 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation, or self-renewal of ES cells. *Nat Genet.* 24:372–376.

Onal P, Grün D, Adamidi C, Rybak A, Solana J, Mastrobuoni G, Wang Y, Rahn HP, Chen W, Kempa S, et al. 2012. Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells. *EMBO J.* 31:2755–2769.

Onichtchouk D. 2012. Pou5f1/oct4 in pluripotency control: insights from zebrafish. *Genesis* 50:75–85.

Phillips K, Luisi B. 2000. The virtuoso of versatility: POU proteins that flex to fit. *J Mol Biol.* 302:1023–1039.

Raible F, Tessmar-Raible K, Osoegawa K, Wincker P, Jubin C, Balavoine G, Ferrier D, Benes V, de Jong P, Weissenbach J, et al. 2005. Vertebrate-type intron-rich genes in the marine annelid *Platynereis dumerilii*. *Science* 310:1325–1326.

Reményi A, Lins K, Nissen LJ, Reinbold R, Schöler HR, Wilmanns M. 2003. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.* 17:2048–2059.

Reményi A, Tomilin A, Pohl E, Lins K, Philippsen A, Reinbold R, Schöler HR, Wilmanns M. 2001. Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. *Mol Cell.* 8:569–580.

Rodda DJ, Chew J-L, Lim L-H, Loh Y-H, Wang B, Ng H-H, Robson P. 2005. Transcriptional regulation of Nanog by OCT4 and SOX2. *J Biol Chem.* 280:24731–24737.

Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protocol.* 5:725–738.

Ryan A, Rosenfeld M. 1997. POU domain family values: flexibility, partnerships, and developmental codes. *Genes Dev.* 11:1207–1225.

Ryan JF, Pang K, Comparative Sequencing Program NISC, Mullikin JC, Martindale MQ, Baxevanis AD. 2010. The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. *EvoDevo.* 1:9.

Ryan JF, Pang K, Schnitzler CE, Nguyen A-D, Travis RM, Moreland RT, Simmons DK, Koch BJ, Francis WR, Havlak P, et al. 2013. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* 342:1242592.

Scully KM, Jacobson EM, Jepsen K, Lunyak V, Viadiu H, Carrière C, Rose DW, Hooshmand F, Aggarwal AK, Rosenfeld MG. 2000. Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science* 290:1127–1131.

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.

Shah D, Aurora D, Lance R, Stuart GW. 2000. POU genes in metazoans: homologs in sea anemones, snails, and earthworms. *DNA Seq.* 11:457–461.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126:663–676.

Tapia N, Reinhardt P, Duemmler A, Wu G, Araúzo-Bravo MJ, Esch D, Greber B, Cojocaru V, Rascon CA, Tazaki A, et al. 2012. Reprogramming to pluripotency is an ancient trait of vertebrate Oct4 and Pou2 proteins. *Nat Commun.* 3:1279.

Theodorou E, Dalembert G, Heffelfinger C, White E, Weissman S, Corcoran L, Snyder M. 2009. A high throughput embryonic stem cell screen identifies Oct-2 as a bifunctional regulator of neuronal differentiation. *Genes Dev.* 23:575–588.

Tichy AL, Ray A, Carlson JR. 2008. A new *Drosophila* POU gene, pdm3, acts in odor receptor expression and axon targeting of olfactory neurons. *J Neurosci.* 28:7121–7129.

Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. 2008. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* 19: 327–335.

Voss JW, Wilson L, Rosenfeld MG. 1991. POU-domain proteins Pit-1 and Oct-1 interact to form a heteromeric complex and can cooperate to induce expression of the prolactin promoter. *Genes Dev.* 5: 1309–1320.