The jellyfish genome sheds light on the early evolution of active predation

Hak-Min Kim[1,2†], Jessica A. Weber[3,4†], Nayoung Lee[5†], Seung Gu Park[1], Yun Sung Cho[1,2,6,7], Youngjune Bhak[1,2], Nayun Lee[5], Yeonsu Jeon[1,2], Sungwon Jeon[1,2], Victor Luria[8], Amir Karger[9], Marc W. Kirschner[8], Ye Jin Jo[5], Seonock Woo[10,11], Kyoungsoon Shin[12], Oksung Chung[6,7], Jae-Chun Ryu[13], Hyung-Soon Yim[11], Jung-Hyun Lee[11], Jeremy S. Edwards[14], Andrea Manica[15], Jong Bhak[1,2,6,7*], and Seungshic Yum[5,10*]

[1]Korean Genomics Industrialization and Commercialization Center (KOGIC), Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea.

[2]Department of Biomedical Engineering, School of Life Sciences, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea.

[3]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

[4]Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA.

[5]South Sea Environment Research Center, Korea Institute of Ocean Science and Technology (KIOST), Geoje 53201, Republic of Korea.

[6]Personal Genomics Institute, Genome Research Foundation, Cheongju 28160, Republic of Korea.

[7]Clinomics Inc., Ulsan 44919, Republic of Korea.

[8]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

[9]IT - Research Computing, Harvard Medical School, Boston, MA 02115, USA.

[10]Faculty of Marine Environmental Science, University of Science and Technology (UST), Geoje 53201, Republic of Korea.

[11]Marine Biotechnology Research Center, Korea Institute of Ocean Science and Technology (KIOST), Busan 49111, Republic of Korea.

[12]Ballast Water Center, Korea Institute of Ocean Science and Technology (KIOST), Geoje 53201, Republic of Korea.

[13]Cellular and Molecular Toxicology Laboratory, Center for Environment, Health and Welfare Research, Korea Institute of Science and Technology (KIST), Seoul 02792, Republic of Korea.

[14]Chemistry and Chemical Biology, UNM Comprehensive Cancer Center, University of New Mexico, Albuquerque, NM 87131, USA.

[15]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

31 *Correspondence: jongbhak@gmail.com; syum@kiost.ac.kr

32 †Equal contributors

33

34 Email addresses:

35 HMK: howmany2@gmail.com

36 JAW: weberj.unm@gmail.com

37 NayoungL: dylee@kiost.ac.kr

38 SGP: seung9park@gmail.com

39 YSC: joys0406@gmail.com

40 YB: youngjune29bhak@gmail.com

41 NayunL: aklee@kiost.ac.kr

42 YJ: brain0106@gmail.com

43 SJ: jsw0061@gmail.com

44 VL: Victor_Luria@hms.harvard.edu

45 AK: amir_karger@hms.harvard.edu

46 MWK: marc@hms.harvard.edu

47 YJJ: ye9302@kiost.ac.kr

48 SW: cwoo@kiost.ac.kr

49 KS: ksshin@kiost.ac.kr

50 OC: okokookk219@gmail.com

51 JCR: ryujc@kist.re.kr

52 HSY: yimh@kiost.ac.kr

53 JHL: jlee@kiost.ac.kr

54 JSE: jeremy.scott.edwards@gmail.com

55 AM: am315@cam.ac.uk

JB: jongbhak@gmail.com

SY: syum@kiost.ac.kr

Abstract

Background: Unique among cnidarians, jellyfish have remarkable morphological and biochemical innovations that allow them to actively hunt in the water column. One of the first animals to become free-swimming, jellyfish employ pulsed jet propulsion and venomous tentacles to capture prey.

Results: To understand these key innovations, we sequenced the genome of the giant Nomura's jellyfish (*Nemopilema nomurai*), the transcriptomes of its bell and tentacles, and transcriptomes across tissues and developmental stages of the *Sanderia malayensis* jellyfish. Analyses of *Nemopilema* and other cnidarian genomes revealed adaptations associated with swimming, marked by codon bias in muscle contraction and expansion of neurotransmitter genes, along with expanded Myosin type II family and venom domains; possibly contributing to jellyfish mobility and active predation. We also identified gene family expansions of *Wnt* and posterior *Hox* genes, and discovered the important role of retinoic acid signaling in this ancient lineage of metazoans, which together may be related to the unique jellyfish body plan (medusa formation).

Conclusions: Taken together, the jellyfish genome and transcriptomes genetically confirm their unique morphological and physiological traits that have combined to make these animals one of the world's earliest and most successful multi-cellular predators.

Keywords: Jellyfish mobility, Medusa structure formation, Scyphozoa, *de novo* genome assembly.

Background

Cnidarians, including jellyfish and their predominantly sessile relatives the coral, sea anemone, and hydra, first appeared in the Precambrian Era and are now key members of aquatic ecosystems worldwide [1]. Between 500 and 700 million years ago, jellyfish developed novel physiological traits that allowed them to become one of the first free-swimming predators. The life cycle of the jellyfish includes a small polypoid, sessile stage which reproduces asexually to form the mobile medusa form that can reproduce both sexually and asexually [2]. The class Scyphozoa, or true jellyfish, are characterized by a predominant medusa life-stage consisting of a bell and venomous tentacles used for hunting and defense [3]. Jellyfish medusae feature a radially symmetric body structure, powered by readily identifiable cell types such as motor neurons and striated muscles that expand and contract to create the most energy-efficient swimming method in the animal kingdom [4, 5]. Over 95% water, jellyfish are osmoconformers that use ion gradients to deliver solutes to cells and tissues where sodium and calcium ions activate the muscle contractions that power their propulsion. Notably, many jellyfish species can survive in habitats with varying levels of salinity and are successful in low-oxygen environments, allowing them to bloom even in dead zones [6]. These innovations have allowed them to colonize aquatic habitats across the globe both in brackish and marine environments, spanning the shallow surface waters to the depths of the seas.

Results and discussion

Jellyfish genome assembly and annotation

Here, we present the first *de novo* genome assembly of a jellyfish (*Nemopilema nomurai*). It resulted in a 213 Mb genome comprised of 255 scaffolds and an N50 length of 2.71 Mb,

containing only 1.48 % gaps (Additional file 1: Tables S2 and S3). The *Nemopilema* hybrid assembly was created using a combination of short and long read sequencing technologies, consisting of 38.2 Gb Pacific Biosciences (PacBio) single molecule real time sequencing (SMRT) reads, along with 98.6 Gb of Illumina short insert, mate-pair, and TruSeq synthetic long reads (Additional file 1: Figures S3-S5; Tables S4-S7). The resulting assembly shows the longest continuity among cnidarian genomes (Additional file 1: Table S9). We predicted 18,962 protein-coding jellyfish genes by combining *de novo* (using medusa bell and tentacle tissue transcriptomes) and homologous gene prediction methods (Additional file 1: Tables S10 and S11). This process recovered the highest number of single-copy orthologous genes [7] among all published non-bilaterian metazoan genome assemblies to date (Additional file 1: Table S12). A total of 21.07% of the jellyfish genome was found to be made up of transposable elements, compared to those of *Acropora digitifera* (9.45%), *Nematostella vectensis* (33.63%), and *Hydra magnipapillata* (42.87%) (Additional file 1: Table S13).

We compared the *Nemopilema* genome to other cnidarian genomes, all of which are from predominantly sessile taxa, to detect unique Scyphozoa function (active mobility), physical structure (medusa bell), and chemistry (venom). We also performed transcriptome analyses of both *Nemopilema nomurai* and the *Sanderia malayensis* jellyfish across three medusa tissue types and four developmental stages.

Evolutionary analysis of the jellyfish

To identify jellyfish-specific evolutionary traits, we examined gene family expansions and contractions across one unicellular holozoan and eleven metazoans using 15,255 orthologous

gene families (see Additional file 1: Section 4.1). Of these, 7,737 were found in *Nemopilema* and 4,156 were shared by all four available cnidarian genomes (*Nemopilema nomurai*, *Hydra magnipapillata* [8], *Acropora digitifera* [9], and *Nematostella vectensis* [10]; Fig. 1a). A phylogeny constructed using these orthologs revealed a monophyletic cnidarian clade that diverged from the metazoan stem prior to the evolution of the bilaterians (Fig. 1b; Additional file 1: Figure S7). To determine how many genes appeared in every evolutionary era in the genome of Nomura's jellyfish, we also evaluated the evolutionary age of the protein-coding genes. Grouping jellyfish genes into 3 broad evolutionary eras, we observed that while the majority (80%) of genes are ancient (older than 741 Mya), a few (~3%) are of an intermediate age (741 - 239 Mya) and some (17%) are young (239 Mya to present; Fig. 1c; Additional file 1: Figure S10). Interestingly, normalizing the number of genes by the age and length of evolutionary era suggests that gene turnover is highest near the present time. In total, the *Nemopilema* genome contained 67 expanded and 80 contracted gene families compared to the common ancestor of *Nemopilema* and *Hydra* (Fig. 1b; see Additional file 1: Section 4.2). Gene Ontology (GO) terms related to sensory perception were under-represented in the Cnidaria lineage compared to Bilateria, accurately reflecting cnidarian's less complex sensory system (Additional file 1: Tables S14 and S15). However, neurotransmitter transport function (GO:0005326, P = 1.66E-16) was significantly enriched in *Nemopilema* compared to other cnidarians (Additional file 1: Tables S16 and S17), likely due to the balance and visual structures, such as the statocyst and ocelli, that are more elaborate in the mobile medusa than in sessile polyps [11].

Genomic context and muscle associated genes

Jellyfish have two primary muscle types: the epitheliomuscular cells, which are the predominant muscle cells found in sessile cnidarians; and the striated muscle cells located in the medusa bell that are essential for swimming. To understand the evolution of active-swimming in jellyfish, we examined their codon bias compared to other metazoans by calculating the guanine and cytosine content at the third codon position (GC3) [12, 13] (Additional file 1: Figure S13). It has been suggested that genes with high level of GC3 are more adaptable to external stresses (e.g., environmental changes) [14]. Among the high-scoring top 100 GC3 biased genes, the regulation of muscle contraction and neuropeptide signaling pathways GO terms were specific to *Nemopilema* (Additional file 4). Calcium plays a key role in the striated muscle contraction in jellyfish, and the calcium signaling pathway (GO:0004020, P = 5.60E-10) showed a high level of GC3 biases specific to *Nemopilema*. *Nemopilema* top 500 GC3 genes were enriched in GO terms associated with homeostasis (e.g. cellular chemical homeostasis and sodium ion transport), which we speculate is essential for the activation of muscle contractions that power the jellyfish's mobile predation (see Additional file 1: Section 5.1).

Since cnidarians have been reported to lack titin and troponin complexes, which are critical components of bilaterian striated muscles, it has been suggested that the two clades independently evolved striated muscles [15]. A survey of genes that encode muscle structural and regulatory proteins in cnidarians showed a conserved eumetazoan core actin-myosin contractile machinery shared with bilaterians (Additional file 1: Table S23). However, like other cnidarians, *Nemopilema* lacks titin and troponin complexes, which are key components of bilaterian striated muscles. Also, γ-syntrophin, a component of the dystroglycan complex, was absent in both *Nemopilema* and *Hydra*. However, *Nemopilema* do possess α/β-Dystrobrevin and α/ε-Sarcoglycan dystroglycan-associated costamere proteins, indicating that several components

of the dystroglycan complex were lost after the Scyphozoa-Hydrozoa split. It was suggested that *Hydra* undergone secondary simplifications relative to *Nematostella*, which has a greater degree of muscle-cell-type specialization [8]. Compared to *Hydra* and *Nematostella*, *Nemopilema* shows intermediate complexity of muscle structural and regulatory proteins between *Hydra* and *Nematostella*.

Medusa bell and tentacle transcriptome profiling

Jellyfish medusa bell and tentacles are morphologically distinct and perform discrete physiological functions [16, 17]. We generated bell and tentacle transcriptomes from *Nemopilema* and the smaller *Sanderia malayensis*, which can be grown in the laboratory, to assess developmental regulation (Additional file 1: Table S20). Enrichment tests of highly expressed genes showed that muscle-associated functional categories (e.g. muscle myosin complex and muscle tissue morphogenesis) were enriched in the bell (Fig. 2a; see Additional file 5). Myosins comprise a superfamily of motor proteins and play a critical role in muscle contraction and are involved in a wide range of motility processes in Eukaryotes. Critically, the Myosin II family proteins, found in cells of both striated muscle tissue and smooth muscle tissue, are responsible for producing contraction in muscle cells [18]. Cnidarians possess both epitheliomuscular cells and striated muscle cells. Striated muscle is a critical component of the subumbrella of the medusa bell, where its fast contractions power the unique propulsion-based swimming of the jellyfish. We found that type II Myosin heavy chain (MYH) and Myosin light chain (MYL) gene families were highly expressed in the bell, and are closely associated with striated and smooth muscle cells [15]. Interestingly, *Nemopilema* also showed the largest copy

numbers of MYH and MYL genes among non-bilaterian metazoans (Fig. 2c; see Additional file 1: Section 5.3), and six of the seven MYH genes and 12 out of 21 MYL genes showed higher expression in the bell than the tentacles with very high ~8.8 and ~17-fold increase on average, respectively (Fig. 2d). These results suggest that the combinations of copy number expansion of type II Myosin gene families and high expression of muscle associated genes confirmed that muscles in medusa bell are an important determinant of jellyfish motility.

Conversely, gene expression analyses of the tentacles revealed high RNA expression levels of neurotransmitter associated functional categories (ion channel complex, postsynapse, and neurotransmitter receptor activity; Fig. 2b); consistent with the anatomy of jellyfish tentacles, which contain the sensory cells and a loose plexus of the neuronal subpopulation at the base of the ectoderm [19].

Body patterning in the jellyfish

There has been much debate surrounding the early evolution of body patterning in the metazoan common ancestor, particularly concerning the origin and expansion of Hox and *Wnt* gene families [20-22]. In total, 83 homeodomains were found in *Nemopilema*, while 41, 120, and 148 of homeodomains were found from *Hydra*, *Acropora*, and *Nematostella*, respectively (Additional file 1: Table S24). Five of the eight Hox genes in *Nemopilema* are of the posterior type that are associated with aboral axis development [22] and clustered with *Nematostella's* posterior Hox genes, *HOXE* and *HOXF* (Additional file 1: Figures S18-S20). Though absent in *Hydra* and *Acropora*, synteny analyses of ParaHox genes in *Nemopilema* show that the *XLOX/CDX* gene is located immediately downstream of *GSX* in the same tandem orientation as those in

Nematostella, suggesting that *XLOX/CDX* was present in the cnidarian common ancestor and subsequently lost in some lineages (Additional file 1: Figure S21). Hox related genes, *EVX* and *EMX*, are also present in *Nemopilema*, although they are absent in *Hydra*. Given the large amount of ancestral diversity in the *Wnt* genes, it has been proposed that Wnt signaling controlled body plan development in the early metazoans [23]. *Nemopilema* possesses 13 *Wnt* orthologs representing 10 *Wnt* subfamilies (Additional file 1: Figure S22; Table S25). *Wnt9* is absent from all cnidarians, likely representing losses in the cnidarian common ancestor. Cnidarians have undergone dynamic lineage specific *Wnt* subfamily duplications, such as *Wnt8* (*Nematostella* and *Acropora*), *Wnt10* (*Hydra*), and *Wnt11*, and *Wnt16* (*Nemopilema*). It has been proposed that a common cluster of *Wnt* genes (*Wnt1–Wnt6–Wnt10*) existed in the last common ancestor of arthropods and deuterostomes [24]. Our analyses of cnidarian and bilaterian genomes revealed that *Acropora* also possess this cluster, while *Nemopilema* and *Hydra* are missing *Wnt6*, suggesting loss of the *Wnt6* gene in the Medusozoa common ancestor (Additional file 1: Figure S23). Taken together, the jellyfish have comparable number of Hox and *Wnt* genes to other cnidarians, but the dynamic repertoire of these gene families suggests that cnidarians have evolved independently to adapt their physiological characteristics and life cycle.

Polyp to medusa transition in jellyfish

The polyp-to-medusa transition is prominent in jellyfish compared to the other sessile cnidarians. To understand the genetic basis of the medusa structure formation in the jellyfish, we compared transcriptional regulation between cnidarians and across jellyfish developmental stages (see Additional file 1: Sections 7.1 and 7.2). We assembled the *Sanderia* transcripts using six pooled samples of transcriptomes (Additional file 1: Table S26). The assembled transcripts had a total

length of 61 Mb and resulted in 58,290 transcript isoforms and 43,541 unique transcripts, with a N50 of 2,325 bp. On average, 87% of the RNA reads were aligned to into the assembled transcripts (Additional file 1: Table S27), indicating that the transcript assembly represented the majority of sequenced reads. Furthermore, the composition of the protein domains contained in the top 20 ranks was quite similar between *Nemopilema* and *Sanderia* (Additional file 1: Table S28). To obtain differentially expressed genes for each stage, we compared each stage with the previous or next stage in the life cycle of the jellyfish. The polyp stage, which represents a sessile stage in the jellyfish life cycle, showed enriched terms related to ion channel activity and energy metabolism (regulation of metabolic process, and amino sugar metabolic process; Additional file 1: Table S29). Active feeding in the polyp stimulates asexual proliferation either into more polyps or metamorphosis to strobila [25]. Since anthozoans do not form a medusa, the strobila asexual reproductive stage is an important stage in which to study the metamorphosis from polyp to medusa. In this stage, GO terms related to amide biosynthetic and metabolic process were highly expressed compared to the polyp stage (Additional file 1: Table S30). It has been reported that RF-amide and LW-amide neuropeptides were associated with metamorphosis in cnidarians [26-28]. However, we could not confirm this finding in our strobila and ephyra stage comparisons. In our system, the gene expression patterns of the two stages are quite similar. In the ephyra, the released mobile stage, GO terms involving amide biosynthetic and metabolic process were also highly expressed compared to the merged medusa stage (Additional file 1: Table S31). In the medusa, extracellular matrix, metallopeptidase activity, and immune system process terms were enriched (Additional file 1: Table S32), consistent with the physiology of their bell, tentacles, and oral arm tissue types.

Polyp-to-medusa metamorphosis was previously shown to be strongly associated with *CL390* and *retinoid X receptor* (*RXR*) genes in the *Aurelia aurita* jellyfish [29]. Interestingly, *CL390* was not found in *Nemopilema* or other published cnidarians, suggesting that it may be an *Aurelia*-specific strobilation inducer gene. However, we confirm that *RXR* is present in *Nemopilema*, and absent from cnidarians without a prominent medusa stage (Additional file 1: Figure S24). Retinoic acid (RA) signaling plays a central role during vertebrate growth and development [30], where it regulates transcription by interacting with the RA receptor (RAR) bound to RA response elements (RAREs) of nearby target genes [31]. Of the genes in the RA signaling pathway, *Nemopilema* possess ADH and RALDH enzymes that metabolize retinol to RA, and *RXR* and RAREs to activate transcription of the target gene (Fig. 3a). We discovered 1,630 *Nemopilema* RAREs regions with an average distance of 13 Kbp to the nearest gene (Fig. 3b; Additional file 1: Tables S33 and S34). Interestingly, four posterior Hox genes of *Nemopilema* were located within ±10 Kbp from RAREs, which is unique among the non-bilaterian metazoans (Fig. 3c). Together these findings suggest that retinoic acid signaling was present in early metazoans for regulating target genes with *RXR* and RAREs, and that *RXR* and RAREs may play a critical role for polyp-to-medusa metamorphosis [29]

Identification of toxin related domains in jellyfish

Jellyfish produce complex mixtures of proteinaceous venoms for active prey capture and defense [32]. We identified abundant toxin domains in *Nemopilema* when compared to the non-bilaterian metazoan gene sets in the Tox-Prot database [33]. In total, 69 out of 136 toxin domains aligned to non-bilaterian metazoans; of these 69 toxin domains, 53 were found in *Nemopilema* (Additional file 1: Table S35). Expectedly, the *Nemopilema* genome contains the largest number

of venom or toxin associated domains of the included non-bilaterian metazoans. These domains include Reprolysin (M12B) family zinc metalloprotease (PF01421), Kazal-type serine protease inhibitor domain (PF07648), phospholipase A_2 (PF05826), and ShK domain-like (PF01549) domains (Fig. 4). Compared to the common ancestor of *Nemopilema* and *Hydra*, *Nemopilema* showed expanded gene families associated with metallopeptidase activities (GO:0008237, P = 1.99E-16). In particular, Reprolysin (M12B) family zinc metalloproteases are enzymes that cleave peptides and comprise most snake venom endopeptidases [34]. Furthermore, it has been reported that serine protease inhibitor and ShK domains were abundantly found in the transcriptomes of both the cannonball jellyfish (*Stomolophus meleagris*), and the box jellyfish (*Chironex fleckeri*)[35, 36], and phospholipase A_2 is well-characterized toxin-related enzyme, which is critical to the production of venom components, found in the class Scyphozoa [37].

Conclusions

A unique branch on the tree of life, jellyfish have evolved remarkable morphological and biochemical innovations that allow them to actively hunt using pulsed jet propulsion and venomous tentacles. While the expansion and contraction of distinct families reflect the adaptation to salinity and predation and the convergent evolution of muscle elements, the *Nemopilema* genome strikes a balance between the conservation of many ancient genes and an innovative potential reflected in significant number of new genes that appeared since *Rhizostomeae* emerged. The *Nemopilema nomurai* genome has provided clues to the genetic basis of the innovative structure, function, and chemistry that have allowed this distinctive early group of predators to colonize the waters of the globe.

304

Methods

Sample preparation

A medusa *Nemopilema nomurai* was collected at the Tongyeong Marine Science Station, KIOST (34.7699 N, 128.3828 E) on Sep. 12, 2013. The *Sanderia malayensis* samples were obtained from Aqua Planet Jeju Hanwha (Seogwipo, Korea) for transcriptome analyses of developmental stages since *Nemopilema* cannot be easily grown in the laboratory. The DNA and RNA preparation of *Nemopilema* and *Sanderia* are described in the Additional file 1: Section 1.1. Species identification of *Nemopilema* was confirmed by comparing the *MT-COI* gene of five species of jellyfish. We aligned *Nemopilema* Illumina short reads (~400 bp insert-size) to the *MT-COI* gene of *Chrysaora quinquecirrha* (NC_020459.1), *Cassiopea frondosa* (NC_016466.1), *Craspedacusta sowerbyi* (NC_018537.1), and *Aurelia aurita* (NC_008446.1) jellyfish with BWA-MEM aligner [38]. Consensus sequences for each jellyfish were generated using SAMtools [39]. The consensus sequence from *C. sowerbyi* was excluded due to low coverage. We conducted multiple sequence alignment using MUSCLE [40] and ran the MEGA v7 [41] neighbor joining phylogenetic tree (gamma distribution) with 1,000 bootstrap replicates. Mitochondrial DNA phylogenetic analyses confirmed the identification of the *Nemopilema* sample as *Nemopilema nomurai*.

Genome sequencing and scaffold assembly

For the *de novo* assembly of *Nemopilema*, PacBio SMRT and five Illumina DNA libraries with various insert sizes (400bp, 5 Kb, 10 Kb, 15 Kb, and 20 Kb) were constructed according manufacturers' protocols. The Illumina libraries were sequenced using a HiSeq2500 with read

length of 100 bp (400 bp, 15 Kb, and 20 Kb) and a HiSeq2000 with read length of 101 bp (5 Kb and 10 Kb). Quality filtered PacBio subreads were assembled into distinct contigs using the FALCON assembler [42] with various read length cutoffs. To extend contigs to scaffolds, we aligned the Illumina long mate-pair libraries (5 Kb, 10 Kb, 15 Kb, and 20 Kb) to contig sets and extended the contigs using SSPACE [43]. Gaps generated by SSPACE were filled by aligning the Illumina short-insert paired-end sequences using GapCloser [44]. We also generated TSLRs using an Illumina HiSeq2000, which were aligned to scaffolds to correct erroneous sequences and to close gaps using an in-house script. Detailed genome sequencing and assembly process are provided in Additional file 1: Section 2.2.

Genome annotation

The jellyfish genome was annotated for protein-coding genes and repetitive elements. We predicted protein-coding genes using a two-step process, with both homology and evidence-based prediction. Protein sequences of the sea anemone, hydra, sponge, human, mouse, and fruit fly from the NCBI database, and Cnidaria protein sequences from the NCBI Entrez protein database were used for homology-based gene prediction. Two tissue transcriptomes from *Nemopilema* were used for evidence-based gene prediction via AUGUSTUS [45]. Final *Nemopilema* protein-coding genes were determined using AUGUSTUS with exon (from homology-based gene prediction) and intron (from evidence-based gene prediction) hints. Repetitive elements were also predicted using Tandem Repeats Finder [46] and RepeatMasker [47]. Details of the annotation process are provided in Additional file 1: Sections 3.1 and 3.2.

Gene age estimation

Phylostratigraphy employs BLASTP-scored sequence similarity to estimate the minimal age of every protein-coding gene. The protein sequence is used to query the NCBI non-redundant database and detect the most distant species in which a sufficiently similar sequence is present, and inferring that the gene is at least as old as the age of the common ancestor [48]. For every species, we use the NCBI taxonomy. The timing of most divergence events is estimated using TimeTree [49] and the Encyclopedia of Life [50]. To facilitate detection of sequence similarity, we use the e-value threshold of 10^{-3}. We evaluate the age of all proteins whose length is equal or greater than 40 amino acids. We count the number of genes in each phylostratum, from most ancient (PS 1) to newest (PS 11). To see broad evolutionary patterns, we aggregate the counts from several phylostrata into 3 broad evolutionary eras: ancient (PS 1-5, cellular organisms to Eumetazoa, 4,204 Mya - 741 Mya), middle (PS 6-7, Cnidaria to Scyphozoa, 741 Mya - 239 Mya) and young (PS 8-11, Rhizostomeae to *Nemopilema nomurai*, 239 Mya to present).

Comparative evolutionary analyses

Orthologous gene clusters were constructed to examine the conservation of gene repertoires among the genomes of the *Nemopilema nomurai, Hydra magnipapillata, Acropora digitifera, Nematostella vectensis, Caenorhabditis elegans, Danio rerio, Drosophila melanogaster, Homo sapiens, Trichoplax adhaerens, Amphimedon queenslandica*, *Mnemiopsis leidyi*, and *Monosiga brevicollis* using OrthoMCL [51]. To infer a phylogeny and divergence times, we used RAxML [52] and MCMCtree [53], respectively. A gene family expansion and contraction analysis was conducted using the Café program [54]. Domain regions were predicted by InterProScan [55]

with domain databases. Details of the comparative analysis are provided in Additional file 1: Sections 4.1-4.4.

Transcriptome sequencing and expression profiling

Illumina RNA libraries from *Nemopilema nomurai* and *Sanderia malayensis* were sequenced using a HiSeq2500 with 100 bp read lengths. Since there is not a reference genome for *S. malayensis*, we *de novo* assembled a pooled six RNA-seq read set using the Trinity assembler [56]. Quality filtered RNA reads from *Nemopilema* and *Sanderia* were aligned to the *Nemopilema* genome assembly and the assembled transcripts, respectively, using the TopHat [57] program. Expression values were calculated by the Fragments Per Kilobase Of Exon Per Million Fragments Mapped (FPKM) method using Cufflinks [57], and differentially expressed genes were identified by DEGseq [58]. Details of the transcriptome analysis are presented in Additional file 1: Sections 5.2 and 7.1.

Hox and ParaHox analyses

We examined the homeodomain regions in *Nemopilema* using the InterProScan program. Hox and ParaHox genes were identified in *Nemopilema* by aligning the homeodomain sequences of human and fruit fly to the identified *Nemopilema* homeodomains. We considered only domains that were aligned to both the human and fruit fly. We also used this process for *Acropora*, *Hydra*, and *Nematostella* for comparison. Additionally, we added one Hox gene for *Acropora* and two Hox genes for *Hydra*, which are absent in NCBI gene set, though they were present in previous

study [21, 59]. Hox and ParaHox genes of *Clytia hemisphaerica*, a hydrozoan species with a medusa stage, were also added based on a previous study [60]. Finally, a multiple sequence alignment of these domains was conducted using MUSCLE, and a FastTree [61] maximum-likelihood phylogeny was generated using the Jones–Taylor–Thornton (JTT) model with gamma option.

Wnt gene subfamily analyses

Wnt genes of *Nematostella* and *Hydra* were downloaded from previous studies [23, 62], and those of *Acropora* were downloaded from the NCBI database. *Wnt* genes in *Nemopilema* were identified using the Pfam database by searching for "wnt family" domains. A multiple sequence alignment of *Wnt* genes was conducted using MUSCLE, and aligned sequences were trimmed using the trimAl program [63] with “gappyout” option. A phylogenetic tree was generated using RAxML with the PROTGAMMAJTT model and 100 bootstraps.

Abbreviations

SMRT, Single molecule real time sequencing; TSLR, TruSeq synthetic long reads;

Declarations

Availability of data and materials

The jellyfish genome project has been deposited at DDBJ/ENA/GenBank under the accession PEDN00000000. The version described in this paper is version PEDN01000000. Raw DNA and RNA sequence reads for *Nemopilema nomurai* and *Sanderia malayensis* have been submitted to the NCBI Sequence Read Archive database (SRA627560). All other data can be obtained from the authors upon reasonable request.

Authors' contributions

JB and SY supervised the project. YSC, JB, and SY planned and coordinated the project. HMK, JAW, YSC, SY, and JB wrote the manuscript. NayoungL, NayunL, YJJ, SW, KS, JCR, HSY, JHL, and SY prepared the samples, performed the experiments, and provided toxinological considerations. VL, AK, and MWK performed the gene evolutionary age analysis. HMK, SGP, YSC, YB, YJ, SJ, OC, JSE, and AM performed in-depth bioinformatics data analyses. All authors reviewed the manuscript and discussed the work.

Ethics approval

This is not applicable.

Competing interests

YSC and OC are employees, and JB is on the scientific advisory board of Clinomics Inc. HMK, YSC and JB have an equity interest in the company. All other coauthors have no conflicts of interest to declare.

Funding

This work was supported by the Genome Korea Project in Ulsan Research Funds (1.180024.01 and 1.180017.01) of Ulsan National Institute of Science & Technology (UNIST). This work was also supported by a grant from the Marine Biotechnology Program (20170305, Development of Biomedical materials based on marine proteins) and the Collaborative Genome Program (20140428) funded by the Ministry of Oceans and Fisheries, Korea. This work was also supported by the Collaborative Genome Program for Fostering New Post-Genome Industry of the National Research Foundation (NRF) funded by the Ministry of Science and ICT (MSIT) (NRF-2017M3C9A6047623 and NRF-2017R1A2B2012541).

Acknowledgements

Korea Institute of Science and Technology Information (KISTI) provided us with Korea Research Environment Open NETwork (KREONET), which is the internet connection service for efficient information and data transfer.

References

1. Park E, Hwang DS, Lee JS, Song JI, Seo TK, Won YJ: **Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record.** *Mol Phylogenet Evol* 2012, **62:**329-345.
2. Arai MN: *A functional biology of Scyphozoa*. Springer Science & Business Media; 2012.
3. Hale G: **The classification and distribution of the class Scyphozoa.** *Eugene: University of Oregon* 1999.
4. Anderson PA, Schwab WE: **The organization and structure of nerve and muscle in the jellyfish Cyanea capillata (Coelenterata; Scyphozoa).** *Journal of Morphology* 1981, **170:**383-399.
5. Gemmell BJ, Costello JH, Colin SP, Stewart CJ, Dabiri JO, Tafti D, Priya S: **Passive energy recapture in jellyfish contributes to propulsive advantage over other metazoans.** *Proc Natl Acad Sci U S A* 2013, **110:**17904-17909.
6. Dong Z, Liu D, Keesing JK: **Jellyfish blooms in China: Dominant species, causes and consequences.** *Mar Pollut Bull* 2010, **60:**954-963.
7. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics* 2015, **31:**3210-3212.
8. Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, Rattei T, Balasubramanian PG, Borman J, Busam D, et al: **The dynamic genome of Hydra.** *Nature* 2010, **464:**592-596.
9. Shinzato C, Shoguchi E, Kawashima T, Hamada M, Hisata K, Tanaka M, Fujie M, Fujiwara M, Koyanagi R, Ikuta T: **Using the Acropora digitifera genome to understand coral responses to environmental change.** *Nature* 2011, **476:**320.
10. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV, et al: **Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization.** *Science* 2007, **317:**86-94.
11. Galliot B, Quiquand M, Ghila L, de Rosa R, Miljkovic-Licina M, Chera S: **Origins of neurogenesis, a cnidarian view.** *Dev Biol* 2009, **332:**2-24.
12. Shen W, Wang D, Ye B, Shi M, Ma L, Zhang Y, Zhao Z: **GC3-biased gene domains in mammalian genomes.** *Bioinformatics* 2015, **31:**3081-3084.
13. Wan X-F, Xu D, Kleinhofs A, Zhou J: **Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes.** *BMC Evolutionary Biology* 2004, **4:**19.
14. Tatarinova TV, Alexandrov NN, Bouck JB, Feldmann KA: **GC3 biology in corn, rice, sorghum and other grasses.** *BMC Genomics* 2010, **11:**308.
15. Steinmetz PR, Kraus JE, Larroux C, Hammel JU, Amon-Hassenzahl A, Houliston E, Worheide G, Nickel M, Degnan BM, Technau U: **Independent evolution of striated muscles in cnidarians and bilaterians.** *Nature* 2012, **487:**231-234.
16. Seipel K, Schmid V: **Evolution of striated muscle: jellyfish and the origin of triploblasty.** *Dev Biol* 2005, **282:**14-26.
17. Cegolon L, Heymann WC, Lange JH, Mastrangelo G: **Jellyfish stings and their management: a review.** *Marine drugs* 2013, **11:**523-550.
18. Leclere L, Rottinger E: **Diversity of Cnidarian Muscles: Function, Anatomy, Development and Regeneration.** *Front Cell Dev Biol* 2016, **4:**157.
19. Watanabe H, Fujisawa T, Holstein TW: **Cnidarians and the evolutionary origin of the nervous system.** *Dev Growth Differ* 2009, **51:**167-183.
20. Kamm K, Schierwater B, Jakob W, Dellaporta SL, Miller DJ: **Axial patterning and diversification in the cnidaria predate the Hox system.** *Curr Biol* 2006, **16:**920-926.

21. Chourrout D, Delsuc F, Chourrout P, Edvardsen RB, Rentzsch F, Renfer E, Jensen MF, Zhu B, de Jong P, Steele RE, Technau U: **Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements.** *Nature* 2006, **442:**684-687.
22. Finnerty JR, Pang K, Burton P, Paulson D, Martindale MQ: **Origins of bilateral symmetry: Hox and dpp expression in a sea anemone.** *Science* 2004, **304:**1335-1337.
23. Kusserow A, Pang K, Sturm C, Hrouda M, Lentfer J, Schmidt HA, Technau U, von Haeseler A, Hobmayer B, Martindale MQ, Holstein TW: **Unexpected complexity of the Wnt gene family in a sea anemone.** *Nature* 2005, **433:**156-160.
24. Nusse R: **An ancient cluster of Wnt paralogues.** *Trends Genet* 2001, **17:**443.
25. Han C-H, Uye S-i: **Combined effects of food supply and temperature on asexual reproduction and somatic growth of polyps of the common jellyfish Aurelia aurita sl.** *Plankton and Benthos Research* 2010, **5:**98-105.
26. Maiorova TD, Kosevich IA, Melekhova OP: **[On some features of embryonic development and metamorphosis of Aurelia aurita (Cindaria, Scyphozoa)].** *Ontogenez* 2012, **43:**333-349.
27. Müller WA, Leitz T: **Metamorphosis in the Cnidaria.** *Canadian Journal of Zoology* 2002, **80:**1755-1771.
28. Schmich J, Trepel S, Leitz T: **The role of GLWamides in metamorphosis of Hydractinia echinata.** *Dev Genes Evol* 1998, **208:**267-273.
29. Fuchs B, Wang W, Graspeuntner S, Li Y, Insua S, Herbst EM, Dirksen P, Bohm AM, Hemmrich G, Sommer F, et al: **Regulation of polyp-to-jellyfish transition in Aurelia aurita.** *Curr Biol* 2014, **24:**263-273.
30. Cunningham TJ, Duester G: **Mechanisms of retinoic acid signalling and its roles in organ and limb development.** *Nat Rev Mol Cell Biol* 2015, **16:**110-123.
31. So EN, Crowe DL: **Characterization of a retinoic acid responsive element in the human ets-1 promoter.** *IUBMB Life* 2000, **50:**365-370.
32. Lassen S, Helmholz H, Ruhnau C, Prange A: **A novel proteinaceous cytotoxin from the northern Scyphozoa Cyanea capillata (L.) with structural homology to cubozoan haemolysins.** *Toxicon* 2011, **57:**721-729.
33. Jungo F, Bougueleret L, Xenarios I, Poux S: **The UniProtKB/Swiss-Prot Tox-Prot program: A central hub of integrated venom protein data.** *Toxicon* 2012, **60:**551-557.
34. Rawlings ND, Barrett AJ: **Evolutionary families of metallopeptidases.** *Methods Enzymol* 1995, **248:**183-228.
35. Li R, Yu H, Xue W, Yue Y, Liu S, Xing R, Li P: **Jellyfish venomics and venom gland transcriptomics analysis of Stomolophus meleagris to reveal the toxins associated with sting.** *J Proteomics* 2014, **106:**17-29.
36. Brinkman DL, Jia X, Potriquet J, Kumar D, Dash D, Kvaskoff D, Mulvenna J: **Transcriptome and venom proteome of the box jellyfish Chironex fleckeri.** *BMC Genomics* 2015, **16:**407.
37. Jouiaei M, Yanagihara AA, Madio B, Nevalainen TJ, Alewood PF, Fry BG: **Ancient Venom Systems: A Review on Cnidaria Toxins.** *Toxins (Basel)* 2015, **7:**2251-2271.
38. Li H: **Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.** *arXiv preprint arXiv:13033997* 2013.
39. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The sequence alignment/map format and SAMtools.** *Bioinformatics* 2009, **25:**2078-2079.
40. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32:**1792-1797.
41. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.** *Mol Biol Evol* 2016, **33:**1870-1874.
42. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, et al: **Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data.** *Nat Methods* 2013, **10:**563-569.

43. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W: **Scaffolding pre-assembled contigs using SSPACE.** *Bioinformatics* 2010, **27:**578-579.
44. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al: **SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler.** *Gigascience* 2012, **1:**18.
45. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B: **AUGUSTUS: ab initio prediction of alternative transcripts.** *Nucleic Acids Res* 2006, **34:**W435-439.
46. Benson G: **Tandem repeats finder: a program to analyze DNA sequences.** *Nucleic acids research* 1999, **27:**573.
47. Tarailo-Graovac M, Chen N: **Using RepeatMasker to identify repetitive elements in genomic sequences.** *Current protocols in bioinformatics* 2009**:**4.10. 11-14.10. 14.
48. Domazet-Loso T, Brajkovic J, Tautz D: **A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages.** *Trends Genet* 2007, **23:**533-539.
49. Kumar S, Stecher G, Suleski M, Hedges SB: **TimeTree: A Resource for Timelines, Timetrees, and Divergence Times.** *Mol Biol Evol* 2017, **34:**1812-1819.
50. Parr CS, Wilson N, Leary P, Schulz KS, Lans K, Walley L, Hammock JA, Goddard A, Rice J, Studer M, et al: **The Encyclopedia of Life v2: Providing Global Access to Knowledge About Life on Earth.** *Biodivers Data J* 2014**:**e1079.
51. Li L, Stoeckert CJ, Jr., Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes.** *Genome Res* 2003, **13:**2178-2189.
52. Stamatakis A: **RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.** *Bioinformatics* 2006, **22:**2688-2690.
53. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24:**1586-1591.
54. Han MV, Thomas GW, Lugo-Martinez J, Hahn MW: **Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3.** *Mol Biol Evol* 2013, **30:**1987-1997.
55. Zdobnov EM, Apweiler R: **InterProScan--an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17:**847-848.
56. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al: **De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.** *Nat Protoc* 2013, **8:**1494-1512.
57. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nat Protoc* 2012, **7:**562-578.
58. Wang L, Feng Z, Wang X, Wang X, Zhang X: **DEGseq: an R package for identifying differentially expressed genes from RNA-seq data.** *Bioinformatics* 2010, **26:**136-138.
59. DuBuc TQ, Ryan JF, Shinzato C, Satoh N, Martindale MQ: **Coral comparative genomics reveal expanded Hox cluster in the cnidarian-bilaterian ancestor.** *Integr Comp Biol* 2012, **52:**835-841.
60. Chiori R, Jager M, Denker E, Wincker P, Da Silva C, Le Guyader H, Manuel M, Queinnec E: **Are Hox genes ancestrally involved in axial patterning? Evidence from the hydrozoan Clytia hemisphaerica (Cnidaria).** *PLoS One* 2009, **4:**e4231.
61. Price MN, Dehal PS, Arkin AP: **FastTree 2--approximately maximum-likelihood trees for large alignments.** *PLoS One* 2010, **5:**e9490.
62. Lengfeld T, Watanabe H, Simakov O, Lindgens D, Gee L, Law L, Schmidt HA, Ozbek S, Bode H, Holstein TW: **Multiple Wnts are involved in Hydra organizer formation and regeneration.** *Dev Biol* 2009, **330:**186-199.
63. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T: **trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses.** *Bioinformatics* 2009, **25:**1972-1973.

Fig. 1 Gene family relationships of cnidarian and metazoan species. **a** Venn diagram of the number of unique and shared gene families among four cnidarian genomes. **b** Gene family expansions and contractions in the *Nemopilema* genome. Numbers designate the number of gene families that have expanded (red, +) and contracted (blue, -) after the split from the common ancestor. **c** The proportion of *Nemopilema* genes in each evolutionary era. While most *Nemopilema* genes (~80%) are ancient (~1,877 Mya), a few (~3%) are of intermediate age (~659 Mya) and a significant fraction (~17%) are relatively young (~147 Mya).

Fig. 2 Gene expression patterns of medusa bell and tentacle tissues and expansion of myosin heavy chain genes in jellyfish. **a** *P*-value heatmap of enriched GO categories using highly expressed genes in medusa bell tissue. Greater than 2-fold and 4-fold higher expression in medusa bell than tentacles are shown in each column. Only shared GO categories between *N. nomurai* and *S. malayensis* are shown. **b** *P*-value heatmap of enriched GO categories using highly expressed genes in tentacle tissue. **c** Unrooted Le-Gascuel model tree of myosin heavy chain genes using BLAST best hit method. **d** Expression pattern of MYH and MYL genes in *Nemopilema*. Genes that are not expressed in both tentacles and medusa bell were excluded.

Fig. 3 Retinoic acid signaling pathway and RAREs in *Nemopilema*. **a** Schematic of the retinoic acid signaling pathway in humans. Blue denotes presence of the gene and/or element in Cnidaria. Red denotes presence only in *Nemopilema* among the published cnidarians. **b** The distribution of distances between the RAREs and the nearest gene. The distance was calculated by identifying its proximity to transcription start site (TSS) of the genes. The gene count was calculated for

each non-overlapping 1 Kb bin across a range of -100 Kb to 100 Kb. **c** The RAREs located nearby posterior Hox genes in *Nemopilema*.

Fig. 4 Phylogenetic analysis of venom related domains in non-bilaterian metazoans. Five venom domains (PF01421, PF01549, PF07648, PF00068, and PF05826) are represented in four circular dendrograms. Two phospholipase A_2 domains (PF00068 and PF05826) were merged into one circular dendrogram (bottom right) and shadings on branches and nodes (sky-blue) in phospholipase A_2 denote the PF05826 domain.

Additional files

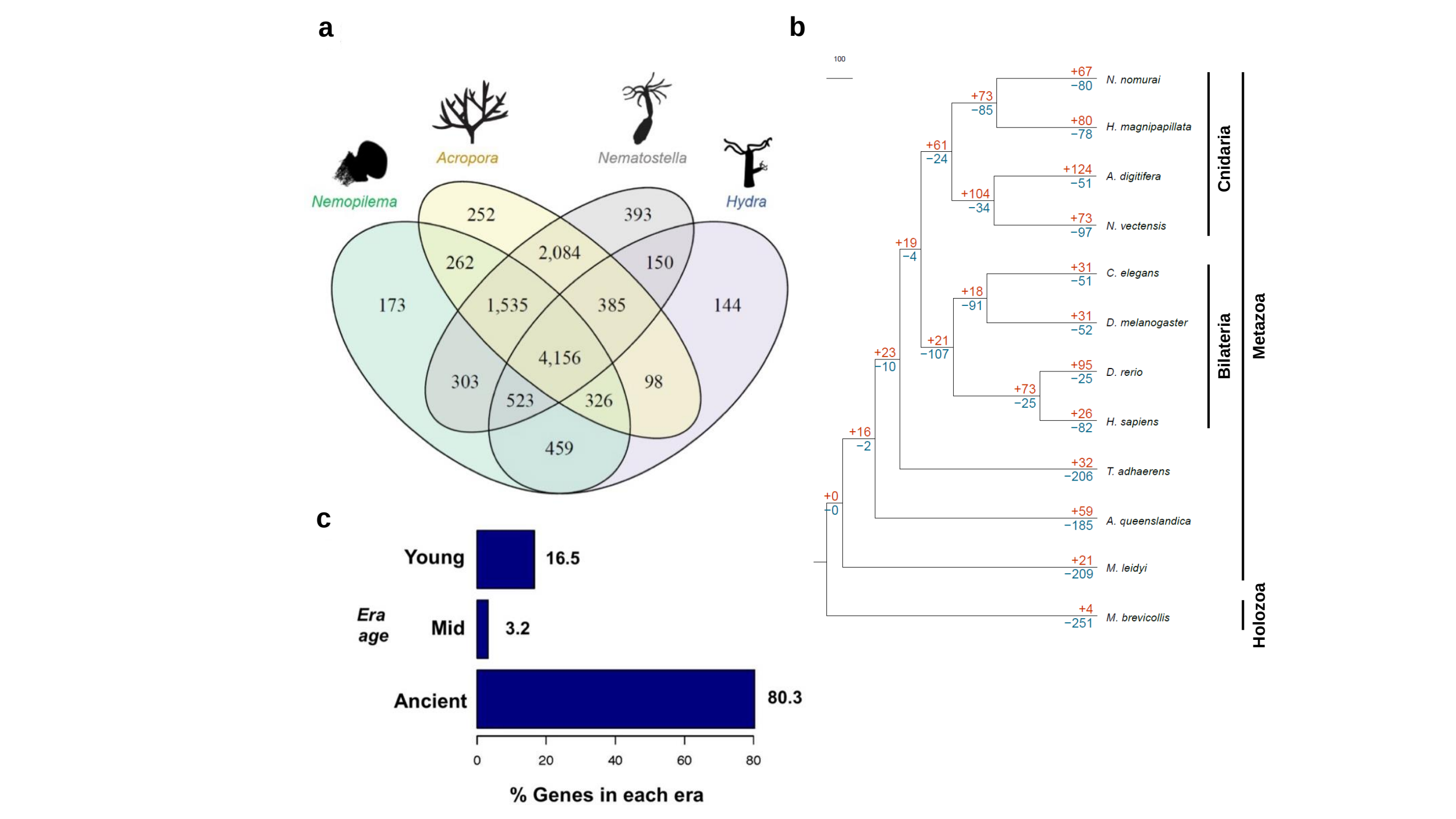
Additional file 1: Supplementary figures, tables, and methods. This document contains additional supporting evidence for this study that are presented in form of supplemental figures and tables.

Additional file 2: Supplementary data. Protein domain annotation statistics.

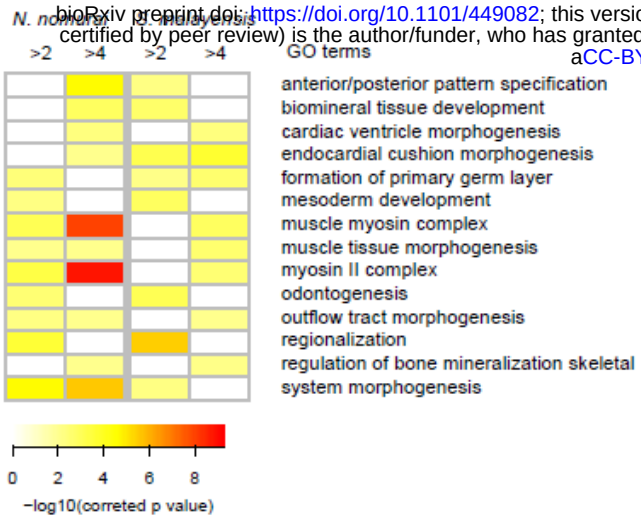
Additional file 3: Supplementary data. List of gene clusters evolving faster in the *Nemopilema nomurai* genome.

Additional file 4: Supplementary data. Gene ontology and KEGG enrichment result of top 100 and 500 GC3 genes in *Nemopilema nomurai*.

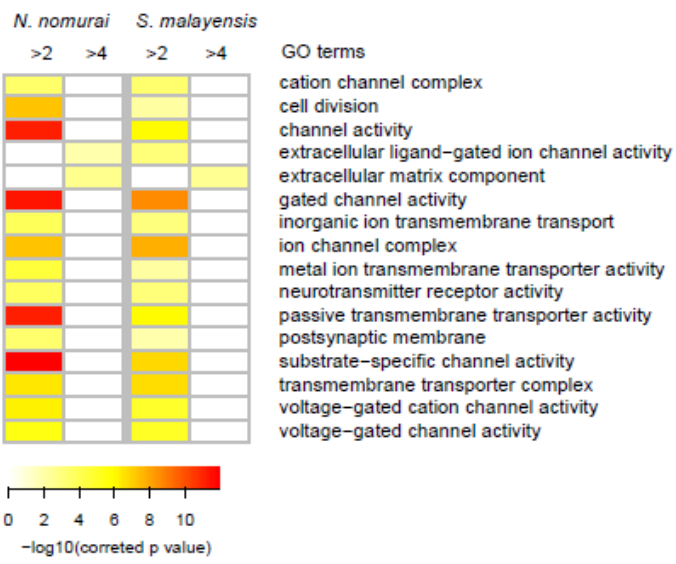
Additional file 5: Supplementary data. Gene ontology enrichment result of highly expressed genes in *Nemopilema nomurai* and *Sanderia malayensis*.


a
Nemopilema
Acropora
Nematostella
Hydra
252
393
262
2,084
150
173
1,535
385
144
4,156
303
523
326
98
459
b
100
+67
−80
N. nomurai
+73
−85
+80
−78
H. magnipapillata
+61
−24
+124
−51
A. digitifera
+104
−34
+73
−97
N. vectensis
+19
−4
+31
−51
C. elegans
+18
−91
+31
−52
D. melanogaster
+21
−107
+23
−10
+95
−25
D. rerio
+73
−25
+26
−82
H. sapiens
+16
−2
+32
−206
T. adhaerens
+0
−0
+59
−185
A. queenslandica
+21
−209
M. leidyi
+4
−251
M. brevicollis
Cnidaria
Bilateria
Metazoa
Holozoa
c
Young
16.5
Era age
Mid
3.2
Ancient
80.3
0
20
40
60
80
% Genes in each era

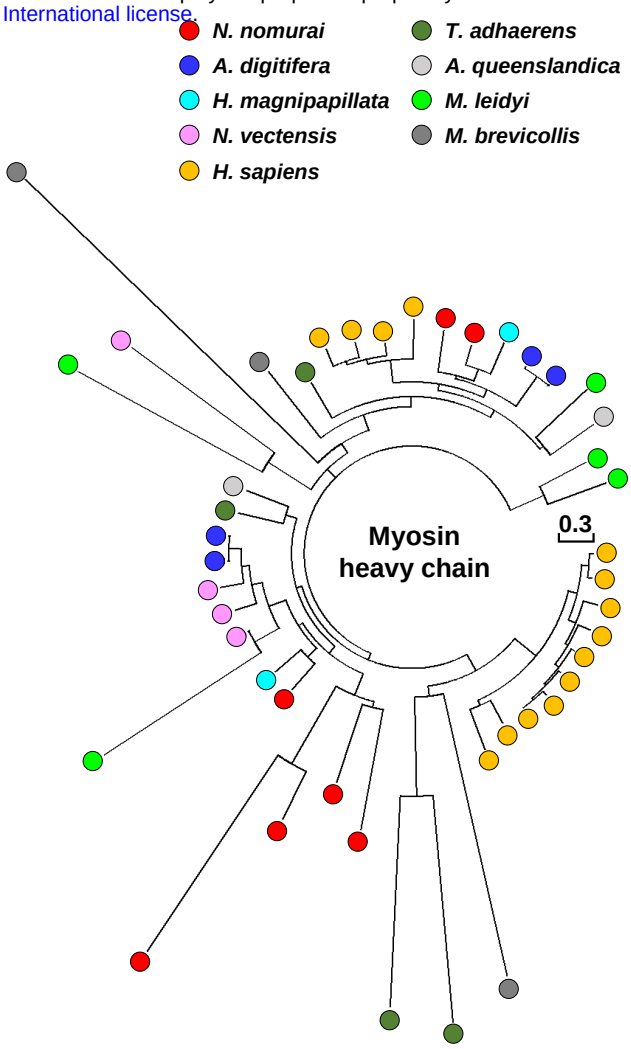
a



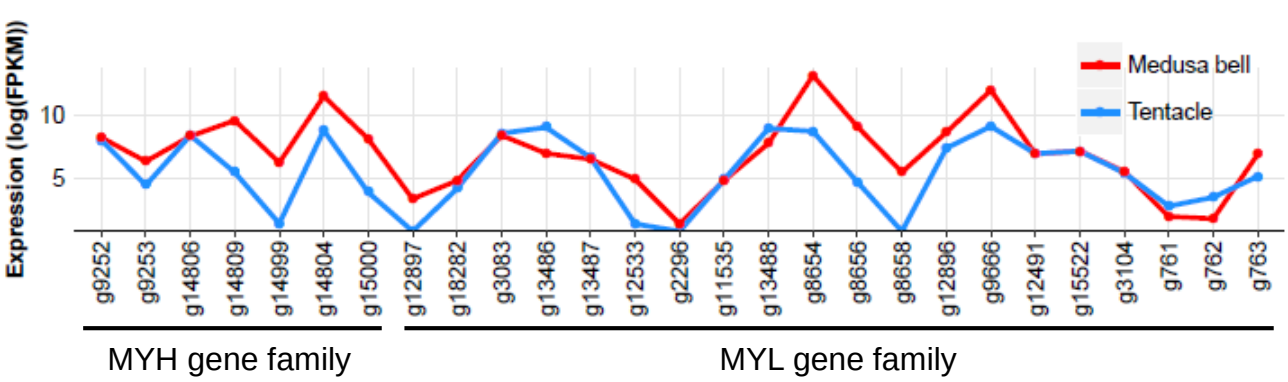
b

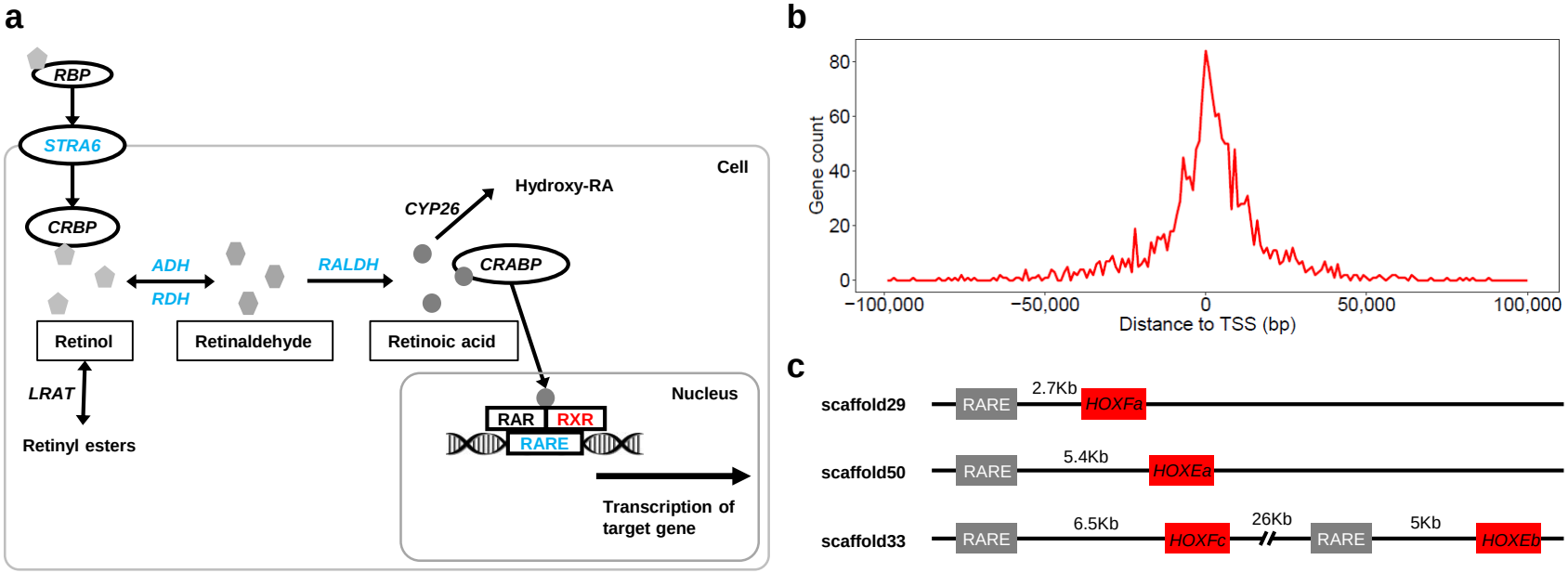


c

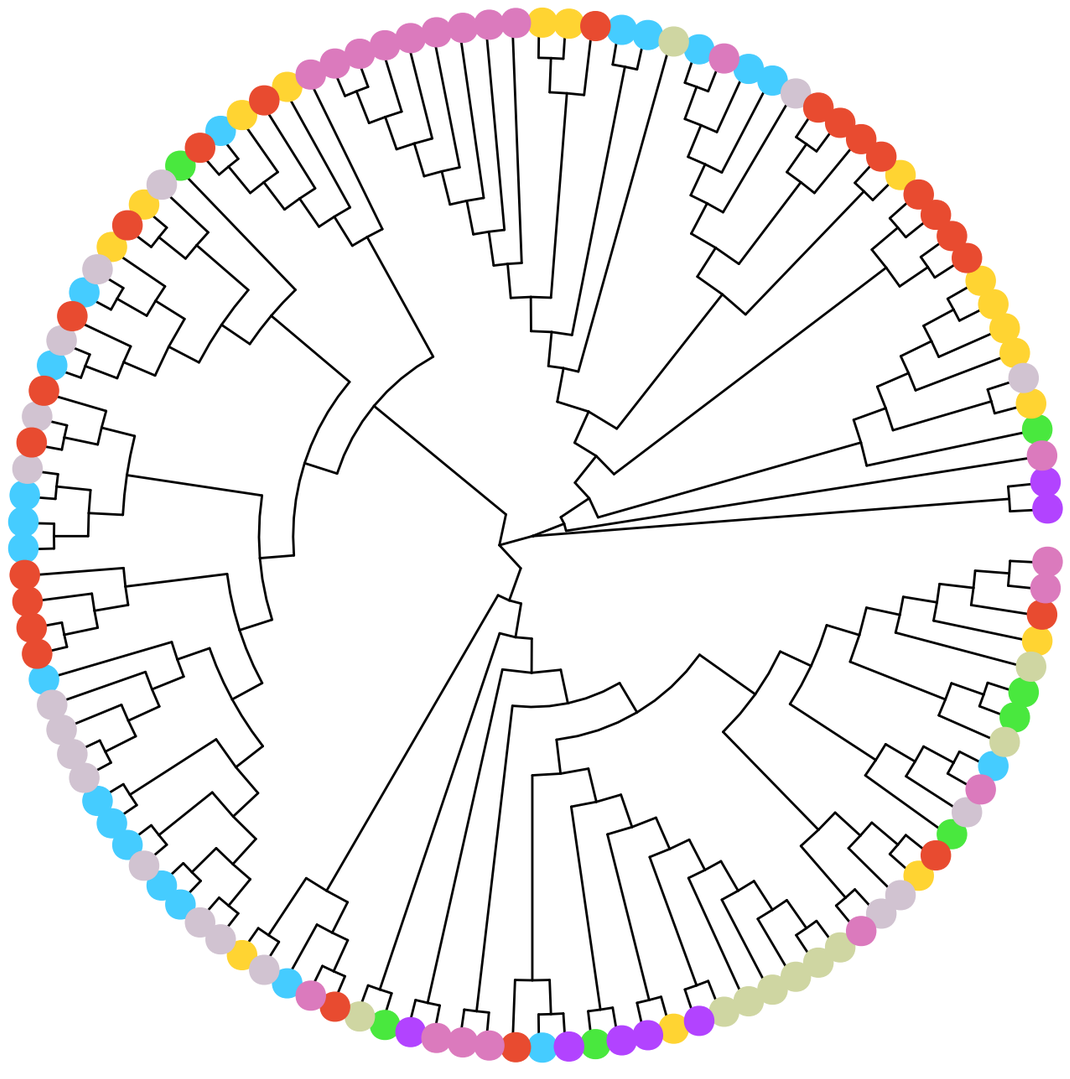


d

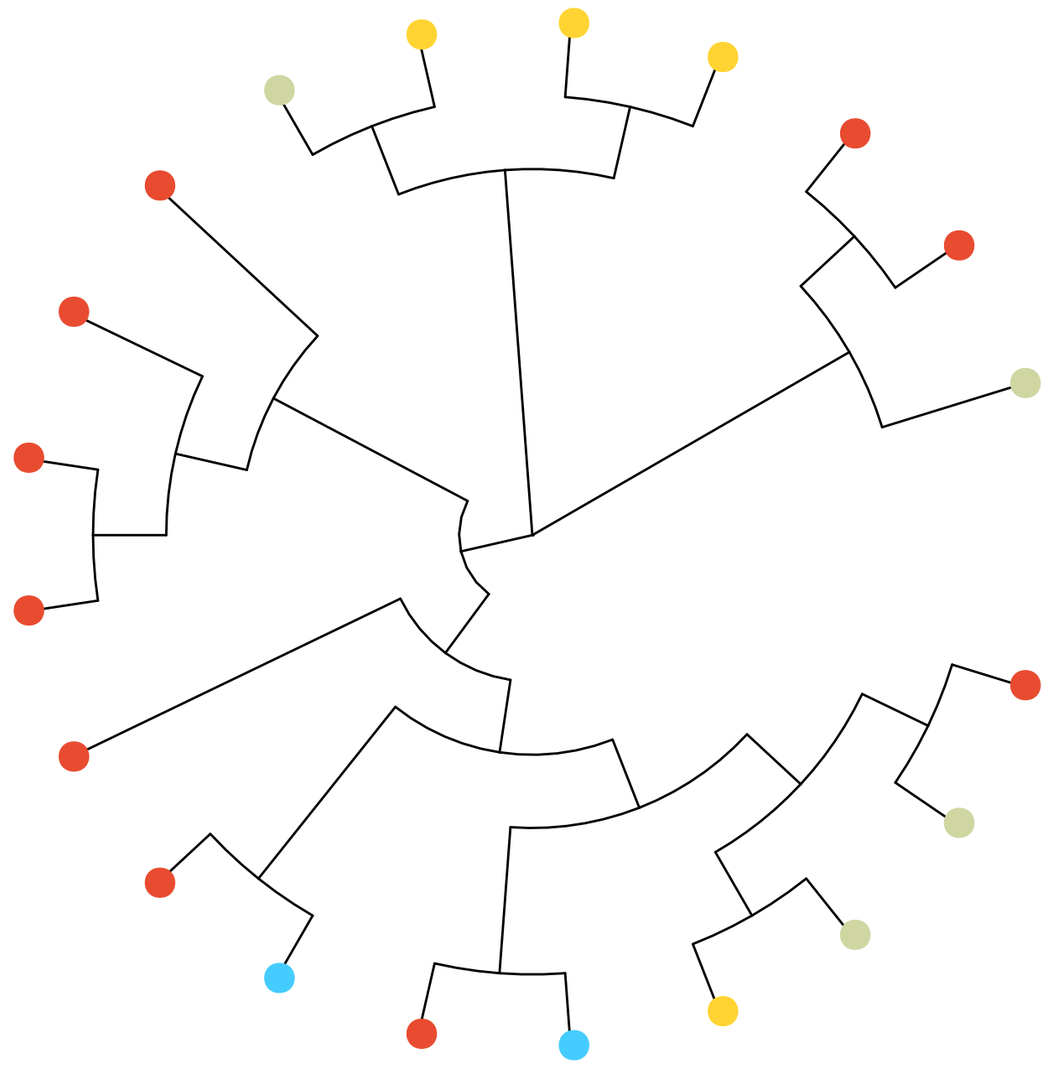



a
RBP
STRA6
CRBP
ADH
RDH
RALDH
CYP26
Hydroxy-RA
CRABP
Cell
Retinol
Retinaldehyde
Retinoic acid
LRAT
Retinyl esters
Nucleus
RAR
RXR
RARE
Transcription of target gene
b
Gene count
80
60
40
20
0
−100,000
−50,000
0
50,000
100,000
Distance to TSS (bp)
c
scaffold29
RARE
2.7Kb
HOXFa
scaffold50
RARE
5.4Kb
HOXEa
scaffold33
RARE
6.5Kb
HOXFc
26Kb
RARE
5Kb
HOXEb

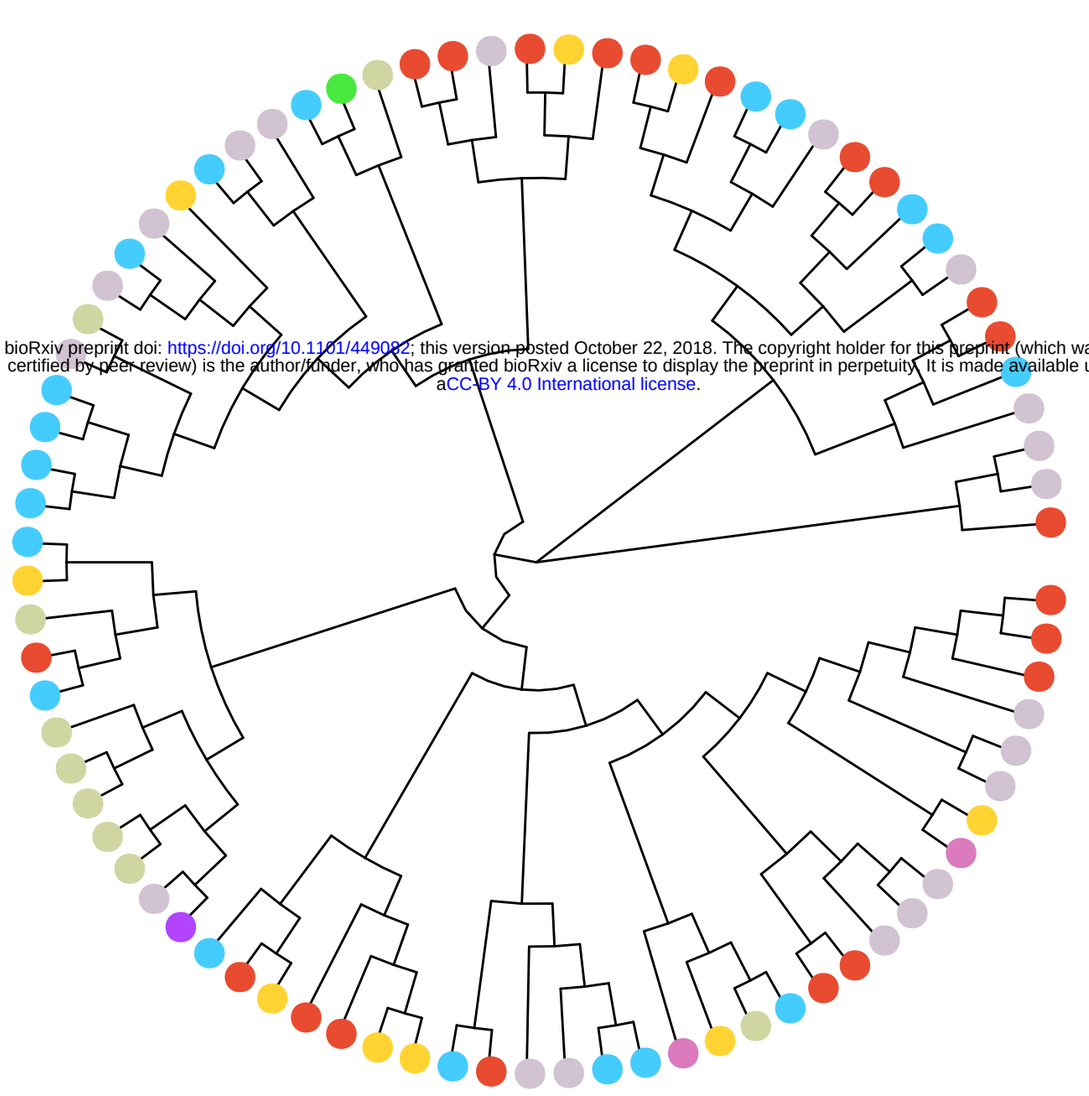
Reprolysin (M12B) family zinc metalloprotease (PF01421)



ShK domain-like (PF01549)



Kazal-type serine protease inhibitor domain (PF07648)



Phospholipase A_2 (PF00068/PF05826)