The genome of the giant Nomura’s jellyfish sheds light on the early evolution of active predation

Hak-Min Kim1,2†, Jessica A. Weber3,4†, Nayoung Lee5†, Seung Gu Park1, Yun Sung Cho1,2,6, Youngjune Bhak1,2, Nayun Lee5, Yeonsu Jeon1,2, Sungwon Jeon1,2, Victor Luria7, Amir Karger8, Marc W. Kirschner7, Ye Jin Jo5, Seonock Woo9,10, Kyungsun Shin11, Oksung Chung6,12, Jae-Chun Ryu13, Hyung-Soon Yim10, Jung-Hyun Lee10, Jeremy S. Edwards14, Andrea Manica15, Jong Bhak1,2,6,12* and Seungshic Yum5,9*

Abstract

Background: Unique among cnidarians, jellyfish have remarkable morphological and biochemical innovations that allow them to actively hunt in the water column and were some of the first animals to become free-swimming. 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, as well as using pulsed jet propulsion for mobility. Here, we present the genome of the giant Nomura’s jellyfish (Nemopilema nomurai) to understand the genetic basis of these key innovations.

Results: We sequenced the genome and transcriptomes of the bell and tentacles of the giant Nomura’s jellyfish as well as transcriptomes across tissues and developmental stages of the Sanderia malayensis jellyfish. Analyses of the 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 Nemopilema jellyfish genome and transcriptomes genetically confirm their unique morphological and physiological traits, which may have contributed to the success of jellyfish as early 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 (Fig. 1a) [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 (Fig. 1c) [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

* Correspondence: jongbhak@gmail.com; syum@kiost.ac.kr
† Hak-Min Kim, Jessica A. Weber and Nayoung Lee contributed equally to this work.
1 Korean Genomics Industrialization Center (KOGIC), Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea
2 Ecological Risk Research Division, Korea Institute of Ocean Science and Technology (KIOST), Geoje 53201, Republic of Korea
Full list of author information is available at the end of the article

© The Author(s). 2019 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
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 in both 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 the Nomura's jellyfish (Nemopilema nomurai; Fig. 1b). 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, Additional files 2 and 3). 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 vulgaris (42.87%) (Additional file 1: Table S13).

We compared the Nemopilema genome to other cnidarian genomes, including the recently published Aurelia aurita [8] and Clytia hemisphaerica genomes [9], 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 13 metazoans using 18,458 orthologous gene families (see Additional file 1: Section 4.1). Of these, 10,434 were found in Nemopilema and 6764 were shared by all three available cnidarian classes (Scyphozoa: Nemopilema nomurai and Aurelia aurita; Hydrozoa: Hydra vulgaris [10], Clytia hemisphaerica; Anthozoa: Acropora digitifera [11] and Nematostella vectensis [12]; Fig. 2a). 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. 2b; 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 three 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. 2c; 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 123 expanded and 164 contracted gene families compared to the common ancestor of *Nemopilema* and *Aurelia* (Fig. 2b; 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 (GO:0006836, \(P = 6.01 \times 10^{-10}\)) was significantly enriched in Scyphozoa lineage compared to the common ancestor of Scyphozoa and Hydrozoa (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 [13]. Compared to the common ancestor of *Nemopilema* and *Aurelia*, *Nemopilema* showed expanded gene families associated with metallopeptidase activities (GO:0008237, \(P = 2.86 \times 10^{-14}\); Additional file 1: Tables S18 and S19). Additionally, we found 1589 orthologous gene families that are specific to Scyphozoa. Enrichment tests of scyphozoan-specific genes showed the terms of sodium ion transport, ion channel activity, and neurotransmitter receptor activity (Additional file 1: Table S20).

**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) [14, 15] (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) [16]. 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: Tables S25 and S26). Calcium plays a key role in the striated muscle contraction in jellyfish, and the calcium signaling pathway (GO:0004020, \(P = 5.60 \times 10^{-10}\)) showed a high level of GC3 biases specific to *Nemopilema* *Nemopilema* and *Aurelia* 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 (Additional file 1: Section 5.1; Additional file 4: Tables S27 and S28).
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 [17]. 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 S32). 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 *Nemopilema, Aurelia*, and *Hydra*. However, *Nemopilema* and *Aurelia* 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* underwent secondary simplifications relative to *Nematostella*, which has a greater degree of muscle-cell-type specialization [10]. Compared to *Hydra* and *Nematostella*, *Nemopilema* and *Aurelia* show 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 [18, 19]. 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 S29). 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. 3a; Additional file 5: Tables a, b).
S30–S33). 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 [20]. Cnidarians possess both epithelio-muscular 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 [17]. Interestingly, Nemopilema and Aurelia showed the largest copy numbers of MYH and MYL genes among non-bilaterian metazoans (Fig. 3c; Additional file 1: Figures S14–S17; Tables S38–40), and six of the seven MYH genes and 12 out of 21 MYL genes in Nemopilema showed higher expression in the bell than the tentacles with very high ~8.8 and ~17-fold increase on average, respectively (Fig. 3d). 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. 3b; Additional file 5: Tables S34–S37); 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 [21].

### 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 S43). 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 2325 bp. On average, 87% of the RNA reads were aligned into the assembled transcripts (Additional file 1: Table S44), 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 S45). 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 S46). Active feeding in the polyp stimulates asexual proliferation either into more polyps or metamorphosis to strobila [27]. 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 S47). It has been reported that RF-amide and LW-amide neuropeptides were associated with metamorphosis in cnidarians [28–30]. 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 S48). In the medusa, extracellular matrix, metallopeptidase activity, and immune system process terms were enriched (Additional file 1: Table S49), 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 [31]. 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 medusa stage (Additional file 1: Figure S24). Retinoic acid (RA) signaling plays a central role during vertebrate growth and development [32], where it regulates transcription by interacting with the RA receptor (RAR) bound to RA response elements (RAREs) of nearby target genes [33]. 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. 4a). We discovered 1630 Nemopilema RARE regions with an average distance of 13 Kbp to the nearest gene (Fig. 4b; Additional file 1: Tables S50 and S51). Interestingly, four posterior Hox genes of Nemopilema and two Hox genes of Aurelia were located within ±10 Kbp from RAREs, which is unique among the non-bilaterian metazoans (Fig. 4c; Additional file 1: Table S52). 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 [31].

Identification of toxin-related domains in jellyfish
Jellyfish produce complex mixtures of proteinaceous venoms for active prey capture and defense [34]. We identified abundant toxin domains in Nemopilema when compared to the non-bilaterian metazoan gene sets in the Tox-Prot database [35]. In total, 67 out of 136 toxin domains aligned to non-bilaterian metazoans; of these 67 toxin domains, 52 were found in Nemopilema (Additional file 1: Table S53). 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), phospholipase A2 (PF05826), and Prokineticin (PF06607) domains (Fig. 5). Also, Nemopilema and Aurelia possess 8 and 11 ShK domain-like (PF01549) domains, respectively, which are the most abundant in these species compared to other non-bilaterians. In particular, Reprolysin (M12B) family zinc metalloproteases are enzymes that cleave peptides and comprise most snake venom endopeptidases [36]. Furthermore, it has been reported that serine protease

![Fig. 4 Retinoic acid signaling pathway and RAREs in Nemopilema.](image-url)
inhibitor and ShK domains were abundantly found in the transcriptomes of both the cannonball jellyfish (*Stomolophus meleagris*) and the box jellyfish (*Chironex fleckeri*) [37, 38], and phospholipase A2 is a well-characterized toxin-related enzyme, which is critical to the production of venom components, found in the class Scyphozoa [39].

**Conclusions**

An interesting 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.

**Methods**

**Sample preparation**

A medusa *Nemopilema nomurai* was collected at the Tongyeong Marine Science Station, KIOST (34.7699 N, 128.3828 E) on Sept. 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 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 [40]. Consensus sequences for each jellyfish were generated using SAMtools [41]. The consensus sequence from *C. sowerbyi* was excluded due to low coverage. We conducted multiple sequence alignment using MUSCLE [42] and ran the MEGA v7 [43] neighbor joining phylogenetic tree (gamma distribution) with 1000 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 (400 bp, 5 Kb, 10 Kb, 15 Kb, and 20 Kb) were constructed according to the manufacturers’ protocols. The Illumina libraries were sequenced using a HiSeq2500 with a read length of 100 bp (400 bp, 15 Kb, and 20 Kb) and a HiSeq2000 with a read length of 101 bp (5 Kb and 10 Kb). Quality filtered PacBio subreads were assembled into distinct contigs using the FALCON assembler [44] 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 [45]. Gaps generated by SSPACE were filled by aligning the Illumina short-insert paired-end sequences using GapCloser [46]. 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: 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 inferring that the gene is at least as old as the age of the common ancestor [50]. For every species, we use the NCBI taxonomy. The timing of most divergence events is estimated using TimeTree [51] and the Encyclopedia of Life [52]. 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 phylustratum, from the most ancient (PS 1) to the newest (PS 11). To see broad evolutionary patterns, we aggregate the counts from several phylustrata into three broad evolutionary eras: ancient (PS 1–5), cellular organisms to Eumetazoa, 4204 Mya to 741 Mya), middle (PS 6–7, Cnidaria to Scyphozoa, 741 Mya to 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*, *Aurelia aurita*, *Hydra vulgaris*, *Clytia hemisphaerica*, *Acropora digitifera*, *Nematostella vectensis*, *Caenorhabditis elegans*, *Danio rerio*, *Drosophila melanogaster*, *Homo sapiens*, *Trichoplax adhaerens*, *Amphimedon queenslandica*, *Mnemiopsis leidyi*, and *Monosiga brevicollis* using OrthoMCL [53]. To infer a phylogeny and divergence times, we used RAxML [54] and MEGA7 program [43], respectively. A gene family expansion and contraction analysis was conducted using the Café program [55]. Domain regions were predicted by InterProScan [56] with domain databases. Details of the comparative analysis are provided in Additional file 1: Sections 4.1–4.3.

**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 [57]. Quality filtered RNA reads from *Nemopilema* and *Sanderia* were aligned to the *Nemopilema* genome assembly and the assembled transcripts, respectively, using the TopHat [58] program. Expression values were calculated by the Fragments Per Kilobase Of Exon Per Million
Fragments Mapped (FPKM) method using Cufflinks [58], and differentially expressed genes were identified by DEGSeq [59]. 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 Nematomella for comparison. Additionally, we added one Hox gene for Acropora and two Hox genes for Hydra, which are absent in the NCBI gene set, though they were present in previous studies [23, 60]. Hox and ParaHox genes of Clytia hemisphaerica, a hydroidoan species with a medusa stage, were also added based on a previous study [61]. Finally, a multiple sequence alignment of these domains was conducted using MUSCLE, and a FastTree [62] maximum likelihood phylogeny was generated using the Jones–Taylor–Thornton (JTT) model with gamma option.

Wnt gene subfamily analyses
Wnt genes of Nematomella and Hydra were downloaded from previous studies [25, 63], and those of Acropora were downloaded from the NCBI database. Wnt genes in Nemopilema and Aurelia 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 [64] with “gappyout” option. A phylogenetic tree was generated using RAxML with the PROTGAMMAJTT model and 100 bootstraps. Further information, including sample preparation, assembly, genome annotation, and evolutionary analyses, can be found in Additional file 1 [65–112].

Additional files

Additional file 1: Additional Methods. Figures S1–S25. Tables S1–S20. S23, S24, S29, S38–S53. Figure S1. Species identification. Figure S2, Table S1. Estimating genome size and complexity. Figures S3–S5. Tables S2–S9. Sequencing and assembly. Figure S6. GC content distributions. Tables S10–S13. Gene and repeat annotations. Figure S7. Phylogenetic tree with divergence times. Tables S14–S20. Functional enrichment of gene family expansion and contraction. Figure S8. Domain expansion. Figure S9, Table S23. Fast-evolving gene. Figure S10. Gene age estimation. Figures S11 and S12. Table S24. Analyses of Otx and POU genes. Figure S13. Genomic context. Table S29. Transcriptome sequencing. Figures S14–S17. Tables S38–S40. Myosin type II gene analyses. Table S41. Homeobox domain. Tables S14–S20. Functional enrichment of gene family expansion and contraction. Figure S8. Domain expansion. Figures S22 and S23. Table S42. Wnt gene analyses. Figures S24–S25. Table S43–S45. Transcriptome assembly. Tables S46–S49. Functional enrichment of development stages. Figures S24 and S25. Tables S50–S52. RXR and RAREs analyses. Table S53. Venom domain analysis. (DOCX 4760 kb)

Additional file 2: Table S21. Domain annotation statistics. (XLSX 11 kb)

Additional file 3: Table S22. List of faster evolving genes. (XLSX 21 kb)

Additional file 4: Tables S25 and S26. Functional enrichment of top 100 and 500 GC3 genes in Nemopilema nomurai. Tables S27 and S28. Functional enrichment of top 100 and 500 GC3 genes in Aurelia aurita. (XLSX 32 kb)

Additional file 5: Tables S30–S33. Functional enrichment of medusa bell in Nemopilema nomurai and Sandenia malayensis. Tables S34–S37. Functional enrichment of tentacles in Nemopilema nomurai and Sandenia malayensis. (XLSX 51 kb)

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

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 (20180430) 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). V.L. and M.W.K. gratefully acknowledge funding support from the National Institutes of Health of USA (R01 HD073104 and R01 HD091846 to MWK).

Availability of data and materials
The jellyfish genome project has been deposited at DDBJ/ENA/GenBank under the accession PEDNO00000000 [113]. The version described in this paper is version PEDNO1000000. Raw DNA and RNA sequence reads for Nemopilema nomurai and Sandenia malayensis have been submitted to the NCBI Sequence Read Archive database (SRA627560) [114]. 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, Nayıurln, YJJ, SW, KS, JCR, HSY, JHL, and SY prepared the samples, performed the experiments, and provided toxicological considerations. VL, AK, and M.W.K. performed the gene evolutionary age analysis. HMK, SGP, YSC, YB, YJ, SJ, OC, JSE, and AM performed the in-depth bioinformatics data analyses. All authors read and approved the final manuscript.

Ethics approval and consent to participate
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 coauthors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1Korean Genomics Industrialization Center (KOGIC), Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea. 2Department of Biomedical Engineering, School of Life Sciences, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea. 3Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. 4Department of Biology, University of New Mexico, Albuquerque,
References

1. Park E, Hwang DS, Lee JS, Song JI, Won YJ. Estimation of divergence times in cnidian evolution based on mitochondrial protein-coding genes and the fossil record. Mol Phylogenet Evol. 2012;62(1):329–45.

2. Arai MN. A functional biology of Scyphozoa: Springer Science & Business Media, 2012.

3. Ijima 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). J Morphol. 1981;170(3):383–99.

5. Gemmell BJ, Costello JH, Colin SP, Stewart CJ, Dabiri JO, Tafti D, Priya S. Cellular and Molecular Toxicology Laboratory, Center for Environment, Health and Welfare Research, Korea Institute of Science and Technology (KIST), Seoul 02792, Republic of Korea. 11Personal Genomics Institute, Genome Research Foundation, Cheongju 28160, Republic of Korea. 14Chemistry and Chemical Biology, UNM Comprehensive Cancer Center, University of New Mexico, Albuquerque, NM 87131, USA. 16Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

Received: 8 January 2019 Accepted: 28 February 2019

Published online: 29 March 2019
46. 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;11(18).

47. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 2006;34(Database issue):W345–9.

48. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(7):573.

49. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2009;4:10.11–14.10.14

50. Domazet-Loso T, Brajovic J, Tautz D. A phylostratigraphic approach to uncover the genomics history of major adaptations in metazoan lineages. Trends Genet. 2007;23(11):533–9.

51. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017;34(7):1812–9.

52. Parr CS, Wilson N, Leary P, Schulz KS, Lans K, Walley L, Hammock JA, Goddard CA, Jager M, Denker E, Wincker P, Da Silva C, Le Guyader H, Manuel M, Ozbek S, Bode H, Holstein TW. Multiple Wnts are involved in Hydra organizer formation and regeneration. Dev Biol. 2009;330(1):186–97.

53. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Couger MB, Eccles D, Li B, Lieber M, et al. De novo transcript sequence assembly from RNA-seq data. Nat Biotechnol. 2012;30(9):511–516.

54. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses of large gene lists using RAxML bioinformatics resources. Nat Protoc. 2009;4(1):44–57.

55. 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 with CASTeL. Nat Biotechnol. 2011;29(8):765–70.

56. Zdobnov EM, Apweiler R. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001;17(9):847–853.

57. Wang L, Feng Z, Wang X, Zhang X. DEGseq: an R package for identification of differentially expressed genes in RNA-seq data. Bioinformatics. 2010;26(1):135–44.

58. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Pires TL, Gibson TJ, Maruyama K, Boyer M, et al. Transcriptome assembly and quantification by RNA-Seq reveals unannotated transcripts and isoforms. Nat Biotechnol. 2010;28(5):511–515.

59. Wang L, Feng Z, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. 2010;26(1):136–41.

60. Thibault S, Zhou P, Liu B, Bao Y, Liu Y, Wu D, Wang S, Wang X, Zhang X. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Bioinformatics. 2011;27(12):1505–6.

61. Chiori R, Jager M, Denker E, Wincker P, Da Silva C, Le Guyader H, Manuel M, Ozbek S, Bode H, Holstein TW. Multiple Wnts are involved in Hydra organizer formation and regeneration. Dev Biol. 2009;330(1):186–97.

62. Seidel U, Bober E, Winter B, Lenz S, Lohse P, Goedde HW, Grzeschik KH, Arnold H. Alkaline myosin light chains in man are encoded by a multigene family that includes the adult skeletal muscle, the embryonic or atrial, and noncardiac isoforms. Hum Genet. 1986;66(1):135–46.

63. Sutterli RA. Do jellyfish have central nervous systems? J Exp Biol. 2011;214(8):1215–23.

64. Lesh-Laurie GE, Suchy PE. Cnidaria: scyphozoa and cubozoa. Microsc Anat Rev. 1991;17(7):422–432.

65. Robinson-Beaudet L, Desjardins CA, Yuen R, Chiocca D, Kim J, Stewart K, Hackett H, Hackett A, Veech M, Finney JS, et al. A comprehensive evaluation of automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.

66. Ellegren H, Gautier N. Determinants of genetic diversity. Nat Rev Genet. 2016;17(7):422–33.

67. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764–70.

68. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011;27(4):578–9.

69. Salmela L, Walve R, Rivals E, Ukkonen E. Accurate self-correction of errors in long reads using de Bruijn graphs. Bioinformatics. 2017;33(6):799–806.

70. McCoy RC, Taylor RW, Blauwkamp T, Kelley JL, Kertesz M, Pushkarev D, Petrov A, Fiston-Lavier AS, Liu B, Estelle J. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform. 2013;14(6):671–83.

71. Sutterli RA. Do jellyfish have central nervous systems? J Exp Biol. 2011;214(8):1215–23.
98. Alonso CR. Hox proteins: sculpting body parts by activating localized cell death. Curr Biol. 2002;12(22):R776–8.
99. Kuhn K, Streit B, Schierwater B. Isolation of Hox genes from the scyphozoan Cassiopeia xamachana: implications for the early evolution of Hox genes. J Exp Zool. 1999;285(1):63–75.
100. Komiya Y, Habas R. Wnt signal transduction pathways. Organogenesis. 2008; 4(2):68–75.
101. Haas B, Papanicolaou A. TransDecoder (find coding regions within transcripts); 2016.
102. Duong V, Rochette-Egly C. The molecular physiology of nuclear retinoic acid receptors. From health to disease. Biochim Biophys Acta. 2011;1812(8): 1023–31.
103. Gutierrez-Mazariegos J, Schubert M, Lauder V. Evolution of retinoic acid receptors and retinoic acid signaling. In: The biochemistry of retinoic acid receptors I: Structure, activation, and function at the molecular level. Dordrecht: Springer; 2014. p. 55–73.
104. Kostrouch Z, Kostrouchova M, Love W, Jannini E, Pietigorsky J, Rall JE. Retinoic acid X receptor in the diploblast, Tripedalia cystophora. Proc Natl Acad Sci U S A. 1998;95(23):13442–7.
105. Chamber P. The nuclear receptor superfamily: a personal retrospect on the first two decades. Mol Endocrinol. 2005;19(6):1418–28.
106. Rhinn M, Dolle P. Retinoic acid signalling during development. Development. 2012;139(5):843–58.
107. Lalevee S, Anns YN, Chatagnon A, Samaran E, Poch O, Lauder V, Benoit G, Lecompte G, Rochette-Egly C. Genome-wide in silico identification of new conserved and functional retinoic acid receptor response elements (direct repeats separated by 5 bp). J Biol Chem. 2011;286(38):33322–34.
108. Mariotti GL, Pane L. Cytotoxic and cytolytic cnidarian venoms. A review on health implications and possible therapeutic applications. Toxins (Basel). 2013;5(1):108–51.
109. Frazao B, Vasconcelos V, Antunes A. Sea anemone (Cnidaria, Anthozoa, Actiniaria) toxins: an overview. Mar Drugs. 2012;10(8):1812–51.
110. Nevalainen TJ, Peuravuori HJ, Quinn RJ, Llewellyn LE, Benzle JA, Fenner PJ, Winkel KD. Phospholipase A2 in cnidaria. Comp Biochem Physiol B Biochem Mol Biol. 2004;139(4):731–5.
111. Talvinen KA, Nevalainen TJ. Cloning of a novel phospholipase A2 from the cnidarian Adamsia carcinopados. Comp Biochem Physiol B Biochem Mol Biol. 2002;133(3):571–8.
112. Johnson LS, Eddy SR, Portugal E. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics. 2010;11:431.
113. Kim H-K, Weber JA, Lee N, Park SG, Cho YS, Bhak Y et al. The genome of the giant Nomura’s jellyfish sheds light on the early evolution of active predation. 2019. BioProject datasets. Available from: https://www.ncbi.nlm.nih.gov/nuccore/PEDN00000000
114. Kim H-K, Weber JA, Lee N, Park SG, Cho YS, Bhak Y et al. The genome of the giant Nomura’s jellyfish sheds light on the early evolution of active predation. 2019. SRA datasets. Available from: https://www.ncbi.nlm.nih.gov/sra/?term=SRA627560