Genomic insights into the evolutionary origin of Myxozoa within Cnidaria

E. Sally Chang[a], Moran Neuhof[b,c], Nimrod D. Rubinstein[d], Arik Diamant[e], Hervé Philippe[f,g], Dorothée Huchon[b,1], and Paulyn Cartwright[a,1]

[a]Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS 66045; [b]Department of Zoology, George S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 6997801, Israel; [c]Department of Neurobiology, George S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 6997801, Israel; [d]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138; [e]National Center for Mariculture, Israel Oceanographic and Limnological Research, Eilat 88112, Israel; [f]CNRS, Station d'Ecologie Expérimentale du CNRS, Moulis 09200, France; and [g]Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Montréal, QC, Canada H3C 3J7

Edited by David M. Hillis, The University of Texas at Austin, Austin, TX, and approved October 16, 2015 (received for review June 12, 2015)

The Myxozoa comprise over 2,000 species of microscopic obligate parasites that use both invertebrate and vertebrate hosts as part of their life cycle. Although the evolutionary origin of myxozoans has been elusive, a close relationship with cnidarians, a group that includes corals, sea anemones, jellyfish, and hydroids, is supported by some phylogenetic studies and the observation that the distinctive myxozoan structure, the polar capsule, is remarkably similar to the stinging structures (nematocysts) in cnidarians. To gain insight into the extreme evolutionary transition from a free-living cnidarian to a microscopic endoparasite, we analyzed genomic and transcriptomic assemblies from two distantly related myxozoan species, *Kudoa iwatai* and *Myxobolus cerebralis*, and compared these to the transcriptome and genome of the less reduced cnidarian parasite, *Polypodium hydriforme*. A phylogenomic analysis, using for the first time to our knowledge, a taxonomic sampling that represents the breadth of myxozoan diversity, including four newly generated myxozoan assemblies, confirms that myxozoans are cnidarians and are a sister taxon to *P. hydriforme*. Estimations of genome size reveal that myxozoans have one of the smallest reported animal genomes. Gene enrichment analyses show depletion of expressed genes in categories related to development, cell differentiation, and cell–cell communication. In addition, a search for candidate genes indicates that myxozoans lack key elements of signaling pathways and transcriptional factors important for multicellular development. Our results suggest that the degeneration of the myxozoan body plan from a free-living cnidarian to a microscopic parasitic cnidarian was accompanied by extreme reduction in genome size and gene content.

Myxozoa | Cnidaria | *Polypodium* | parasite | genome evolution

Obligate parasitism can lead to dramatic reduction of body plans and associated morphological structures (1, 2). One of the most spectacular examples is the microscopic Myxozoa, which spend most of their parasitic life cycle as just a few cells in size (3). The most conspicuous myxozoan cell type houses a polar capsule, which is a complex structure with an eversible tube (or filament) that is thought to facilitate attachment to the host. The polar capsule bears remarkable similarity to the stinging structures (nematocysts) of cnidarians (corals, sea anemones, jellyfish, and hydroids), suggesting that nematocysts and polar capsules are homologous and that myxozoans are related to cnidarians (4, 5).

Myxozoa are a diverse group of obligate endoparasites that comprise over 2,180 species (6). The vast majority of myxozoan species alternate between a fish and annelid host. In *Myxobolus cerebralis* (7), the causative agent for whirling disease in rainbow trout (8), the annelid host *Tubifex tubifex* releases infective actinospores (Fig. 1*A*), which subsequently anchor onto the fish, injecting the sporoplasm into the host tissue (9). Infective myxospores develop within the fish and are eventually ingested by the annelid where they develop into an actinospore, thereby completing the parasitic life cycle (Fig. 1*A*). Although the vast majority of myxozoan species are composed of just a few cells, some malacosporean myxozoans, such as *Buddenbrockia plumatellae*, have a complex vermiform life cycle stage (10), which is a derived trait that has been lost and regained several times within this particular lineage (11).

Another obligate cnidarian endoparasite, *Polypodium hydriforme*, in contrast to myxozoans, does not have a degenerate body form and instead displays conventional cnidarian-like features, including tentacles, a gut, and a mouth (Fig. 1*B*). *P. hydriforme* lies dormant as a binucleate cell within the oocytes of female acipenseriform fishes (paddlefish and sturgeon) (Fig. 1*B*), eventually developing into an elongated stolon, which emerges from the host's oocyte upon spawning. Once freely living, the stolon begins to fragment into multiple individuals, each developing a mouth with which to feed (Fig. 1*B*). Adult *P. hydriforme* infect juvenile female fish to repeat the life cycle (12).

Classification of Myxozoa and *P. hydriforme* has been controversial (13–15). From the first descriptions in the 1880s until relatively recently myxozoans were considered to be protists, largely due to their highly reduced, microscopic construction (16). Unlike Myxozoa, the placement of the monotypic species *P. hydriforme* as a cnidarian has long been proposed based on morphology (17, 18). With the advent of molecular phylogenetics, it was discovered that myxozoans are not protists, but instead are metazoans (19, 20). The first studies, based mainly on analyses of 18S rDNA, usually recovered Myxozoa as the sister taxon to *P. hydriforme*. However, the

Significance

Myxozoans are a diverse group of microscopic parasites that infect invertebrate and vertebrate hosts. The assertion that myxozoans are highly reduced cnidarians is supported by the presence of polar capsules, which resemble cnidarian stinging structures called "nematocysts." Our study characterizes the genomes and transcriptomes of two distantly related myxozoan species, *Kudoa iwatai* and *Myxobolus cerebralis*, and another cnidarian parasite, *Polypodium hydriforme*. Phylogenomic analyses that use a broad sampling of myxozoan taxa confirm the position of myxozoans within Cnidaria with *P. hydriforme* as the sister taxon to Myxozoa. Analyses of myxozoan genomes indicate that the transition to the highly reduced body plan was accompanied by massive reduction in genome size, including depletion of genes considered hallmarks of animal multicellularity.

Author contributions: D.H. and P.C. designed research; E.S.C., D.H., and P.C. performed research; E.S.C., A.D., D.H., and P.C. contributed new reagents/analytic tools; E.S.C., M.N., N.D.R., H.P., D.H., and P.C. analyzed data; and D.H. and P.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Sequence Read Archive database (accession nos. SRX704259, SRX702459, SRX554567, SRX1034928, SRX1034914, SRX570527, and SRX687102).

[1]To whom correspondence may be addressed. Email: pcart@ku.edu or huchond@post.tau.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1511468112/-/DCSupplemental.

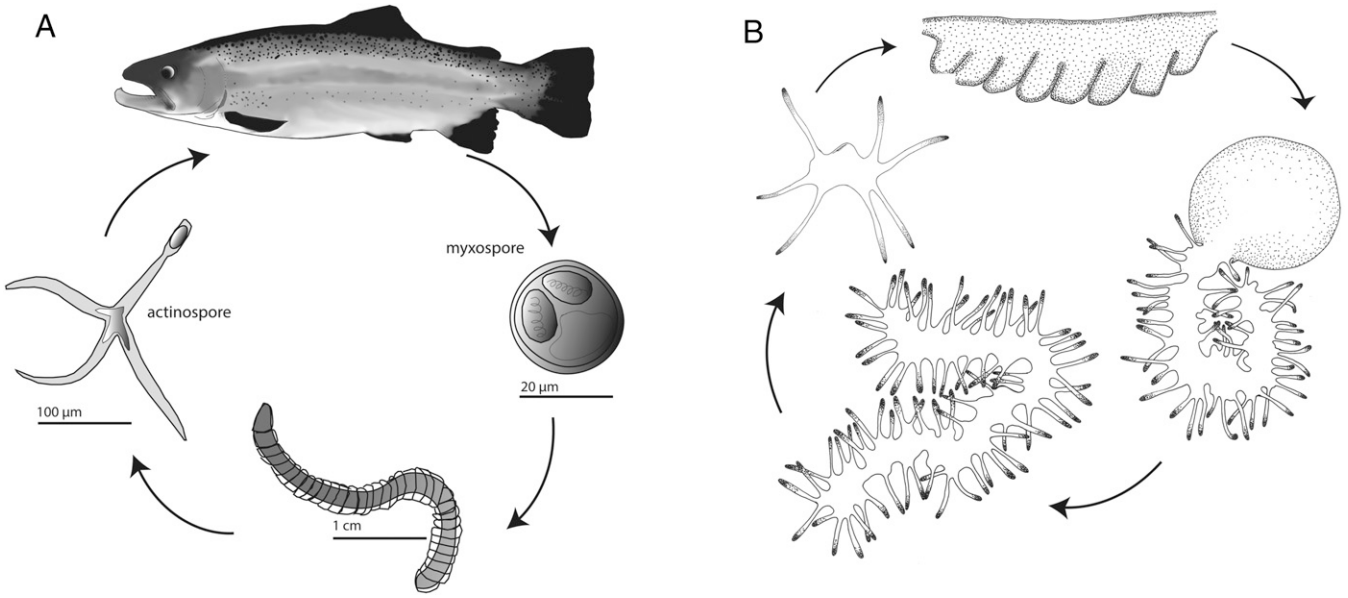


Fig. 1. Life cycles of *M. cerebralis* and *P. hydriforme*. (*A*) *M. cerebralis* alternates its development between a fish (salmonid) host and an annelid (*T. tubifex*) host. The myxospore is produced in the fish (*Right*), and the actinospore is produced in the annelid (*Left*). Both stages consist of just a few cells, including those housing polar capsules. (*B*) In *P. hydriforme*, the stolon stage (*Top*) develops inside the ovaries of its host (acipenceriform fish). Upon host spawning, *P. hydriforme* emerges from the host's oocyte (*Right*), fragments, and lives as a free-living stage with a mouth (*Left*) before infecting its host.

position of this clade was unstable. It was either placed as the sister clade to Bilateria or nested within Cnidaria, depending on taxon sampling, alignment, optimization method, and the characters considered (13, 19–23). Recent phylogenomic studies support a position of Myxozoa within Cnidaria, as the sister clade to Medusozoa (24–26). However, in these studies *P. hydriforme* and representatives of major lineages of myxozoan and nonmyxozoan cnidarians were notably absent. Thus, the precise phylogenetic position within Cnidaria remains uncertain.

In recent studies, myxozoans were found to possess the cnidarian-specific minicollagen and nematogalectin genes (24, 27, 28), each of which has been shown to play important roles in the nematocyst structure in *Hydra* (29, 30). These studies support previous morphology-based assertions that myxozoan polar capsules are homologous to cnidarian nematocysts (4, 5, 31, 32) and thus indirectly suggest a close evolutionary relationship between these two groups.

For this study, we analyzed genomic and transcriptomic assemblies from two distantly related myxosporean myxozoans, *Kudoa iwatai* and *M. cerebralis*, as well as the cnidarian parasite *P. hydriforme*, to gain insight into the evolutionary transition to parasitism and extreme reduction of body plans from a free-living cnidarian. First, we used these newly generated data, in conjunction with publicly available data, to determine the phylogenetic relationships between all major lineages of myxozoans (33), *P. hydriforme*, and other cnidarians to reconstruct the evolutionary history of endoparasitism in Cnidaria. Second, we compared genome size, gene number, gene content, and enrichment of expressed genes between myxozoans, *P. hydriforme*, and published cnidarian genomes to determine if the degeneration of the cnidarian body plan displayed in Myxozoa (but not in *P. hydriforme*) was accompanied by genome reduction and gene loss. Our findings reaffirm a cnidarian origin for myxozoans and recover them as the sister group to *P. hydriforme*. Analysis of genome and transcriptome assemblies reveal that the highly degenerate body plan of myxozoans coincided with extreme reduction in genome size and gene loss while retaining some genes necessary to function as an obligate parasite. By contrast, *P. hydriforme*, which displays many cnidarian-like morphological features, has a genome size and gene content similar to that of published cnidarian genomes.

Results

Phylogenetic Position. A phylogenomic analysis was performed using the newly generated transcriptome assemblies from *K. iwatai*, *M. cerebralis*, and *P. hydriforme* as well as genomic data of *Enteromyxum leei* and *Sphaeromyxa zaharoni* in conjunction with published sequences from three additional myxozoans (*Buddenbrockia plumatellae*, *Tetracapsuloides bryosalmonae*, and *Thelohanellus kitauei*), altogether encompassing 22 cnidarians, 38 representatives of the Metazoan diversity, and 9 unicellular opistokont taxa. Both Bayesian analyses using the CAT model (34) and a maximum-likelihood (ML) analysis using the GTR model recovered *P. hydriforme* as sister to a monophyletic Myxozoa with maximal support [Bayesian posterior probability (PP) of 1.0, ML bootstrap percentage (BP) of 100]. Within a monophyletic Cnidaria (PP = 1.0, BP = 100), the Myxozoa + *P. hydriforme* clade was recovered as sister to the medusozoan clade with maximal support (PP = 1.0, BP = 100) (Fig. 2). The Bayesian and ML topologies differed only in the position of two taxa (*Porites* and *Strigamia*). Several analyses were conducted to evaluate the robustness of the position of the Myxozoa + *P. hydriforme* clade within Cnidaria. Because it has been claimed that ribosomal genes can contain a different signal from nonribosomal genes (35), phylogenetic analyses were conducted on a dataset of 41,237 amino acids, excluding ribosomal genes (Fig. S1). Additionally, because taxon sampling can affect phylogenetic inferences, phylogenetic reconstructions were performed either with only cnidarian taxa (Fig. S2) or after removing either Myxozoa or *P. hydriforme*. None of these analyses affected the position of Myxozoa and *P. hydriforme* or its position as the sister clade to Medusozoa.

Estimation of the Completeness of Genome and Transcriptome Assemblies. RNA libraries from *M. cerebralis*, *K. iwatai*, and *P. hydriforme* and DNA libraries from the latter two were sequenced using a short-read Illumina platform. Data were deposited in the National Center for Biotechnology Information (NCBI) archives (Table S1). Previously generated *M. cerebralis* genomic data (26) were downloaded from the NCBI (SRX208206). Assembly statistics are shown in Table 1, and size distribution of the transcriptome sequences is shown in Fig. S3. Completeness of the genome and transcriptome assemblies was estimated by determining the presence of the 248 ultra-conserved core eukaryotic

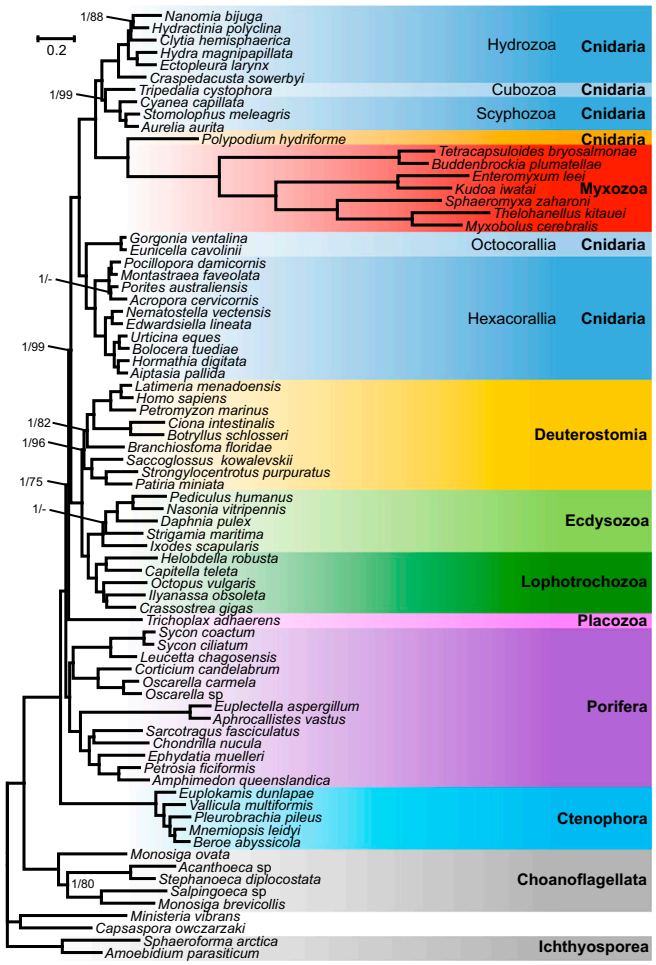


Fig. 2. Phylogenetic tree generated from a matrix of 51,940 amino acid positions and 77 taxa using Bayesian inference under the CAT model. Support values are indicated only for nodes that did not received maximal support. Bayesian posterior probabilities/ML bootstrap supports under the PROTGAMMAGTR model are given near the corresponding node. A minus sign ("–") indicates that the corresponding node is absent from the ML bootstrap consensus tree.

genes (CEGs), obtained from the Core Eukaryotic Genes Mapping Approach (CEGMA) database (36) (Table 1). The *K. iwatai* genome and transcriptome assemblies recovered over 70% of the CEGs with over 1,000× estimated mean base-pair coverage. The *M. cerebralis* transcriptome was less complete, recovering only 39% of the CEGs. We were unable to recover any CEGs for *M. cerebralis* from its published genomic data, most likely due to their low coverage. Although the *P. hydriforme* genome assembly recovered very few CEGs due to low coverage, its transcriptome assembly recovered 90% of complete CEGs.

General Characteristics of Genomes. Genome size estimates based on overall coverage of known individual genes are shown in Table 2. These estimates suggest that the myxozoan genome of *K. iwatai* (22.5 Mb) is one of the smallest reported animal genomes, comparable to the genome of the recently reported parasitic nematode (~20 Mb) (37). The *K. iwatai* genome is more than 20-fold smaller than the estimated size of the *P. hydriforme* genome (561 Mb) and the published genome of the cnidarian *Nematostella vectensis* (450 Mb) (38) and more than 40-fold smaller than the published estimated genome size of the cnidarian *Hydra magnipapillata* (1,005 Mb) (39). Although the published estimated genome size of the myxozoan *Thelohanellus kitauei* (188.5 Mb), which was based on K-mer distribution (40), is eightfold larger than our estimated *K. iwatai* genome size, it is still significantly smaller than the nonmyxozoan cnidarian genomes. As an independent test of the accuracy of genome-size estimation, we compared genome size based on the overall assembly coverage of the *K. iwatai* genome from two independent sequencing runs (*SI Materials and Methods*). This revealed a very similar genome-size estimate (23.5 Mb). Due to the low coverage of the published *M. cerebralis* genomic read data, it was not possible to estimate its genome size.

The number of protein-coding genes and average intron and exon sizes were estimated from the genome assemblies using the Maker2 (*SI Materials and Methods*). These analyses revealed that the number of protein-coding genes in the *K. iwatai* genome (5,533) is less than 30% of those estimated in *P. hydriforme* (17,440), *H. magnipapillata* (16,839), and *N. vectensis* (18,000) (Table 2). In addition, the myxozoan genome appears to be much more compact, with a mean intron size of 82 bp in *K. iwatai*, compared with 1,163 bp in *P. hydriforme*, 2,673 bp in *H. magnipapillata*, and 799 bp in *N. vectensis* (Table 2).

Characteristics of Transcriptomes: Comparisons of Gene Ontology and Gene Orthology. To identify the biological pathways that have gained or lost a significant fraction of expressed genes in the myxozoans *M. cerebralis*, *K. iwatai*, and *P. hydriforme* compared with the cnidarian model species *H. magnipapillata* and *N. vectensis*, Fisher's exact tests were used to infer enrichment and depletion in the proportion of genes present in 112 Gene Ontology (GO) categories as defined by the GOSlim list of CateGOrizer (41) (Table 3 and Dataset S1). Because the GO terms of *P. hydriforme* were more similar to *N. vectensis* and *H. magnipapillata* than to the myxozoans *M. cerebralis* and *K. iwatai*, the most informative comparison was between *M. cerebralis* and *K. iwatai* versus *P. hydriforme*, *H. magnipapillata*, and *N. vectensis* (Dataset S1, Tab 2). Of the top 20 GO categories with the highest occurrences of GO terms (Fig. 3), the expressed myxozoan genes appear to be significantly depleted (by comparison with other cnidarians) in categories that are related to development, cell differentiation, and cell-to-cell communication (Table 3), consistent with lack of a complex multicellular body in myxosporean myxozoans. By contrast, myxozoan-expressed genes have an abundance of categories such as cellular function, for which the number of genes does not differ significantly

Table 1. Assembly statistics for sequenced genomes and transcriptomes

	K. iwatai		*M. cerebralis*		*P. hydriforme*	
	Genome	Transcriptome	Genome*	Transcriptome	Genome	Transcriptome
Raw reads	167,917,062	154,253,215	NA	312,202,378	229,917,588	467,431,688
Contigs	1,637	6,528	NA	52,972	83,415	24,523
N50	40,195	1,662	NA	994	3,865	1,475
CEGs (c), %†	179/72	190/77	0/0	97/39	14/6	223/90
CEGs (p), %†	188/76	208/84	0/0	164/66	56/23	232/94

*Published ESTs.
†Number/percentage of 248 ultra-conserved CEGs. c, complete; p, partial.

Table 2. Estimated genome characteristics

	Kudoa	*Polypodium*	*Hydra* (39)	*Nematostella* (38)
Genome size (Mb)	22.5	561	1,005	450
No. protein genes	5,533	17,440	16,839	18,000
GC content, %	28	47	29	41
Mean intron size, bp	82	1,163	2,673	799
Mean exon size, bp	102	216	218	208

from the number observed in cnidarians (e.g., a similar number of nucleoplasm genes was found in both) (Dataset S1, Tab 2). Although these analyses were from transcriptomes, the general patterns likely reflect overall genome content as multiple life cycle stages are represented in the combined transcriptome of *K. iwatai* and *M. cerebralis* (*SI Materials and Methods*). To confirm this, we also performed a GO comparison analysis of the genes predicted based on genomic sequences, which revealed the same general patterns for *K. iwatai*, but not for *P. hydriforme* whose transcriptome assembly was of better quality than its genome assembly (Dataset S1 and Fig. S4).

Using the OrthoMCL database (42), we determined the number of orthologous groups (OG) that could be identified from our transcriptome assemblies of *P. hydriforme* and *K. iwatai*, compared with published predicted proteins from *H. magnipapillata* (39) and *N. vectensis* (38) (Fig. S5). A total of 8,021 unique OGs were recovered from *H. magnipapillata*, 11,162 from *N. vectensis*, 5,451 from *P. hydriforme*, and 2,735 from *K. iwatai*. Although there was more overlap in the number of OGs between *H. magnipapillata*, *N. vectensis*, and *P. hydriforme* than between *K. iwatai* and the other three, this is consistent with the significantly lower number of total OGs in *K. iwatai* (Fig. S5).

Analyses of Gene Pathways and Candidate Genes in the Transcriptome and Genome Assemblies. We searched for several candidate genes and gene pathways that have been previously characterized as important for cnidarian cell signaling and development in the transcriptome and genome assemblies of *M. cerebralis*, *K. iwatai*, and *P. hydriforme*. Using BLAST searches and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses we determined presence/absence of representatives within a gene family, but not precise orthology, within the particular families. Myxozoan genomes appear to lack key genes and signaling pathways that are present in *P. hydriforme* and other cnidarians (Table 4). Specifically, the conserved transcriptional factors belonging to the Hox and Runx gene families, which have been shown to be important for cnidarian patterning and cellular differentiation, respectively (43, 44), were found in neither the genomes nor transcriptomes of myxozoans, but were nevertheless present in *P. hydriforme* (Table 4 and Dataset S2). In addition, myxozoans appear to have lost the ligands, receptors, and most downstream elements of the Wnt- and Hedgehog-signaling pathways, which have been shown to be important for axial patterning (45) and cell signaling (46), respectively, whereas nearly all components of these pathways were recovered in *P. hydriforme* (Table 4 and Dataset S3). By contrast, myxozoans and *P. hydriforme* possess orthologs to the stem-cell markers FoxO (47) and Piwi (48), and *P. hydriforme* and *K. iwatai* appear to have the gene Hap2, which was shown to be involved in gamete fusion in *Hydra* (49) (Table 4 and Dataset S2). The two myxozoans and *P. hydriforme* were also found to have key elements of the Notch-signaling pathway, and *M. cerebralis* possessed some elements of the TGFβ pathway (Table 4 and Dataset S3). Notch is reported to have an important role in differentiation of stem-cell lineages (50), whereas TGFβ appears to play a more general role in cell signaling (51).

Discussion

Our analyses of transcriptomic and genomic assemblies of myxozoans have yielded significant insight into the evolution of these microscopic parasites from free-living cnidarians. We report, for the first time to our knowledge, a broad phylogenomic sampling of myxozoans, including representatives from the malacosporean clade and the freshwater and marine myxosporean clades (33), as well as the only phylogenomic study to date to include *P. hydriforme*. In addition, we have a more comprehensive sampling of cnidarians than previous phylogenomic studies, addressing the placement of myxozoans (24–26). We recover *P. hydriforme* as the sister taxon to Myxozoa and can confirm, with an increased sampling and thus a higher degree of confidence, the placement of this clade as the sister taxon to medusozoan cnidarians. These results are consistent with those of other molecular phylogenetic studies (19, 23), although these have been criticized as possible artifacts of long-branch attraction (21). The monophyly of Myxozoa + *P. hydriforme* is also supported by endoparasitism in fish, a unique cell-within-cell developmental stage, possession of a single similar type of nematocyst (19, 23), and similarity in minicollagen sequences (28). This phylogenomic pattern suggests that endoparasitism in Cnidaria was a single event that occurred at the base of Myxozoa + *P. hydriforme*, but that the dramatic reduction in body plan occurred following the divergence of *P. hydriforme* from myxozoans, as *P. hydriforme* retains many cnidarian features.

The Myxozoa represent an extreme example of degeneration of body plans due to parasitism. Genome and transcriptome analyses reveal that this degeneration was accompanied by massive genome reduction, with myxozoans having one of the smallest reported animal genomes. Genome size reduction included loss of many genes considered hallmarks of metazoan development, yet retention of genes necessary to function as obligate parasites, such as nematocyst-specific genes. In contrast to myxozoans, *P. hydriforme* has a genome similar in size, gene number, and gene content to the model system *Hydra*. This finding is not surprising given that, although *P. hydriforme* is an obligate parasite, it has maintained its cnidarian-like body plan, including epithelia, mouth, gut, and tentacles. Our study provides a robust phylogenetic hypothesis for myxozoan placement within Cnidaria, as the sister taxon to *P. hydriforme*, and a framework for comparative genomic studies, which should be valuable for future phylogenetic and genomic investigations of Cnidaria sensu lato.

Materials and Methods

Phylogenetic Reconstruction. Curated sequence alignments of 200 protein markers (52) were augmented with sequences from the four myxozoan species generated for this study, from *P. hydriforme*, and from the NCBI databases (www.ncbi.nlm.nih.gov) [GenBank and Sequence Read Archive (SRA)] as described (52). After removing positions with unreliable alignment, the final dataset included 51,940 sites and 12% missing data (Dataset S4). Bayesian tree reconstructions were conducted under the Bayesian CAT model (34) as implemented in Phylobayes MPI vs.1.5 (53) (*SI Materials and Methods*). The ML analysis was conducted under the PROTGAMMAGTR model, as implemented in RAxML 8.1.3 (54). Alignments and Bayesian tree

Table 3. Transcriptome-based Gene Ontology categories showing depletion in myxozoans compared with other cnidarians

Category	Myxozoa, %*	Cnidaria, %†
Cell differentiation	548/0.0234	1763/0.0302
Development	1052/0.0450	3892/0.0667
Morphogenesis	456/0.0195	1693/0.0290
Receptor activity	46/0.00197	310/0.00531
Signal transducer activity	49/0.0002	328/0.0056

Myxozoan total number = 23,388. Cnidarian total number = 53,354. All categories had a *P* value ≤0.0001.
*Myxozoa = *K. iwatai* + *M. cerebralis*.
†Cnidaria = *H. magnipillata* + *N. vectensis* + *P. hydriforme*.

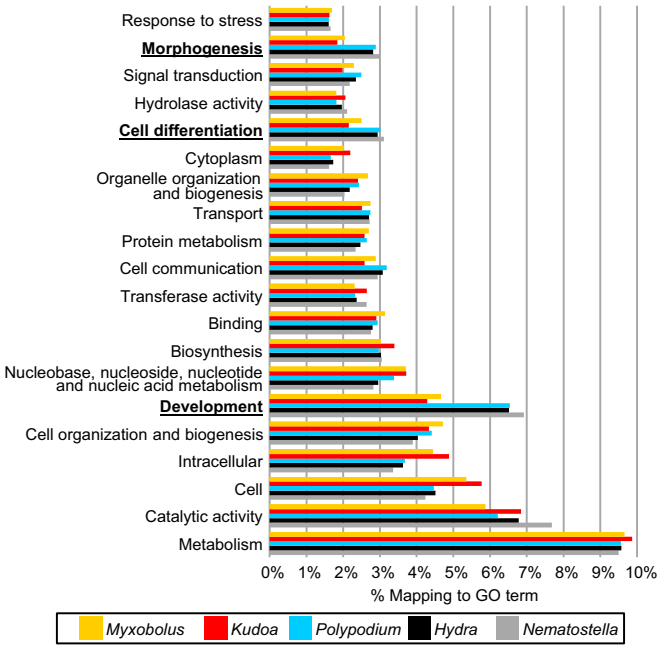


Fig. 3. GO annotation of unigenes in transcriptomes. The top 20 GO categories are shown as a percentage of total GO terms from the transcriptome assemblies of *M. cerebralis*, *K. iwatai*, *P. hydriforme*, and the published protein sequences of *H. magnipapillata* and *N. vectensis*. Categories for which myxozoans present significantly fewer GO terms than other cnidarians are indicated by boldface type.

have been deposited in the TreeBASE repository (55) (purl.org/phylo/treebase/phylows/study/TB2:S17743?format=html).

Specimen Collection. Collection of specimens of *P. hydriforme*, *K. iwatai* plasmodia, *E. leei* trophozoites, and *S. zaharoni* plasmodia for both genomic and transcriptomic purposes was carried out as described in ref. 28 and *SI Materials and Methods*. Flash-frozen *M. cerebralis* actinospores were kindly provided by Ron Hedrick (University of California at Davis).

Illumina Sequencing. For the *M. cerebralis* trancriptome assembly, RNA extraction, library preparation, and sequencing was performed for the *P. hydriforme* transcriptome as described (28). Library preparations and genome sequencing of *P. hydriforme* was carried out at the Genome Sequence Facility at the University of Kansas Medical School. *P. hydriforme* gDNA was sheared to a size of 350 bp, and 100 bp paired-end (PE) sequencing was performed on an Illumina HiSEq 2000. Library preparation and Illumina 100-bp PE HiSEq 2000 sequencing of *K. iwatai* (transcriptome and genome), *S. zaharoni* (genome), and *E. leei* (genome) was described (28). In addition to the Illumina HiSeq sequencing, the *K. iwatai* genomic library was also independently sequenced with an Illumina Genome Analyzer IIx platform, which produced 95,434,687 paired reads (76 bp).

Genome and Transcriptome Assemblies. The *M. cerebralis* transcriptome was filtered for read quality and assembled following protocols described for the *P. hydriforme* transcriptome (28). Genome de novo assemblies for *P. hydriforme*, *E. leei*, *S. zaharoni*, and *K. iwatai* were performed with ABySS v. 1.3.6 (56), and the transcriptome de novo assembly of *K. iwatai* was performed with Trinity (57). Contigs shorter than 500 and 300 bp were removed from the genomic and transcriptomic assemblies, respectively. Filtering methods for host contaminants are described in *SI Materials and Methods*. The accession of the different assemblies and short read data are indicated in Table S1.

Analysis of Assembly Completeness. For each transcriptome and genome assembly, relative completeness was assessed using CEGMA that searches for the presence of 248 ultra-conserved CEGs (58). *P. hydriforme* assemblies were run using default settings. Because our evidence indicates that myxozoans have unusually small intron sizes (see below), the –max_intron_size parameter for *M. cerebralis* and *K. iwatai* genomic CEGMA runs was adjusted to match the maximum intron size of the *M. cerebralis* intron size distribution (2,630 bp).

Table 4. Presence (+) or absence (–) of genes and KEGG pathways that have been characterized in other cnidarians

	K. iwatai	*M. cerebralis*	*Polypodium*
Genes			
Hox-like	–	–	+
Runx	–	–	+
Piwi	+	+	+
FoxO	+	+	+
Hap2	+	–	+
KEGG pathways	–	–	–
Wnt	–	–	+
Hedge	–	–	+
TGFβ	–	+	+
Notch	+	+	+

Estimation of Genome Size and Content. Output from the CEGMA runs was used for coverage-based estimates of the genome size for *P. hydriforme* and *K. iwatai*. We used the "dna" output files from the genomic CEGMA runs, which include the raw sequence for each region identified by CEGMA as a partial or complete core gene, as well as the 2,000 bp of sequence on each side. Because CEGs were chosen to minimize in-paralogy and therefore should be largely single-copy, mapping reads to these regions provide a simple unbiased estimate of genome coverage. For each of the genomes, the complete set of raw reads was mapped to the CEGMA output files using the Burrows-Wheeler Aligner, BWA-MEM (v.0.7.8), with default settings (59). Coverage was calculated using QualiMap 2 (60). Under the assumption that the coverage estimated from the conserved CEGMA-identified regions is representative of genome-wide coverage, the number of base pairs used in the whole read set for each assembly was divided by the calculated coverage for that species, thus providing an estimate of the genome size for each species (Table 2). In addition, an independent approach based on coverage of all contigs was used to estimate the genome size of *K. iwatai* (*SI Materials and Methods*). Genome annotation was conducted using MAKER2 (*SI Materials and Methods*). The OrthoMCL database (42) was used to determine the number of orthologous groups identified in the transcriptome assemblies of *K. iwatai* and *P. hydriforme*, compared with published predicted proteins in *H. magnipapillata* (*SI Materials and Methods*).

Gene Enrichment Analysis of Transcriptomes. Before conducting the gene enrichment analysis, duplicate sequences and alternative transcripts were removed from the transcriptomes of *P. hydriforme* and *M. cerebralis* using methods described in *SI Materials and Methods*. No redundant contigs were found in the transcriptome of *K. iwatai*. We used the Trinotate pipeline (*SI Materials and Methods*) to annotate the *K. iwatai*, *M. cerebralis*, and *P. hydriforme* unigenes. In particular, Trinotate searched the GO database (61) and recovered for each transcript its relevant GO terms. The GO terms of each transcriptome were reduced using the GOSlim list in CateGOrizer (41). The database of *H. magnipapillata* and *N. vectensis* protein sequences was downloaded from the Metazome website (metazome.net/). GO terms were also assigned to *H. magnipapillata* and *N. vectensis* using Trinotate (*SI Materials and Methods*). The GO terms categories of Myxozoa and nonmyxozoan Cnidaria were compared for depletion or enrichment using Fisher's exact tests. The significance level was corrected for multiple testing using a Bonferroni correction (specifically, $\alpha = 0.05$ was corrected to $\alpha = 0.000446$). Gene enrichment analysis of genomes is described in *SI Materials and Methods*.

Analysis of Gene Pathways and Candidate Genes. We searched assembled genomes and transcriptomes for genes and gene pathways that are considered important for cnidarian development. A multistep BLAST-based approach (62) was used for identifying candidate genes (*SI Materials and Methods*). All BLAST-search results are found in Dataset S2. We also assessed the completeness of candidate signaling pathways in our assemblies using KEGG (63) (*SI Materials and Methods*).

ACKNOWLEDGMENTS. We thank the Oklahoma Paddlefish Research Center for help with collecting *P. hydriforme*; Ron Hedrick for providing frozen tissue from *M. cerebralis*; A. Shaver for illustrations; and M. Shcheglovitova, Liran Dray, and Tamar Feldstein for technical assistance. Genome sequencing services were provided by the University of Kansas Medical School Genome Sequencing Center, the Technion Genome Center, and the Duke Center for Genomic and Computational Biology Sequencing Core Facility. Computing support was provided by the University of Kansas Information and Telecommunication Technology Center. We thank K. Jensen, J. Kelly, J. Ryan, S. Sanders, C. Schnitzler, O. Simakov, and R. Steele for discussions and two anonymous reviewers for helpful comments and suggestions. This work was supported by the National Science Foundation

(NSF)-Division of Environmental Biology 0953571 (to P.C.), the Israel Science Foundation (663/10 to D.H.), and the Binational Science Foundation (BSF)-NSF Joint Funding Program in International Collaborations in Organismal Biology (ICOB): NSF-IOS 1321759 (to P.C.) and BSF 2012768 (to D.H.).

1. Hoeg JT (1995) The biology and life cycle of the Rhizocephala (Cirripedia). *J Mar Biol Assoc UK* 75(3):517–550.
2. Kobayashi M, Furuya H, Holland PWH (1999) Dicyemids are higher animals. *Nature* 401(6755):762–762.
3. Kent ML, et al. (2001) Recent advances in our knowledge of the Myxozoa. *J Eukaryot Microbiol* 48(4):395–413.
4. Štolc A (1899) Actinomyxidies, nouveau groupe de Mesozoaires parent des Myxosporidies. *Bull Intl Acad Sci Boheme* 22:1–12.
5. Weill R (1938) L'interpretation des Cnidosporidies et la valeur taxonomique de leur cnidome. Leur cycle comparé à la phase larvaire des Narcomeduses Cuninides. *Travaux de la Station Zoologique de Wimereaux* 13:727–744.
6. Lom J, Dyková I (2006) Myxozoan genera: Definition and notes on taxonomy, life-cycle terminology and pathogenic species. *Folia Parasitol (Praha)* 53(1):1–36.
7. Wolf K, Markiw ME (1984) Biology contravenes taxonomy in the myxozoa: New discoveries show alternation of invertebrate and vertebrate hosts. *Science* 225(4669):1449–1452.
8. Hedrick RP, el-Matbouli M, Adkison MA, MacConnell E (1998) Whirling disease: Re-emergence among wild trout. *Immunol Rev* 166:365–376.
9. El-Matbouli M, Hoffmann RW, Mandok C (1995) Light and electron-microscopic observations on the route of the triactinomyxon-sporoplasm of Myxobolus cerebralis from epidermis into rainbow trout cartilage. *J Fish Biol* 46(6):919–935.
10. Monteiro AS, Okamura B, Holland PWH (2002) Orphan worm finds a home: Buddenbrockia is a myxozoan. *Mol Biol Evol* 19: 968–971.
11. Hartikainen H, Gruhl A, Okamura B (2014) Diversification and repeated morphological transitions in endoparasitic cnidarians (Myxozoa: Malacosporea). *Mol Phylogenet Evol* 76:261–269.
12. Raikova EV, Suppes VC, Hoffman GL (1979) The parasitic coelenterate, Polypodium hydriforme USSOV, from the eggs of the American acipenseriform Polyodon spathula. *J Parasitol* 65(5):804–810.
13. Evans NM, Holder MT, Barbeitos MS, Okamura B, Cartwright P (2010) The phylogenetic position of Myxozoa: Exploring conflicting signals in phylogenomic and ribosomal datasets. *Mol Biol Evol* 27(12):2733–2746.
14. Foox J, Siddall ME (2015) The road to Cnidaria: History of phylogeny of the Myxozoa. *J Parasitol* 101(3):269–274.
15. Okamura B, Gruhl A (2015) *Myxozoan Affinities and Route to Endoparasitism. Myxozoan Evolution, Ecology and Development* (Springer, Berlin), pp 23–44.
16. Bütschli O (1881) Myxosporidien. *Zoologischer Jahrbericht für* 1:162–164.
17. Lipin A (1925) Geschlechtliche Form, Phylogenie und systematische Stellung von Polypodium hydriforme Ussov. *Zool Jahrb Anat* 47:541–635.
18. Raikova EV (1988) *On the systematic position of Polypodium hydriforme Ussov (Coelenterata).* Porifera and Cnidaria: Contemporary State and Perspectives of Investigations, eds Koltum VM, Stepanjants SD (Zoological Institute of Academy of Sciences of USSR, Leningrad), pp 116–122.
19. Siddall ME, Martin DS, Bridge D, Desser SS, Cone DK (1995) The demise of a phylum of protists: Phylogeny of Myxozoa and other parasitic cnidaria. *J Parasitol* 81(6):961–967.
20. Smothers JF, von Dohlen CD, Smith LH, Jr, Spall RD (1994) Molecular evidence that the myxozoan protists are metazoans. *Science* 265(5179):1719–1721.
21. Evans NM, Lindner A, Raikova EV, Collins AG, Cartwright P (2008) Phylogenetic placement of the enigmatic parasite, Polypodium hydriforme, within the Phylum Cnidaria. *BMC Evol Biol* 8(1):139.
22. Siddall ME, Whiting MF (1999) Long-branch abstractions. *Cladistics* 15(1):9–24.
23. Zrzavý J, Hypša V (2003) Myxozoa, Polypodium, and the origin of the Bilateria: The phylogenetic position of "Endocnidozoa" in the light of the rediscovery of Buddenbrockia. *Cladistics* 19(3):164–169.
24. Feng J-M, et al. (2014) New phylogenomic and comparative analyses provide corroborating evidence that Myxozoa is Cnidaria. *Mol Phylogenet Evol* 81(0):10–18.
25. Jiménez-Guri E, Philippe H, Okamura B, Holland PWH (2007) Buddenbrockia is a cnidarian worm. *Science* 317(5834):116–118.
26. Nesnidal MP, Helmkampf M, Bruchhaus I, El-Matbouli M, Hausdorf B (2013) Agent of whirling disease meets orphan worm: Phylogenomic analyses firmly place Myxozoa in Cnidaria. *PLoS One* 8(1):e54576.
27. Holland JW, Okamura B, Hartikainen H, Secombes CJ (2011) A novel minicollagen gene links cnidarians and myxozoans. *Proc R Soc Lond B Biol Sci* 278(1705):546–553.
28. Shpirer E, et al. (2014) Diversity and evolution of myxozoan minicollagens and nematogalectins. *BMC Evol Biol* 14(1):205.
29. Adamczyk P, et al. (2008) Minicollagen-15, a novel minicollagen isolated from Hydra, forms tubule structures in nematocysts. *J Mol Biol* 376(4):1008–1020.
30. Hwang JS, et al. (2010) Nematogalectin, a nematocyst protein with GlyXY and galectin domains, demonstrates nematocyte-specific alternative splicing in Hydra. *Proc Natl Acad Sci USA* 107(43):18539–18544.
31. Reft AJ, Daly M (2012) Morphology, distribution, and evolution of apical structure of nematocysts in hexacorallia. *J Morphol* 273(2):121–136.
32. Okamura B, Gruhl A, Reft AJ (2015) *Cnidarian Origins of the Myxozoa: Myxozoan Evolution, Ecology and Development* (Springer, Berlin), pp 45–68.
33. Fiala I, Bartosová P (2010) History of myxozoan character evolution on the basis of rDNA and EF-2 data. *BMC Evol Biol* 10(1):228.
34. Lartillot N, Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 21(6):1095–1109.
35. Nosenko T, et al. (2013) Deep metazoan phylogeny: When different genes tell different stories. *Mol Phylogenet Evol* 67(1):223–233.
36. Parra G, Bradnam K, Ning Z, Keane T, Korf I (2009) Assessing the gene space in draft genomes. *Nucleic Acids Res* 37(1):289–297.
37. Burke M, et al. (2015) The plant parasite Pratylenchus coffeae carries a minimal nematode genome. *Nematology* 17(6):621–637.
38. Putnam NH, et al. (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317(5834):86–94.
39. Chapman JA, et al. (2010) The dynamic genome of Hydra. *Nature* 464(7288):592–596.
40. Yang Y, et al. (2014) The genome of the myxosporean Thelohanellus kitauei shows adaptations to nutrient acquisition within its fish host. *Genome Biol Evol* 6(12):3182–3198.
41. Zhi-Liang H, Jie B, Reecy JM (2008) CateGOrizer: A web-based program to batch analyze gene ontology classification categories. *Online J Bioinform* 9(2):108–112.
42. Chen F, Mackey AJ, Stoeckert CJ, Jr, Roos DS (2006) OrthoMCL-DB: Querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res* 34(Database issue, suppl 1):D363–D368.
43. Ryan JF, et al. (2006) The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: Evidence from the starlet sea anemone, Nematostella vectensis. *Genome Biol* 7(7):R64.
44. Sullivan JC, et al. (2008) The evolutionary origin of the Runx/CBFbeta transcription factors: Studies of the most basal metazoans. *BMC Evol Biol* 8(1):228.
45. Hensel K, Lotan T, Sanders SM, Cartwright P, Frank U (2014) Lineage-specific evolution of cnidarian Wnt ligands. *Evol Dev* 16(5):259–269.
46. Matus DQ, Magie CR, Pang K, Martindale MQ, Thomsen GH (2008) The Hedgehog gene family of the cnidarian, Nematostella vectensis, and implications for understanding metazoan Hedgehog pathway evolution. *Dev Biol* 313(2):501–518.
47. Boehm A-M, et al. (2012) FoxO is a critical regulator of stem cell maintenance in immortal Hydra. *Proc Natl Acad Sci USA* 109(48):19697–19702.
48. Juliano CE, et al. (2014) PIWI proteins and PIWI-interacting RNAs function in Hydra somatic stem cells. *Proc Natl Acad Sci USA* 111(1):337–342.
49. Steele RE, Dana CE (2009) Evolutionary history of the HAP2/GCS1 gene and sexual reproduction in metazoans. *PLoS One* 4(11):e7680.
50. Käsbauer T, et al. (2007) The Notch signaling pathway in the cnidarian Hydra. *Dev Biol* 303(1):376–390.
51. Hobmayer B, Rentzsch F, Holstein TW (2001) Identification and expression of HySmad1, a member of the R-Smad family of TGFbeta signal transducers, in the diploblastic metazoan Hydra. *Dev Genes Evol* 211(12):597–602.
52. Philippe H, et al. (2011) Acoelomorph flatworms are deuterostomes related to Xenoturbella. *Nature* 470(7333):255–258.
53. Lartillot N, Rodrigue N, Stubbs D, Richer J (2013) PhyloBayes MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol* 62(4):611–615.
54. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.
55. Piel WH, et al. (2009) TreeBASE v. 2: A database of phylogenetic knowledge.
56. Simpson JT, et al. (2009) ABySS: A parallel assembler for short read sequence data. *Genome Res* 19(6):1117–1123.
57. Grabherr MG, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29(7):644–652.
58. Parra G, Bradnam K, Korf I (2007) CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23(9):1061–1067.
59. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.
60. García-Alcalde F, et al. (2012) Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics* 28(20):2678–2679.
61. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.
62. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
63. Kanehisa M, et al. (2014) Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Res* 42(Database issue, D1):D199–D205.
64. Diamant A, Ucko M, Paperna I, Colorni A, Lipshitz A (2005) Kudoa iwatai (Myxosporea: Multivalvulida) in wild and cultured fish in the Red Sea: Redescription and molecular phylogeny. *J Parasitol* 91(5):1175–1189.
65. Dray L, Neuhof M, Diamant A, Huchon D (August 8, 2014) The complete mitochondrial genome of the devil firefish Pterois miles (Bennett, 1828) (Scorpaenidae). *Mitochondrial DNA*, 10.3109/19401736.2014.945565.
66. Dray L, Neuhof M, Diamant A, Huchon D (2014) The complete mitochondrial genome of the gilthead seabream Sparus aurata L. (Sparidae). *Mitochondrial DNA*, 10.3109/19401736.2014.928861.
67. Li B, Dewey CN (2011) RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12(1):323.
68. Haas BJ, et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8(8):1494–1512.
69. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9(9):868–877.
70. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.
71. Holt C, Yandell M (2011) MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.
72. Min XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: Predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Res* 33(Web Server issue, suppl 2):W677–W680.
73. Bardou P, Mariette J, Escudié F, Djemiel C, Klopp C (2014) jvenn: An interactive Venn diagram viewer. *BMC Bioinformatics* 15(1):293.

Supporting Information

Chang et al. 10.1073/pnas.1511468112

SI Materials and Methods

Specimen Collection. Locality information is given in Table S1. Actinospores of *M. cerebralis* (kindly provided by R. Hedrick, University of California, Davis, CA) were collected and flash-frozen as they emerged from the annelid host (Fig. 1). For *K. iwatai*, plasmodia were collected from the fish host. *K. iwatai* plasmodia form pseudocysts encapsulated by host cells. Inside the plasmodia, the myxospores are at various stages of maturation (64). Because our RNA extractions were based on several cysts pooled together, we can assume that our RNA data represent all *Kudoa* life stages present in the fish. Unfortunately, the annelid host of *K. iwatai* is unknown. *P. hydriforme* was collected and flash-frozen 3–5 d after emerging from the host's oocytes, after it has fragmented into free-living individuals (Fig. 1).

Host Contaminant Filtering. To filter the assemblies from fish contaminants, genomic sequences were obtained for *Sparus aurata* (*E. leei* and *K. iwatai* host) and *Pterois miles* (*S. zaharoni* host) as described (65, 66). Blast searches were conducted to eliminate contaminating fish sequences from the genomic assemblies. Specifically BLASTN (version 2.2.27+) searches were performed for each of the three myxozoans, using the genomic assembly sequences as query against a database of their respective fish host DNA contigs. Sequences of *S. aurata* available in the NCBI dbEST (www.ncbi.nlm.nih.gov/dbEST/) were also included. The BLASTN parameters that were used are: -e-value 1e-75 and -perc_identity 85 (62). All sequences that passed this threshold were considered to be contaminants. Furthermore, we performed additional BLASTN searches against the NCBI nonredundant nucleotide database (e.g., to remove other contaminant such as bacterial sequences).

To filter the *K. iwatai* RNA assembly from contaminants (either host RNA or other sample contaminants), we ran RSEM (67) with the default parameters and filtered the low abundance transcripts using the filter_fasta_by_rsem_values.pl script supplied by Trinity (68) with default parameters (–fpkm_cutoff = 1200–isopct_cutoff = 1.00). We then ran BLASTN with -e-value 1e-75 and -perc_identity 85 against the sequences of the contaminant database described above (mainly, *S. aurata* DNA contigs and ESTs available in the NCBI database). We also added the sequence of the *S. aurata* mitochondrial genome (65). All sequences identified were removed. We then performed BLASTN against the NCBI nucleotide database on the remaining sequences with -e-value 1e-75 and -perc_identity 80. We removed all of the contigs that matched any euteleost sequence with over 80% identity and other taxa (e.g., fungi, *Drosophila*) with over 90% identity. Finally, we then performed a BLASTN search against the two filtered *K. iwatai* DNA assemblies (HiSeq and GIIx assemblies) with an –e-value 1e-5 threshold and removed sequences that could not align to any of the DNA assemblies. *P. hydriforme* genomic and transcriptomic sequences were filtered using sequence material from the paddlefish (*Polyodon spathula*) oocyte transcriptome as described (28).

Phylogenetic Reconstruction. Phylogenetic reconstructions based on the Bayesian and maximum-likelihood criteria were performed for different gene and species combinations: (*i*) a dataset that includes 77 species representative of the animal diversity with their closest outgroups and 200 ribosomal and nonribosomal protein genes (51,940 amino acids); (*ii*) a dataset that includes 77 species representative of the animal diversity with their closest outgroups and 128 nonribosomal protein genes (41,237 amino acids); (*iii*) a dataset that includes 30 cnidarian species and 200 ribosomal and nonribosomal protein genes (51,940 amino acids); and (*iv*) a dataset that includes 30 cnidarian species and 128 nonribosomal protein genes (41,237 amino acids). Additional analyses were also performed excluding either Myxozoa or *P. hydriforme*. For all datasets, Bayesian tree reconstructions were conducted under the CAT model (34) as implemented in Phylobayes MPI vs.1.5 (53). For the third dataset, the CAT-GTR model, which is more computationally intensive, was also used. For the analyses of datasets 1 and 3, two independent chains were run for 10,000 cycles, and trees were saved every 10 cycles. The first 2,000 trees were discarded (burn-in). For the analysis of the second dataset, the two chains were run for 6,000 generations and sampled every 10 trees, and the first 2,000 trees were discarded. For the analysis of the fourth dataset, the two chains were run for 20,000 generations and sampled every 10 trees, and the first 5,000 trees were discarded. Chain convergence was evaluated with the bpcomp and tracecomp programs of the Phylobayes software. The maximum and average differences observed at the end of each run were lower than 0.0005 for all analyses. Similarly, the effsize and rel_diff parameters were always higher than 30 and lower than 0.3, respectively, which indicates a correct chain convergence because our analyses investigate the topology rather than branch length and all relevant posterior probabilities = 1. The ML analyses were conducted for each dataset under the PROTGAMMAGTR as implemented in RAxML 8.1.3 (54). Bootstrap support was computed after 250 rapid bootstrap replicates. Alignments and Bayesian tree have been deposited in the TreeBASE repository (55) (purl.org/phylo/treebase/phylows/study/TB2:S17743?format=html).

Genome Size Estimation. Independent estimate of *K. iwatai* genome size based on assembly coverage using reads and contig sequences from two independent sequencing runs. We ran CAP3 (69) with the parameters -o 300 -p 90 (300 bp overlap between contigs and 90% identity in the overlapping sequence) on the GIIx *K. iwatai* DNA assembly to remove redundant contigs. A total of 952 redundant contigs were removed using CAP3. To filter the *K. iwatai* HiSeq DNA reads from contaminants, we created a contaminant database consisting of the *Kudoa* sequences identified as contaminants, the *S. aurata* DNA contigs obtained, and all *S. aurata* ESTs available in the NCBI EST database on May 2014. Bowtie2 was used with default settings to align the *K. iwatai* HiSeq DNA reads to the contaminant database. The –unconc flag was used to save the reads that did not map to the contaminant database to separate paired-end FASTQ files. A total of 164,284,209 reads remained after this step. We then used Bowtie2 (70) to align the filtered paired-end HiSeq reads to the GIIx assembly with default parameters. The coverage was calculated using bedtools genomecov -d –ibam on the output of Bowtie2 (70). The average coverage-per-position of the GIIx assembly was estimated (1391.6; SD: 1095.4; SE: 0.2577). The genome size was then estimated by dividing the number of base pairs sequenced (filtered reads) by the coverage according to the following formula: (number of paired reads) × (read length) × 2)/(average coverage-per-position) = (163,897,028 × 100 × 2)/1,391.6 = 23,555,192. Using this method, the genome size was thus estimated to be ∼23.5 Mbp.

Genome Annotation. Genome annotation was conducted using MAKER2 (71), incorporating the Semi-HMM based Nucleic Acid Parser (SNAP) gene predictor software to assess gene content of the *K. iwatai* and *P. hydriforme* genome assemblies. For each assembly, MAKER2 was first run using the assembled transcriptome for each species (EST evidence), a file of the core

CEGMA proteins, and a random precompiled eukaryotic HMM profile to train SNAP. The output of this training was a species-specific HMM profile for each assembly created by SNAP. In the next round, MAKER2 was used to annotate the *K. iwatai* and *P. hydriforme* genomes by using EST evidence, protein evidence for each species, and the species-specific HMM files generated in the training round. The number of genes found by MAKER2 and other annotation statistics were tabulated using the gene-stats function of the SNAP package (Table 2). Mean intron and exon sizes for *H. magnipapillata* and *N. vectensis* were calculated from the annotated scaffolds from the Joint Genome Institute. An independent estimate of intron size for *K. iwatai* was performed by mapping the RNA contigs onto the DNA contigs (a total of 23,393 introns were evaluated). This method revealed a similar mean intron length estimate (i.e., 85.4 bp).

Gene Enrichment Analysis of Predicted Genes from Annotated Genomes. The gene contigs longer than 300 bp, predicted by MAKER2, for the *K. iwatai* and *P. hydriforme* assembly were annotated using the Trinotate pipeline. The GO terms provided by the Trinotate annotation (68) were then analyzed and compared with those assigned to the transcriptome of *K. iwatai* and *P. hydriforme* and to the *H. magnipapillata* and *N. vectensis* protein sequences, as described for the gene enrichment analysis of transcriptomes.

OrthoMCL Analysis. The OrthoMCL database (42) was used to determine the number of orthologous groups identified in the transcriptome assemblies of *K. iwatai* and *P. hydriforme*, compared with published predicted proteins in *H. magnipapillata* and *N. vectensis*. ORFs were predicted and sequences translated using the OrfPredictor server (72). Predicted proteins for *H. magnipapillata* were downloaded from NCBI. Each protein FASTA file was uploaded to the OrthoMCL Groups web server (www.orthomcl.org/orthomcl/proteomeUpload.do). OrthoMCL analysis results in lists of OrthoGroup assignments for each protein in an assembly. These were parsed using R to create lists of unique ortholog group IDs (OGs) found in each assembly. These lists were used as input for the jvenn web-server (bioinfo.genotoul.fr/jvenn/example.html) (73), which calculated the overlaps between all combinations of the lists of OGs and created a four-way Venn Diagram to visualize these overlaps (Fig. S3).

Analysis of Gene Pathways and Candidate Genes. For each candidate gene, a sequence from *H. magnipapillata* (Dataset S2) was used as a query for performing a tblastx search (e-value cutoff: 1e-03) against the genome and transcriptome assemblies. To confirm their cnidarian identity, assembly sequences with significant hits were then BLAST-searched against the NCBI NR sequence database using the blastx algorithm. We also assessed the completeness of candidate signaling pathways in our assemblies using KEGG. Genomic and transcriptomic materials were combined into one file per species and sent through the KEGG Automatic Annotation Server (KAAS) for ortholog assignment and pathway mapping (www.genome.jp/tools/kaas/). The KAAS assigned KEGG orthology (KO) terms for each species dataset using the single-directional best-hit method against a representative eukaryotic dataset. After assignment of KO terms, completeness of the candidate pathways compared with their canonical pathway was assessed and visualized in each species using the KEGG Mapper tool (www.genome.jp/kegg/tool/map_pathway1.html).

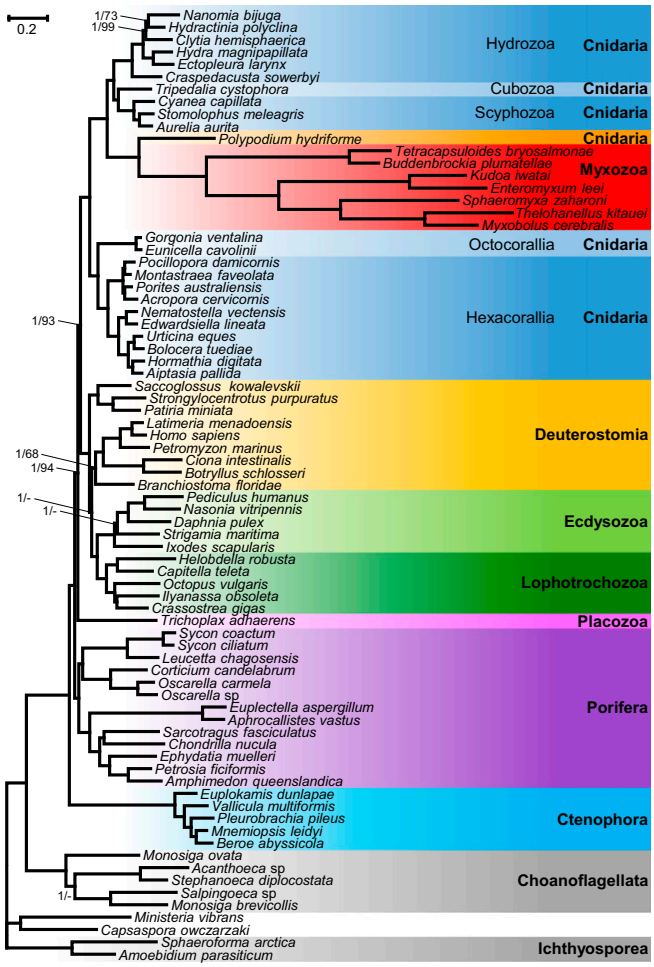


Fig. S1. Phylogenetic tree generated from a matrix of 41,237 amino acid positions, which excludes ribosomal genes, and 77 taxa using Bayesian inference under the CAT model. Support values are indicated only for nodes that did not receive maximal support. Bayesian posterior probabilities/ML bootstrap supports under the PROTGAMMAGTR are given near the corresponding node. A minus sign (''–'') indicates that the corresponding node is absent from the ML bootstrap consensus tree.

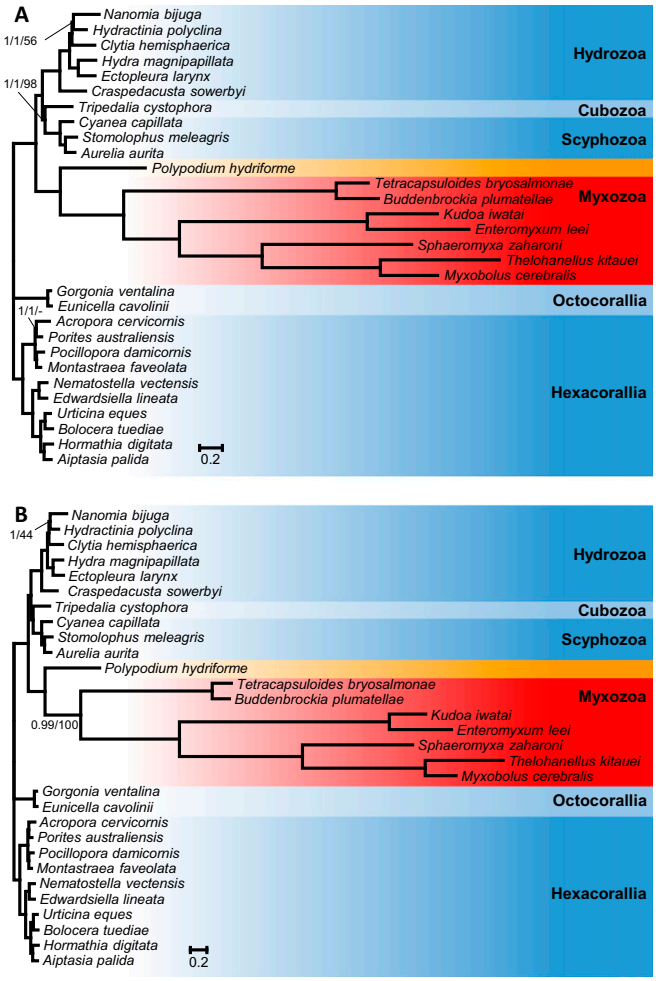


Fig. S2. (*A*) Phylogenetic tree generated from a matrix of 51,940 amino acid sequences and 30 cnidarian taxa using Bayesian inference under the CAT model. Support values are indicated only for nodes that did not receive maximal support. Bayesian posterior probabilities under the CAT model/Bayesian posterior probabilities under the CAT-GTR model/ML bootstrap supports under the PROTGAMMAGTR are given near the corresponding node. A minus sign ("–") indicates that the corresponding node is absent from the ML bootstrap consensus tree. (*B*) Phylogenetic tree generated from a matrix that excludes ribosomal genes, comprising 41,237 amino acid sequences and 30 cnidarian taxa using Bayesian inference under the CAT model. Support values are indicated only for nodes that did not receive maximal support. Bayesian posterior probabilities/ML bootstrap supports under the PROTGAMMAGTR are given near the corresponding node.

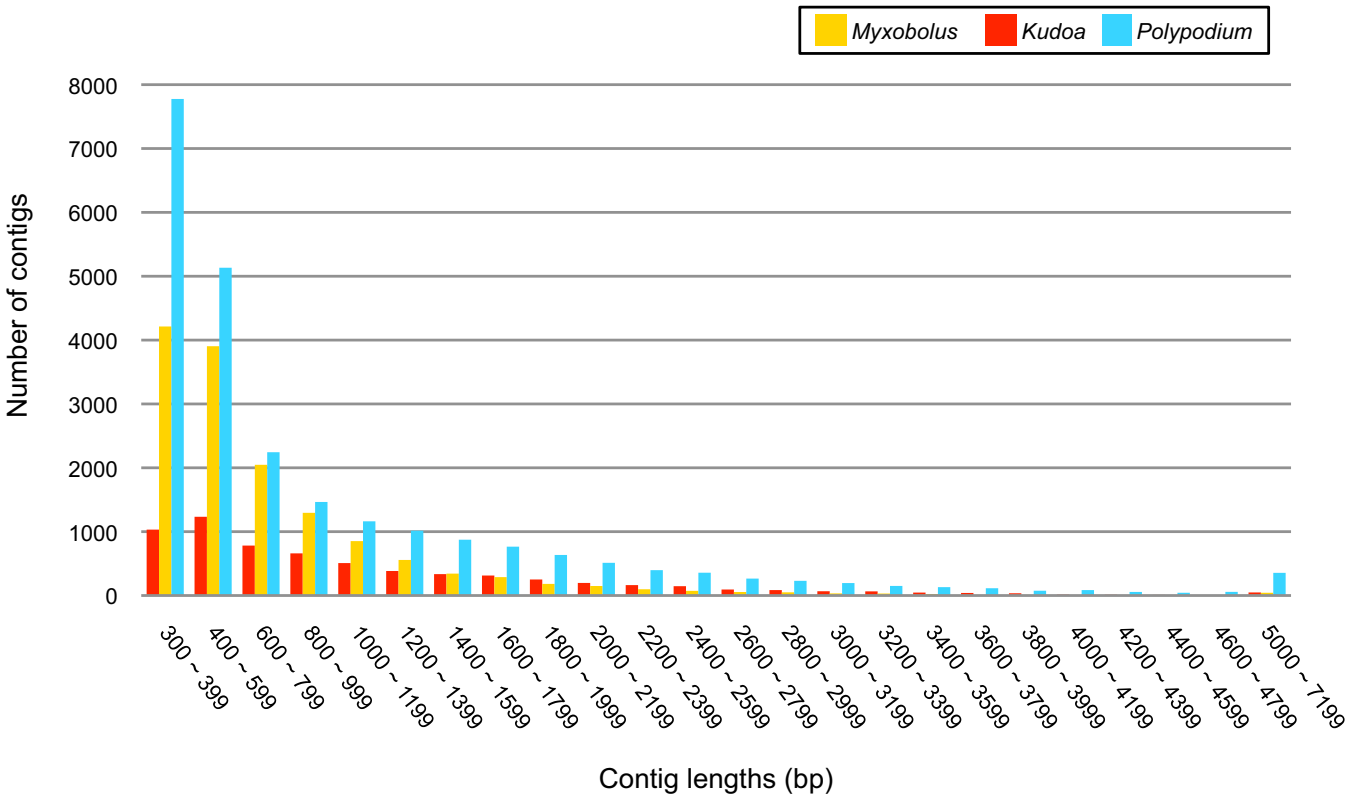


Fig. S3. Sequence size distribution of the assembled transcriptome sequences.

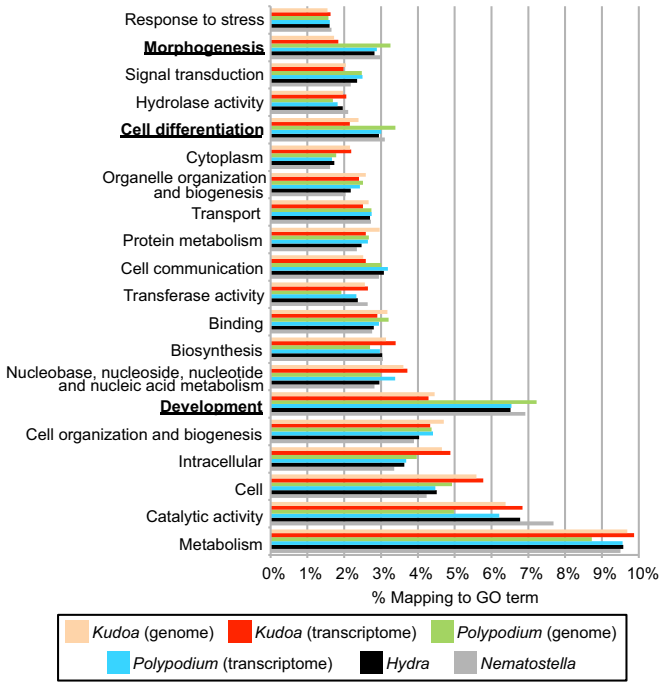


Fig. S4. GO annotation of unigenes in genomes and transcriptomes. The top 20 GO categories are shown as a percentage of total GO terms from the assemblies of *K. iwatai*, *P. hydriforme*, and the published protein sequences of *H. magnipapillata* and *N. vectensis*. Categories for which *K. iwatai* presents significantly fewer GO terms than other cnidarians are indicated in boldface type.

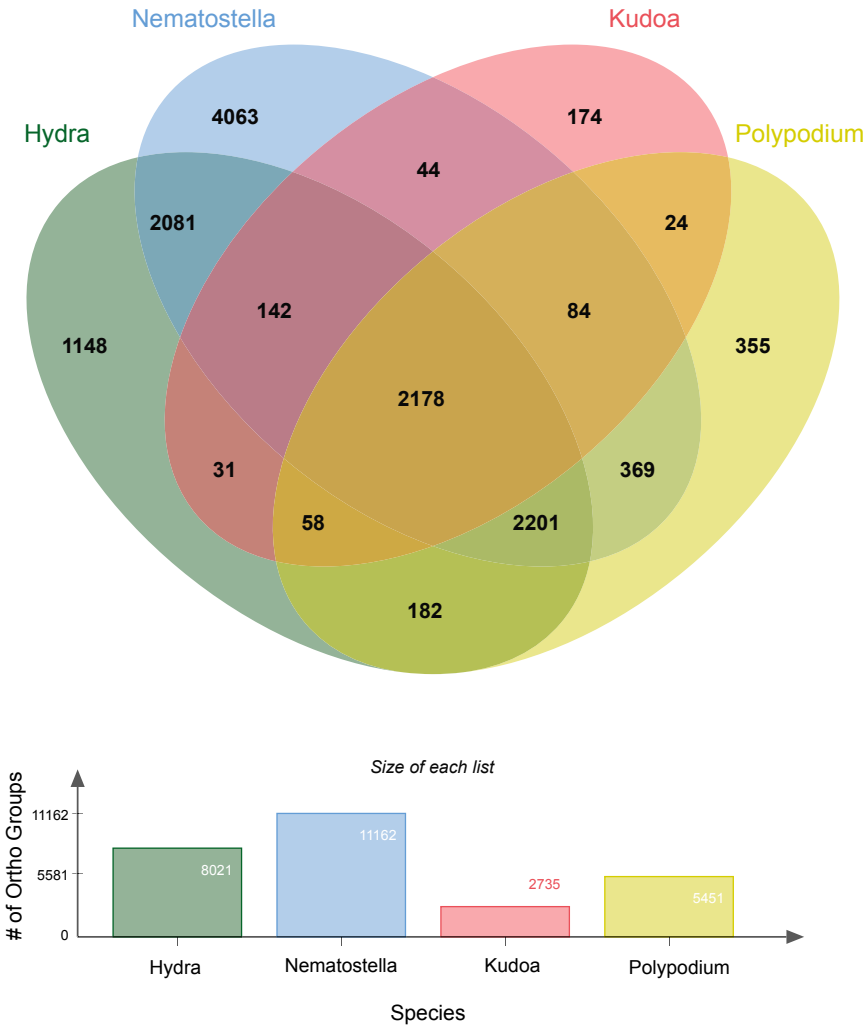


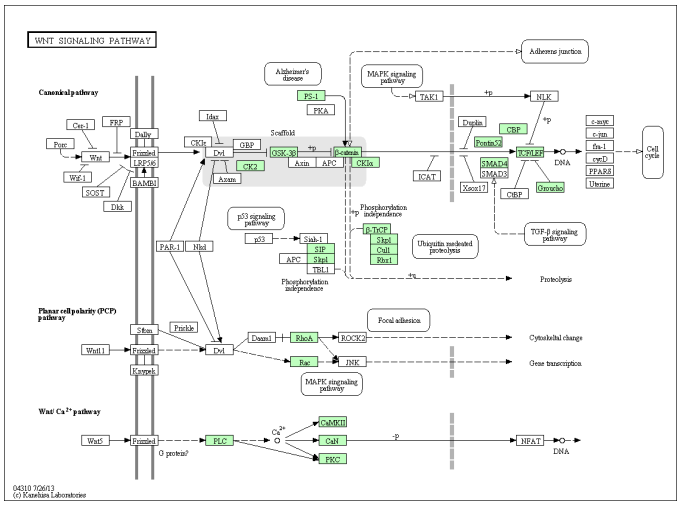
Fig. S5. Comparison of OGs in myxozoan and other cnidarian transcriptomes. VENN diagram comparing OGs for the OrthoMCL database from transcriptome assemblies of *K. iwatai* and *P. hydriforme* and published predicted proteins in *H. magnipapillata* and *N. vectensis*.

Table S1. Sample information and accession numbers

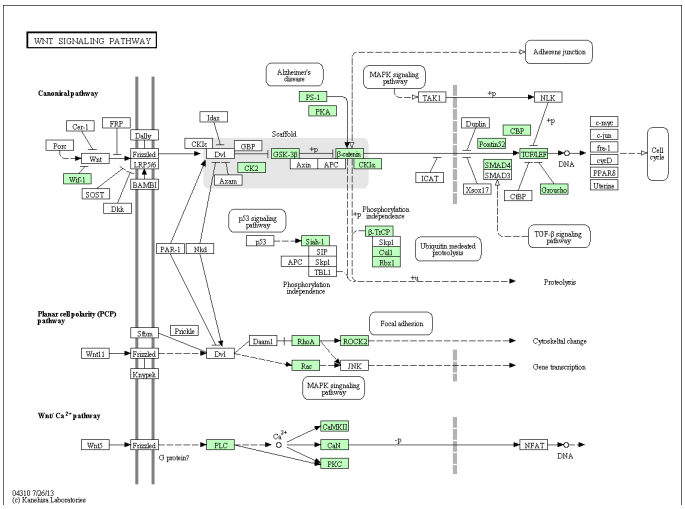
Species	DNA/RNA	Platform	Reads (bp PE)	BioProject	BioSample	SRA experiments	Genome/transcriptome shotgun assembly	Locality
K. iwatai	DNA	G2x	76	PRJNA261422	SAMN03068681	SRX704259	JRUY00000000	Red Sea, Eilat
K. iwatai	DNA	HiSEq. 2000	100	PRJNA261052	SAMN03068681	SRX702459	JRUX00000000	Red Sea, Eilat
K. iwatai	RNA	HiSEq. 2000	100	PRJNA248713	SAMN02800925	SRX554567	GBGI00000000	Red Sea, Eilat
E. leei	DNA	HiSEq. 2000	100	PRJNA284325	SAMN03701405	SRX1034928	LDNA00000000	Red Sea, Eilat
S. zaharoni	DNA	HiSEq. 2000	100	PRJNA284326	SAMN03701400	SRX1034914	LDMZ00000000	Red Sea, Eilat
M. cerebralis	RNA	HiSEq. 2000	100	PRJNA258474	SAMN02998096		GBKL00000000	California
P. hydriforme	RNA	HiSEq. 2000	100	PRJNA251648	SAMN02837860	SRX570527	GBGH00000000	Grand Lake State Park, OK
P. hydriforme	DNA	HiSEq. 2000	100	PRJNA259515	SAMN02998112	SRX687102		Grand Lake State Park, OK

Other Supporting Information Files

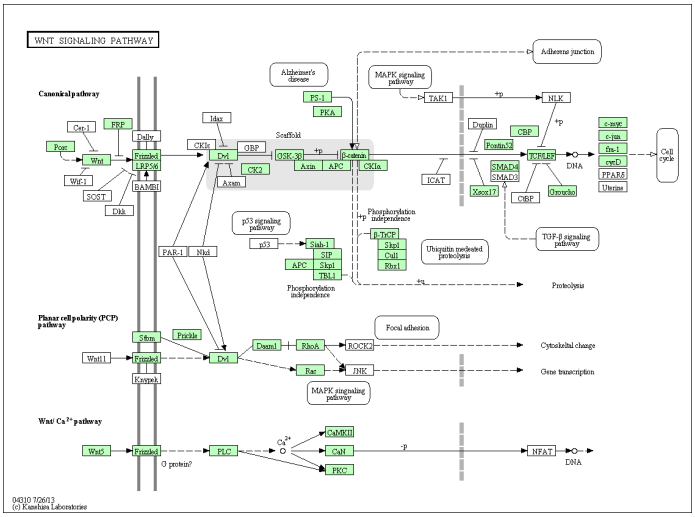
Dataset S1 (XLSX)
Dataset S2 (XLSX)
Dataset S3 (PDF)
Dataset S4 (XLSX)



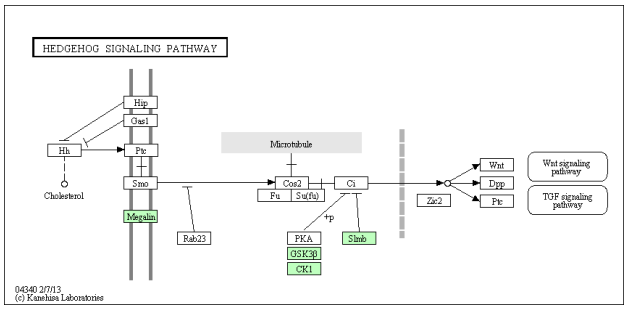
Wnt pathway components identified in the genome and transcriptome of *Kudoa iwatai* through assignment of KEGG orthology terms. Components identified in this species are highlighted in green.



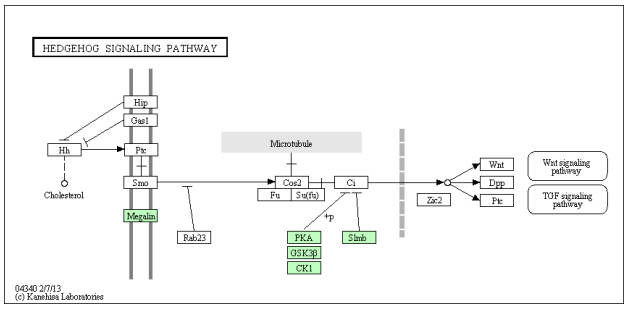
Wnt pathway components identified in the genome and transcriptome of *Myxobolus cerebralis* through assignment of KEGG orthology terms. Components identified in this species are highlighted in green



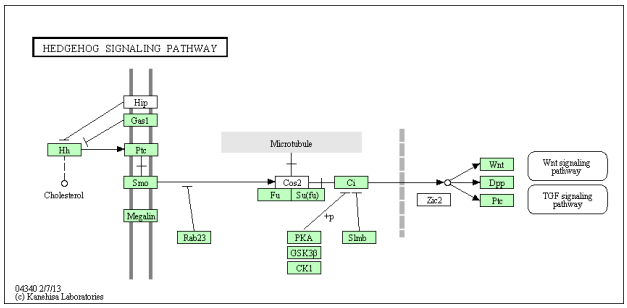
Wnt pathway components identified in the genome and transcriptome of *Polypodium hydriforme* through assignment of KEGG orthology terms. Components identified in this species are highlighted in green.



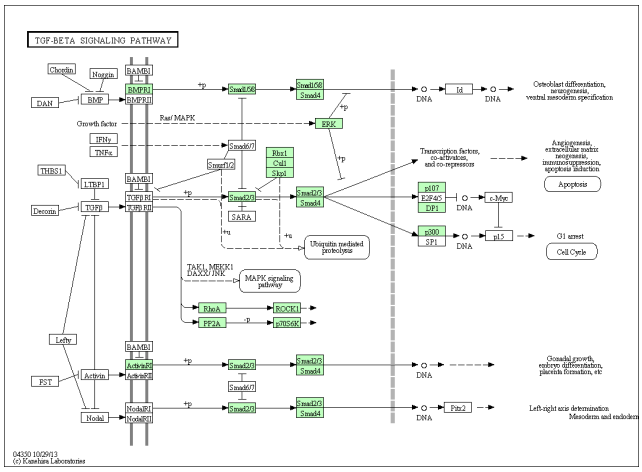
Hedgehog signaling pathway components identified in the genome and transcriptome of *Kudoa iwatai* through assignment of KEGG orthology terms. Components identified in this species are highlighted in green.



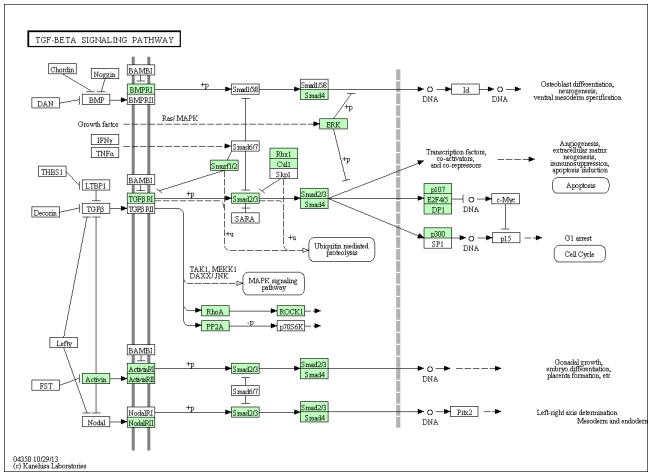
Hedgehog signaling pathway components identified in the genome and transcriptome of *Myxobolus cerebralis* through assignment of KEGG orthology terms. Components identified in this species are highlighted in green.



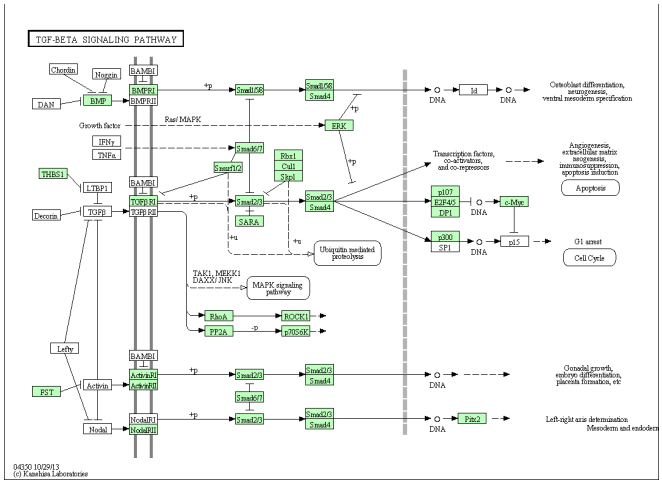
Hedgehog signaling pathway components identified in the genome and transcriptome of *Polypodium hydriforme* through assignment of KEGG orthology terms. Components identified in this species are highlighted in green.



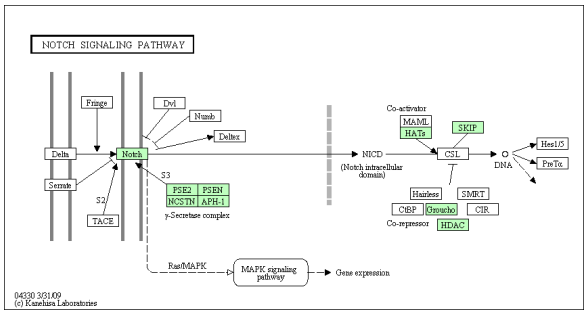
TGF-β signaling pathway components identified in the genome and transcriptome of *Kudoa iwatai* through assignment of KEGG orthology terms. Components identified in this species are highlighted in green.



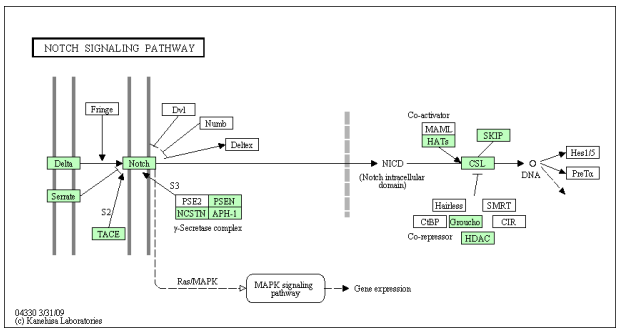
TGF-β signaling pathway components identified in the genome and transcriptome of *Myxobolus cerebralis* through assignment of KEGG orthology terms. Components identified in this species are highlighted in green.



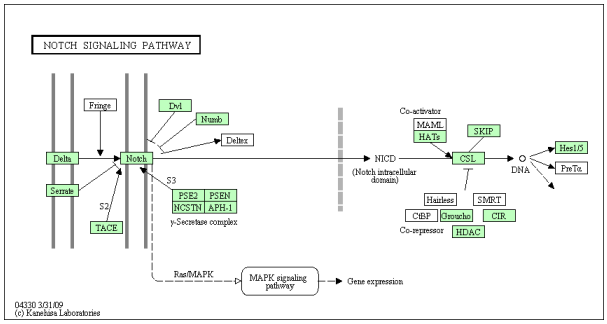
TGF-β signaling pathway components identified in the genome and transcriptome of *Polypodium hydriforme* through assignment of KEGG orthology terms. Components identified in this species are highlighted in green.



Notch signaling pathway components identified in the genome and transcriptome of *Kudoa iwatai* through assignment of KEGG orthology terms. Components identified in this species are highlighted in green.



Notch signaling pathway components identified in the genome and transcriptome of *Myxobolus cerebralis* through assignment of KEGG orthology terms. Components identified in this species are highlighted in green.



Notch signaling pathway components identified in the genome and transcriptome of *Myxobolus cerebralis* through assignment of KEGG orthology terms. Components identified in this species are highlighted in green.