The Genome of the Ctenophore *Mnemiopsis leidyi* and Its Implications for Cell Type Evolution

Joseph F. Ryan, Kevin Pang, Christine E. Schnitzler, Anh-Dao Nguyen, R. Travis Moreland, David K. Simmons, Bernard J. Koch, Warren R. Francis, Paul Havlak, NISC Comparative Sequencing Program, Stephen A. Smith, Nicholas H. Putnam, Steven H. D. Haddock, Casey W. Dunn, Tyra G. Wolfsberg, James C. Mullikin, Mark Q. Martindale, Andreas D. Baxevanis*

Introduction: An understanding of ctenophore biology is critical for reconstructing events that occurred early in animal evolution. The phylogenetic relationship of ctenophores (comb jellies) to other animals has been a source of long-standing debate. Until recently, it was thought that Porifera (sponges) was the earliest diverging animal lineage, but recent reports have instead suggested Ctenophora as the earliest diverging animal lineage. Because ctenophores share some of the same complex cell types with bilaterians (such as neural and mesodermal cells), the phylogenetic position of ctenophores affects how we think about the early evolution of these cell types.

Methods: We have sequenced, annotated, and analyzed the 150-megabase genome of the ctenophore *Mnemiopsis leidyi*. We have performed detailed phylogenetic analyses on these new data using both sequence matrices and information on gene content. We conducted extensive genomic inventories on signaling pathway components and genes known to be critical to neural and mesodermal cell types, among others.

Results: Our phylogenetic analyses suggest that ctenophores are the sister group to the rest of the extant animals. We find that the sets of neural components present in the genomes of *Mnemiopsis* and the sponge *Amphimedon queenslandica* are quite similar, suggesting that sponges have the necessary genetic machinery for a functioning nervous system but may have lost these cell types. We also find that, although *Mnemiopsis* has most of the genes coding for structural components of mesodermal cells, they lack many of the genes involved in bilaterian mesodermal specification and, therefore, may have independently evolved these cell types.

Discussion: These results present a newly supported view of early animal evolution that accounts for major losses and/or gains of sophisticated cell types, including nerve and muscle cells. This evolutionary framework, along with the comprehensive genomic resources made available through this study, will yield myriad discoveries about our most distant animal relatives, many of which will shed light not only on the biology of these extant organisms but also on the evolutionary history of all animal species, including our own.

READ THE FULL ARTICLE ONLINE
http://dx.doi.org/10.1126/science.1242592

Cite this article as J. F. Ryan *et al.*, *Science* **342**, 1242592 (2013).
DOI: 10.1126/science.1242592

FIGURES IN THE FULL ARTICLE

Fig. 1. *M. leidyi* life history and anatomy.

Fig. 2. Previously proposed relationships of the five deep clades of animals.

Fig. 3. Tree produced by maximum-likelihood analysis of the EST set.

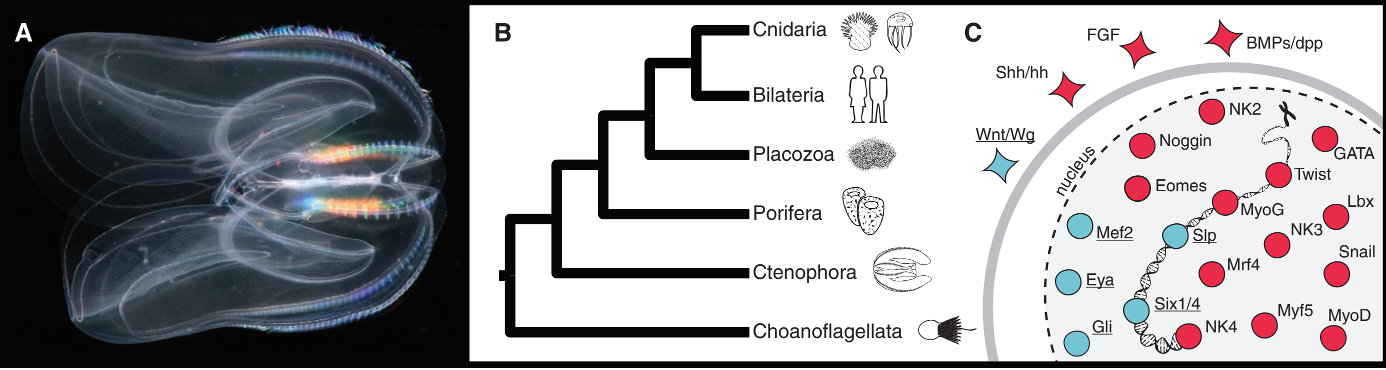
Fig. 4. Tree produced by maximum-likelihood analysis of gene content.

Fig. 5. The origin of postsynaptic genes.

Fig. 6. Inventory of myogenic components in *M. leidyi*.

SUPPLEMENTARY MATERIALS

Materials and Methods
Figs. S1 to S10
Tables S1 to S31
References



The phylogenetic position of the ctenophore *Mnemiopsis leidyi* and its implications regarding the origin of mesodermal cell types. (**A**) Adult *M. leidyi*. (**B**) Summary of the relationships of the five main branches of animals and the outgroup Choanoflagellata. (**C**) Inventory of myogenic specification genes in *Mnemiopsis*. Components present in the *Mnemiopsis* genome are in blue, and names are underlined. Absent components are in red. The lack of many of these factors in *Mnemiopsis* indicates that ctenophore mesodermal cell types are specified differently than in bilaterians, suggesting that they perhaps evolved independently in these two lineages.

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: andy@mail.nih.gov

Published by AAAS

The Genome of the Ctenophore *Mnemiopsis leidyi* and Its Implications for Cell Type Evolution

Joseph F. Ryan,[1,2] Kevin Pang,[2] Christine E. Schnitzler,[1] Anh-Dao Nguyen,[1] R. Travis Moreland,[1] David K. Simmons,[3] Bernard J. Koch,[1] Warren R. Francis,[4] Paul Havlak,[5] NISC Comparative Sequencing Program,[6] Stephen A. Smith,[7,8] Nicholas H. Putnam,[5] Steven H. D. Haddock,[4] Casey W. Dunn,[7] Tyra G. Wolfsberg,[1] James C. Mullikin,[1,6] Mark Q. Martindale,[3] Andreas D. Baxevanis[1]*

An understanding of ctenophore biology is critical for reconstructing events that occurred early in animal evolution. Toward this goal, we have sequenced, assembled, and annotated the genome of the ctenophore *Mnemiopsis leidyi*. Our phylogenomic analyses of both amino acid positions and gene content suggest that ctenophores rather than sponges are the sister lineage to all other animals. *Mnemiopsis* lacks many of the genes found in bilaterian mesodermal cell types, suggesting that these cell types evolved independently. The set of neural genes in *Mnemiopsis* is similar to that of sponges, indicating that sponges may have lost a nervous system. These results present a newly supported view of early animal evolution that accounts for major losses and/or gains of sophisticated cell types, including nerve and muscle cells.

The phylogenetic position of ctenophores presents a challenge to our understanding of early animal evolution, especially as it relates to complex features such as cell types. The stark difference between the body plans of ctenophores and that of all other animals makes comparisons inherently difficult. Genomic sequencing of animals (*1–4*) and their closest relatives (*5*) provides invaluable insight into the molecular innovations contributing to the morphological diversity exhibited among modern-day animals. The vast majority of sequenced animal genomes are from Bilateria, the clade that includes most animal species (including humans and traditional model systems). Three of the four nonbilaterian metazoan lineages—Porifera (sponges), Placozoa, and Cnidaria (for example, sea anemones, corals, hydroids, and jellyfish)—have at least one species with a sequenced genome. The absence of a complete genome sequence from the fourth nonbilaterian metazoan lineage, Ctenophora (or comb jellies), has made it difficult to resolve the earliest evolutionary events in the animal tree of life and reconstruct the likely gene inventory of the most recent common ancestor of animals.

Ctenophores are gelatinous marine animals characterized by eight longitudinal rows of ciliated comb plates that run along their oral-aboral axis (Fig. 1, A to C). Their bodies consist of an inner gastrodermal layer and an outer epidermal layer separated by a mesoglea. The muscular system, deployed in discrete regions of the body (for example, in the body wall, pharynx, and tentacles), is composed almost exclusively of smooth muscle cells; however, sarcomeric muscles have been reported in a single ctenophoran species (*6*). The ctenophore nervous system includes the apical sensory organ, a peripheral subepithelial nerve net, neurons that run through the mesoglea, and nerves associated with the tentacles. Most ctenophores, unlike all other animals, have specialized adhesive cells called colloblasts that are involved in prey capture. Most species are hermaphroditic and capable of self-fertilization. Fertilized eggs undergo a highly stereotyped ctenophore-specific cleavage program (Fig. 1, D to M), with embryogenesis in most species leading to a free-swimming cydippid stage that displays most of the features of the adult body plan (that is, development is direct).

Mnemiopsis leidyi is a lobate ctenophore native to the coastal waters of the western Atlantic Ocean. This species has recently invaded the Black, Caspian, and North Seas, causing major economic and ecological impact to native species in those areas. *M. leidyi* has been used effectively to study regeneration (*7*), axial patterning (*8*, *9*), and bioluminescence (*10–12*). In addition, a cell lineage fate map (*13–15*), as well as resources for collecting and spawning, has been established (*16*), promoting *M. leidyi* as a leading model for evolutionary and developmental studies.

The phylogenetic relationship of ctenophores to other animals has been a source of long-standing debate. The group lacks a reliable fossil record, and, on the basis of morphological features, ctenophores have been assigned various positions in animal phylogeny, including as sister to cnidarians in a clade called Coelenterata (sometimes called Radiata) (Fig. 2A) and as sister to Bilateria (Fig. 2B). Phylogenetic analyses of ribosomal RNA show little or no support uniting ctenophores with cnidarians or bilaterians and have tended to place ctenophores sister to a clade that includes all animals besides Porifera (Fig. 2C). Phylogenomic studies have also produced conflicting results, with a series of multigene analyses placing ctenophores sister to all other metazoans (Fig. 2D) (*17*, *18*), and another, based primarily on ribosomal proteins, supporting the Coelenterata hypothesis (Fig. 2A) (*19*). Yet another study, also based primarily on ribosomal characters but with expanded taxon sampling, upheld the relationship of ctenophores as sister to all metazoans except Porifera (similar to Fig. 2C) (*20*). On the basis of its simple morphology, it has been suggested that Placozoa is the sister group to all animals (Fig. 2E) (*21*). Ctenophores have also been placed in a clade of nonbilaterian animals called "Diploblastica," on the basis of a curated set of nuclear and mitochondrial proteins and a small morphological matrix (Fig. 2F) (*22*). The most recent analyses of the placement of sponges and ctenophores indicated that supermatrix analyses of the publicly available data are sensitive to gene selection, taxon sampling, model selection, and other factors (*23*). The inconsistency of reports about the phylogenetic position of ctenophores (table S1) has made it difficult to evaluate morphological, developmental, and experimental data involving these animals in an evolutionary context, complicating efforts to understand the early evolution of animals.

Genome Sequencing and Assembly

Genomic DNA was isolated from the embryos of two self-fertilized adult *M. leidyi* collected in Woods Hole, Massachusetts. DNA from one embryo pool was used to construct a library for Roche 454 sequencing. We generated 7.3 million raw reads, which yielded 2.5 Gb of sequence. Using the Phusion assembler (*24*), we assembled these data into 24,884 contigs, constituting 150 Mb of sequence and providing roughly 12-fold coverage of the genome. DNA from the other embryo pool was used to create two mate-pair libraries for Illumina GA-II sequencing, one with a 3-kb insert and the other with a 4-kb insert. After duplicate read-pairs were removed, 4.2 million and 2.6 million pairs remained for the 3- and 4-kb insert libraries, respectively. These reads were used to construct scaffolds of the original set of Roche 454 contigs. The final assembly consists of 5100 scaffolds, resulting in 160-fold physical coverage and an N50 of 187 kb (supplementary materials). To test the accuracy and completeness of our assembly, we aligned 99.4% of 15,752 public expressed sequence tags (ESTs) to our assembly. The average coverage of each alignable EST,

[1]Genome Technology Branch, Division of Intramural Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA. [2]Sars International Centre for Marine Molecular Biology, University of Bergen, 5008 Bergen, Norway. [3]Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, FL 32080, USA. [4]Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA. [5]Department of Ecology and Evolutionary Biology, Rice University, Houston, TX 77098, USA. [6]NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Rockville, MD 20852, USA. [7]Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA. [8]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA.

*Corresponding author. E-mail: andy@mail.nih.gov

Fig. 1. *M. leidyi* life history and anatomy. (**A**) Adult *M. leidyi* (about 10 cm long). (**B**) Close-up view of comb rows. (**C**) Aboral view of cydippid stage. (**D**) One-celled fertilized embryo. (**E** to **H**) Early cleavage stages. (**I**) Gastrula stage. (**J** to **M**) Later development of *M. leidyi* embryo shown oral side down. Embryos are about 200 μm. See the supplementary materials for a more detailed description of the ctenophore body plan. [Photo credit for (A): courtesy of Bruno Vellutini]

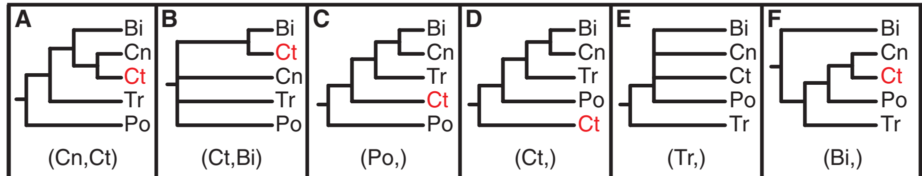


Fig. 2. Previously proposed relationships of the five deep clades of animals. The label at the bottom of each pane corresponds to the header of Table 1. (**A**) Coelenterata hypothesis. (**B**) Ctenophora as sister to Bilateria. (**C**) Porifera as sister group to the rest of Metazoa. (**D**) Ctenophora as sister group to the rest of Metazoa. (**E**) Placozoa as sister group to the rest of Metazoa. (**F**) Diploblastica hypothesis. We see no support in any of our analyses for the hypotheses in (A), (E), and (F) and very little support for (B) (see Table 1). Ct, Ctenophora; Po, Porifera; Tr, Placozoa; Cn, Cnidaria; Bi, Bilateria.

as determined by baa.pl (*25*), was 98.2%. In 94.8% of cases, a single EST mapped completely to a single scaffold. These numbers suggest that the assembly is both complete and accurately assembled.

Characteristics of the *M. leidyi* Genome

The *M. leidyi* genome is among the smallest 7% of genomes when compared with those cataloged in the Animal Genome Size Database (*26*) and is densely packed with gene sequences. It encodes 16,548 predicted protein-coding loci, which make up 58% of the genome, and we conservatively assign 44% of these gene predictions into homology groups with non-ctenophores. The average length of an unspliced *M. leidyi* transcript is 5.8 kb. Eight percent of predicted genes are embedded within other genes. This number of nested intronic genes is high compared to other genomes (table S2), but may be inflated owing to a subset of these being alternatively expressed exons. The level of repetitive sequence in the *M. leidyi* genome is low to moderate, as compared to other metazoans (tables S3 and S4); this has made it possible to produce a high-quality genome assembly based on paired-end and mate-pair sequencing alone. Additional characteristics of this genome are presented in tables S5 to S10.

Phylogenetic Position of *M. leidyi*

The availability of the complete genome of *M. leidyi* has allowed us to improve on the ctenophore sampling used in previous phylogenomic analyses of gene sequence evolution. We assessed two data matrices that differ in breadth of taxon sampling and fraction of missing data: a "Genome Set" that includes only data from complete genomes (13 animals, 19.6% missing data) and an "EST Set" that includes partial genomic data from many taxa (58 animals, 64.9% missing data). We analyzed both matrices by using maximum-likelihood [with the GTR+Γ model as implemented in RAxML (*27*)] and Bayesian [with the CAT model as implemented in PhyloBayes (*28*)] methods. To understand the effect of outgroup selection on our ingroup topology, we included four different sets of nonmetazoan outgroups (table S11) in each combination of method and matrix. This multifactorial strategy yielded a total of 16 analyses (Table 1).

We found no support in any of these analyses for Coelenterata (Cn,Ct), Diploblastica (Bi,), or Placozoa being the sister lineage to the rest of animals (Tr,) (Table 1 and fig. S1). We recovered broad support for a sister relationship between Cnidaria and Bilateria (Cn,Bi) and for a clade of Placozoa, Cnidaria, and Bilateria (Tr,Cn,Bi). Maximum-likelihood analyses support the placement of Ctenophora as sister group to all other

Table 1. Support for various hypotheses across 16 phylogenetic analyses. Two amino acid matrices (Genome Set and EST Set) were analyzed with two different method/model combinations [ML indicates maximum-likelihood with the GTR+Γ model using RAxML (*27*) and Bayes is Bayesian with the CAT model using PhyloBayes (*28*)], using four different sets of nonmetazoan outgroups for each analysis (Opisthokonta are fungi, amoeboids, and choanoflagellates; Holozoa, amoeboids and choanoflagellates; Choanimalia, choanoflagellates; and Animalia, no outgroups). Columns represent support for tested hypotheses, and most hypotheses are represented as trees in Fig. 2. In the absence of nonanimal outgroups, *(Ct,) and (Po,) are concordant with all possible topologies and **(Ct,Po) is the same as (Bi,Cn,Tr). ***Despite an average run time of 205 days per run, none of the Bayesian analyses on the EST data set converged; convergence was monitored by using the maxdiff statistic generated by the bpcomp program within PhyloBayes (>0.3).

			Position of Ctenophora				Robustness of Cnidaria + Bilateria and Parahoxozoa	
			(Ct,)	(Po,)	(Ct,Po)	(Cn,Ct)	(Cn,Bi)	(Bi,Cn,Tr)
GENOME SET 104,840 cols 13 animals 80.4% occup.	ML	Opisthokonta	47	0	53	0	18	90
		Holozoa	31	0	69	0	27	90
		Choanimalia	100	0	0	0	41	100
		Animalia	*	*	**	0	92	100
	Bayes	Opisthokonta	0	0	100	0	100	100
		Holozoa	0	0	100	0	100	100
		Choanimalia	0	7	93	0	100	100
		Animalia	*	*	**	0	100	100
EST SET 88,384 cols 58 animals 35.1% occup.	ML	Opisthokonta	96	0	0	0	100	82
		Holozoa	96	0	0	0	100	83
		Choanimalia	93	0	0	0	100	62
		Animalia	*	*	**	0	100	35
	Bayes***	Opisthokonta	71	29	0	0	73	72
		Holozoa	13	64	0	0	33	30
		Choanimalia	2	98	0	0	100	99
		Animalia	*	*	**	0	100	97

Metazoa (Ct,) regardless of data matrix used (Fig. 3). The Bayesian analysis of the genome data set strongly supports a clade of Ctenophora and Porifera (Ct,Po) as the sister group to all other Metazoa. This relationship also receives some support in our maximum likelihood trees, and we suspect that the result is due to poor taxon sampling in the Genome Set. However, until there are more complete genomes available to test this hypothesis, this relationship cannot be completely dismissed. Despite an average run time of 205 days per run, none of the Bayesian analyses on the EST data set converged (maxdiff > 0.3). The lack of convergence in these analyses suggests that the application of this method to this data set is insufficient to resolve this relationship.

The analyses run without nonmetazoan outgroups show strong support for a monophyletic clade of Cnidaria and Bilateria (Table 1). This evidence contradicts the idea that long-branch attraction between ctenophores and the outgroup is masking a close relationship between ctenophores and cnidarians (*19*). Another common misconception, based on the extremely high evolutionary rates in the mitochondrial genomes of ctenophores (*29*, *30*), is that the phylogenetic placement of these animals is essentially random because of equally extreme rates of evolution in the nuclear genomes of ctenophores. We have found instead that the branch lengths in the phylogenetic analyses of our concatenated protein matrices show *M. leidyi* branches to be of similar length to those of *Drosophila melanogaster*, therefore exhibiting high (but not extreme) amino acid replacement rates (tables S12 and S13).

The conflict between the maximum-likelihood and Bayesian analyses of the amino acid matrix makes it difficult to determine from these analyses whether Ctenophora or Porifera is the sister group to the rest of the Metazoa, but there is substantial support for ctenophore as the sister group to the rest of animals (Table 1). Furthermore, our results strongly show that Placozoa, Cnidaria, and Bilateria (that is, ParaHoxozoa) are monophyletic. Given the sensitivity of the molecular sequence evolution analyses to taxon sampling and inference method, consistent with other recent analyses (*23*), we also examined the evolution of gene content.

We clustered genes by using default parameters in OrthoMCL (*31*) and used these clusters to construct a gene presence/absence matrix. By using RAxML with a GTR+Γ model, we conducted a weighted likelihood-based analysis on this matrix. We then calibrated sites on the basis of the congruence of columns to known bilaterian relationships with the "-f u" parameter in RAxML. The result of this analysis was a tree supporting Ctenophora as the sister group to all other animals (Ct,) (Fig. 4) and the rejection of all other alternative topologies (in Fig. 2) at the 5% confidence level by likelihood-based statistical hypothesis testing (table S14). The pattern of presence and absence of gene families and signaling pathway components seen in previous studies is consistent with these results (*32*–*36*). Our reanalysis of an expanded set of near intron pairs (*37*) was also consistent with these results (fig. S2).

Cell Signaling Components in *M. leidyi*

Across Bilateria, there are seven major cell signaling pathways that play important roles during embryological development: Wnt, transforming growth factor–β (TGF-β), receptor tyrosine kinase (RTK), Notch, nuclear receptor, Hedgehog, and Janus kinase (JAK)/signal transducers and activators of transcription (STAT) (*38*). Comparisons of nonbilaterian (*2*–*4*) and nonmetazoan genomes (*5*, *39*) show that some of these signaling pathways evolved before the evolution of animal multicellularity, others are specific to metazoan evolution, and some were lineage-specific innovations. The cell signaling components present in the *M. leidyi* genome include the RTK family, which predates the origin of Metazoa (*40*); the TGF-β signaling pathway (*33*), thought to have evolved in the metazoan common ancestor (*39*); and the canonical Wnt signaling pathway (*34*). Notably absent from both the TGF-β and Wnt pathways are the major bilaterian antagonists; members of the Wnt/PCP (planar cell polarity) pathway, such as Flamingo and Strabismus, are not present. Relatively few components of the Notch pathway (tables S15 and S16) are present, and many of those lack key diagnostic domains. *M. leidyi* also lacks most of the major genes necessary for Hedgehog signaling [for example, the Hedgehog ligand, the smoothened receptor, and SUFU (suppressor of fused)]. Last, the JAK/STAT pathway is most likely a bilaterian innovation because there are no true JAK orthologs in *M. leidyi* or any other nonbilaterians reported to date.

Neural Components in *M. leidyi*

Ctenophores have a nervous system consisting of a nerve net, mesogleal fibers, and tentacular nerves (*41*). In contrast to the cnidarian nervous system, which contains an ectodermal and endodermal nerve net, the nerve nets of ctenophores consist of polygonal nerve cords spread under the ectodermal epithelium; these nerve nets show high levels of regional specialization and concentrations associated with the apical sensory organ/polar fields and tentacle bulbs, structures without clear homologs in any other animal groups (*42*). Unlike in cnidarians and bilaterians, immunological investigations have failed to detect the presence of serotonin in ctenophores (*43*). Ctenophore nervous systems are also unique in their abundance

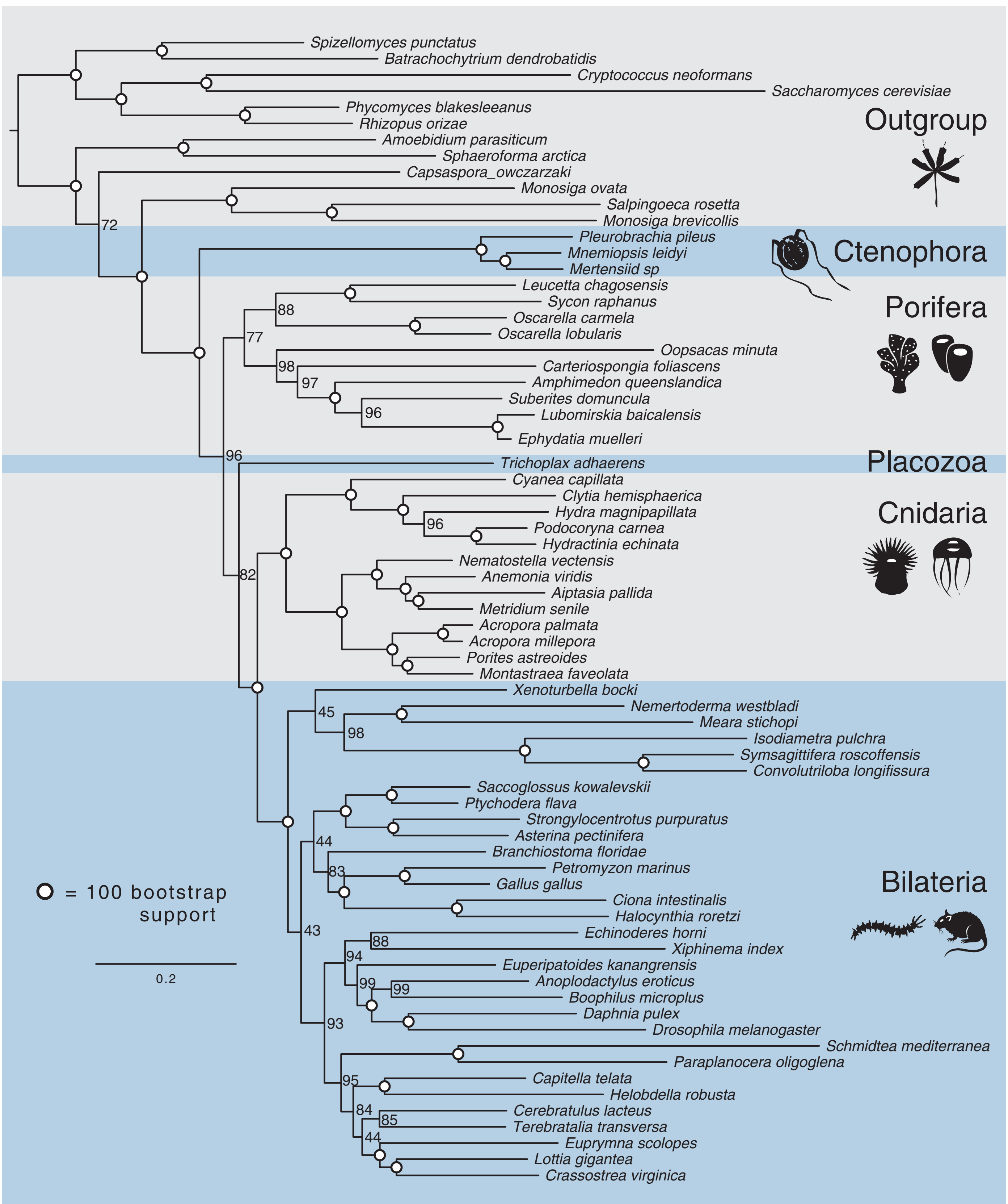


Fig. 3. Tree produced by maximum-likelihood analysis of the EST Set. The tree was produced from a matrix consisting of 242 genes and 104,840 amino acid characters. Circles on nodes indicate 100% bootstrap support. Support placing ctenophores as sister to the rest of Metazoa is 96% of 100 bootstrap replicates.

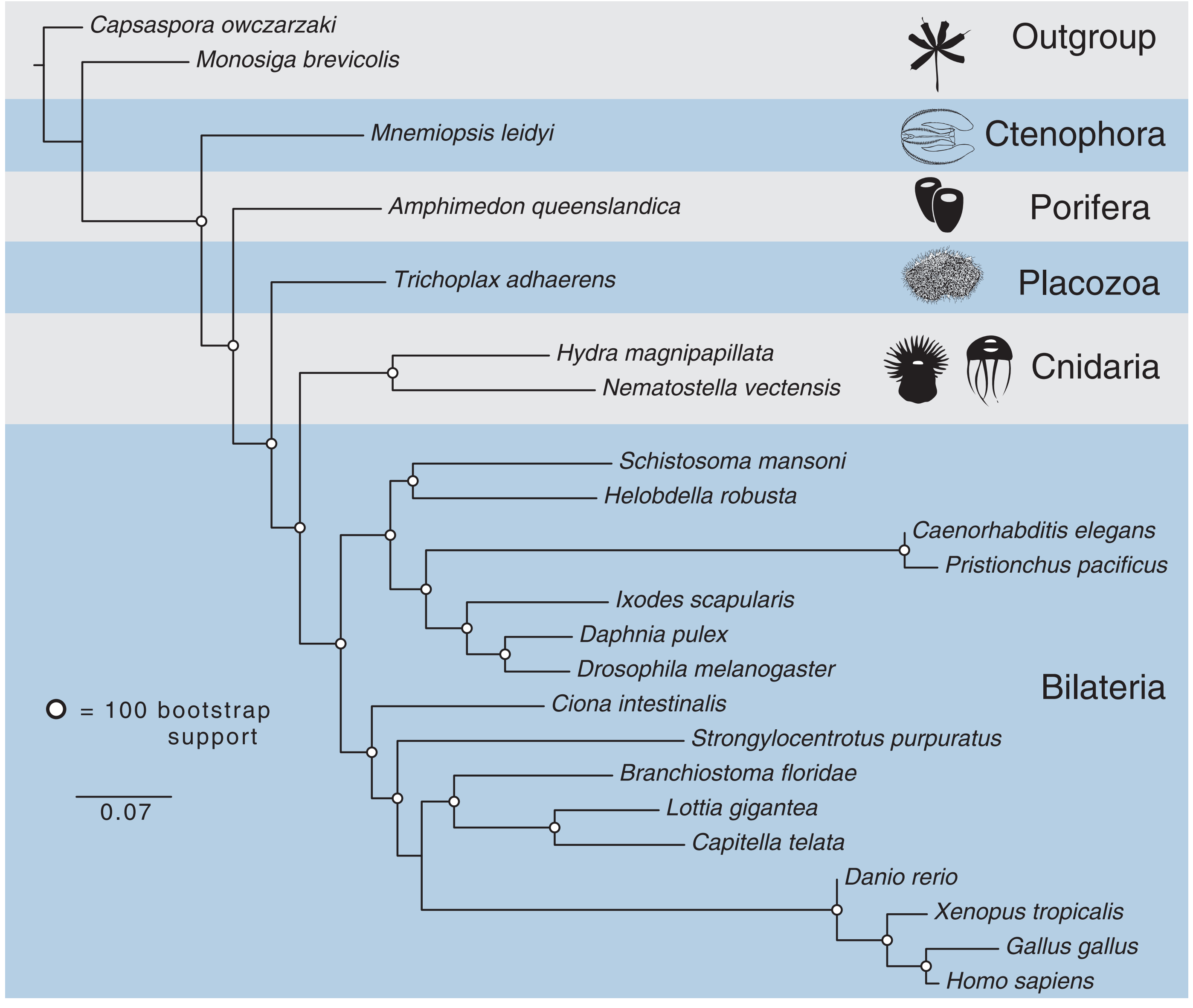


Fig. 4. Tree produced by maximum-likelihood analysis of gene content. The tree was produced from a matrix consisting of 23,910 binary characters indicating the presence or absence of a particular species within a cluster of genes. Clusters were produced with default settings of OrthoMCL. Columns consistent with known relationships within Bilateria were up-weighted, whereas conflicting characters were down-weighted. The matrix was analyzed with RAxML under the GTR-Γ model of rate heterogeneity. All nodes received 100% bootstrap support. Constraining known relationships did not affect the position of Ctenophora (fig. S4).

of synaptic connections and their presynaptic morphology (*44*).

Many of the genes known to be critical to the nervous system of bilaterians and cnidarians are present in the sponge *Amphimedon queenslandica*, an animal without a nervous system. It has been hypothesized that the origin of the nervous system in nonsponges coincided with the origin of a few neural components that are absent from *A. queenslandica* (*4*, *45*), but our phylogenetic results and the absence of these same components in *M. leidyi* challenge this hypothesis. Both *A. queenslandica* and *M. leidyi* contain orthologs of transcription factors involved in bilaterian and cnidarian neural development, including lhx (*46*), bHLH (basic helix-loop-helix), six, gli, and sox (classes B, C, E, and F) genes. The neural differentiation RNA binding genes ELAV and Musashi, as well as the axon guidance genes neurexin, semaphorin, plexin, and an ephrin receptor, are all present in both *A. queenslandica* and *M. leidyi*. However, netrin, slit, and unc-5, involved in axon guidance, are absent from both genomes.

Many of the genes involved in the formation of bilaterian synapses and neural differentiation are present in both *A. queenslandica* and *M. leidyi*—but again, sponges and ctenophores lack a similar set of synaptic scaffolding genes (tables S17 and S18), all of which are present in cnidarians and bilaterians (Fig. 5). The pattern of presence and absence of these scaffolding genes is consistent with these genes being primitively absent in sponges and ctenophores. Almost all of the enzymes involved in the biosynthesis of dopamine and other catecholamine neurotransmitters are also absent in both *A. queenslandica* and *M. leidyi* (table S19). An exception to this shared pattern with sponges is the presence of two definitive opsin genes in *M. leidyi*, but not *A. queenslandica*, that are expressed in photocytes (light-producing cells), as well as in putative photosensory cells in the apical sense organ (*12*).

Mesoderm Components in *M. leidyi*

Ctenophores have several cell types (such as distinct muscle cells and mesenchymal cells) that, in bilaterians, are characteristically derived from mesodermal tissues. Cell lineage studies (*14*) have indicated that these cells are derived from a true endomesoderm because mesodermal cells are generated from precursors that also give rise to the endodermal portions of the gut; this is similar to the

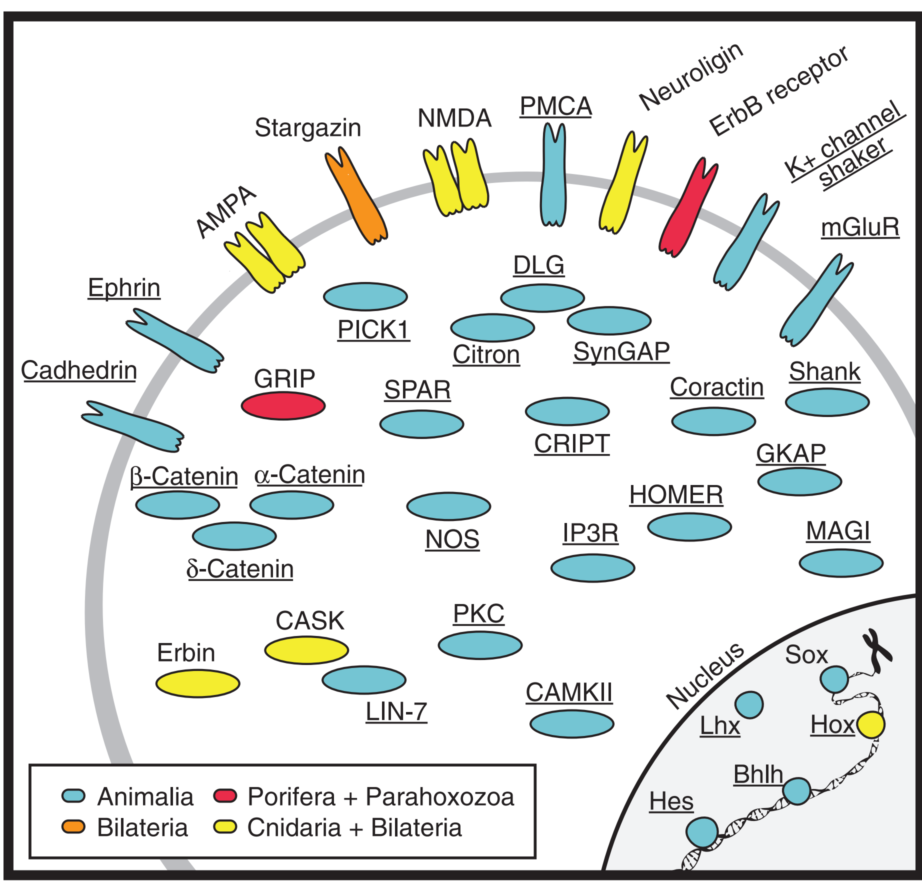


Fig. 5. Origin of postsynaptic genes. A possible configuration for postsynaptic genes. Genes are colored by their node of origin (**inset**). Names of genes present in the *M. leidyi* genome are underlined. Accession numbers of *M. leidyi* genes are given in table S16.

endomesodermal origins of mesoderm in virtually all bilaterians. However, screening the *M. leidyi* genome reveals a surprising result in that almost none of the genes involved in bilaterian mesoderm development can be found (Fig. 6 and tables S20 and S21). Functional components of the fibroblast growth factor, Notch, Hedgehog, and the nodal (TGF-β superfamily) pathways, all of which are important in the segregation of mesoderm in different bilaterian forms, are also not observed. Other genes known to be involved in bilaterian mesoderm development, such as gli/glis genes, are expressed in neural (but not mesodermal) cells in *M. leidyi* (47).

Mesoderm and Neural Components Also Absent from Other Ctenophores

To test whether these absences from the *M. leidyi* genome were true for other ctenophores, we searched the deeply sequenced transcriptomes of seven other ctenophore species (*Bathyctena chuni*, *Beroe forskalii, Charistephane fugiens*, *Euplokamis dunlapae, Hormiphora californensis*, *Lampea lactea*, and *Thalassocalyce inconstans*) for FGF (fibroblast growth factor), Hedgehog, nodal, twist, snail, Lbx, NK4, NK3, NK2, Myf5, Noggin, Mrf4, Myogenin, Eomesoderm, GATA, MyoD, and troponin. We were able to identify putative snail genes in *T. inconstans* and *E. dunlapae* and putative GATA genes in five of the seven species. We were unable to identify the other 15 missing genes in any of these ctenophore transcriptomes (tables S22 and S23). A phylogenetic analysis of ionotropic glutamate receptor sequences from *M. leidyi* and these ctenophore transcriptomes suggests that the ctenophore receptors form a sister clade to the bilaterian glutamate receptors

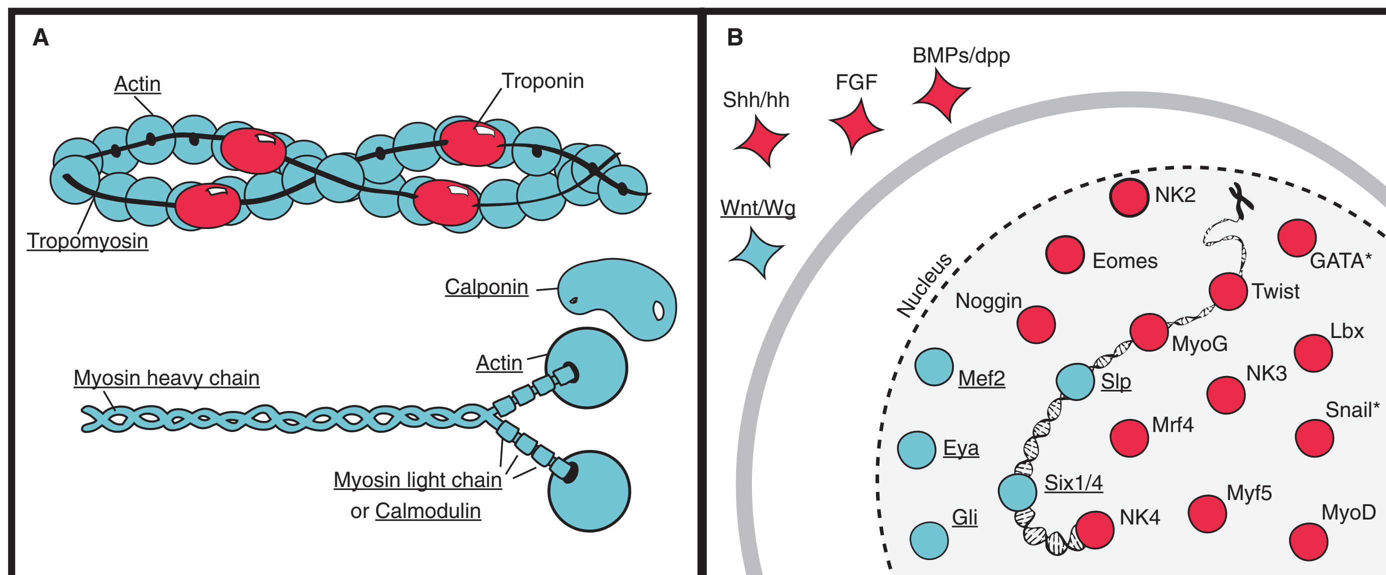


Fig. 6. Inventory of myogenic components in *M. leidyi*. Components present in the *M. leidyi* genome are in blue, and names are underlined. Absent components are in red. (**A**) The main structural components of smooth muscle are present in the *M. leidyi* genome. All structural components are present except for troponin (in red). (**B**) The majority of signaling molecules and transcription factors involved in specifying and differentiating the mesoderm of bilaterian animals are absent from the genome of *M. leidyi*. The asterisks next to Snail and GATA indicate that these components have been identified in the transcriptomes of other ctenophores.

(fig. S3). Ionotropic glutamate receptors are absent from *A. queenslandica* but are present in the transcriptomes of eight other sponges (*48*). The tree topology suggests that the ctenophore sequences descended from an ancestral glutamate receptor that differentiated into AMPA, NMDA (*N*-methyl-D-aspartate), kainate-type, and delta2-like glutamate receptors after ctenophores diverged from the rest of animals. These results indicate that, within ctenophores, the majority of absences are not specific to the *M. leidyi* lineage, but that there are some intriguing differences in gene content between ctenophores themselves.

Discussion and Conclusion

The sequence of the *M. leidyi* genome has given rise to multiple categories of evidence that support the placement of ctenophores as the sister group to all other animals, a conclusion supported by phylogenetic analysis of amino acid matrices from concatenated protein sequences. However, these analyses are sensitive to taxon sampling and phylogenetic methods and, therefore, provide some support for alternative hypotheses. With a ctenophore genome in hand, we show that gene content data support Ctenophora as the sister group to all other animals and statistically reject competing hypotheses. It will be important to test this result once more genomic data are available from other ctenophores, sponges, and other relevant groups. Nevertheless, this result is congruent with the structure and inventory of a variety of gene families and signaling pathways, as well as genes essential to neural and mesodermal cell types.

It appears that much of the genetic machinery necessary for a nervous system was present in the ancestor of all extant animals. This pattern suggests that a less elaborate nervous system was present in the metazoan ancestor and was secondarily reduced in placozoans and sponges. The alternative is that neural cell types arose independently in both the ctenophore lineage and the lineage that led to cnidarians and bilaterians, which might explain some of the unique aspects of the ctenophore nervous system. Resolving these alternative hypotheses will require functionally characterizing the nervous system–related genes in ctenophores and sponges.

Like the nervous system, the mesoderm appears to have had a complex evolutionary history. Our results are consistent with several alternative hypotheses. One possibility is that the mesoderm was present in the most recent common ancestor of ctenophores and bilaterians but was lost in sponges, placozoans, and cnidarians. However, given the absence of the majority of genes involved in the specification and differentiation of the bilaterian mesoderm from the *M. leidyi* genome, it appears more likely that ctenophores independently evolved mesodermal cell types after they diverged from the rest of animals. This interpretation is compatible with a recent report that striated musculature evolved independently in bilaterians, cnidarians, and in the ctenophore *E. dunlapae* (*49*).

The implications of these findings go well beyond the rearrangement of the branches of the metazoan tree of life, arguing for a new way of thinking regarding the emergence and/or conservation of what heretofore were considered to be unique and indispensible biological features. Likewise, theories on the evolution of animal multicellularity have to be reevaluated. This evolutionary framework, along with the comprehensive genomic resources made available through this study, will undoubtedly yield myriad new discoveries about our most distant animal relatives, many of which will shed new light not only on the biology of these extant organisms but also on the evolutionary history of all animal species, including our own.

Methods

Genome Sequencing and Assembly

We isolated genomic DNA from the embryos of a self-fertilized adult and sequenced this DNA with Roche 454 sequencing. We generated another pool of genomic DNA from the embryos of a second self-fertilized adult and sequenced this DNA using Illumina GA-II mate-pair sequencing. These data were assembled using the Phusion assembler (*24*). We have deposited the assembly at GenBank under the project accession AGCP00000000.

Transcript Sequencing and Assembly

We isolated RNA from mixed-stage *M. leidyi* embryos and sequenced this material using Illumina GA-II sequencing. We assembled these data into transcripts using Cufflinks (*50*) and Trinity (*51*). Assembled transcripts are available through the *Mnemiopsis* Genome Project Portal (http://research.nhgri.nih.gov/mnemiopsis/).

Gene Prediction

We generated gene model predictions using a range of gene prediction programs and then used EVidenceModeler (EVM) (*52*) to combine models, transcripts, and sequence similarity to other protein data sets into a final set of protein-coding gene predictions. These are available through the *Mnemiopsis* Genome Project Portal (http://research.nhgri.nih.gov/mnemiopsis/).

Phylogenetic Analysis of Concatenated Gene Matrices

We analyzed two matrices constructed from concatenated protein sequences. One consisted of *M. leidyi* amino acids added to a genome-based data matrix that was reported in the *A. queenslandica* genome paper (*4*). The second used a phenetic sequence clustering method as described previously (*18*). We generated maximum-likelihood trees with the GTR+Γ model using RAxML (*27*) and Bayesian trees with the CAT model using PhyloBayes (*28*). All alignments and trees are available at http://research.nhgri.nih.gov/manuscripts/Baxevanis/science2013_supplement/.

Phylogenetic Analysis of Gene Content

We assembled a presence/absence matrix of gene clusters and analyzed these data with RAxML under the GTR-gamma model of rate heterogeneity. We used known bilaterian relationships to generate a weight matrix in RAxML. We used per-site log likelihoods generated in RAxML as input to CONSEL (*53*) to generate *P* values for alternative hypotheses.

References and Notes

1. J. A. Chapman *et al*., The dynamic genome of *Hydra*. *Nature* **464**, 592–596 (2010). doi: 10.1038/nature08830; pmid: 20228792
2. N. H. Putnam *et al*., Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* **317**, 86–94 (2007). doi: 10.1126/science.1139158; pmid: 17615350
3. M. Srivastava *et al*., The *Trichoplax* genome and the nature of placozoans. *Nature* **454**, 955–960 (2008). doi: 10.1038/nature07191; pmid: 18719581
4. M. Srivastava *et al*., The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**, 720–726 (2010). doi: 10.1038/nature09201; pmid: 20686567
5. N. King *et al*., The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783–788 (2008). doi: 10.1038/nature06617; pmid: 18273011
6. G. O. Mackie, C. E. Mills, C. L. Singla, Structure and function of the prehensile tentilla of *Euplokamis* (Ctenophora, Cydippida). *Zoomorphology* **107**, 319–337 (1988). doi: 10.1007/BF00312216
7. J. Q. Henry, M. Q. Martindale, Regulation and regeneration in the ctenophore *Mnemiopsis leidyi*. *Dev. Biol.* **227**, 720–733 (2000). doi: 10.1006/dbio.2000.9903; pmid: 11071786
8. M. Q. Martindale, J. R. Finnerty, J. Q. Henry, The Radiata and the evolutionary origins of the bilaterian body plan. *Mol. Phylogenet. Evol.* **24**, 358–365 (2002). doi: 10.1016/S1055-7903(02)00208-7; pmid: 12220977
9. K. Pang, M. Q. Martindale, Developmental expression of homeobox genes in the ctenophore *Mnemiopsis leidyi*. *Dev. Genes Evol.* **218**, 307–319 (2008). doi: 10.1007/s00427-008-0222-3; pmid: 18504608
10. M. Anctil, Ultrastructure of the luminescent system of the ctenophore *Mnemiopsis leidyi*. *Cell Tissue Res.* **242**, 333–340 (1985). doi: 10.1007/BF00214545
11. G. Freeman, G. T. Reynolds, The development of bioluminescence in the ctenophore *Mnemiopsis leidyi*. *Dev. Biol.* **31**, 61–100 (1973). doi: 10.1016/0012-1606(73)90321-7; pmid: 4150750
12. C. E. Schnitzler *et al*., Genomic organization, evolution, and expression of photoprotein and opsin genes in *Mnemiopsis leidyi*: A new view of ctenophore photocytes. *BMC Biol.* **10**, 107 (2012). doi: 10.1186/1741-7007-10-107; pmid: 23259493
13. M. Q. Martindale, J. Q. Henry, Reassessing embryogenesis in the Ctenophora: The inductive role of e_1 micromeres in organizing ctene row formation in the "mosaic" embryo, *Mnemiopsis leidyi*. *Development* **124**, 1999–2006 (1997). pmid: 9169846
14. M. Q. Martindale, J. Q. Henry, Intracellular fate mapping in a basal metazoan, the ctenophore *Mnemiopsis leidyi*, reveals the origins of mesoderm and the existence of indeterminate cell lineages. *Dev. Biol.* **214**, 243–257 (1999). doi: 10.1006/dbio.1999.9427; pmid: 10525332
15. G. Reverberi, G. Ortolani, On the origin of the ciliated plates and mesoderm in the Ctenophore. *Acta Embryol. Morphol. Exp.* **6**, 175–199 (1963).
16. K. Pang, M. Q. Martindale, *Mnemiopsis leidyi* spawning and embryo collection. *CSH Protoc.* **2008**, pdb.prot5085 (2008). pmid: 21356725
17. C. W. Dunn *et al*., Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745–749 (2008). doi: 10.1038/nature06614; pmid: 18322464
18. A. Hejnol *et al*., Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc. Biol. Sci.* **276**, 4261–4270 (2009). doi: 10.1098/rspb.2009.0896; pmid: 19759036
19. H. Philippe *et al*., Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19**, 706–712 (2009). doi: 10.1016/j.cub.2009.02.052; pmid: 19345102

20. K. S. Pick *et al.*, Improved phylogenomic taxon sampling noticeably affects non-bilaterian relationships. *Mol. Biol. Evol.* **27**, 1983–1987 (2010). doi: 10.1093/molbev/msq089; pmid: 20378579
21. B. Schierwater, My favorite animal, *Trichoplax adhaerens*. *Bioessays* **27**, 1294–1302 (2005). doi: 10.1002/bies.20320; pmid: 16299758
22. B. Schierwater *et al.*, Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. *PLOS Biol.* **7**, e20 (2009). doi: 10.1371/journal.pbio.1000020; pmid: 19175291
23. T. Nosenko *et al.*, Deep metazoan phylogeny: When different genes tell different stories. *Mol. Phylogenet. Evol.* **67**, 223–233 (2013). doi: 10.1016/j.ympev.2013.01.010; pmid: 23353073
24. J. C. Mullikin, Z. Ning, The phusion assembler. *Genome Res.* **13**, 81–90 (2003). doi: 10.1101/gr.731003; pmid: 12529309
25. J. F. Ryan, Baa.pl: A tool to evaluate de novo genome assemblies with RNA transcripts (2013), http://arxiv.org/abs/1309.2087.
26. T. R. Gregory, Animal Genome Size Database, www.genomesize.com.
27. A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006). doi: 10.1093/bioinformatics/btl446; pmid: 16928733
28. N. Lartillot, T. Lepage, S. Blanquart, PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288 (2009). doi: 10.1093/bioinformatics/btp368; pmid: 19535536
29. A. B. Kohn *et al.*, Rapid evolution of the compact and unusual mitochondrial genome in the ctenophore, *Pleurobrachia bachei*. *Mol. Phylogenet. Evol.* **63**, 203–207 (2012). doi: 10.1016/j.ympev.2011.12.009; pmid: 22201557
30. W. Pett *et al.*, Extreme mitochondrial evolution in the ctenophore *Mnemiopsis leidyi*: Insight from mtDNA and the nuclear genome. *Mitochondrial DNA* **22**, 130–142 (2011). doi: 10.3109/19401736.2011.624611; pmid: 21985407
31. L. Li, C. J. Stoeckert Jr., D. S. Roos, OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003). doi: 10.1101/gr.1224503; pmid: 12952885
32. B. J. Liebeskind, D. M. Hillis, H. H. Zakon, Evolution of sodium channels predates the origin of nervous systems in animals. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9154–9159 (2011). doi: 10.1073/pnas.1106363108; pmid: 21576472
33. K. Pang, J. F. Ryan, A. D. Baxevanis, M. Q. Martindale, Evolution of the TGF-β signaling pathway and its potential role in the ctenophore, *Mnemiopsis leidyi*. *PLOS One* **6**, e24152 (2011). doi: 10.1371/journal.pone.0024152; pmid: 21931657
34. K. Pang, J. F. Ryan; NISC Comparative Sequencing Program, J. C. Mullikin, A. D. Baxevanis, M. Q. Martindale, Genomic insights into Wnt signaling in an early diverging metazoan, the ctenophore *Mnemiopsis leidyi*. *Evodevo* **1**, 10 (2010). doi: 10.1186/2041-9139-1-10; pmid: 20920349
35. A. M. Reitzel *et al.*, Nuclear receptors from the ctenophore *Mnemiopsis leidyi* lack a zinc-finger DNA-binding domain: Lineage-specific loss or ancestral condition in the emergence of the nuclear receptor superfamily? *Evodevo* **2**, 3 (2011). doi: 10.1186/2041-9139-2-3; pmid: 21291545
36. J. F. Ryan, K. Pang, NISC Comparative Sequencing Program, J. C. Mullikin, M. Q. Martindale, A. D. Baxevanis, The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. *Evodevo* **1**, 9 (2010). doi: 10.1186/2041-9139-1-9; pmid: 20920347
37. J. Lehmann, P. F. Stadler, V. Krauss, Near intron pairs and the metazoan tree. *Mol. Phylogenet. Evol.* **66**, 811–823 (2013). doi: 10.1016/j.ympev.2012.11.012; pmid: 23201572
38. A. Pires-daSilva, R. J. Sommer, The evolution of signalling pathways in animal development. *Nat. Rev. Genet.* **4**, 39–49 (2003). doi: 10.1038/nrg977; pmid: 12509752
39. A. Sebe-Pedros, A. J. Roger, F. B. Lang, N. King, I. Ruiz-Trillo, Ancient origin of the integrin-mediated adhesion and signaling machinery. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10142–10147 (2010). doi: 10.1073/pnas.1002257107; pmid: 20479219
40. N. King, S. B. Carroll, A receptor tyrosine kinase from choanoflagellates: Molecular insights into early animal evolution. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 15032–15037 (2001). doi: 10.1073/pnas.261477698; pmid: 11752452
41. M. Jager *et al.*, New insights on ctenophore neural anatomy: Immunofluorescence study in *Pleurobrachia pileus* (Müller, 1776). *J. Exp. Zool. B Mol. Dev. Evol.* **316B**, 171–187 (2011). pmid: 21462312
42. G. R. Harbison, On the classification and evolution of the Ctenophora, in *The Origins and Relationships of Lower Invertebrates*, S. Conway Morris, J. D. George, R. Gibson, H. M. Platt, Eds. (Oxford Univ. Press, Oxford, 1985), pp. 78–100.
43. A. Hay-Schmidt, The evolution of the serotonergic nervous system. *Proc. Biol. Sci.* **267**, 1071–1079 (2000). doi: 10.1098/rspb.2000.1111; pmid: 10885511
44. M. L. Hernandez-Nicaise, The nervous system of ctenophores. III. Ultrastructure of synapses. *J. Neurocytol.* **2**, 249–263 (1973). doi: 10.1007/BF01104029; pmid: 9224490
45. O. Sakarya *et al.*, A post-synaptic scaffold at the origin of the animal kingdom. *PLOS One* **2**, e506 (2007). doi: 10.1371/journal.pone.0000506; pmid: 17551586
46. D. K. Simmons, K. Pang, M. Q. Martindale, Lim homeobox genes in the Ctenophore *Mnemiopsis leidyi*: The evolution of neural cell type specification. *Evodevo* **3**, 2 (2012). doi: 10.1186/2041-9139-3-2; pmid: 22239757
47. M. J. Layden, N. P. Meyer, K. Pang, E. C. Seaver, M. Q. Martindale, Expression and phylogenetic analysis of the *zic* gene family in the evolution and development of metazoans. *Evodevo* **1**, 12 (2010). doi: 10.1186/2041-9139-1-12; pmid: 21054859
48. N. Farrar, A. Riesgo, S. Leys, paper presented at the 2013 Society for Integrative and Comparative Biology Annual Meeting, San Francisco, CA, 3 to 7 January 2013.
49. P. R. H. Steinmetz *et al.*, Independent evolution of striated muscles in cnidarians and bilaterians. *Nature* **487**, 231–234 (2012). doi: 10.1038/nature11180; pmid: 22763458
50. C. Trapnell *et al.*, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010). doi: 10.1038/nbt.1621; pmid: 20436464
51. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011). doi: 10.1038/nbt.1883; pmid: 21572440
52. B. J. Haas *et al.*, Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008). doi: 10.1186/gb-2008-9-1-r7; pmid: 18190707
53. H. Shimodaira, M. Hasegawa, CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246–1247 (2001). doi: 10.1093/bioinformatics/17.12.1246; pmid: 11751242

Acknowledgments: This research was supported by the Intramural Research Program of the National Human Genome Research Institute (NHGRI), NIH. This work was also supported in part by NASA and NSF grants to M.Q.M. J.F.R. received additional support from the Sars International Centre for Marine Molecular Biology and the University of Bergen. We thank A. Young, B. Schmidt, N. Gurson, R. Legaspi, B. Novotny, and R. Blakesley, who were responsible for the sequencing performed at the NIH Intramural Sequencing Center (NISC); A. Prasad, D. Gildea, N. Trivedi, A. Yun, K. Siewert, D. Leja, S. Bond, and G. Bouffard at NHGRI; A. Hejnol, D. Chourrout, L. Leclère, G. Richards, F. Rentzsch, C. Martin, H. Hausen, S. Henriet, S. Church, and S9 at Sars; B. Vellutini for the photo of *M. leidyi* used in Fig. 1A; W. Browne for access to additional RNA-seq data from early developmental stages of *M. leidyi*; J. Lehmann for supplying an updated near intron pair matrix; C. Trapnell for help with the Bowtie short-read aligner and Cufflinks transcript assembly program; M. Srivastava for phylogenetic data sets and advice; I. Ruiz-Trillo and the Origins of Multicellularity Sequencing Project, Broad Institute of Harvard and MIT for use of their genomic data; D. Rokhsar and the Joint Genome Institute for use of their genomic data; and other researchers whose publicly available sequence data were used in this study. The views expressed in this paper do not necessarily reflect the views of those acknowledged. We dedicate this manuscript to the pioneering work of the late Sebastian Beroe of the Stazione Zoologica in Naples. The authors declare no competing financial interests. The genome sequence data can be accessed from GenBank as project accession AGCP00000000 and from http://research.nhgri.nih.gov/mnemiopsis/. Contributions are as follows: genome and RNA-seq sequencing and assembly: NISC and J.C.M.; annotation: J.F.R., K.P., C.E.S., P.H., N.H.P., A.-D.N., R.T.M., B.J.K., and T.G.W.; analysis: J.F.R., K.P., C.E.S., D.K.S., B.J.K., P.H., N.H.P., M.Q.M., and A.D.B.; additional ctenophore data: S.H.D.H. and W.R.F.; phylogenetics: J.F.R., S.A.S., and C.W.D.; writing: J.F.R., K.P., C.E.S., D.K.S., R.T.M., C.W.D., M.Q.M., and A.D.B.; project design and coordination: J.F.R., M.Q.M., and A.D.B.

Supplementary Materials

www.sciencemag.org/content/342/6164/1242592/suppl/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S31
References

1 July 2013; accepted 28 October 2013
10.1126/science.1242592