The *Amphimedon queenslandica* genome and the evolution of animal complexity

Mansi Srivastava[1][2], Oleg Simakov[3]; Jarrod Chapman[4], Bryony Fahey[5], Marie E.A. Gauthier[5][6], Therese Mitros[1], Gemma S. Richards[5][11]; Cecilia Conaco[7], Michael Dacre[8], Uffe Hellsten[4], Claire Larroux[5], Nicholas H. Putnam[9], Mario Stanke[10], Maja Adamska[5][11], Aaron Darling[12], Sandie M. Degnan[5], Todd H. Oakley[13], David C. Plachetzki[13], Yufeng Zhai[8]; Marcin Adamski[5][11], Andrew Calcino[5], Scott F. Cummins[5], David M. Goodstein[4], Christina Harris[5], Daniel J. Jackson[5][14], Sally P. Leys[15], Shengqiang Shu[4], Ben J. Woodcroft[5]; Michel Vervoort[16], Kenneth S. Kosik[7], Gerard Manning[8], Bernard M. Degnan[5] and Daniel S. Rokhsar[1,4]

1. Center for Integrative Genomics and Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA.
2. Current Address: Whitehead Institute for Biomedical Research, Cambridge, MA 02138, USA.
3. Molecular Evolution Genomics, University of Heidelberg, Heidelberg, Germany.
4. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA.
5. School of Biological Sciences, The University of Queensland, Brisbane Queensland 4072, Australia.
6. Current Address: Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstr. 190, CH-8057, Zurich, Switzerland
7. Neuroscience Research Institute, University of California, Santa Barbara, Santa Barbara, California 93106, USA.
8. Razavi Newman Center for Bioinformatics, Salk Institute for Biological Studies, La Jolla, CA 92037, USA.
9. Department of Ecology and Evolutionary Biology, Rice University, 6100 Main Street, Houston TX 77005, USA.
10. Institut für Mikrobiologie und Genetik, Abteilung für Bioinformatik, Goldschmidtstr. 1, 37077 Göttingen, Germany.
11. Current Address: Sars International Centre for Marine Molecular Biology, Bergen, Norway.
12. Genome Center, University of California-Davis, Davis, CA 95616, USA.
13. Department of Ecology, Evolution and Marine Biology, University of California Santa Barbara, Santa Barbara, California 93106, USA.
14. Current address: Courant Research Centre Geobiology, Georg-August University of Göttingen, Goldschmidtstr.3, 37077 Göttingen, Germany.
15. Department of Biological Sciences, University of Alberta, Edmonton, AB, T6G 2E9, Canada.
16. Development and Neurobiology program Institut Jacques Monod, UMR 7592 CNRS/Université Paris Diderot-Paris 7, 75205 Paris Cedex 13, France.

Sponges are an ancient group of animals that diverged from other metazoans over 600 million years ago. Here we present the draft genome sequence of *Amphimedon queenslandica,* a demosponge from the Great Barrier Reef, and show that it is remarkably similar to other animal genomes in content, structure and organization. Comparative analysis enabled by the sponge sequence reveals genomic events linked to the origin and early evolution of animals, including the appearance, expansion, and diversification of pan-metazoan transcription factor, signaling pathway, and structural genes. This diverse "toolkit" of genes correlates with critical aspects of all metazoan body plans, and comprises cell cycle control and growth, development, somatic and germ cell specification, cell adhesion, innate immunity, and allorecognition. Notably, many of the genes associated with the emergence of animals are also implicated in cancer, which arises from defects in basic processes associated with metazoan multicellularity.

The emergence of multicellular animals from single-celled ancestors over 600 My ago required the evolution of mechanisms for coordinating cell division, growth, specialization, adhesion, and death. Dysfunction of these mechanisms drives diseases such as cancers, in which social controls on multicellularity fail, and autoimmune disorders, in which distinctions between self and non-self are disrupted. The hallmarks of metazoan multicellularity are therefore intimately related to those of cancer[1] and immunity,[2] relying on oncogenes, tumor suppressors, and cell-surface and signaling components.

Sponges play a critical role in the search for the origins of metazoan multicellular processes,[3] as they are generally recognized as the oldest surviving metazoan phyletic lineage. Although the kinship of sponges to other animals was recognized by the nineteenth century[4], the

absence of a gut and nervous system had relegated sponges to the "Parazoa,"[5] a grade below the "Eumetazoa" or "true animals" (*e.g.*, cnidarians, ctenophores, and bilaterians)[6]. Nevertheless, sponges share key adhesion and signaling genes[7-11] with eumetazoans, as well as other genes important in body plan patterning such as developmental transcription factors[12-15]; sponge embryos and larvae (Fig. 1) are readily comparable to those of other animals.[12,16]. Sponges are diverse and their phylogeny is poorly resolved,[17-19] allowing for the possibility that sponges are paraphyletic[20], which implies that other animals evolved from sponge-like ancestors.

Here we report on the genome of *Amphimedon queenslandica*, a haplosclerid demosponge whose adult organization and lifestyle is typical for sponges, feeding on microbes and particulate organic matter filtered by flagellated collar cells that resemble choanoflagellates. Although the diversity of sponges, and their uncertain phylogeny, makes it doubtful that any single species can reveal the intricacies of early animal evolution, comparison of the *A. queenslandica* draft genome with sequences from other species can provide a conservative estimate of the genome of the common ancestor of all animals and the timing and nature of the genomic events that led to the origin and early evolution of animal lineages.

The *A. queenslandica* genome harbors an extensive repertoire of developmental signaling and transcription factor genes, suggesting that the metazoan ancestor had a developmental "toolkit" similar to that in modern complex bilaterians. The origins of many of these and other genes specific to animal processes such as cell adhesion, and social control of cell proliferation, death and differentiation can be traced to genomic events (gene birth, subfamily expansions, intron gain/loss, *etc.*) that occurred in the lineage that led to the metazoan ancestor, after animals diverged from their unicellular cousins. In addition to possessing a wide range of metazoan-specific genes, the *Amphimedon* draft genome is missing some genes that are

conserved in other animals, suggesting gene origin and expansion in eumetazoans after their divergence from the demosponge lineage and/or gene loss in *Amphimedon*.

Genome Sequencing and Annotation

Amphimedon queenslandica is a hermaphroditic spermcast spawner, and cannot be readily inbred in the lab (Fig. 1a-c; Supplementary Note S1).[21] Adult sponges also harbor many commensal microbes. To minimize allelic variation and microbial contamination we sequenced genomic DNA from multiple embryos and larvae from a single mother. This DNA contains four dominant parental haplotypes (~3% polymorphism), although a single brood may have multiple fathers (Supplementary Notes S2.1 and S3). We used ~9-fold whole genome Sanger shotgun coverage to produce a ~167 Mbp assembly that typically represents each locus once rather than splitting alleles (Supplementary Notes S2 and S3) and captures ~97% of protein-coding gene content (Supplementary Note S2.5). We also recovered an alpha-proteobacterial genome that is likely a vertically transmitted commensal microbe of *Amphimedon* embryos (Supplementary Note S2.7).

The assembled *A. queenslandica* genome encodes ~30,000 predicted protein-coding loci (Supplementary Note S4). This is an overestimate of the true gene number due to overprediction, unrecognized transposable elements, and gene fragmentation at contig or scaffold boundaries. Nevertheless, 18,693 (63%) have identifiable homologs in other organisms in the Swiss-Prot database; there are no doubt novel or rapidly evolving sponge genes unknown in other species. CpG dinucleotides are depleted, and TpG and CpA dinucleotides augmented, relative to overall G+C composition, which is suggestive of germline cytosine methylation in

the *Amphimedon* genome. This is consistent with the presence of a DNMT3-related putative *de novo* methytransferase as well as proteins with predicted methyl CpG binding domains.

Analysis of the *Amphimedon* gene set reveals striking conservation of gene structure (intron phase and position) and genome organization (synteny) relative to other animals (Supplementary Notes S5 and S6). In *Amphimedon*, intragenic position and phase are retained for 84% of the introns inferred for the metazoan ancestor, comparable to the 76% and 88% retention in human and sea anemone, respectively.[22,23] The organization of genes shows conserved synteny (*i.e.*, conserved linkage without necessarily requiring colinearity) relative to other animals. In particular, 83 of the 153 longest *Amphimedon* scaffolds (*i.e.*, those that contain genes from more than ten distinct metazoan gene families, sufficient for synteny to be assessed) show segments of conserved synteny with other animals (Supplementary Note S6). This suggests that portions of the fifteen ancestral linkage groups inferred for the cnidarian-bilaterian ancestor[22,24] were already in place in the demosponge-eumetazoan ancestor. No such conserved synteny was detected between animals and the choanoflagellate *M. brevicollis*.

Animal relationships

We addressed the controversial phyletic branching of early animal lineages by comparing sets of orthologous genes in *A. queenslandica* and a diverse sampling of eighteen complete genomes (Supplementary Note S7). Our analyses support the grouping of placozoans, cnidarians, and bilaterians into a eumetazoan clade, with demosponge as an earlier-branching lineage,[25] and rejects the diploblast-triploblast phylogeny[17] in favor of a more conventional "sponges first" tree[19,20] (Fig. 1d). In our discussion below we therefore refer to descendants of the placozoan-cnidarian-bilaterian last common ancestor as Eumetazoa, and reserve "Eumetazoa *sensu*

stricto (*s.s.*)" for the more limited clade defined by descendants of the cnidarian-bilaterian ancestor.

Our analysis emphasizes the quantitative divergence between metazoans and their closest living unicellular relatives. For example, 28% of the amino acid substitutions between humans and their last common ancestor with choanoflagellates occurred on the metazoan stem lineage (bold line in Fig. 1d), prior to the divergence of sponges from other animals. This pre-metazoan period can be crudely estimated to be ~150-200 My (Supplementary Note S7.6),

The zootype and origin of metazoan genes

With multiple animal genomes now in hand, we can extend the "zootype" concept[26] to include other shared derived genomic characteristics of animals. Out of 4,670 pan-metazoan gene families defined by clustering sponge and eumetazoan peptides, 1,286 (27%) appear to be metazoan-specific (see Supplementary Note 9.2). Similarly, there are eumetazoan, eumetazoan *s.s.*, and bilaterian genomic synapomorphies, as well as sponge-specific gene families (*e.g.*, kinases, see Supplementary Note S8). Due to residual incompleteness of the sponge genome draft, and possible gene losses in the *Amphimedon* lineage, this analysis provides a conservative estimate.

Nearly three quarters of the 1,286 animal-specific gene families arose by gene duplication on the metazoan stem (Supplementary Note S9). These include the early duplication of transcription factor families such as homeodomains and basic helix-loop-helix domains.[13,14,27] Additional gene duplication and divergence in eumetazoans further increased transcription factor gene family number, which in general are 2 to 34 times larger in eumetazoans than in *Amphimedon*. In contrast, substantial diversification of kinase gene families occurred before

the divergence of the sponge and eumetazoan lineages (see below).[28] We can assess the role of tandem duplication in the creation of these families by seeking evidence for linkages among anciently diverged paralogs (Supplementary Note S10). A significant fraction remain linked (up to 30%, as found in *Trichoplax*, $p<0.0001$, with lower levels in other contemporary metazoan genomes), indicating that (1) many gene family expansions originally occurred as tandem or proximal duplications, and (2) these genomically local duplications have remained linked over time. This is consistent with overall preservation of relict linkages observed here and in other basal metazoan genomes.[22,24,25].

We find 235 animal-specific protein domains, and 769 animal-specific domain combinations also evolved along the metazoan stem (Supplementary Note S9). Additionally, lineage-specific changes to these animal domain architectures occurred in early metazoan evolution [16,29,30] For example, new combinations of domains in death-fold domain proteins and laminins possibly allow for the modification of protein interactions and pathways involved in programmed cell death and cell adhesion, respectively (Supplementary Note S9.3), and the co-option of sponge-, eumetazoan- or bilaterian-specific architectures into novel functions.

Kinases. The 705 *Amphimedon* kinases represent the largest reported metazoan kinome, and include members of >70% of human kinase classes (compared with 59% in choanoflagellate, 83% in sea anemone, 70% in *C. elegans*, and 77% in fruit fly; see Supplementary Note S8.7). *Amphimedon* has single copies of most metazoan kinase classes, but has several expansions of over 50 genes per class. The largest expansions are in the tyrosine kinase and tyrosine kinase-like groups, and include over 150 likely receptor tyrosine kinases (RTKs). Unlike *Monosiga*, where RTKs could not be classified into metazoan families,[28] *Amphimedon* has kinase domains from

six known animal families (EGFR, Met, DDR, ROR, Eph, and Sevenless). The EGFR and some Eph extracellular domain architectures are as in their eumetazoan counterparts, but many other RTKs have unique extracellular domains. For instance, DDRs have immunoglobulin repeats, and sushi domains are found in some members of the expanded Eph and Met families. This suggests that the activating ligands, presumably found largely in the external environment, may be quite distinct.

Six hallmarks of animal multicellularity

The *A. queenslandica* genome allows us to systematically assess the origin of the six hallmarks of metazoan multicellularity: (1) regulated cell cycling and growth, (2) programmed cell death, (3) cell-cell and cell-matrix adhesion, (4) developmental signaling and gene regulation, (5) allorecognition and innate immunity, and (6) specialization of cell types. These cardinal features of metazoan multicellularity have their origins on the metazoan stem and often are the result of metazoan gene novelties combining with more ancient factors. A recurring theme is the overlap of these core "multicellularity" genes with genes perturbed in cancer, a disease of aberrant multicellularity (see oncogenes and tumor suppressors in Figs. 2 and 3).

Regulated cell cycling and growth

While the core machinery of the animal cell cycle traces back to the early eukaryotes (Fig. 2a; Supplementary Note S8.2), some critical metazoan regulatory mechanisms emerged more recently. For example, while the p53/p63/p73 tumor suppressor family is holozoan-specific,[31] the HIPK kinase that phosphorylates p53 in the presence of DNA breaks is metazoan-specific, and the MDM2 ubiquitin ligase that regulates p53 appears as an eumetazoan feature.

Thus the p53-mediated response to DNA damage emerged prior to the divergence of eumetazoans. Intramolecular regulation also has evolved, as illustrated by the Myc oncogene, which is found in the unicellular *Monosiga*[31] but lacks the N-terminal 'DCMW' motif present in *Amphimedon* and other animal Mycs. Since mutation of this motif disrupts Myc function in vertebrates, it likely plays an important role in all animals.

Tumor suppressors encoded by two classes of cyclin-dependent kinase (CDK) inhibitors mediate growth-factor-dependent regulation of the cell cycle. Although the INK4/CDKN2 class (p15/p16/p18/p19) regulates the eumetazoan-specific CDK4/6-cyclin D kinase and is chordate-specific, the Cip/Kip/CDKN1 class (p21/p27/p57) is more general, regulating many CDKs, and appears to have arisen on the eumetazoan stem. In bilaterians Cip/Kip genes integrate external growth signals, and are regulated transcriptionally and post-transcriptionally by the major growth pathways (see below). The emergence of this class of CDK inhibitors on the eumetazoan stem suggests a central regulatory role even in early animals.

While cell growth and cell division are tightly coupled in unicellular species, they can be separately regulated in multicellular organisms. In bilaterians, growth is regulated by six major signaling pathways (Receptor tyrosine kinase (RTK) signaling via Ras, insulin signaling via the PI3K pathway, Rheb/Tor, cytokine-JAK/STAT, Warts/Hippo, and the Myc oncogene) that also modulate the cell cycle (Supplementary Note S8.2). While the Rheb/Tor pathway dates back to early eukaryotes, the other pathways contain several genes that are holozoan and metazoan innovations. For example, the insulin receptor substrate and phosphotyrosine binding proteins GAB1/GAB2 emerged on the metazoan stem after the divergence of choanoflagellates, suggesting that an insulin-signaling-like pathway may have been a key regulator of growth in early animals by tying into the ancient PDK1 and Akt kinases (Fig. 2b). However, since p21, p27

and Mdm2 are all eumetazoan novelties, this pathway may not have acquired the ability to regulate cell proliferation until after the divergence of sponges.

Programmed cell death

In contrast to the cell cycle machinery, most of the apoptotic circuitry is unique to animals, increasing in complexity along metazoan, eumetazoan, and bilaterian stems (Fig. 2c; Supplementary Note S8.3). Both intrinsic and extrinsic programmed cell death pathways require caspases, a metazoan-specific family of cysteine aspartyl proteases. *Amphimedon* encodes initiator caspases with the characteristic caspase recruitment and death effector domains, as well as an expanded repertoire of effector capases.

The intrinsic pathway drives cell death by permeabilization of the outer mitochondrial membrane and is regulated by the Bcl2 oncogene family of pro- and anti-apoptotic factors. The pro-apoptotic protein Bak arose in the metazoan lineage, while Bax and Bok appear to be eumetazoan-specific. Bcl2/Bcl-X are anti-apoptotic and metazoan-specific. Mitochondrial permeabilization releases proteins of varying evolutionary origin, including the ancient apotosis-inducing factor that contributes to caspase-independent apoptosis, metazoan-specific apoptotic protease activating factor 1, and eumetazoan *s.s.*-specific caspase-activated DNase (CAD) and its regulator ICAD/DFF.

The extrinsic apoptotic pathway is activated by external signals through transmembrane tumor necrosis factor receptors (TNFRs) whose intracellular death domain interacts with downstream adaptors. *Amphimedon* encodes a nerve growth factor receptor (NGFR) p75-like protein, though it lacks the crucial death domain that is seen in *Nematostella* and bilaterians (see Robertson et al. 2006[32]); other death TNFRs (*i.e.*, Fas, DR4, DR5, and TNFR1) are vertebrate-

specific.[32,33] Since the intrinsic cascade is composed of components which predate the metazoans, it is likely to be the original mechanism for inducing apoptosis.

Cell-cell and cell-matrix adhesion

The diagnostic domains of two major cell-cell adhesion superfamilies, the cadherins and the immunoglobulins (Ig), are present in *Monosiga* within the extracellular region of putative transmembrane proteins[31,34] (Supplementary Note S8.8). *Amphimedon* cadherins differ from those of *Monosiga* in having proteins with domain architectures diagnostic for the metazoan-specific classical cadherin and seven pass transmembrane cadherin subfamilies.[31,35] A considerable expansion of Ig-like domain containing proteins occurred on the metazoan stem, with 218 predicted in *Amphimedon vs.* 5 in *Monosiga,*[31]. The combination of an N-terminal Ig domain with C-terminal FN3 repeats is found only in metazoans.

Similarly, metazoan extracellular matrix (ECM) proteins use domains that evolved on the holozoan stem. For example, *Monosiga* encodes proteins with collagen triple helix repeats and other genes with fibrillar collagen C-terminal domains, but these domains only appear together in metazoans.[30,31] Thrombospondin domain architectures are found in *Amphimedon*, however agrin, netrin, and perlecan appear to be eumetazoan innovations. The ECM receptors, α and β integrin, also appear to be specific to metazoans (Fig. 3a).

Developmental signaling and transcription

Components of the major metazoan developmental signaling pathways as well as classes of transcription factors are mostly present in *Amphimedon* and absent from *Monosiga* and other non-metazoan genomes[13,14,16,27,29], suggesting that ontogenetic development, including primary

germ cell formation (Supplementary Note S8.4), originated on the metazoan stem.[3,11,12] Although *Amphimedon* possesses a characteristically metazoan repertoire of transcription factor families (Supplementary Note S8.6),[13,14,27,31] in general these families are further expanded in eumetazoans[13]. Some differences between sponges and eumetazoans correlate with morphological complexity. For example, sponges do not appear to have a mesoderm and accordingly *Amphimedon* lacks transcription factors involved in mesoderm development (Fkh, Gsc, Twist, Snail). In contrast, sponges possess several transcription factors involved in determination or differentiation of muscles and nerves despite lacking a neuromuscular system (PaxB, Lhx genes, SoxB, Msx, Mef2, Irx and bHLH neurogenic factors).[13,14,27] *Amphimedon* lacks Hox genes and some other transcription factor subfamilies that are involved in specifying and patterning bilaterian nervous systems and body plans.[13,14,27,36,37]

Signaling cascades, such as the Wnt, TGF-beta, Notch, and Hedgehog pathways, pattern embryos by specifying cellular identity and coordinating morphogenetic events. The ligands and receptors of all of these cascades are metazoan innovations at the cell surface (Supplementary Note S8.5), apart from the eumetazoan *s.s.*-specific Hedgehog ligand.[29] The transcription factors specific to these pathways are also metazoan-specific (Tcf/Lef, Smads, CSL, Gli) whereas the cytosolic signal transducers generally have more ancient origins. This pattern suggests that these pathways arose by the engagement of novel ligands and receptors with already active signaling mechanisms, enabling multicellular communication.

Amphimedon also has fewer ligands and receptors in each pathway compared to eumetazoans (3 Wnt and 2 Fzd, 8 TGF-beta ligands and 5 TGF-beta receptors, 1 Notch and 5 Deltas) (Supplementary Note S8.5) as observed for many transcription factor families. In contrast to transcription factors,[13,14,27] however, these proteins generally can not be assigned to

eumetazoan subfamilies or are obvious recent sponge-specific duplications. This lack of phylogenetic resolution may reflect a period of rapid evolution and diversification of ligand/receptor molecules in sponge and eumetazoan lineages. Perhaps as a consequence, the inhibitors that interact with ligands and receptors to modulate pathway activity also appear to be lineage specific. In particular, inhibitors described from bilaterians were not found in *Amphimedon* (*e.g.*, Chordin, Numb, I-Smads, Wif).

Allorecognition and innate immunity

The transition to multicellularity was accompanied by mechanisms to defend against invading pathogens and to prevent the fusion of genetically distinct conspecifics.[2] Although some metazoan immunity genes originated early in eukaryotic evolution, many are restricted to animals, as illustrated by the signaling cascades shared by the Toll-like receptor (TLR) and the Interleukin1 receptor (IL1R) (Supplementary Note S8.10). An ancestral form belonging to this receptor superfamily was likely present in the last common metazoan ancestor and independently diversified in poriferan and cnidarian lineages. Nuclear factor kappa B (NF-kB), Tollip, and ECSIT genes are present in holozoans, however most TLR/IL1R pathway proteins are either composed of metazoan-specific domains (*e.g.*, Pellino) or architectures *(e.g.*, the death domain with TIR and protein kinase domains in MyD88 and IRAKs, respectively). Immune effector systems also consist largely of metazoan innovations, such as the macrophage-expressed gene 1 that participates directly in pathogen elimination.[38] Likewise all animals share specific antiviral defense factors such as MDA5-like RNA helicases, and interferon regulatory factor-like proteins, although other systems (*e.g.*, RNAi) have more ancient origins.[39] A

primordial complement pathway appears to have evolved exclusively on the eumetazoan (*s.s.*) stem and further diversified in bilaterians.[40]

Amphimedon and other demosponges encode unique extracellular Calx-beta domain-containing proteoglycans called aggregation factors (AFs), which promote cell adhesion and may also be involved in allorecognition.[41] The presence of a cluster of AF-related genes in the *Amphimedon* genome suggests that allorecognition could be under the control of a multigene family.

Specialized cell types

Polarized epithelia. Sponge cells adhere to form tissue-like layers, but a true epithelial cell layer, characterized by aligned cell polarity, belt-form junctions and underlying basal lamina, is thought to be a eumetazoan innovation. *Amphimedon* possesses all the main components of the Par, Crumbs and Discs Large complexes, a set of interacting proteins that are largely metazoan specific and determine polarity in epithelial cells (Fig. 3a) (Supplementary Note S8.8). The main proteins comprising bilaterian spot-form and zonula adherens junctions are also present in *Amphimedon* and appear to be metazoan-specific.[34,42] By contrast, septate junction and basal lamina proteins appear to be largely eumetazoan innovations (Fig. 3a); *Amphimedon* does possess several genes with laminin-like domain architectures (Supplementary Note S9.3).

Sensory systems and the neuron. Sponges can sense and respond to their environment, although nerve cells appear to be restricted to eumetazoans *s.s.*[43,44] However, the expression of orthologs of post-synaptic structural and proneural regulatory proteins in *Amphimedon* larval globular cells suggests an evolutionary connection with an ancestral

protoneuron.[36,42] *Amphimedon* possesses homologs of bilaterian proteins involved in nervous system development (*e.g.*, elav- and musashi-like RNA-binding proteins, neural transcription factors), pre-and post-synaptic organization (*e.g.*, synaptotagmin)[42], endogenous and exogenous signaling (*e.g.*, GPCRs), and neuroendocrine secretion, although bilaterian peptide hormones are not detected (Supplementary Note S8.9). Some key synaptic genes are conspicuously missing from *Amphimedon* (Fig. 3b; Supplementary Note S8.9), including the ionotropic glutamate receptor family;[42] neuronal-type metabotropic glutamate, dopamine and serotonin receptors are present. *Amphimedon* has a homolog of the ephrin receptor, an axon guidance protein, although the ephrin ligand and developmental genes involved in axon guidance (e.g., slit, netrin, unc-5, and robo) are not present. *Amphimedon* also possesses over 200 GPCRs, which includes a large lineage-specific expansion of rhodopsin-related GPCRs that are encoded largely by clusters of single exon genes as observed in other metazoans (Supplementary Note S8.9). From these observations we infer that the metazoan ancestor possessed a complex sensory system, and many of the molecular requirements for neural development and nerve cell function. This suggests that exaptation was critical for the genesis of the first nerve cell, with eumetazoan-specific gene innovations providing the regulatory and structural requirements to connect these protoneural components into a functional neuron (Fig. 3b).

Molecular correlates of morphological complexity

With a diverse sample of genomes in hand, we sought differences in gene repertoire that are associated with gross morphological complexity. Fig. 4 shows molecular function categories that are significantly enriched ($p<1E-10$) in one or more metazoan complexity groups, with the relative frequencies of genes with these functions in each species shown by color code. Here we

have defined broad groupings representing three grades of morphological complexity, guided by the number of described cell types,[45] including non-bilaterian (or "basal") metazoans (*Nematostella*, *Trichoplax*, *Amphimedon*; ~5-15 cell types), invertebrate bilaterians (*Drosophila*, *C. elegans*, sea urchin; ~50-100 cell types), and vertebrates (~225 cell types, represented by the human genome), with a selection of non-animals as an outgroup (Supplementary Note S11). Similarly, using a principal component analysis, we also identified suites of molecular functions that are associated with complexity (Supplementary Figure S11.2). The first component differentiates between metazoans and non-metazoans; the second component partly differentiates between metazoan complexity groups.

Included amongst the functional categories that correlate with increase in metazoan morphological complexity are (Fig. 4; Supplementary Table S11.1.1): GPCRs, ion channels, cell adhesion proteins, and defense and immunity proteins, which are enriched in basal metazoans relative to non-animals; homeobox transcription factors and gap junction proteins, which are enriched in bilaterians relative to non bilaterian animals; and immunoglobulin receptor family members, immunoglobulins, MHC antigens, and cytokine receptors, which are enriched in vertebrates relative to invertebrate bilaterians. These broad associations with complexity are evidently superimposed on notable lineage-specific variation as seen in Fig. 4 (***e.g.***, serine protease gene loss in *C. elegans*, and voltage-gated ion channel expansion in *Paramecium*). Similar functional categories contribute to principal components (Supplementary Table S11.2.1).

Conclusions

The *Amphimedon* genome, combined with recently sequenced genomes of diverse invertebrates and a choanoflagellate, identifies innovations that underlie the emergence and early diversification of the Metazoa. These genomic comparisons resurrect a common animal ancestor of remarkable complexity. Metazoans can now be defined by a long list of genomic synapomorphies -- gene content, intron-exon structure and syntenies -- as well as characteristics common to all animal life such as sex, development, controlled cellular proliferation, differentiation and growth, and immunity. To what extent the ancestral functioning of this gene set is reflected in modern poriferans is unclear, although studies of both sponge development, which yields a highly patterned larva with axial polarity[12], and sponge immunity provide points of direct comparison with the eumetazoan condition.

While the eumetazoan lineage produced a wide diversity of body forms, the sponge body plan has been stable for over 600 My. What can explain this disparity in evolved morphological complexity? Although we have seen that sponges and eumetazoans share many common pathways related to morphogenesis and cell-type specification, there are of course notable genomic differences, including different miRNA assemblages,[46] lineage-specific domains and domain architectures, and the differential expansions of gene families. While there has been minimal characterization of cis-regulatory architectures in non-bilaterians, we note that since most classes of bilaterian transcription factors are also present in sponges, cnidarians, and placozoans, it may be that quantitative rather than qualitative differences in cis-regulatory mechanisms were needed to produce more diverse body plans.

The sexually-reproducing, heterotrophic metazoan ancestor had the capacity to sense, respond to, and exploit the surrounding environment while maintaining multicellular homeostasis. Although sponges lack some of the cell types found in eumetazoans, including

neurons and muscles, they share with all other animals genes that are essential for the form and function of integrated multicellular organisms,. With these genomic innovations enabling the regulation of cellular proliferation, death, differentiation, and cohesion, metazoans transcended their microbial ancestry.

Methods Summary

Detailed methods are described in Supplementary Information. The genome assembly, gene model sequences, predicted proteins and EST clusters and sequences have been deposited with DDBJ/EMBL/GenBank as project accession ACUQ00000000.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements. This study was supported by funds from the Australian Research Council (B.M.D., M.A), US Department of Energy Joint Genome Institute (B.M.D., D.S.R., S.P.L.) Harvey Karp (K.S.K.), NSF (T.H.O.), NIH/NHGRI (G.M.), University of Queensland Postdocotral Fellowship (M.A., S.F.C), Sars International Centre for Marine Molecular Biology (M.A.), DFG (M.St.), ANR (M.V.), CNRS (M.V.), Gordon and Betty Moore Foundation (D.S.R.) and Richard Melmon (D.S.R.). We thank John Huelsenbeck and Iswar Hariharan for generous help with phylogenetic analyses and growth pathways respectively.

Author Contributions. Genome and EST sequencing, assembly, annotation and analysis: J.C., T.M., U.H., N.H.P., M.St., A.D., Y.Z., M.A., A.C., D.M.G., D.J., S.S., B.J.W. and D.S.R.

Phylogenetics: M.S. and D.S.R. Gene family and biological process analyses: M.S., B.F., M.G., G.S.R.; C.C., M.D., C.L.; M.A., S.M.D., T.H.O., D.P., S.F.C., C.H., M.V., K.S.K., G.M., B.M.D. and D.S.R. Clustering, novelty, domain content and complexity analyses: O.S. and D.S.R. Gene family expansion analyses: M.S., O.S., D.S.R. Writing: M.S., B.M.D., D.S.R., O.S., J.C., B.F., M.G., G.S.R., G.M., K.S.K, M.V., C.L., S.M.D., N.H.P., A.D., C.C., M.A., T.H.O. and S.P.L. Project design and coordination: B.M.D and D.S.R.

Author Information. The genome sequence data can be accessed from DDBJ/EMBL/GenBank as project accession ACUQ00000000. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. Correspondence and requests for materials should be addressed to M.S. (mansi@wi.mit.edu), B.M.D. (b.degnan@uq.edu.au) and D.S.R. (dsrokhsar@gmail.com).

References

1. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
2. Muller, W.E. & Muller, I.M. Origin of the metazoan immune system: identification of the molecules and their functions in sponges. *Integr Comp Biol* **43**, 281-292 (2003).
3. Muller, W.E. et al. Bauplan of urmetazoa: basis for genetic complexity of metazoa. *Int Rev Cytol* **235**, 53-92 (2004).
4. Grant, R.E. Animal Kingdom. in *The Cyclopaedia of Anatomy and Physiology*, Vol. 1 (ed. Todd, R.B.) (Sherwood-Gilbert-Piper, London, 1836).
5. Sollas, W.J. Report on the Tetractinellida. in *Report on the Scientific Results of the Voyage of H.M.S. Challenger*, Vol. 25 (1888).
6. Hyman, L.H. *The Invertebrates. Volume 1, Protozoa through Ctenophora*, (McGraw-Hill Book Company, Inc., New York and London, 1940).
7. Muller, W.E. Origin of metazoan adhesion molecules and adhesion receptors as deduced from cDNA analyses in the marine sponge *Geodia cydonium*: a review. *Cell Tissue Res* **289**, 383-95 (1997).
8. Muller, W.E. & Schacke, H. Characterization of the receptor protein-tyrosine kinase gene from the marine sponge *Geodia cydonium*. *Prog Mol Subcell Biol* **17**, 183-208 (1996).
9. Suga, H., Katoh, K. & Miyata, T. Sponge homologs of vertebrate protein tyrosine kinases and frequent domain shufflings in the early evolution of animals before the parazoan-eumetazoan split. *Gene* **280**, 195-201 (2001).
10. Skorokhod, A. et al. Origin of insulin receptor-like tyrosine kinases in marine sponges. *Biol Bull* **197**, 198-206 (1999).
11. Nichols, S.A., Dirks, W., Pearse, J.S. & King, N. Early evolution of animal cell signaling and adhesion genes. *Proc Natl Acad Sci U S A* **103**, 12451-6 (2006).
12. Larroux, C. et al. Developmental expression of transcription factor genes in a demosponge: insights into the origin of metazoan multicellularity. *Evol Dev* **8**, 150-73 (2006).
13. Larroux, C. et al. Genesis and expansion of metazoan transcription factor gene classes. *Mol Biol Evol* **25**, 980-96 (2008).
14. Simionato, E. et al. Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. *BMC Evol Biol* **7**, 33 (2007).
15. Gazave, E. et al. NK homeobox genes with choanocyte-specific expression in homoscleromorph sponges. *Dev Genes Evol* **218**, 479-89 (2008).
16. Adamska, M. et al. Wnt and TGF-beta expression in the sponge *Amphimedon queenslandica* and the origin of metazoan embryonic patterning. *PLoS ONE* **2**, e1031 (2007).
17. Schierwater, B. et al. Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. *PLoS Biol* **7**, e20 (2009).
18. Dunn, C.W. et al. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745-9 (2008).
19. Pick, K.S. et al. Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. *Mol Biol Evol* (2010).
20. Sperling, E.A., Peterson, K.J. & Pisani, D. Phylogenetic-signal dissection of nuclear housekeeping genes supports the paraphyly of sponges and the monophyly of Eumetazoa. *Mol Biol Evol* **26**, 2261-74 (2009).

21. Degnan, B. et al. The demosponge *Amphimedon queenslandica*: Reconstructing the ancestral metazoan genome and deciphering the origin of animal multicellularity. in *Emerging Model Organisms: A Laboratory Manual, Volume 1* (Cold Spring Harbor Laboratory Press, 2009).
22. Putnam, N.H. et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86-94 (2007).
23. Sullivan, J.C., Reitzel, A.M. & Finnerty, J.R. A high percentage of introns in human genes were present early in animal evolution: evidence from the basal metazoan Nematostella vectensis. *Genome Inform* **17**, 219-29 (2006).
24. Putnam, N.H. et al. The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064-71 (2008).
25. Srivastava, M. et al. The *Trichoplax* genome and the nature of placozoans. *Nature* **454**, 955-60 (2008).
26. Slack, J.M., Holland, P.W. & Graham, C.F. The zootype and the phylotypic stage. *Nature* **361**, 490-2 (1993).
27. Larroux, C. et al. The NK homeobox gene cluster predates the origin of Hox genes. *Curr Biol* **17**, 706-10 (2007).
28. Manning, G., Young, S.L., Miller, W.T. & Zhai, Y. The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proc Natl Acad Sci U S A* **105**, 9674-9 (2008).
29. Adamska, M. et al. The evolutionary origin of hedgehog proteins. *Curr Biol* **17**, R836-7 (2007).
30. Exposito, J.Y. et al. Demosponge and sea anemone fibrillar collagen diversity reveals the early emergence of A/C clades and the maintenance of the modular structure of type V/XI collagens from sponge to human. *J Biol Chem* **283**, 28226-35 (2008).
31. King, N. et al. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783-8 (2008).
32. Robertson, A.J. et al. The genomic underpinnings of apoptosis in *Strongylocentrotus purpuratus*. *Dev Biol* **300**, 321-34 (2006).
33. Huang, S. et al. Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res* **18**, 1112-26 (2008).
34. Abedin, M. & King, N. The premetazoan ancestry of cadherins. *Science* **319**, 946-8 (2008).
35. Tepass, U., Truong, K., Godt, D., Ikura, M. & Peifer, M. Cadherins in embryonic and neural morphogenesis. *Nat Rev Mol Cell Biol* **1**, 91-100 (2000).
36. Richards, G.S. et al. Sponge genes provide new insight into the evolutionary origin of the neurogenic circuit. *Curr Biol* **18**, 1156-61 (2008).
37. Srivastava, M. et al. Evolution of the LIM homeobox gene family in basal metazoans. *BMC Biology* **8**, 8 (2010).
38. Wiens, M. et al. Innate immune defense of the sponge *Suberites domuncula* against bacteria involves a MyD88-dependent signaling pathway. Induction of a perforin-like molecule. *J Biol Chem* **280**, 27949-59 (2005).
39. de Jong, D. et al. Multiple dicer genes in the early-diverging metazoa. *Mol Biol Evol* **26**, 1333-40 (2009).

40. Kimura, A., Sakaguchi, E. & Nonaka, M. Multi-component complement system of Cnidaria: C3, Bf, and MASP genes expressed in the endodermal tissues of a sea anemone, *Nematostella vectensis*. *Immunobiology* **214**, 165-78 (2009).
41. Fernandez-Busquets, X. & Burger, M.M. Circular proteoglycans from sponges: first members of the spongican family. *Cell Mol Life Sci* **60**, 88-112 (2003).
42. Sakarya, O. et al. A post-synaptic scaffold at the origin of the animal kingdom. *PLoS ONE* **2**, e506 (2007).
43. Pavans de Ceccatty, M. Coordination in sponges. The foundations of integration. *Am. Zool.* **14**, 895-903 (1974).
44. Leys, S.P. & Degnan, B.M. The cytological basis of photoresponsive behavior in a sponge larva. *Biol Bull* **201**, 323-38 (2001).
45. Valentine, J.W. Late Precambrian bilaterians: grades and clades. *Proc Natl Acad Sci U S A* **91**, 6751-7 (1994).
46. Grimson, A. et al. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**, 1193-7 (2008).
47. Boute, N. et al. Type IV collagen in sponges, the missing link in basement membrane ubiquity. *Biol Cell* **88**, 37-44 (1996).

Figure Legends

Figure 1: (a) *Amphimedon queenslandica* adult; scale bar, 5 cm. **(b)** Embryos in a brood chamber; scale bar, 1 mm. **(c)** Larva; scale bar, 100 um. **(d)** Animal phylogeny based on whole-genome data. This unrooted tree is inferred from 229 concatenated nuclear protein-coding genes with 44,616 amino acids using Bayesian inference. All clades are supported with a posterior probability of 1. Colored boxes mark the nodes for which origins of genes are inferred in Figs. 3 and 4. The same topology is supported by the nuclear gene datasets generated by alternative methods as well as by other inference methods (Supplementary Note S7). The metazoan stem leading to the animal radiation is shown in bold. Contrary to the current consensus of eukaryotic relationships, Amoebozoa are not sister-group to Opisthokonta in this tree (Supplementary Note S7).

Figure 2: Origins of vertebrate/bilaterian pathways. Reference pathways from human and other vertebrates are depicted here for comparative purposes. Gene products are colored in by their node of origin as per Fig. 1. White text denotes known oncogenes or tumor suppressor genes. Genes with eumetazoan origin are found in either *Nematostella* or *Trichoplax* or both. **(a)** Cell cycle. **(b)** Cell growth. **(c)** Apoptosis. Dashed lines indicate cases where proteins could not be affiliated to a subtype (see Supplementary Note S8.3).

Figure 3: Origins of complexes and pathways of bilaterian cell types. Reference cellular structures from human and other vertebrates are depicted here for comparative purposes. Gene products are colored in by their node of origin as per Fig. 1. White text denotes known oncogenes or tumor suppressor genes. Genes with eumetazoan origin are found in either

Nematostella or *Trichoplax* or both. **(a)** Cell adhesion and polarity in epithelia. *Collagen IV genes have not been found in the *Amphimedon* genome but have been reported as present in the homoscleromorph sponge *Pseudocorticium jarrei.*[47] **(b)** Synaptic and signaling elements in neurons.

Figure 4: Molecular functions enriched in various complexity groups. Molecular function categories that show significant enrichment (1E-10) in Fisher's exact tests were selected (Supplementary Note S11). Significance of enrichments (grey background) and depletions (white background) for the three "metazoan complexity groups" (non-bilaterian metazoans; invertebrate bilaterians; vertebrates) are indicated in the columns to the left of the heatmap. The heatmap shows normalized gene counts of PANTHER molecular function categories for the species in the analysis. Ath, *Arabidopsis thaliana*; Ddi, *Dictyostelium discoideum*; Pte, *Paramecium tetraurelia*; Ncr, *Neurospora crassa*; Mbr, *Monosiga brevicollis*; Aqu, *Amphimedon queenslandica*; Tad, *Trichoplax adhaerens*; Nve, *Nematostella vectensis*; Hma, *Hydra magnipapillata*; Dme, *Drosophila melanogaster*; Cel, *Caenorhabditis elegans*; Spu, *Strongylocentrotus purpuratus*; Hsa, *Homo sapiens*.

Online Methods

A detailed description of methods used in this study can be found in the supplementary information available online.

Genome Sequencing

Genomic DNA was sheared and cloned into plasmid and fosmid vectors for whole genome shotgun sequencing as described.[48] The data were assembled using a custom approach described in the Supplementary Information. The *Amphimedon* 9X assembly and the preliminary data analysis has been deposited at DDBJ/EMBL/GenBank as project accession ACUQ00000000.

Gene Prediction and Annotation

Protein coding genes were annotated using homology-based methods (Augustus[49], Genomescan[50]) and one *ab initio* method (SNAP[51]). Protein-coding gene predictions are deposited in DDBJ/EMBL/GenBank as accession ACUQ00000000.

Phylogenetic Methods

Three datasets of orthologous genes from seventeen genomes were aligned using default parameters using CLUSTALW[52] and poorly aligned regions were excluded using Gblocks.[53] Phylogenetic analyses were conducted using Bayesian inference and maximum likelihood with bootstrap using mrbayes,[54,55] and PHYML[56] respectively. Alternative likelihood topologies were tested using TREEPUZZLE[57] and CONSEL.[58] Bayesian analysis using site-heterogeneous models were done using aamodel (Huelsenbeck, unpublished) and PhyloBayes.[59,60]

Identification of *Amphimedon* Orthologs of Specific Bilaterian Genes

Putative orthologs of genes involved in various processes in bilaterians were identified by reciprocal BLAST of human, mouse, or *Drosophila* genes against the *Amphimedon* gene models

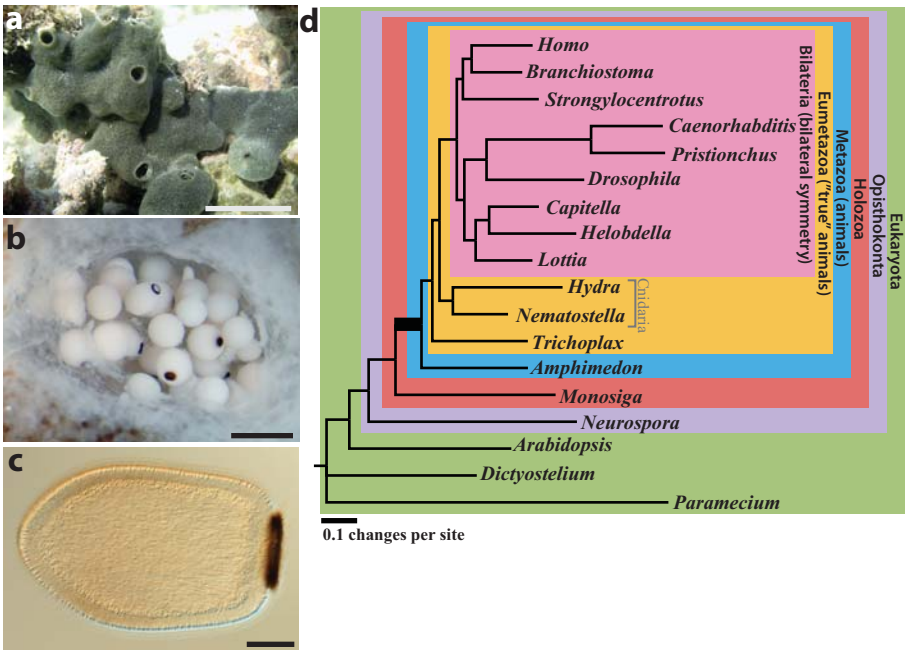
(blastp) or the assembly (tblastn). PFAM[61] domain composition, assignment of PANTHER HMMs[62,63] and phylogenetic trees were used to determine orthology. Trees were built using the neighbor-joining method in Phylip[64] with one hundred bootstrap replicates.

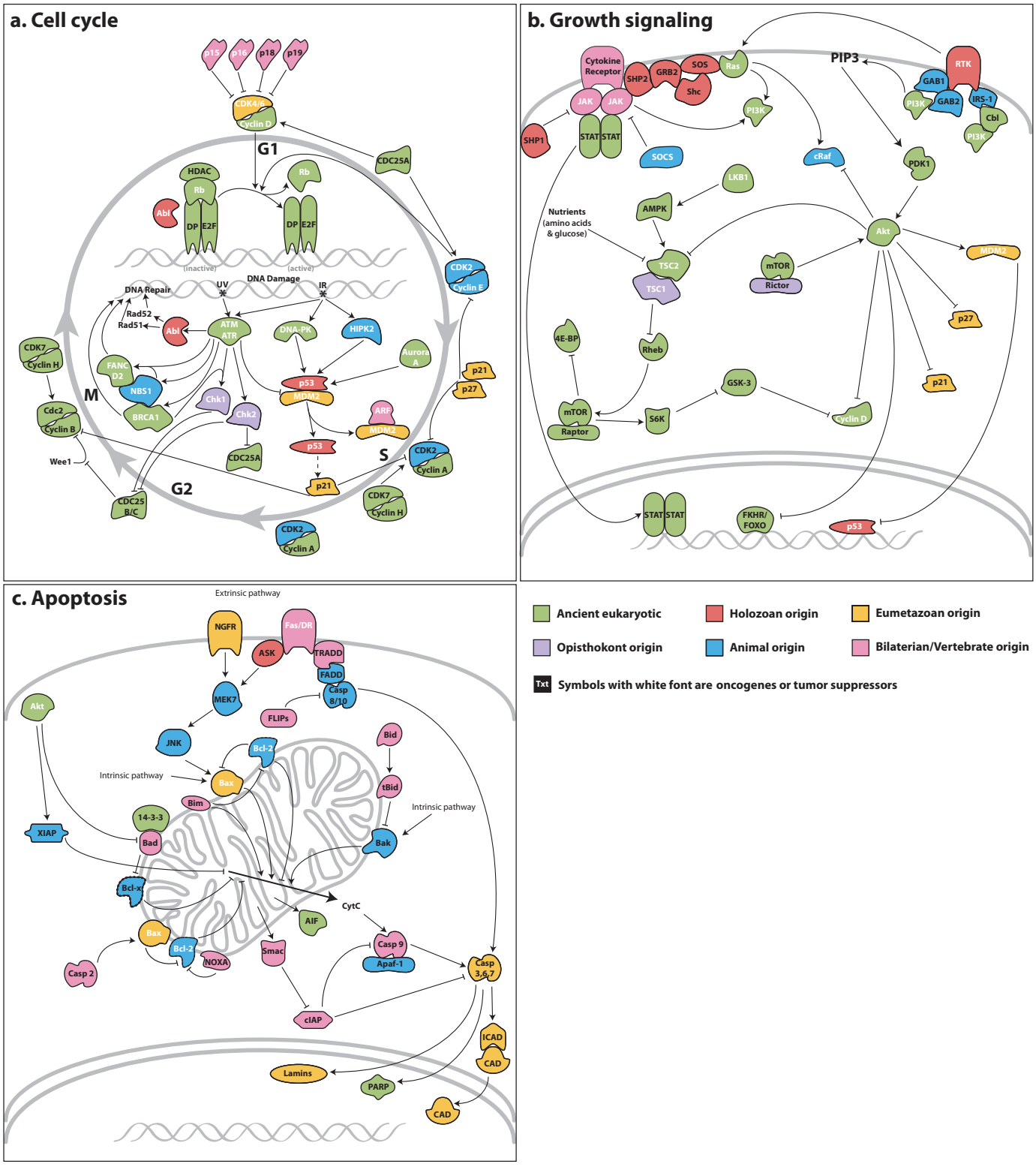
Molecular Function Enrichments and Correlation of Complexity

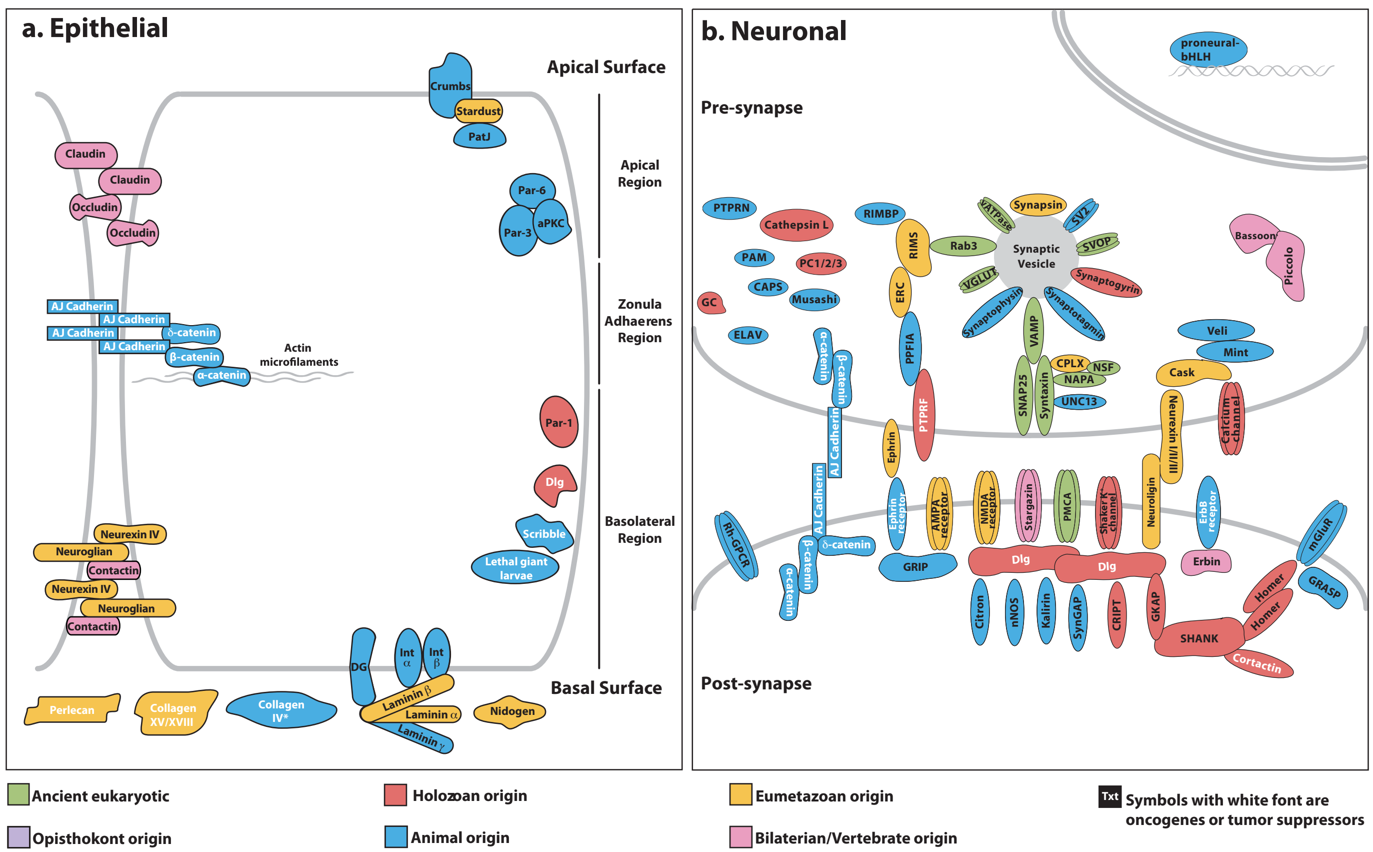
Metazoan gene families were assigned molecular functions using PANTHER[62] annotations. Fisher's exact test as implemented in R[65] was run to test for enrichment or depletion of numbers of gene families for each molecular function category in the novel versus ancestral gene sets. Numbers of genes (not gene families) for various molecular function categories were tested for enrichment between different pairs of four eukaryotic complexity groups (vertebrate, non-vertebrate bilaterian, basal metazoan, non-animal) to identify molecular function families that correlate with the differences in complexity. Principal components analysis was used to identify the contribution of each molecular function category to a eukaryotic complexity group.

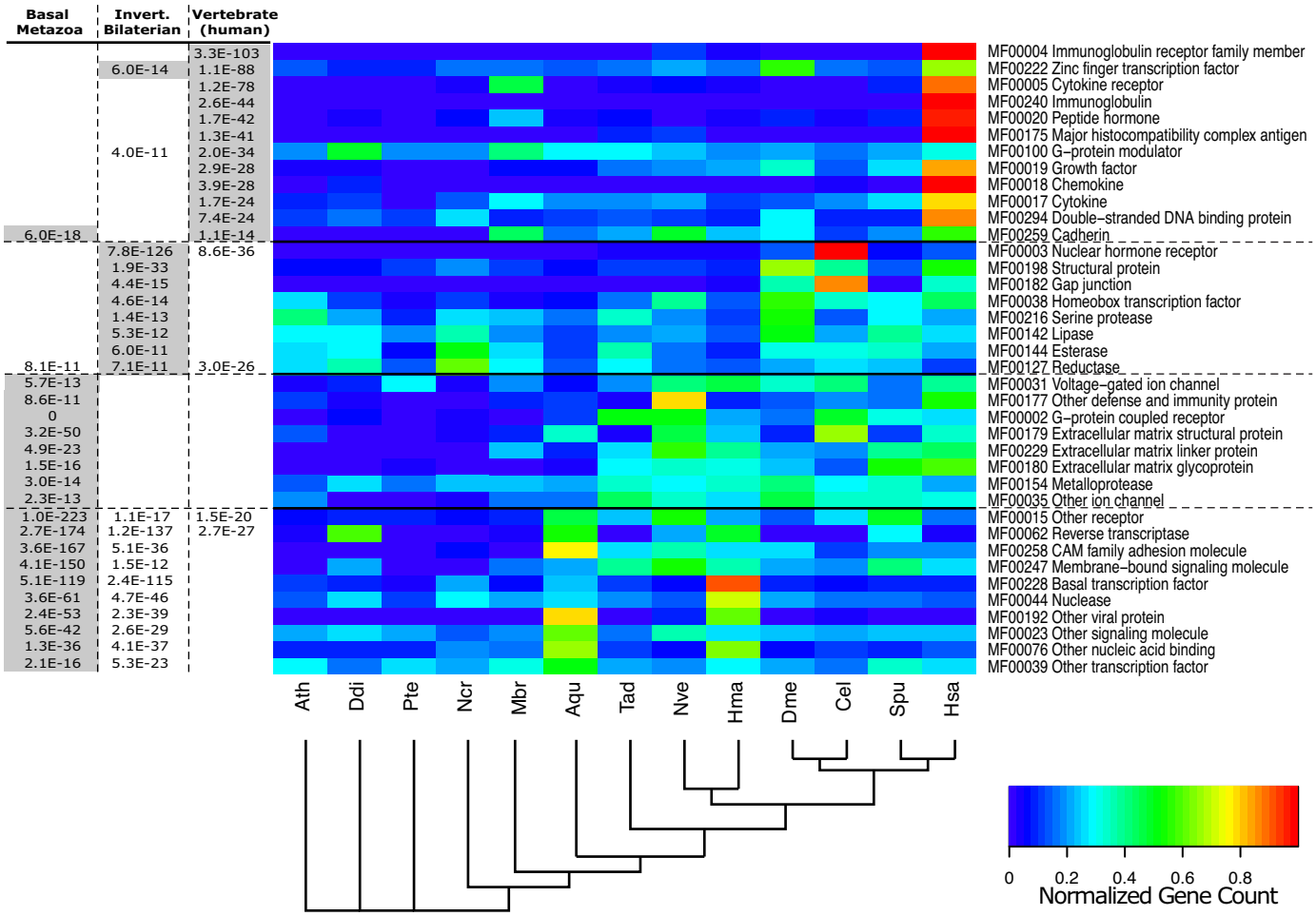
48. Aparicio, S. et al. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. *Science* **297**, 1301-10 (2002).
49. Stanke, M., Tzvetkova, A. & Morgenstern, B. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol* **7 Suppl 1**, S11 1-8 (2006).
50. Yeh, R.F., Lim, L.P. & Burge, C.B. Computational inference of homologous gene structures in the human genome. *Genome Res* **11**, 803-16 (2001).
51. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
52. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-80 (1994).
53. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**, 540-52 (2000).
54. Ronquist, F. & Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-4 (2003).
55. Huelsenbeck, J.P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-5 (2001).
56. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704 (2003).

57. Schmidt, H.A., Strimmer, K., Vingron, M. & von Haeseler, A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502-4 (2002).
58. Shimodaira, H. & Hasegawa, M. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246-7 (2001).
59. Lartillot, N., Brinkmann, H. & Philippe, H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol* **7 Suppl 1**, S4 (2007).
60. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* **21**, 1095-109 (2004).
61. Bateman, A. et al. The Pfam protein families database. *Nucleic Acids Res* **32**, D138-41 (2004).
62. Thomas, P.D. et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* **13**, 2129-41 (2003).
63. Thomas, P.D. et al. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res* **31**, 334-41 (2003).
64. Felsenstein, J. PHYLIP -- Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164-166 (1989).
65. Team, R.D.C. *R: A language and environment for statistical computing.* (R Foundation for Statistical Computing, Vienna, Austria, 2009).


a
b
c
d
Homo
Branchiostoma
Strongylocentrotus
Caenorhabditis
Pristionchus
Drosophila
Capitella
Helobdella
Lottia
Hydra
Nematostella
Trichoplax
Amphimedon
Monosiga
Neurospora
Arabidopsis
Dictyostelium
Paramecium
Cnidaria
Bilateria (bilateral symmetry)
Eumetazoa ("true" animals)
Metazoa (animals)
Holozoa
Opisthokonta
Eukaryota
0.1 changes per site


a. Cell cycle
p15
p16
p18
p19
CDK4/6
Cyclin D
G1
CDC25A
HDAC
Rb
Abl
DP
E2F
(inactive)
(active)
DNA Damage
DNA Repair
UV
IR
Rad52
Rad51
ATM
ATR
DNA-PK
HIPK2
CDK2
Cyclin E
CDK7
Cyclin H
FANC D2
NBS1
BRCA1
M
Chk1
Chk2
p53
MDM2
ARF
Aurora A
p21
p27
Cdc2
Cyclin B
Wee1
CDC25 B/C
S
CDK2
Cyclin A
G2
b. Growth signaling
Cytokine Receptor
JAK
STAT
SHP2
GRB2
SOS
Ras
Shc
PIP3
RTK
GAB1
GAB2
PI3K
IRS-1
Cbl
SHP1
SOCS
cRaf
PDK1
LKB1
AMPK
Nutrients (amino acids & glucose)
Akt
TSC2
TSC1
mTOR
Rictor
MDM2
p27
4E-BP
Rheb
GSK-3
p21
mTOR
Raptor
S6K
cyclin D
FKHR/ FOXO
p53
c. Apoptosis
Extrinsic pathway
NGFR
Fas/DR
ASK
TRADD
FADD
Casp 8/10
MEK7
FLIPs
Akt
JNK
Bcl-2
Bid
Intrinsic pathway
Bax
tBid
Bim
14-3-3
XIAP
Bad
Bak
Bcl-x
CytC
AIF
Casp 9
Apaf-1
Casp 3,6,7
Smac
Casp 2
NOXA
cIAP
ICAD
CAD
Lamins
PARP
Ancient eukaryotic
Holozoan origin
Eumetazoan origin
Opisthokont origin
Animal origin
Bilaterian/Vertebrate origin
Txt
Symbols with white font are oncogenes or tumor suppressors


a. Epithelial
Apical Surface
Crumbs
Stardust
PatJ
Apical Region
Par-6
aPKC
Par-3
Claudin
Claudin
Occludin
Occludin
Zonula Adhaerens Region
AJ Cadherin
AJ Cadherin
AJ Cadherin
AJ Cadherin
δ-catenin
β-catenin
α-catenin
Actin microfilaments
Par-1
Dlg
Basolateral Region
Scribble
Lethal giant larvae
Neurexin IV
Neuroglian
Contactin
Neurexin IV
Neuroglian
Contactin
DG
Int α
Int β
Laminin β
Laminin α
Laminin γ
Basal Surface
Perlecan
Collagen XV/XVIII
Collagen IV*
Nidogen
b. Neuronal
proneural-bHLH
Pre-synapse
PTPRN
Cathepsin L
RIMBP
PAM
PC1/2/3
CAPS
Musashi
GC
ELAV
RIMS
Rab3
ERC
PPFIA
PTPRF
Ephrin
Synapsin
VATPase
SV2
SVOP
Synaptic Vesicle
VGLUT
Synaptogyrin
Synaptophysin
Synaptotagmin
VAMP
CPLX
NSF
NAPA
UNC13
SNAP25
Syntaxin
Bassoon
Piccolo
Veli
Mint
Cask
Neurexin I/II/III
Calcium channel
α-catenin
β-catenin
AJ Cadherin
AJ Cadherin
δ-catenin
β-catenin
α-catenin
Rh-GPCR
Ephrin receptor
AMPA receptor
NMDA receptor
Stargazin
PMCA
Shaker K+ channel
Neuroligin
ErbB receptor
GRIP
Dlg
Dlg
Erbin
mGluR
GRASP
Homer
Homer
Citron
nNOS
Kalirin
SynGAP
CRIPT
GKAP
SHANK
Cortactin
Post-synapse
Ancient eukaryotic
Holozoan origin
Eumetazoan origin
Opisthokont origin
Animal origin
Bilaterian/Vertebrate origin
Txt Symbols with white font are oncogenes or tumor suppressors


Basal Metazoa
Invert. Bilaterian
Vertebrate (human)
3.3E-103
6.0E-14 1.1E-88
1.2E-78
2.6E-44
1.7E-42
1.3E-41
4.0E-11 2.0E-34
2.9E-28
3.9E-28
1.7E-24
7.4E-24
6.0E-18 1.1E-14
7.8E-126 8.6E-36
1.9E-33
4.4E-15
4.6E-14
1.4E-13
5.3E-12
6.0E-11
8.1E-11 7.1E-11 3.0E-26
5.7E-13
8.6E-11
0
3.2E-50
4.9E-23
1.5E-16
3.0E-14
2.3E-13
1.0E-223 1.1E-17 1.5E-20
2.7E-174 1.2E-137 2.7E-27
3.6E-167 5.1E-36
4.1E-150 1.5E-12
5.1E-119 2.4E-115
3.6E-61 4.7E-46
2.4E-53 2.3E-39
5.6E-42 2.6E-29
1.3E-36 4.1E-37
2.1E-16 5.3E-23
MF00004 Immunoglobulin receptor family member
MF00222 Zinc finger transcription factor
MF00005 Cytokine receptor
MF00240 Immunoglobulin
MF00020 Peptide hormone
MF00175 Major histocompatibility complex antigen
MF00100 G−protein modulator
MF00019 Growth factor
MF00018 Chemokine
MF00017 Cytokine
MF00294 Double−stranded DNA binding protein
MF00259 Cadherin
MF00003 Nuclear hormone receptor
MF00198 Structural protein
MF00182 Gap junction
MF00038 Homeobox transcription factor
MF00216 Serine protease
MF00142 Lipase
MF00144 Esterase
MF00127 Reductase
MF00031 Voltage−gated ion channel
MF00177 Other defense and immunity protein
MF00002 G−protein coupled receptor
MF00179 Extracellular matrix structural protein
MF00229 Extracellular matrix linker protein
MF00180 Extracellular matrix glycoprotein
MF00154 Metalloprotease
MF00035 Other ion channel
MF00015 Other receptor
MF00062 Reverse transcriptase
MF00258 CAM family adhesion molecule
MF00247 Membrane−bound signaling molecule
MF00228 Basal transcription factor
MF00044 Nuclease
MF00192 Other viral protein
MF00023 Other signaling molecule
MF00076 Other nucleic acid binding
MF00039 Other transcription factor
Ath
Ddi
Pte
Ncr
Mbr
Aqu
Tad
Nve
Hma
Dme
Cel
Spu
Hsa
0
0.2
0.4
0.6
0.8
Normalized Gene Count