

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of November 5, 2009):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/317/5834/86>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/cgi/content/full/317/5834/86/DC1>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/317/5834/86#related-content>

This article **cites 65 articles**, 23 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/317/5834/86#otherarticles>

This article has been **cited by** 137 article(s) on the ISI Web of Science.

This article has been **cited by** 50 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/317/5834/86#otherarticles>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

14. We experimented with various power ratings from 5 to 75 W.
15. W. A. Edson, *Vacuum-Tube Oscillators* (Wiley, New York, 1953).
16. Note that  $E \neq c\mu_0 H$ , and that the fields are out of phase and not necessarily perpendicular because we are not in a radiative regime.
17. See supporting material on Science Online.
18. *IEEE Std C95.1—2005 IEEE Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 kHz to 300 GHz* (IEEE, Piscataway, NJ, 2006).
19. J. B. Pendry, *Science* **306**, 1353 (2004).
20. We thank J. Pendry for suggesting the use of magnetic resonances and M. Grossman and I. Čelanović for technical assistance. Supported by NSF Materials Research Science and Engineering Center grant DMR 02-13282, U.S. Department of Energy grant DE-FG02-99ER45778, and the Army Research Office through the Institute for Soldier Nanotechnologies under contract DAAD-19-02-D0002.

**Supporting Online Material**

www.sciencemag.org/cgi/content/full/1143254/DC1  
SOM Text  
Figs. S1 to S5

30 March 2007; accepted 21 May 2007

Published online 7 June 2007;

10.1126/science.1143254

Include this information when citing this paper.

# Sea Anemone Genome Reveals Ancestral Eumetazoan Gene Repertoire and Genomic Organization

Nicholas H. Putnam,<sup>1</sup> Mansi Srivastava,<sup>2</sup> Uffe Hellsten,<sup>1</sup> Bill Dirks,<sup>2</sup> Jarrod Chapman,<sup>1</sup> Asaf Salamov,<sup>1</sup> Astrid Terry,<sup>1</sup> Harris Shapiro,<sup>1</sup> Erika Lindquist,<sup>1</sup> Vladimir V. Kapitonov,<sup>3</sup> Jerzy Jurka,<sup>3</sup> Grigory Genikhovich,<sup>4</sup> Igor V. Grigoriev,<sup>1</sup> Susan M. Lucas,<sup>1</sup> Robert E. Steele,<sup>5</sup> John R. Finnerty,<sup>6</sup> Ulrich Technau,<sup>4</sup> Mark Q. Martindale,<sup>7</sup> Daniel S. Rokhsar<sup>1,2\*</sup>

Sea anemones are seemingly primitive animals that, along with corals, jellyfish, and hydras, constitute the oldest eumetazoan phylum, the Cnidaria. Here, we report a comparative analysis of the draft genome of an emerging cnidarian model, the starlet sea anemone *Nematostella vectensis*. The sea anemone genome is complex, with a gene repertoire, exon-intron structure, and large-scale gene linkage more similar to vertebrates than to flies or nematodes, implying that the genome of the eumetazoan ancestor was similarly complex. Nearly one-fifth of the inferred genes of the ancestor are eumetazoan novelties, which are enriched for animal functions like cell signaling, adhesion, and synaptic transmission. Analysis of diverse pathways suggests that these gene “inventions” along the lineage leading to animals were likely already well integrated with preexisting eukaryotic genes in the eumetazoan progenitor.

All living tissue-grade animals, or eumetazoans, are descended from the last common ancestor of bilaterians (flies, worms, snails, and humans), cnidarians (anemones, jellyfish, and hydra), and ctenophores (comb jellies) (1, 2). This eumetazoan ancestor lived perhaps 700 million years ago. Although it is not preserved in the fossil record (3), we can infer many of its characteristics—flagellated sperm, development through a process of gastrulation, multiple germ layers, true epithelia lying upon a basement membrane, a lined gut (enteron), a neuromuscular system, multiple sensory systems, and fixed body axes—because these conserved features are retained by its modern descendants.

Similarly, we can characterize the genome of this long-extinct eumetazoan progenitor by comparing modern DNA and protein sequences and

identifying conserved ancestral features that have an intrinsically slow rate of change and/or are preserved by selective pressures. Comparisons (4–6) between fruit fly, nematode, and vertebrate genomes reveal greater genomic complexity in the vertebrates [and other deuterostomes (7, 8)] as measured by gene content and structure, but at the same time show that many genes and networks are shared across bilaterians. Probing the ancestral eumetazoan genome requires sequences from even deeper branches of the animal tree, comparing bilaterian and nonbilaterian phyla.

In comparison with bilaterians, cnidarians appear morphologically simple. The phylum is defined (2) by a sac-like body plan with a single oral opening, two epithelial tissue layers, the presence of numerous tentacles, a nerve net, and the characteristic stinging cells (cnidocytes, literally translated as “nettle cells”) that give the phylum its name (fig. S1.1G). The class Anthozoa (“flower animals”) includes diverse anemones, corals, and sea pens, all of which lack a medusa stage. The other cnidarian classes are united by their pelagic medusae and characteristically linear mitochondrial genomes (9) into the Medusozoa, including *Hydra* and related hydroids, jellyfish, and box jellies. The disparate bilaterian phyla of the early Cambrian suggest a Precambrian divergence of the cnidarian lineage from the bilaterian stem, and indeed some of the oldest animal body and embryo fossils are plausibly relics of stem cnidarians [reviewed in (10, 11)].

Among Anthozoan cnidarians, the starlet sea anemone *Nematostella vectensis* is an emerging model system (12, 13). This estuarine burrowing anemone is found on the Atlantic and Pacific coasts of North America, as well as the coast of southeast England (14). *Nematostella* cultures are easily maintained in the laboratory; with separate sexes, inducible spawning, and external fertilization (12, 15), embryos are available throughout the year.

Although cnidarians are often characterized as simple or primitive, closer study of *Nematostella* and its relatives has revealed considerable molecular (16–19) and morphological complexity (13). Based on expressed sequence tag (EST) analyses (17, 18) and the targeted study of specific gene families [reviewed in (13, 16, 20–22)], signaling pathways and transcription factors involved in the early patterning and development of bilaterians are present in cnidarian genomes and are active in development (13, 23–28), indicating that these pathways and regulatory mechanisms predate the eumetazoan radiation. Perhaps most notably, genes that establish the main body axes in bilaterian embryos are also expressed asymmetrically in *Nematostella* development, even though cnidarians are conventionally viewed as radial animals [for a critical discussion, see (29)].

Here, we report the draft genome of the starlet sea anemone and use its gene repertoire and genome organization to reconstruct features of the ancestral eumetazoan genome. Analysis of the *Nematostella* genome in the context of sequences from other eukaryotes reveals the genomic complexity of this last common cnidarian-bilaterian ancestor. The emerging picture from the genome and EST studies (17, 18) is one of extensive conservation in gene content, structure, and organization between *Nematostella* and vertebrates. We show that even chromosome-scale linkage has been preserved between *Nematostella* and vertebrates. These are the most ancient conserved linkages known outside of prokaryotic operons. In contrast, the fruit fly and nematode model systems have experienced extensive gene loss (18), intron loss (30), and genome rearrangement. Thus, from a genomic perspective, the eumetazoan ancestor more closely resembled modern vertebrates and sea anemones.

***Nematostella* Genome Assembly and Gene Set**

The draft sequence of the *Nematostella* genome was produced with the use of a random shotgun

<sup>1</sup>Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA. <sup>2</sup>Center for Integrative Genomics and Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. <sup>3</sup>Genetic Information Research Institute, 1925 Landings Drive, Mountain View, CA 94043, USA. <sup>4</sup>Sars International Centre for Marine Molecular Biology, University of Bergen, Thormøhlensgt 55, 5008, Bergen, Norway. <sup>5</sup>Department of Biological Chemistry and the Developmental Biology Center, University of California, Irvine, CA 92697, USA. <sup>6</sup>Department of Biology, Boston University, Boston, MA 02215, USA. <sup>7</sup>Kewalo Marine Laboratory, University of Hawaii, Honolulu, HI 96813, USA.

\*To whom correspondence should be addressed. E-mail: dsroksar@lbl.gov

strategy (31) from approximately 6.5-fold redundant paired-end sequence coverage from several shotgun libraries of a range of insert sizes derived from a single mating pair with ~0.8% allelic variation. [For a detailed discussion of polymorphism, see supporting online material (SOM) text (32)]. The total assembly spans ~357 megabases (Mb), with half of this sequence in 181 scaffolds longer than ~470 kb. Metaphase spreads indicate a diploid chromosome number of  $2n = 30$  (fig. S2.4). Currently, there are no physical or genetic maps of *Nematostella*, so we could not reconstruct the genome as chromosomes. Nevertheless, because half of the predicted genes are in scaffolds containing 48 or more genes, the present draft assembly is sufficiently long-range to permit useful analysis of synteny with other species. The typical locus in the draft genome is in a contiguous gap-free stretch of nearly 20 kb. Comparison of the assembled sequence with open reading frames derived from ESTs shows that the assembly captures ~95% of the known protein-coding content (32). Although approximately one-third of the shotgun sequences were not assembled, they could typically be characterized as derived from long (>100 kb) tandem-repetitive minisatellite arrays suggestive of heterochromatin, implying a total genome size of ~450 Mb (32).

We estimated that the *Nematostella* genome contains ~18,000 bona fide protein-coding genes, comparable to gene counts in other animals. Combining homology-based and ab initio methods with sequences from more than 146,000 ESTs, we predicted ~27,000 complete or partial

protein-coding transcripts in the genome (32). More than 12,000 of these are found in robust eumetazoan gene families and are therefore supported as orthologs of genes in other animals. Whereas ~22,000 of all predicted genes have a significant alignment [Basic Local Alignment Search Tool (BLAST)  $e$  value  $< 10^{-10}$ ] to known proteins in SwissProt/Trembl and therefore have some homology support, analysis of a random sampling of genes suggests that some of these appear to be gene fragments, possible pseudogenes, relics of transposable elements, or allelic variants, leading to a discounting of the true gene count to ~18,000 (32). More than 25% of the genome is composed of repetitive elements that are mutated inactive transposable elements, including DNA transposons and both long terminal repeat (LTR) and non-LTR retrotransposons (table S2.3).

### The Ancestral Eumetazoan Gene Set

By comparing the gene complement of *Nematostella* with other metazoans, we attempted to reconstruct the gene repertoire of the eumetazoan (i.e., cnidarian-bilaterian) ancestor and to infer the gains, losses, and duplications that occurred both before and after the eumetazoan radiation. To approximate the gene repertoire of the eumetazoan ancestor, we constructed 7766 putatively orthologous gene families that are anchored by reciprocal best-scoring BLAST alignments (33) between genes from anemone and one or more of fly, nematode, human, frog, or pufferfish (32). Each family thus represents a single gene in the eumetazoan ancestor whose descendants survive in recognizable form as mod-

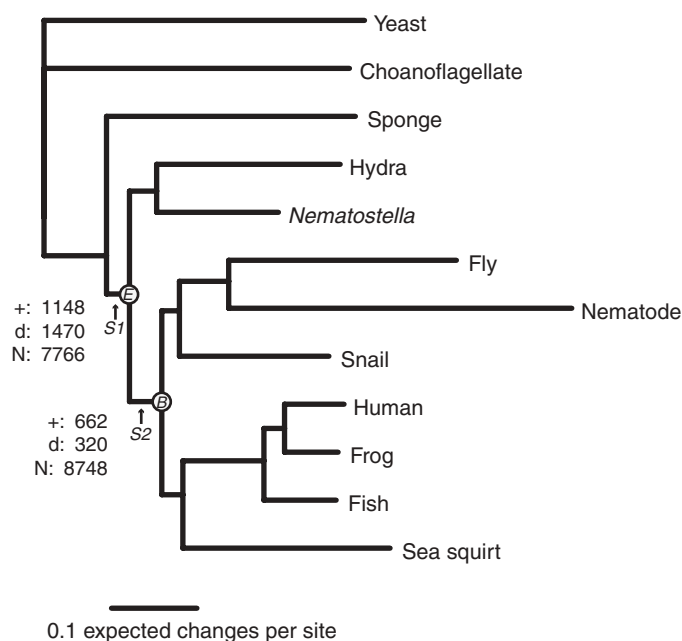
ern genes in both cnidarians and bilaterians. These families account for a substantial fraction of genes in modern animals: We estimated that nearly two-thirds of human genes (13,830) are descended from these progenitors through subsequent gene family expansions along the human lineage, and a comparable number (12,319) of predicted *Nematostella* genes arose by independent diversifications along the cnidarian branch, but only 7309 (~50%) and 7261 (~40%) were found in *Drosophila* and *Caenorhabditis elegans*, respectively. Given that we cannot capture genes that were present in the eumetazoan progenitor but became highly diverged or lost in one or more sequenced descendants, our reconstructed ancestral gene set is necessarily incomplete, but it nevertheless provides a starting point for further analysis.

Of the 7766 ancestral eumetazoan gene families, only 72% (5626) are represented in the complete genomes of all three major modern eumetazoan lineages: cnidarians (i.e., *Nematostella*), protostomes (i.e., *Drosophila* and/or *C. elegans*), and deuterostomes (requiring presence of at least two of pufferfish, frog, and human). We found 1292 eumetazoan gene families that had detectable descendants in anemone and at least two of the three vertebrates, but that appeared to be absent in both fruit fly and soil nematode. This indicates that they were either lost or highly diverged in both of these model protostomes, extending the list of such genes found in EST studies (17, 18). The forthcoming genome sequences of crustaceans, annelids, and mollusks will help address which of these genes survived in the protostome lineage but were convergently lost in flies and nematodes. In contrast, only 33 genes were found in *Nematostella* and both *Drosophila* and *C. elegans*, but not in any vertebrate. These results represent putative deuterostome or vertebrate loss, indicating a much lower degree of gene loss in the vertebrates than in the ecdysozoan model systems. We found 673 gene families that were represented in model protostomes and vertebrates but not in *Nematostella*. These are candidates for bilaterian novelties, but some will no doubt turn out to be losses or divergent sequences in *Nematostella*.

### Molecular Evolution of the Eumetazoa

To address evolutionary relationships between animals, we inferred the phylogeny of Metazoa by combining *Nematostella* data with available genomic sequences from diverse animals, using a subset of 337 single-copy genes suitable for deep phylogenetic analysis (32). In Fig. 1, relative branch lengths represent the accumulation of amino acid substitutions in each lineage across this set of proteins. Our whole-genome analysis groups the fruit fly with the soil nematode, in support of the superphylum Ecdysozoa, a major element of the “new animal phylogeny” (34), in contrast with other whole-genome-based studies that support an early branching acoelomate clade that includes *C. elegans* (35, 36). As expected,

**Fig. 1.** Bayesian phylogeny of Metazoa. Bayesian analysis infers metazoan phylogeny and rate of amino acid substitution from sequenced genomes based on 337 single-copy genes in *Ciona intestinalis* (sea squirt), *Takifugu rubripes* (fish), *Xenopus tropicalis* (frog), human, *Lottia gigantea* (snail), *Drosophila melanogaster* (fly), *C. elegans* (nematode), *Hydra magnipapillata* (hydra), *Nematostella*, *Amphimedon queenslandica* (sponge), *Monosiga brevicollis* (choanoflagellate), and *Saccharomyces cerevisiae* (yeast). All nodes were resolved as shown in 100% of sampled topologies in Bayesian analysis. The scale bar indicates



the expected number of amino acid substitutions per aligned amino acid position. E, the eumetazoan (cnidarian-bilaterian) ancestor; B, the bilaterian (protostome-deuterostome) ancestor. The number of new genes (+), genes created by gene duplication (d), and the total number of reconstructed ancestral genes of the recent common ancestor (N) are labeled for S1 and S2, the eumetazoan and bilaterian stems, respectively (32).

the two cnidarians *Nematostella* and *Hydra* form a monophyletic group that branched off the metazoan stem before the radiation of bilaterians. The depth of the *Nematostella-Hydra* split (comparable to the protostome-deuterostome divergence) emphasizes the distant relationship between anthozoans and hydrozoans. This supports the paleontological evidence that the radiation of the cnidarian phylum is quite ancient (37) and suggests that substantial variation in gene content and gene-family diversity may be found when the anemone genome is compared with that of the hydrozoan *Hydra*. For convenience, here we refer to the last common ancestor of cnidarians and bilaterians as the eumetazoan ancestor, although the precise phylogenetic placement of ctenophores may revise this designation.

Long branch lengths, indicating increased levels of sequence divergence, were found along the fly, nematode, and sea squirt lineages, consistent with systematic trends observed in BLAST-based analyses of ESTs (17, 18). The sea anemone sequences, however, appear to be evolving at a rate comparable to, or even somewhat slower than, vertebrates. Although accelerated rates of molecular evolution have been documented in flies and echinoderms (38) relative to vertebrates, our analysis does not support the extrapolation of these higher rates to all invertebrates. With the use of our branch lengths, a very crude molecular clock interpolation based on the eukaryotic time scales of Douzery *et al.* (39) suggests that the

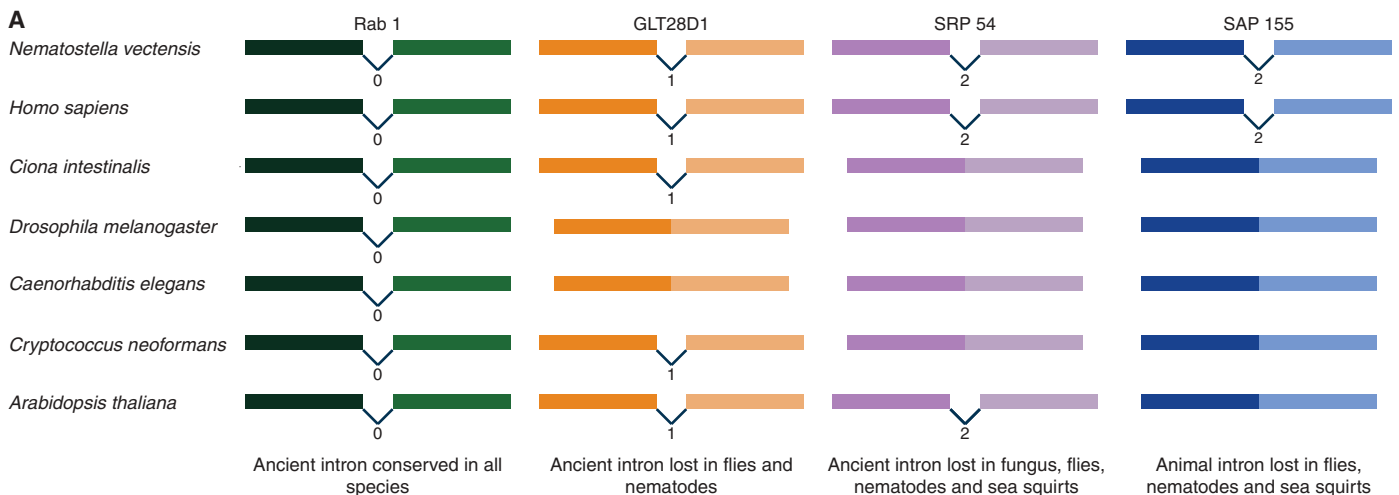
eumetazoan ancestor lived ~670 to 820 million years ago (32). This very rough estimate has numerous caveats—most notably that there is no guarantee that the rate of protein evolution was constant on the eumetazoan stem—but provides a rough time scale for the eumetazoan radiation.

### Conservation of Ancient Eumetazoan Introns

Comparison of *Nematostella* genes to those of other animals reveals that the ancestral eumetazoan genome must have been intron-rich, with gene structures closely resembling those of modern vertebrate and anemone genes. Introns that are shared between *Nematostella* and vertebrates and/or other bilaterians are most parsimoniously interpreted as conserved ancient eumetazoan introns (40). Not only are the numbers of exons per gene similar between *Nematostella* and vertebrates, but the precise location and phase (i.e., the positioning of the splice sites relative to codon boundaries) of introns are also highly conserved between the anemone and human (Fig. 2A). Within alignable regions, nearly 81% of human introns are found in the same position and phase in *Nematostella*; conversely, 82% of the anemone introns are found in orthologous positions in human genes (32). Whereas intron conservation between the annelid *Platynereis* and vertebrates implies that the Protostome-Deuterostome ancestor was intron-rich (30), the analysis of *Nematostella* extends this result to the eumetazoan ancestor.

Using whole-genome data sets, we estimated the tempo of intron evolution across metazoan genomes (32). Figure 2B shows intron gain and loss events inferred by weighted parsimony analysis of 2645 intron positions that lie within highly conserved protein sequence in five representative animals, the flowering plant *Arabidopsis*, and the relatively intron-rich fungus *Cryptococcus neoformans* (32). Although fungi and animals are phylogenetically closer to each other than either group is to plants, fungi are not by themselves a sufficient outgroup for characterizing the history of eumetazoan introns, given that there are putative ancient eukaryotic introns shared by modern animals and plants that have evidently been lost in fungi (41).

Although many eumetazoan introns are evidently of ancient eukaryotic origin (41)—for example, nearly 26% of human and *Nematostella* introns are conserved with *Arabidopsis*, and 24% with *Cryptococcus*—the remainder appear to be shared only by animals. These animal introns are most parsimoniously accounted for as gains on the eumetazoan stem, as shown by the long “gain” branch in Fig. 2B. We cannot rule out the possibility, however, that such apparently animal-specific introns were indeed present in the last common ancestor of plants, fungi, and animals, but were convergently lost in both plants and fungi. Within animals, intron gains range from 8 to 22% relative to the content of the eumetazoan ancestor. Thus, assuming ~8 introns per ancestral gene, ~1 novel intron has been introduced in a typical



**Fig. 2.** Patterns of intron evolution in eukaryotes. **(A)** Examples of different patterns of intron gain and loss. Bars of the same color represent conserved regions across all species. Chevrons indicate introns and the number below the chevron shows the phase of the intron. **(B)** Branch lengths proportional to the number of inferred intron gains (left), and intron losses (right) under the Dollo parsimony assumption that introns with conserved position and phase were gained only once in evolution. The bottom scale indicates the change in intron number for gains (left) and losses (right), relative to the inferred introns of the eumetazoan ancestor. Based on a sample of 5175 introns at highly conserved protein sequence positions from *Arabidopsis thaliana* (plant), *Cryptococcus neoformans* (fungus), *C. elegans* (nematode), *D. melanogaster* (fly), *C. intestinalis* (sea squirt), *Homo sapiens* (human), and *Nematostella* (32).

modern animal gene since the eumetazoan radiation, a rate of approximately  $\sim 10^{-9}$  introns per gene per year, which is comparable to the rate of gene duplication per locus per year (42).

In contrast to intron gains, which seem to occur more or less uniformly across animal phyla, some lineages appear to have experienced extensive intron loss, notably the fly, nematode, and sea squirt, which have each discarded 50 to 90% of inferred ancestral eumetazoan introns. It remains to be seen whether the introns absent in both fly and nematode are the result of ancient loss in the ecdysozoan stem lineage (the most parsimonious explanation, shown in Fig. 2B) or are convergent (independent) losses in flies and nematodes. We can rule out ancient loss on the protostome stem on the basis of the results of Raible *et al.* (30) for the annelid *Platynereis*, which showed that the ancestral protostome genome was intron-rich.

### Conservation of Ancient Eumetazoan Linkage Groups

Conserved linkage groups representing ancestral vertebrate chromosomes can be defined by comparing fish and mammalian genomes and genetic maps, despite the presence of only modest segments of conserved gene order (43, 44). Similarly, limited conservation of synteny is recognizable within insects [such as between flies and bees (45)]. Between animal phyla, however, no large-scale conserved synteny has been identified, suggesting that signals of the ancestral eumetazoan genome organization were erased by sub-

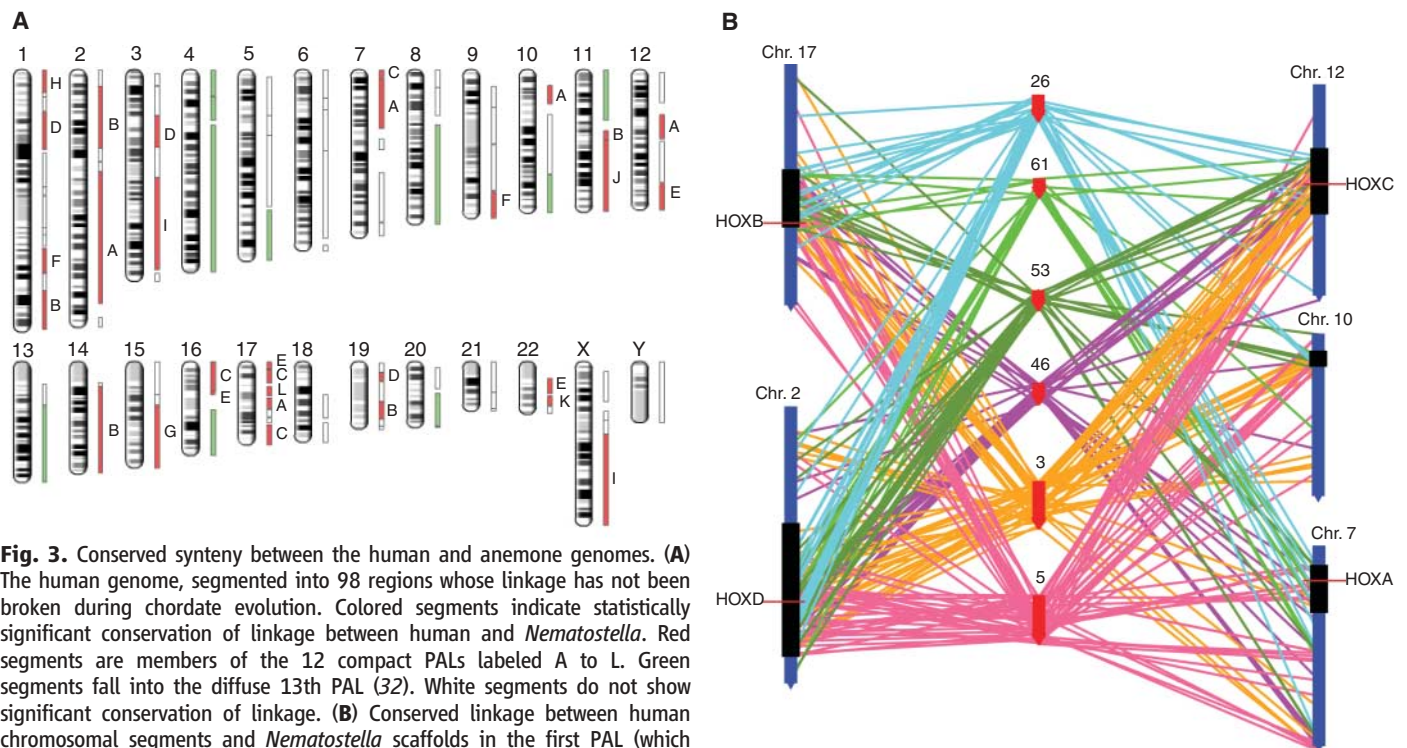
sequent chromosomal breaks and translocations along the various lineages. Despite extensive local scrambling of gene order, we find extensive conservation of synteny between the *Nematostella* and vertebrate genomes, allowing the identification of ancient eumetazoan linkage groups.

Reasoning that the prevalence of intrachromosomal inversions and rearrangements (46) might scramble local gene order yet preserve linkage, we searched for large-scale conserved synteny—that is, sets of orthologous genes on the same chromosomal segment in their respective genomes, regardless of gene order. To remove confounding signals from recent rearrangements, we used comparisons with the genomes of other chordates to identify 98 human segments that do not appear to have undergone recent breaks or fusions (Fig. 3A and fig. S7.1) (32). These segments span 89% of the human genome. The human genome was selected as a reference because it is known to have a slow rate of chromosome evolution relative to other mammals (46) and has preserved chromosomal segments relative to teleost fish (43). To search for ancient conserved linkages across eumetazoa, we then compared these human genome segments to the assembled *Nematostella* scaffolds, using a statistical test for distinguishing significant enrichment for genes linked in both species.

For every scaffold-segment pair, we tabulated the number of predicted ancestral eumetazoan genes with descendants found in both the *Nematostella* scaffold and human segment. This

number of shared orthologous genes was compared to a null model in which the scaffolds and segments have gene content independently drawn from the ancestral set. The “Oxford grid” shown in Table 1 shows not only that there are many scaffold-segment pairs with a significant excess of shared ancestral genes, but that the anemone scaffolds and human chromosome segments can be grouped into classes, such that scaffold-segment pairs drawn from the same class are likely to have a significant excess of shared ancestral genes (32). Each class is most easily interpreted as collecting together segments of the present-day *Nematostella* and human genomes that descend from the same chromosome of the eumetazoan ancestor, and therefore defines a putative ancestral eumetazoan linkage group (PAL). The complete Oxford grid showing all 13 eumetazoan PALs is shown in table S7.2.

The conserved linkage is extensive, and it accounts for a large fraction of the ancestral eumetazoan set. Of the 4402 ancestral eumetazoan gene families represented in the largest anemone scaffolds and human segments (i.e., in the genomic regions large enough to permit statistically significant analysis and therefore eligible for consideration in our analysis), more than 30% (1336) participate in a conserved linkage group. This is a lower bound on the true extent of the remnant ancient linkage groups because the length of the *Nematostella* scaffolds and the use of conservative statistical criteria limit our analysis. A more sensitive approach can



**Fig. 3.** Conserved synteny between the human and anemone genomes. **(A)** The human genome, segmented into 98 regions whose linkage has not been broken during chordate evolution. Colored segments indicate statistically significant conservation of linkage between human and *Nematostella*. Red segments are members of the 12 compact PALs labeled A to L. Green segments fall into the diffuse 13th PAL (32). White segments do not show significant conservation of linkage. **(B)** Conserved linkage between human chromosomal segments and *Nematostella* scaffolds in the first PAL (which includes the human Hox clusters). *Nematostella* scaffolds 26, 61, 53, 46, 3, and 5 (red arrows) and human chromosomes 17, 12, 10, 7, and 2 (blue arrows) are shown with length proportional to the number of genes descended from the inferred ancestral set. Lines, color-coded by *Nematostella* scaffold, join the positions of orthologous *Nematostella* and human genes. The five segments of the human genome that are grouped into PAL A are indicated by black boxes. Red lines indicate the positions of the four human Hox clusters.

assign more than twice as many ancestral genes to a PAL (32). The 40 human segments that show conserved synteny with *Nematostella* cover half of the human genome. Within such human segments, typically 40 to 50% of eumetazoan-derived genes have counterparts in syntenic *Nematostella* segments, and vice versa. This is a notable total, given that any chromosomal fusions and subsequent gene order scrambling on either the human or *Nematostella* lineage during their ~700 million years of independent evolution would attenuate the signal for linkage.

The observation of conserved linkage groups is most easily explained as the remnants of large ancestral chromosomal segments containing hundreds of genes that have evolved without obvious constraint on gene order within each block. Seven of the PALs link anemone scaffolds to multiple regions of the human genome in a manner consistent with multiple large-scale duplication events along the vertebrate lineage [reviewed in (47)]. These seven PALs represent the ancestral (preduplication) linkage of these regions. The extent of this conserved linkage suggests either that the neutral rate of interchromosomal translocations is low (on the order of a few breaks or fusions per chromosome since the eumetazoan ancestor, excluding intrachromosomal rearrangements) or that selection has acted to maintain linkage of large groups of genes, perhaps constrained by higher-level chromosomal organization (48) and/or long-range gene regulation (49).

An ancestral linkage group of particular interest includes the human Hox clusters of homeobox transcription factors that regulate anterior-posterior identity in bilaterians. Putative Hox genes in *Nematostella* and other cnidarians are also expressed in spatial patterns consistent with an ancient role in embryonic development (50–52). Tetrapods have four Hox clusters that arose by duplication on the vertebrate stem—HoxA (human chromosome 7p15.2), HoxB (17q21.32),

HoxC (12q13.13), and HoxD (2q31.1)—which all appear in the same eumetazoan PAL, linked to eight *Nematostella* scaffolds (Fig. 3B), defining the ancestral genomic context for Hox genes. *Nematostella* has several clusters of homeobox genes (52–54), but only those on scaffolds 3 and 61 are embedded within the ancestral eumetazoan Hox context, providing independent support for the assignment of these homeobox genes as bona fide *Nematostella* Hox genes (50, 52, 53, 55). There is an extensive block of 225 ancestral genes (table S7.3) that were linked to Hox in the eumetazoan ancestor and have retained that linkage in both the modern human and anemone genomes.

**Origins of Eumetazoan Genes**

Where did the eumetazoan gene repertoire come from? Nearly 80% (6182 out of 7766) of the ancestral eumetazoan genes have clearly identifiable relatives (i.e., proteins with significant sequence homology and conserved domain architecture) outside of the animals, including fungi, plants, slime molds, ciliates, or other species available from public data sets (32). These are evidently members of ancient eukaryotic gene families that were already established in the unicellular ancestors of the Metazoa and are involved in core eukaryotic cellular functions. Although these eumetazoan gene families are conserved with other eukaryotes, animals have a unique complement due to family expansion and contraction on the eumetazoan stem. The eumetazoan genes of ancient eukaryotic ancestry are themselves descended from ~5148 eukaryotic progenitors by nearly 1000 gene duplications along the eumetazoan stem—i.e., after the early radiation of eukaryotes ~1100 to 1500 million years ago (56) but before the divergence of cnidarians and bilaterians (32).

The remaining 20% (1584) of the ancestral eumetazoan gene set comprises animal novelties that were apparently “invented” along the eumeta-

zoan stem. The mechanism for the creation of “new” genes is obscure (57) but may involve gene duplication followed by bursts of rapid sequence divergence (thus masking the similarity with sister sequences) and/or de novo recruitment of gene and/or noncoding fragments into functional transcription units. We classified these eumetazoan novelties into three categories based on their origin (Fig. 4A).

The first and largest group (type I novelty) comprises animal genes that have no identifiable relatives (with BLAST) outside of animals in the available sequence data sets, and accounts for 15% (1186) of ancestral eumetazoan genes. These include important signaling factors, such as the secreted wingless (Wnt) and fibroblast growth factor (FGF) families, and transcription factors, including the T-box and mothers-against-decapentaplegic (SMAD) families (Table 2). Not only were these genes present in the eumetazoan ancestor, but they had already duplicated and diversified on the eumetazoan stem to establish the subfamilies that, nearly 700 million years later, are still maintained in modern vertebrates. [See for example the Wnt family (58).]

Type II novelties (2% of the eumetazoan complement, or 158 genes) incorporate animal-only domains in combination with ancient eukaryotic sequence. The ancestry of these genes can be traced back to the eukaryotic radiation through their ancient domains, but the novel domains they contain were evidently invented (or evolved into their recognizable animal form) and coupled to more ancient domains on the eumetazoan stem. For example, Notch proteins have two Notch domains found only in metazoans in addition to ancient eukaryotic ankyrin and epidermal growth factor (EGF) domains; focal adhesion kinase (FAK) is targeted to focal adhesions in eumetazoans because of the addition of an animal-specific focal adhesion-targeting domain to the ancient kinase domain.

**Table 1.** Detail of the “Oxford grid” which tabulates the number of ancestral gene clusters shared between the 22 *Nematostella* scaffolds (columns) and 14 segments of the human genome (rows) that are

assigned to PALs A, B and C. Cell symbols indicate Bonferroni-corrected *P* value < 0.01 (\*), < 0.05 (†), < 0.5 (‡). Detailed methods, and the complete Oxford grid can be found in the SOM text.

PAL:		A								B								C						
Nematostella scaffold		3	5	46	26	53	61	44	144	7	74	18	88	52	42	156	89	10	8	34	118	91	191	
Human Chromosome Segment	2q11.2-35	25*	32*	17*	16*	17*	9*	12*	7*	1	2			2	1	1		1	1	2				
	12q12-14.3	16*	14*	9*	5	8*	3	6‡	5‡			1									1	1		
	17q12-21.32	12*	8‡	4	10*	6‡	4	3	1			1	1		1					1				
	7p11.2-21.3	4	10*	3	3	2	7*	1	2			1	1					2	1	1	1	1	1	
A	10p11.22-13	8*	6‡	1	1	2	1	4‡	1															
	B	14q12-32.33	10	3	2	4	5	3	3	2	23*	12*	13*	11*	17*	11*	9*	8*	1	2				1
		11q12.1-13.1	4					2	2	1	12*	7*	1	6*	6‡	1		4			2			
		1q32.2-44	6	4		1	1	2		1	11*	6*	6‡	3	6‡	2	2	1						1
		19q13.11-13.33	4	1	2	2	1	2	1		5	8*	8*	4	6‡	2		3	2	3		1	1	
2p13.2-24.3	5	2	2	1	2	1	2	1	8‡	5	10*	5	3	1	5‡	3	2		1					
C	17q23.3-25.3	1	2						1	2	2	1						19*	10*	12*	8*	7*	3	
	16p11.2-13.3	1			1	1						1	1	1				17*	19*	9*	5‡	6‡	6*	
	7p22.1-22.3											1		1	1			6*	3	3	5*	2		
	17p11.2-13.1				1					1				1		1		6*	2		1	3		

Finally, type III novelties (3%, or 240 gene families) consist of animal genes whose domains are all ancient (i.e., each found in other eukaryotes) but that occur in apparently unique combination in eumetazoa relative to known nonanimal genes (32) because of gene fusions and/or domain-shuffling events on the eumetazoan stem. For example, both the LIM (lin-11, islet, mec-3) protein-protein interaction and homeobox DNA binding domains are found in nonanimal eukaryotes, but only animals have the LIM-homeobox combination. Although such domain-shuffling (57) events are relatively rare, they are disproportionately involved in characterized biochemical pathways, perhaps by bringing together existing catalytic capabilities, localization, and regulatory domains into the same protein (table S8.1).

### Eumetazoan Networks and Pathways

How are the genes that were invented along the eumetazoan stem related to the organismal novelties associated with Eumetazoa? Satisfyingly, but perhaps not surprisingly, we found that the novel genes were significantly enriched for signal transduction, cell communication and adhesion, and developmental processes (32). The eumetazoan ancestor was the progenitor of all extant animals with nervous systems, and genes with neuronal activities are abundant among its novelties (Table 3). It is at first glance surprising that genes known to be involved in mesoderm development in bilaterians are also enriched among eumetazoan novelties, given that the textbook picture of cnidarians is that

they lack mesoderm. Yet we know that many of these genes are associated either with basic patterning functions and/or the regulation of cell migration and fate. The precise deployment and interaction of these genes in the ancestral eumetazoan is therefore still a matter of debate (26, 27, 59–61). Experiments in cnidarians, however, in combination with information about mesodermal networks in bilaterians, could, in principle, constrain the ancestral genetic network and address whether or not the ancestor deployed these genes to generate this key germ layer.

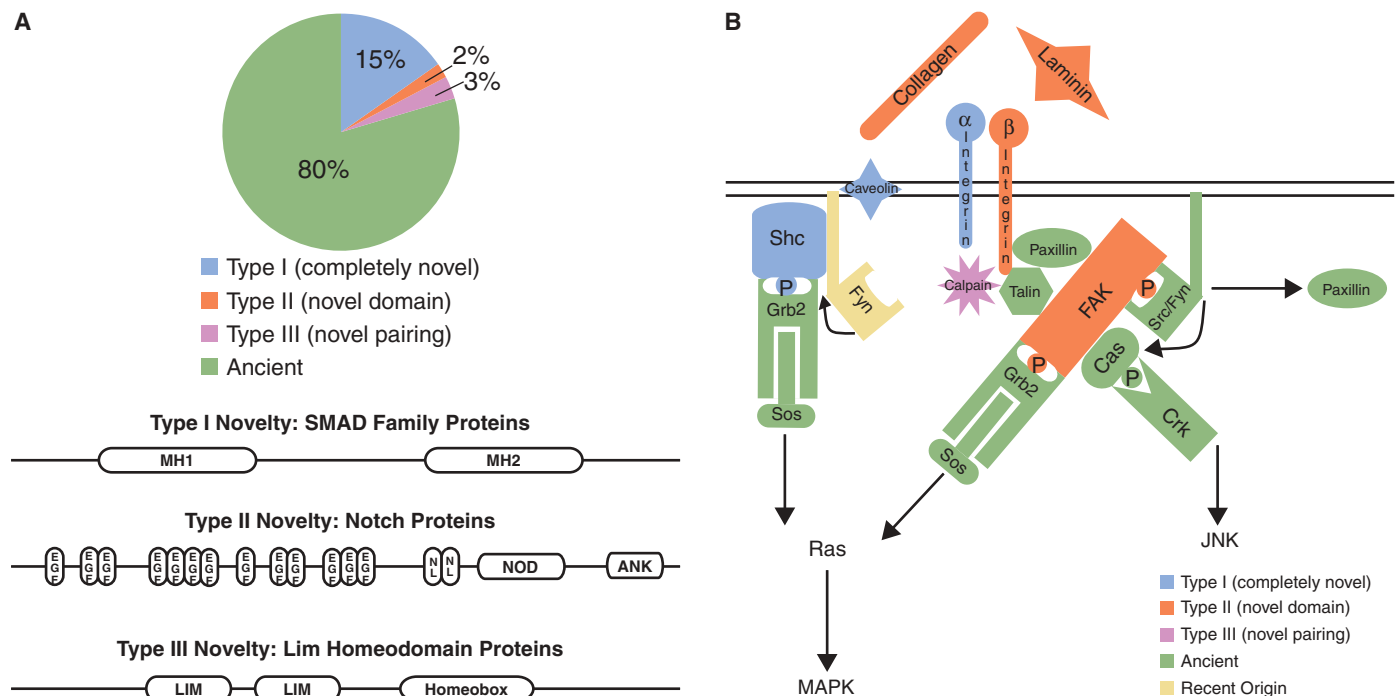
Individual “new” genes are by themselves unlikely to bring about the suite of features needed to evolve animal characteristics from unicellular organisms. Rather, we expect that to generate organismal novelty, such new genes must be integrated with other novel and existing genes to evolve expanded or modified biochemical pathways and/or regulatory networks. Given the reconstructed eumetazoan genome and its various types of novel genes, we conclude by briefly considering selected eumetazoan pathways and processes to see how novel animal genes were incorporated into cellular and organismal functions.

**Cell adhesion.** In Bilateria, the integrin pathway mediates signaling from the extracellular matrix (ECM) that elicits various responses to modulate cell adhesion, motility, and the cell cycle (62). A detailed look at integrin signaling (Table 3 and Fig. 4B) reveals that most of the core components of the FAK and Fyn/Shc pathways were present in the eumetazoan an-

cestor. Various ancient cytosolic proteins (Talin, Paxillin, Grb2, Sos, and Crk) have been brought under the control of two novel receptors, integrin- $\alpha$  and integrin- $\beta$  (the former being a type I novelty and the latter a type II novelty). FAK is a cytosolic component that appears as a type II novelty in eumetazoans, and calpain—a protease that regulates the aggregation of talin, paxillin, and FAK around the receptor—appears as a novel domain combination of ancient domains. Caveolin, a membrane adapter that couples the integrin- $\alpha$  subunit to Fyn is present in the *Nematostella* genome and is a type I novelty. Fyn itself is a more recent invention derived on the tetrapod stem by gene duplication.

Cell-cell adhesion mediated by cell-ECM interactions is a hallmark of animal multicellularity (63). Basement membrane proteins such as collagen and laminin arose as type II novelties along the stem leading to the Eumetazoa, whereas others such as nidogen are novel pairings of ancient domains (Table 1). Matrix metalloproteases also were invented as type II novelties, whereas guidance cues such as netrin and semaphorin that mediate adhesion are novelties with no evident homology to ancient eukaryotic proteins.

**Signaling pathways.** Animals rely on cell-cell signaling for cellular coordination during and after development (64). Various components of the Wnt and transforming growth factor- $\beta$  (TGF $\beta$ ) signaling pathways in the genome of *Nematostella* have been reported (18, 27, 58, 65–68). In both pathways, the secreted ligands and their antagonists [such as Wnt, SFRP, bone morphogenetic



**Fig. 4.** Origins of eumetazoan genes. **(A)** Pie chart showing the percentages of genes in the eumetazoan ancestors according to their origin: type I novelties with no homology to proteins in nonanimal outgroups (blue), type II novelties with novel animal domains paired with ancient domains (orange), type III

novelties with new pairings of ancient domains (pink), and ancient genes (green). **(B)** A schematic representation of the FAK and Shc/Fyn pathways in integrin signaling. The proteins are color-coded to reflect their ancestry, as in (A). JNK, c-Jun N-terminal kinase.

protein (BMP), and chordin] are novelties (Fig. 4B). Some, such as Wnt, secreted frizzled-related protein (SFRP), Dpp/BMP, activin, and chordin are type I novelties with no homology to proteins from outgroups; some are type II novelties (dickkopf), and others (such as tolloid) are novel pairings of ancient domains (type III). The receptor in the Wnt pathway, frizzled, also arose as a type I eumetazoan novelty. Transcription factors that are activated downstream of Wnt signaling are ancient, but the ones involved in TGF $\beta$  signaling are novel. Type I receptors of the TGF $\beta$  pathway arose as a pairing of novel animal domains with ancient domains (type II novelties) and type II receptors turn out to be ancient eukaryotic kinase genes that were co-opted for this function.

The presence of essentially complete signal transduction pathways in the common gene set of cnidarians and bilaterians suggests that the integration of novel eumetazoan genes into these systems was largely complete in the eumetazoan ancestor. A general trend in the evolution of signaling pathways may have been the co-option of cytosolic signaling components

into pathways that could be regulated by newly invented ligands and receptors. For example, in the case of FGF signaling, the interactions of ancient cytosolic components [such as Grb2, Sos, and mitogen-activated protein kinase (MAPK)] could be elaborated with the addition of novel proteins (such as FGF and Shc) or of novel domains added to old proteins (such as Raf homolog) or novel pairings of old domains (such as FGF receptor and phospholipase C- $\gamma$ ).

*Emergence of the neuromuscular system.* Cnidarians and ctenophores are the earliest branching metazoan phyla that have a nervous system, although they lack overt centralization of the kind observed in bilaterians. Genes with neural functions in the Bilateria have been implicated in the cnidarian nervous system (69, 70). Numerous genes known to be involved in neurogenesis, such as members of the homeobox and basic helix-loop-helix (bHLH) transcription factor families (Emx, Otp, Otx, and achaete-scute), can be traced to ancient eukaryotic genes with these signature domains. Some are novel pairings of ancient domains (such as neuropilin and LIM-homeobox

genes), some are pairings of old domains with novel animal-specific domains (such as Dsh, Arx, and neuralized) and others are novel animal genes (such as Hes, Gcm, netrin, semaphorins, and dachshund). Certain enzymes important in synaptic transmission (such as 3,4-dihydroxy-L-phenylalanine (DOPA)- $\beta$  monoxygenase) and some vesicular trafficking proteins (such as synaptophysin) appear as novel (type I) eumetazoan proteins. Regulatory subunits for ion channels important in nerve conduction and muscular function can be type I novelties (such as voltage-dependent calcium channel  $\beta$  subunit and potassium large-conductance calcium-activated channel) or type III novelties (such as voltage-dependent calcium channel  $\alpha 2/\delta$  subunit). Various components of the dystrophin-associated protein complex (DPC) in the sarcolemma such as dystrophin, syntrophin,  $\beta$ -dystrobrevin, and  $\beta$ -sarcoglycan are type I novelties. Other sarcomere proteins are type II novelties (such as nebulin and tropomodulin). This diversity of origins of genes with roles in the neuromuscular system suggests that tracing the evolution of nerves and muscle will require

**Table 2.** Origins of developmental signaling pathway components inferred in the eumetazoan ancestor. ERK, extracellular signal-regulated kinase; MEK, MAPK kinase; GSK3, glycogen synthase kinase 3; APC, anaphase-promoting complex; TCF/LEF, T cell factor/lymphoid enhancer factor; ATF, activating transcription factor; ACVR2, activin receptor, type II; ADAM10, a disintegrin and metalloprotease

domain 10; PEN2, presenilin enhancer 2; SYK, spleen tyrosine kinase; IGF, insulin-like growth factor; PTEN, phosphatase and tensin homolog; GTPase, guanosine triphosphatase; SOCS, suppressor of cytokine signaling; REL/NF $\kappa$ B, reticuloendotheliosis viral oncogene/nuclear factor  $\kappa$ B; NFAT, nuclear factor of activated T cells; STAT5, signal transducers and activators of transcription 5.

Pathway	Type I novelty	Type II novelty	Type III novelty	Ancient gene
Integrin signaling	Integrin-alpha; caveolin	Collagen; Integrin- $\beta$ ; Fak; Jun	Calpain	Talin,; vinculin; paxillin; Ras; Grb2; SoS; Rap; ERK; MEK, Crk
Wnt signaling	Wnt; secreted frizzled related factors; frizzled; strabismus/van gogh	Dickkopf; arrow; dishevelled; axin		$\beta$ -catenin; GSK3; APC; TCF/LEF; groucho
TGF $\beta$ signaling	Dpp/BMP; activin; gremlin; chordin; follistatin; R-SMAD; I-SMAD; co-SMAD	Type I receptors: TGF $\beta$ R1, BMPR1A; ATF/JunB; snoN	Tolloid/BMP1	Type II receptors: ACVR2, BMPR2
Notch signaling	Numb; hairy/E(spl)	Notch		Jagged; deltex; fringe; presenilin; ADAM10; nicastrin; furin; Aph1; PEN2; mastermind
Ephrin signaling		Ephrin; Fak	Eph (receptor)	Abl/SYK
Insulin signaling	Insulin	Insulin receptor substrate; phosphoinositide-3-kinase, catalytic	Insulin receptor/IGF; phosphoinositide-3-kinase, class 2	phosphoinositide-3-kinase, class 3; phosphoinositide-3-kinase, regulatory subunit; 3-phosphoinositide-dependent protein kinase-1; PTEN
FGF signaling	FGF; Shc	Raf homolog serine/threonine-protein kinase; Ras GTPase activating protein	FGFR; RAS protein activator; phospholipase C- $\gamma$ ; phosphoinositide-3-kinase, class 2; protein kinase C $\iota$	MAPK; phosphoinositide-3-kinase, class 3; Grb2; Protein kinase C; SoS; Rac
Cytokine signaling	Inositol 1,4,5-triphosphate receptor; SOCS; arrestin; guanine nucleotide binding protein $\gamma$ ; regulator of G-protein signaling; REL/NF $\kappa$ B; NFAT	Adenylate cyclase 5/6; STAT5; ATF/Jun	CDC42 binding protein kinase	MAPK; Rho kinase; Rho



detailed studies of the functions of these genes in organisms at the base of the metazoan tree.

### Concluding Remarks

Modern animal genomes retain features inherited from the eumetazoan ancestor that have been elaborated on, and sometimes overwritten by, subsequent evolutionary elaborations and simplifications. Here, we compared the genomes of the sea anemone with diverse bilaterians, both to infer the content and organization of the genome of the eumetazoan ancestor and to trace the origins of uniquely animal features. In many ways, the ancestral genome was not so different from ours; it was intron-rich and contained nearly complete toolkits for animal biochemistry and development, which can now be recognized as pan-eumetazoan,

as well as the core gene set required to execute sophisticated neural and muscular function. The ancestor had blocks of linked genes that remain together in the modern human and anemone genomes—the oldest known conserved synteny outside of prokaryotic operons. Whereas fruit flies and soil nematodes have proven to be exquisite model systems for dissecting the genetic underpinnings of metazoan development and physiology, their genomes are relatively poor models for the ancestral eumetazoan genome, having lost introns, genes, and gene linkages.

The eumetazoan ancestor possessed more than 1500 genes that are apparently novel relative to other eukaryotic kingdoms. Some are the result of domain shuffling, bringing together on the animal stem new combinations of domains that are shared

with other eukaryotes. But many animal-specific genes contain sequences with no readily recognizable counterparts outside of animals; these may have arisen by sequence divergence from ancient eukaryotic genes, but the trail is obscured by deep time. Although we can crudely assign the origins of these genes to the eumetazoan stem, this remains somewhat unsatisfying. The forthcoming genomes of sponges, placozoans, and choanoflagellates will allow more precise dating of the origins and diversification of modern eumetazoan gene families, but this will not directly reveal the mechanisms for new gene creation. Presumably, many of these novelties will ultimately be traced back, through deep sequence or structural comparisons, to ancient genes that underwent extreme “tinkering” (71).

**Table 3.** Origins of selected metazoan processes inferred in the eumetazoan ancestor. CREB, cyclic adenosine 3',5'-monophosphate response element-binding protein; HIF, hypoxia-inducible factor; CES, carboxylesterase; cGMP, guanosine 3',5'-monophosphate; TNF, tumor necrosis factor; BOK, B cell leukemia/lymphoma 2-related ovarian killer; GULP, engulfment adaptor PTB domain containing; CRADD, caspase 2 and receptor-interacting serine-threonine kinase domain-containing adaptor with death domain; FMR, fragile X mental

retardation syndrome; CARD, caspase recruitment domain family; SRGAP, Slit-Robo Rho GTPase activating protein; TNFRSF, TNF receptor superfamily; TRAF, TNF receptor-associated factor; SUMO, small ubiquitin-related modifier; L3MBT, Lethal(3)malignant brain tumor protein homolog; SKI, sarcoma viral oncogene homolog; AP-2, activating protein 2; MAF, musculoaponeurotic fibrosarcoma oncogene homolog; CBP, CREB-binding protein; ETO/MTG8, eighty twenty one/myeloid translocation gene 8.

Process	Type I novelty	Type II novelty	Type III novelty	Ancient gene
Neurogenesis	Hes; Gcm; Ephrin; netrin; semaphoring; dachsund; ski oncogene	Notch; NGFR; Dsh; Arx; CREB/ATF; neuralized	Neuropilin; Lhx; ephrin receptor	Single-minded/HIF; achaete-scute; elav; Emx; Otp; Jagged; Deltex; Irx; Gli, Otx/Phox; stonal/neuroD/neuroG; reticulon
Synaptic transmission	Nitric oxide synthase (neuronal) adapter protein; DOPA- $\beta$ monoxygenase; calcium channel voltage-dependent $\beta$ ; syntrophin; synaptophysin; dystrophin; potassium large conductance calcium-activated channel, subfamily M $\beta$	Cholinergic receptor, nicotinic; neurexin	K-voltage gated channel; discs large	Glutamate receptor; synaptotagmin; intersectin; synapsin; neuroligin/CES; syntaxin; glutamate transporter
Extracellular matrix	Netrin; dermatopontin; semaphorin; glypican; stereocilin	Collagen; spondin; laminin	Nidogen; stabilin; neuropilin; matrix metalloprotease; thrombospondin	Leprecan; microfibrillar-associated protein
Cell junction	par-6	Tight junction protein		Salvador
Muscle contraction	Voltage-dependent calcium channel $\beta$ , $\beta$ -sarcoglycan, $\beta$ -dystrobrevin	Cholinergic receptor, nicotinic; nebulin; tropomyosin; calponin/transgelin	Voltage-dependent calcium channel $\alpha$ 2/ $\delta$ subunit; inositol triphosphate receptor; calcium activated potassium channel slowpoke	Phosphorylase kinase; myosin light chain cytoplasmic; calcium channel alpha subunit; cGMP-dependent protein kinase; calcium/calmodulin-dependent kinase II; myosin regulatory light chain
Apoptosis	TNFS/10/11; Bcl2; BOK; GULP; CRADD; caspase 8/10; growth arrest and DNA-damage-inducible; DNA fragmentation factor 40-kD subunit; interleukin enhancer-binding factor 3; FMR	Neuronal apoptosis inhibitory protein; CARD9/11	NGFR; SRGAP; calpain	TNFRSF; TRAF; scavenger receptor class B; huntingtin interacting protein; programmed cell death 1/5; Bcl2-associated athanogene; Akt; SUMO; defender against cell death 1; apoptosis-inducing factor-like mitochondrion-associated inducer of death; death-associated protein kinase
Transcription factors	L3MBT; T-Box; Nuclear hormone receptor; SMAD; dachsund; gcm; NFAT; nuclear respiratory factor; SKI family; sprouty; AP-2; oncut; MAF-related	CBP/p300; ETO/MTG8/Nervy; groucho; Jun; Myt1; runt; STAT	Hairless; nuclear protein 95; LIM homeobox; CCAAT enhancer binding; aryl hydrocarbon receptor related	Zic; Gli; homeobox; bHLH; achaete-scute; sox; retinoblastoma binding protein 5/8; NFKB-related; Krueppel C2H2 type zinc finger; Irx; Deltex; ataxin

The eumetazoan progenitor was more than just a collection of genes. How did these genes function together within the ancestor? Unfortunately, we cannot read from the genome the nature of its gene- and protein-regulatory interactions and networks. This is particularly vexing as it is becoming clear—especially given the apparent universality of the eumetazoan toolkit—that gene regulatory changes can also play a central role in generating novelties, allowing co-option of ancestral genes and networks to new functions (49). Of particular interest are the processes that give rise to body axes, germ layers, and differentiated cell types such as nerve and muscle, as well as the mechanisms that maintain these cells and their interactions through the growth and repair of the organism. *Nematostella* and its genome provide a platform for testing hypotheses about the nature of ancestral eumetazoan pathways and interactions, with the use of the basic principle of evolutionary developmental biology: Processes that are conserved between living species were likely functional in their common ancestor.

#### References and Notes

- E. Haeckel, *Generelle Morphologie der Organismen*, vols. 1 and 2 (Georg Reimer, Berlin, 1866).
- C. Nielsen, in *Animal Evolution* (Oxford Univ. Press, Oxford, ed. 2, 2001), pp. 51–58.
- J. W. Valentine, in *On the Origin of Phyla* (Univ. Chicago Press, Chicago, 2004), pp. 153–196.
- G. M. Rubin *et al.*, *Science* **287**, 2204 (2000).
- G. Ruvkun, O. Hobert, *Science* **282**, 2033 (1998).
- C. I. Bargmann, *Science* **282**, 2028 (1998).
- P. Dehal *et al.*, *Science* **298**, 2157 (2002).
- Sea Urchin Genome Sequencing Consortium *et al.*, *Science* **314**, 941 (2006).
- D. Bridge, C. W. Cunningham, B. Schierwater, R. DeSalle, L. W. Buss, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8750 (1992).
- G. Narbonne, *Annu. Rev. Earth Planet. Sci.* **33**, 421 (2005).
- S. Conway Morris, *Philos. Trans. R. Soc. London Ser. B* **361**, 1069 (2006).
- C. Hand, K. Uhlinger, *Biol. Bull.* **182**, 169 (1992).
- J. A. Darling *et al.*, *Bioessays* **27**, 211 (2005).
- C. Hand, K. Uhlinger, *Estuaries* **17**, 501 (1994).
- J. H. Fritzenwanker, U. Technau, *Dev. Genes Evol.* **212**, 99 (2002).
- R. E. Steele, *Dev. Biol.* **248**, 199 (2002).
- R. D. Kortschak, G. Samuel, R. Saint, D. J. Miller, *Curr. Biol.* **13**, 2190 (2003).
- U. Technau *et al.*, *Trends Genet.* **21**, 633 (2005).
- D. J. Miller, E. E. Ball, U. Technau, *Trends Genet.* **21**, 536 (2005).
- B. Galliot, *Curr. Opin. Genet. Dev.* **10**, 629 (2000).
- K. Seipel *et al.*, *Dev. Dyn.* **231**, 303 (2004).
- S. Reber-Muller *et al.*, *Int. J. Dev. Biol.* **50**, 377 (2006).
- M. Q. Martindale, K. Pang, J. R. Finnerty, *Development* **131**, 2463 (2004).
- J. R. Finnerty, D. Paulson, P. Burton, K. Pang, M. Q. Martindale, *Evol. Dev.* **5**, 331 (2003).
- J. H. Fritzenwanker, M. Saina, U. Technau, *Dev. Biol.* **275**, 389 (2004).
- C. B. Scholz, U. Technau, *Dev. Genes Evol.* **212**, 563 (2003).
- A. H. Wikramanayake *et al.*, *Nature* **426**, 446 (2003).
- C. G. Extavour, K. Pang, D. Q. Matus, M. Q. Martindale, *Evol. Dev.* **7**, 201 (2005).
- M. Q. Martindale, J. R. Finnerty, J. Q. Henry, *Mol. Phylogenet. Evol.* **24**, 358 (2002).
- F. Raible *et al.*, *Science* **310**, 1325 (2005).
- J. L. Weber, E. W. Myers, *Genome Res.* **7**, 401 (1997).
- Materials and methods, as well as additional text, are available as supporting material on Science Online.
- R. L. Tatusov, E. V. Koonin, D. J. Lipman, *Science* **278**, 631 (1997).
- A. Adoutte *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4453 (2000).
- Y. I. Wolf, I. B. Rogozin, E. V. Koonin, *Genome Res.* **14**, 29 (2004).
- F. D. Ciccarelli *et al.*, *Science* **311**, 1283 (2006).
- J. Y. Chen *et al.*, *Dev. Biol.* **248**, 182 (2002).
- K. J. Peterson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6536 (2004).
- E. J. Douzery, E. A. Snell, E. Bapteste, F. Delsuc, H. Philippe, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15386 (2004).
- J. C. Sullivan, A. M. Reitzel, J. R. Finnerty, *Genome Informatics* **17**, 219 (2006).
- I. B. Rogozin, Y. I. Wolf, A. V. Sorokin, B. G. Mirkin, E. V. Koonin, *Curr. Biol.* **13**, 1512 (2003).
- J. S. Conery, M. Lynch, *Pac. Symp. Biocomput.* **2001**, 167 (2001).
- J. H. Postlethwait *et al.*, *Genome Res.* **10**, 1890 (2000).
- O. Jaillon *et al.*, *Nature* **431**, 946 (2004).
- The Honeybee Genome Sequencing Consortium, *Nature* **443**, 931 (2006).
- G. Bourque, E. M. Zdobnov, P. Bork, P. A. Pevzner, G. Tesler, *Genome Res.* **15**, 98 (2005).
- L. G. Lundin, D. Larhammer, F. Hallbook, *J. Struct. Funct. Genomics* **3**, 53 (2003).
- K. J. Meaburn, T. Misteli, *Nature* **445**, 379 (2007).
- M. Levine, R. Tjian, *Nature* **424**, 147 (2003).
- J. R. Finnerty, K. Pang, P. Burton, D. Paulson, M. Q. Martindale, *Science* **304**, 1335 (2004).
- K. Kamm, B. Schierwater, W. Jakob, S. L. Dellaporta, D. J. Miller, *Curr. Biol.* **16**, 920 (2006).
- J. F. Ryan *et al.*, *Genome Biol.* **7**, R64 (2006).
- D. Chourrout *et al.*, *Nature* **442**, 684 (2006).
- K. Kamm, B. Schierwater, *J. Exp. Zool. B. Mol. Dev. Evol.* (Jul 12, 2006).
- J. R. Finnerty, M. Q. Martindale, *Evol. Dev.* **1**, 16 (1999).
- F. Delsuc, S. F. Viczaino, E. J. Douzery, *BMC Evol. Biol.* **4**, 11 (2004).
- M. Long, E. Betran, K. Thornton, W. Wang, *Nat. Rev. Genet.* **4**, 865 (2003).
- A. Kusserow *et al.*, *Nature* **433**, 156 (2005).
- K. Seipel, V. Schmid, *Int. J. Dev. Biol.* **50**, 589 (2006).
- K. Seipel, V. Schmid, *Dev. Biol.* **282**, 14 (2005).
- C. A. Byrum, M. Q. Martindale, in *Gastrulation: From Cells to Embryos*, C. D. Stern, Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004), pp. 33–50.
- F. G. Giacotti, E. Ruoslahti, *Science* **285**, 1028 (1999).
- S. Tyler, *Int. Comp. Biol.* **43**, 55 (2003).
- A. Pires-daSilva, R. J. Sommer, *Nat. Rev. Genet.* **4**, 39 (2003).
- D. Q. Matus *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11195 (2006).
- D. Q. Matus, G. H. Thomsen, M. Q. Martindale, *Curr. Biol.* **16**, 499 (2006).
- P. N. Lee, K. Pang, D. Q. Matus, M. Q. Martindale, *Semin. Cell Dev. Biol.* **17**, 157 (2006).
- F. Rentzsch *et al.*, *Dev. Biol.* **296**, 375 (2006).
- M. Miljkovic-Licina, D. Gauchat, B. Galliot, *Biosystems* **76**, 75 (2004).
- D. Gauchat *et al.*, *Dev. Biol.* **275**, 104 (2004).
- F. Jacob, *Science* **196**, 1161 (1977).
- This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract no. DE-AC02-05CH11231, and Los Alamos National Laboratory under contract no. DE-AC02-06NA25396 and was supported by NIH–National Heart, Lung, and Blood Institute grant HL007279F. Genetic Information Research Institute is under the NIH grant 5 P41 LM006252-09. D.S.R., M.S., and W.D. gratefully acknowledge the support of the Gordon and Betty Moore Foundation. We thank H. Marlow, D. Matus, K. Pang, P. Lee, and C. Magie for contributions to fig. S1; and E. Begovic, E. Edsinger-Gonzales, D. Goodstein, M. Carpenter, C. David, M. Levine, J. Gerhart, and J. Valentine for useful conversations.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/317/5834/86/DC1

Materials and Methods

SOM Text

Figs. S1.1 to S7.4

Tables S1.1 to S8.1

References

21 December 2006; accepted 4 June 2007

10.1126/science.1139158

## Dentate Gyrus NMDA Receptors Mediate Rapid Pattern Separation in the Hippocampal Network

Thomas J. McHugh,<sup>1,2\*</sup> Matthew W. Jones,<sup>1,\*†</sup> Jennifer J. Quinn,<sup>3‡</sup> Nina Balthasar,<sup>4†</sup> Roberto Coppari,<sup>4§</sup> Joel K. Elmquist,<sup>4§</sup> Bradford B. Lowell,<sup>4</sup> Michael S. Fanselow,<sup>3</sup> Matthew A. Wilson,<sup>1</sup> Susumu Tonegawa<sup>1,2||</sup>

Forming distinct representations of multiple contexts, places, and episodes is a crucial function of the hippocampus. The dentate gyrus subregion has been suggested to fulfill this role. We have tested this hypothesis by generating and analyzing a mouse strain that lacks the gene encoding the essential subunit of the *N*-methyl-D-aspartate (NMDA) receptor NR1, specifically in dentate gyrus granule cells. The mutant mice performed normally in contextual fear conditioning, but were impaired in the ability to distinguish two similar contexts. A significant reduction in the context-specific modulation of firing rate was observed in the CA3 pyramidal cells when the mutant mice were transferred from one context to another. These results provide evidence that NMDA receptors in the granule cells of the dentate gyrus play a crucial role in the process of pattern separation.

The hippocampus is crucial for the formation of memories of facts and episodes (1–4). To allow similar episodes to be distinguished, it must rapidly form distinct represen-

tations of the temporal and spatial relationships comprising events (pattern separation), and because specific episodes are rarely replicated in full, the hippocampus must also be capable of