**RESEARCH ARTICLE**                                                                          **Open Access**

# The genome of the giant Nomura's jellyfish 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], Youngjune Bhak[1,2], Nayun Lee[5], Yeonsu Jeon[1,2], Sungwon Jeon[1,2], Victor Luria[7], Amir Karger[8], Marc W. Kirschner[7], Ye Jin Jo[5], Seonock Woo[9,10], Kyoungsoon Shin[11], Oksung Chung[6,12], Jae-Chun Ryu[13], Hyung-Soon Yim[10], Jung-Hyun Lee[10], Jeremy S. Edwards[14], Andrea Manica[15], Jong Bhak[1,2,6,12*] and Seungshic Yum[5,9*]

## 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

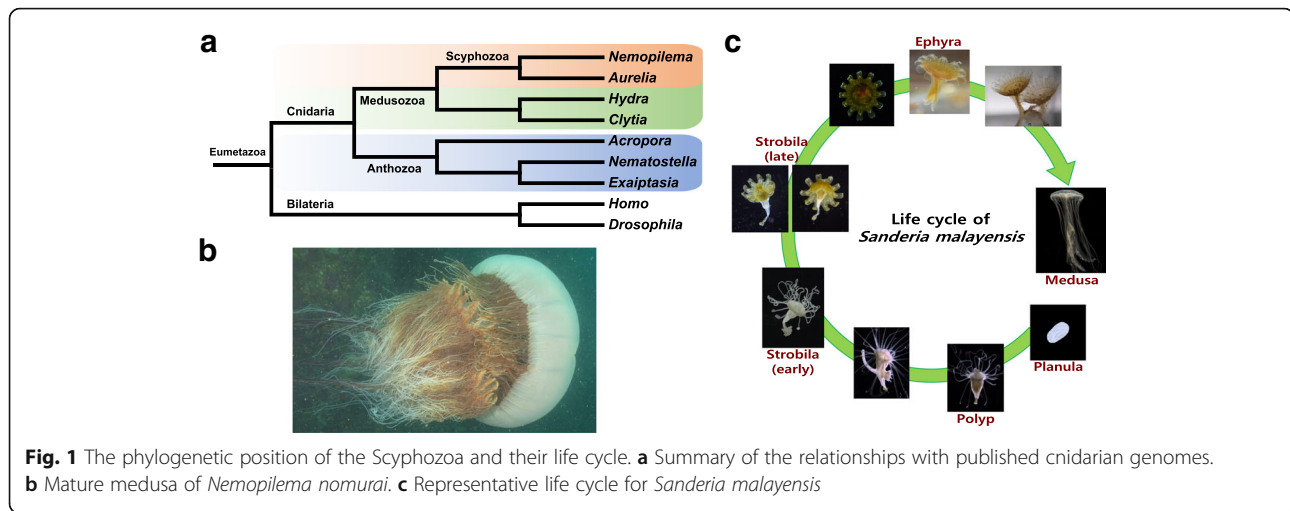* Correspondence: jongbhak@gmail.com; syum@kiost.ac.kr
†Hak-Min Kim, Jessica A. Weber and Nayoung Lee contributed equally to this work.
1Korean Genomics Industrialization Center (KOGIC), Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Republic of Korea
5Ecological Risk Research 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

**Fig. 1** The phylogenetic position of the Scyphozoa and their life cycle. **a** Summary of the relationships with published cnidarian genomes. **b** Mature medusa of *Nemopilema nomurai*. **c** Representative life cycle for *Sanderia malayensis*

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
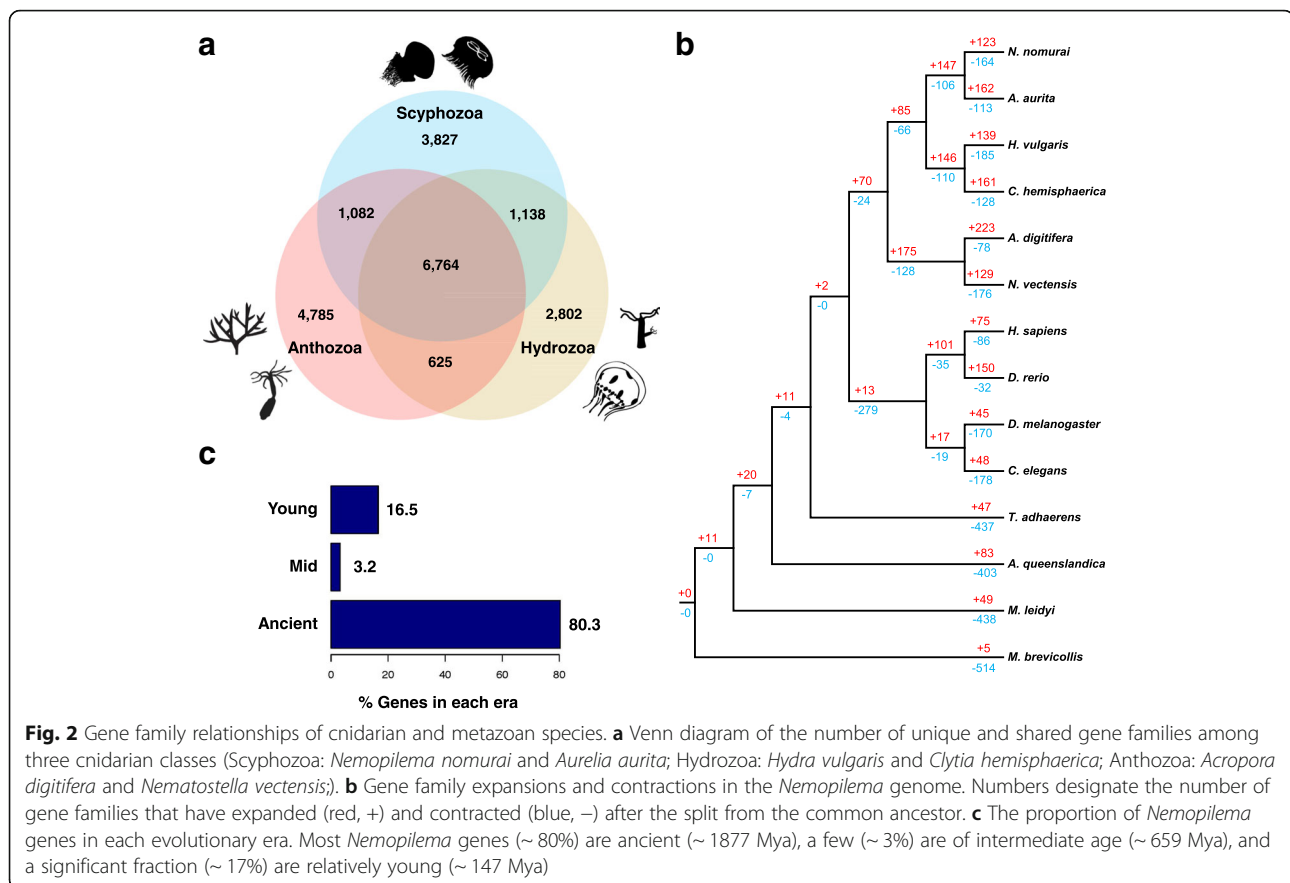### Jellyfish genome assembly and annotation
Here, we present the first de novo genome assembly of the Nomura's jellyfish (*Nemopilema nomurai*; Fig. 1b). It resulted in a 213-Mb genome comprised of 255 scaffolds and an N50 length of 2.71 Mb, containing only 1.48% gaps (Additional file 1: Tables S2 and 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

**Fig. 2** Gene family relationships of cnidarian and metazoan species. **a** Venn diagram of the number of unique and shared gene families among three cnidarian classes (Scyphozoa: *Nemopilema nomurai* and *Aurelia aurita*; Hydrozoa: *Hydra vulgaris* and *Clytia hemisphaerica*; Anthozoa: *Acropora digitifera* and *Nematostella vectensis*;). **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. Most *Nemopilema* genes (~ 80%) are ancient (~ 1877 Mya), a few (~ 3%) are of intermediate age (~ 659 Mya), and a significant fraction (~ 17%) are relatively young (~ 147 Mya)

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, $P = 6.01E–10$) 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, $P = 2.86E–14$; 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, $P = 5.60E–10$) 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* undergone 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
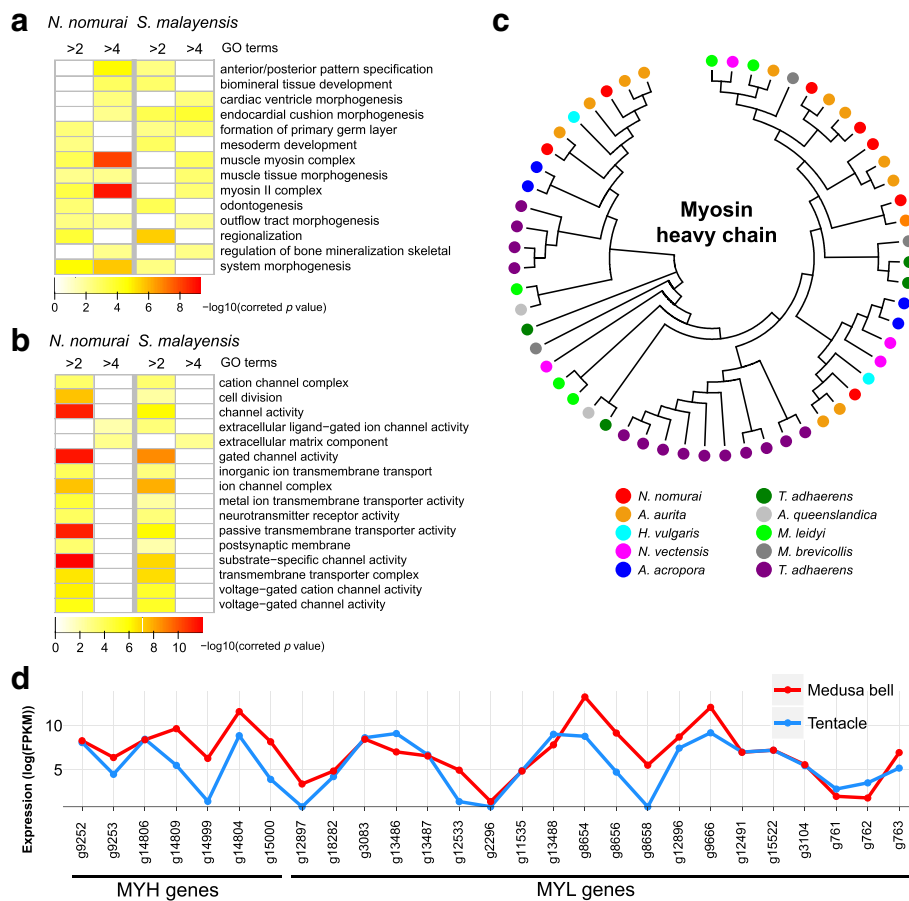


**Fig. 3** Gene expression patterns of medusa bell and tentacle tissues and expansion of myosin heavy chain genes in jellyfish. **a** *P* value heatmap of enriched GO categories using highly expressed genes in medusa bell tissue. Greater than twofold and fourfold higher expression in medusa bell than 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 JTT 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

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 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 [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].

### 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 [22–24]. In total, 83 homeodomains were found in *Nemopilema*, while 82, 41, 120, and 148 of homeodomains were found from *Aurelia*, *Hydra*, *Acropora*, and *Nematostella*, respectively (Additional file 1: Table S41). Five of the eight Hox genes in *Nemopilema* are of the posterior type that are associated with aboral axis development [24] and clustered with *Nematostella's* posterior Hox genes, HOXE and HOXF (Additional file 1: Figures S18–S20). *Aurelia* has six posterior type Hox genes, but does not have the HOXB, C, and D type (HOX2 in humans). Though absent in *Hydra* and *Acropora*, synteny analyses of Para-Hox 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* and *Aurelia*, 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 [25]. *Nemopilema* possesses 13 *Wnt* orthologs representing 10 *Wnt* subfamilies (Additional file 1: Figure S22; Table S42). *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*, *Acropora*, and *Aurelia*), *Wnt10* (*Hydra*), and *Wnt11*, and *Wnt16* (*Nemopilema* and *Aurelia*). It has been proposed that a common cluster of *Wnt* genes (*Wnt1–Wnt6–Wnt10*) existed in the last common ancestor of arthropods and deuterostomes [26]. Our analyses of cnidarian and bilaterian genomes revealed that *Acropora* also possess this cluster, while *Nemopilema*, *Aurelia*, 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 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 to 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
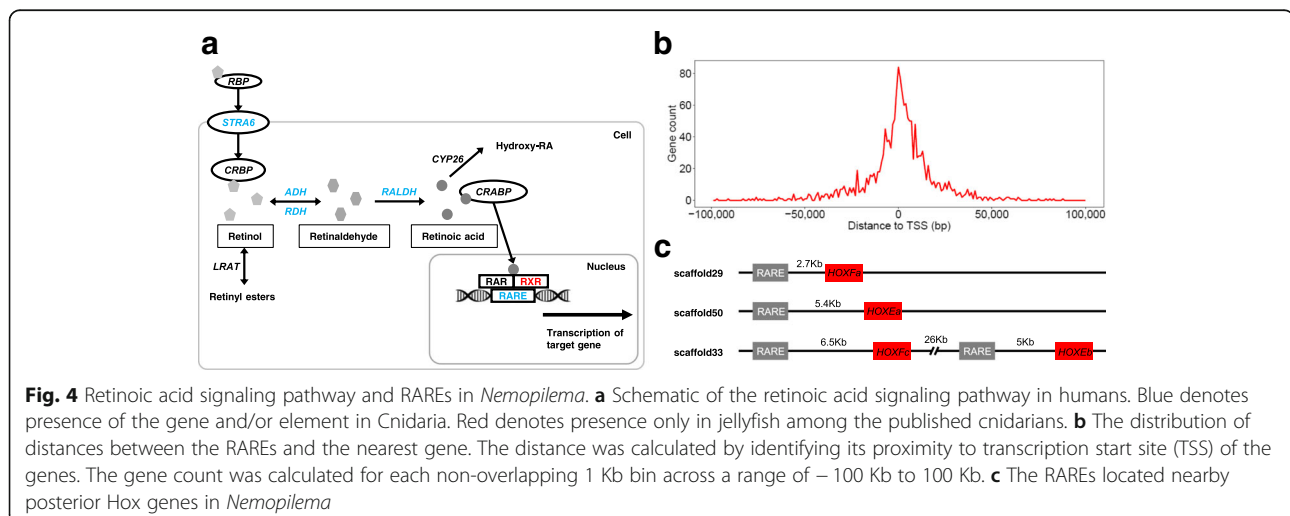
process; Additional file 1: Table S46). Active feeding in the polyp stimulates asexual proliferation either into more polyps or metamorphosis to strobila [27]. Since anthozoans do not form a medusa, the strobila asexual reproductive stage is an important stage in which to study the metamorphosis from 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.
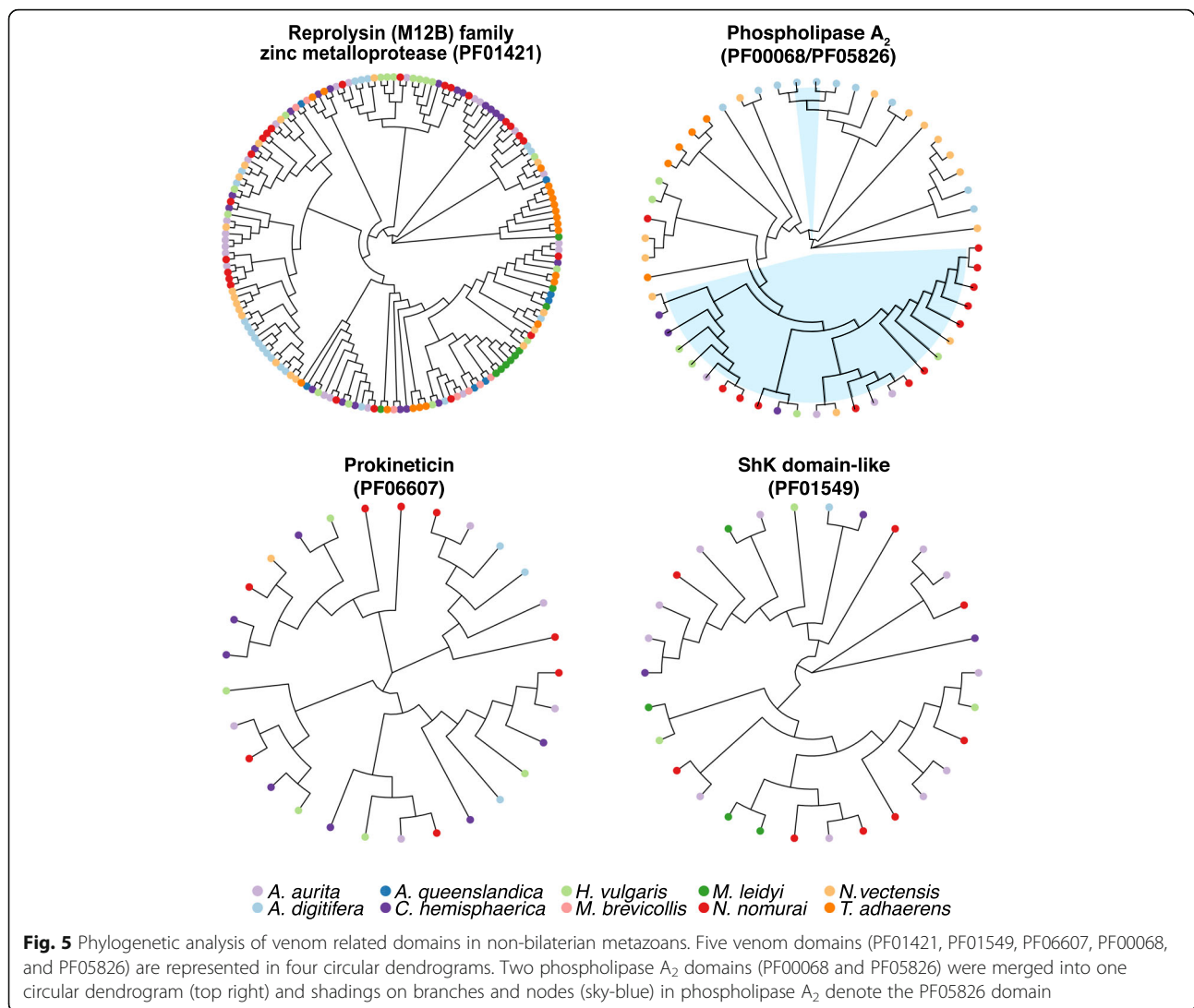
Polyp-to-medusa metamorphosis was previously shown to be strongly associated with *CL390* and *retinoid X receptor* (*RXR*) genes in the *Aurelia aurita* jellyfish [31]. Interestingly, *CL390* was not found in *Nemopilema* or other published cnidarians, suggesting that it may be an *Aurelia*-specific 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 $A_2$ (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*. **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 jellyfish 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. 5** Phylogenetic analysis of venom related domains in non-bilaterian metazoans. Five venom domains (PF01421, PF01549, PF06607, PF00068, and PF05826) are represented in four circular dendrograms. Two phospholipase A$_2$ domains (PF00068 and PF05826) were merged into one circular dendrogram (top right) and shadings on branches and nodes (sky-blue) in phospholipase A$_2$ denote the PF05826 domain

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 A$_2$ 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: 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 [47]. Final *Nemopilema* protein-coding genes were determined using AUGUSTUS with exon (from the homology-based gene prediction) and intron (from the evidence-based gene prediction) hints. Repetitive elements were also predicted using Tandem Repeats Finder [48] and RepeatMasker [49]. 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 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 phylostratum, from the most ancient (PS 1) to the newest (PS 11). To see broad evolutionary patterns, we aggregate the counts from several phylostrata 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

Kim *et al. BMC Biology*       (2019) 17:28

Page 9 of 12

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 *Nematostella* 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 hydrozoan 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 *Nematostella* 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. **Figures S18–S21.** Hox gene analyses. **Figures S22 and S23, Table S42.** Wnt gene analyses. **Tables 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 *Sanderia malayensis*. **Tables S34–S37.** Functional enrichment of tentacles in *Nemopilema nomurai* and *Sanderia malayensis*. (XLSX 51 kb)

### Availability of data and materials
The jellyfish genome project has been deposited at DDBJ/ENA/GenBank under the accession PEDN00000000 [113]. 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) [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, 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 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 other 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
[1]Korean Genomics Industrialization 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,

Kim *et al. BMC Biology*          (2019) 17:28

Page 10 of 12

NM 87131, USA. ⁵Ecological Risk Research Division, Korea Institute of Ocean Science and Technology (KIOST), Geoje 53201, Republic of Korea. ⁶Clinomics Inc., Ulsan 44919, Republic of Korea. ⁷Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA. ⁸IT - Research Computing, Harvard Medical School, Boston, MA 02115, USA. ⁹Faculty of Marine Environmental Science, University of Science and Technology (UST), Geoje 53201, Republic of Korea. ¹⁰Marine Biotechnology Research Center, Korea Institute of Ocean Science and Technology (KIOST), Busan 49111, Republic of Korea. ¹¹Ballast Water Center, Korea Institute of Ocean Science and Technology (KIOST), Geoje 53201, Republic of Korea. ¹²Personal Genomics Institute, Genome Research Foundation, Cheongju 28160, Republic of Korea. ¹³Cellular and Molecular Toxicology Laboratory, Center for Environment, Health and Welfare Research, Korea Institute of Science and Technology (KIST), Seoul 02792, Republic of Korea. ¹⁴Chemistry and Chemical Biology, UNM Comprehensive Cancer Center, University of New Mexico, Albuquerque, NM 87131, USA. ¹⁵Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

## 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(1):329–45.
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). J Morphol. 1981;170(3):383–99.
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(44):17904–9.
6. Dong Z, Liu D, Keesing JK. Jellyfish blooms in China: dominant species, causes and consequences. Mar Pollut Bull. 2010;60(7):954–63.
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(19):3210–2.
8. Gold DA, Katsuki T, Li Y, Yan X, Regulski M, Ibberson D, Holstein T, Steele RE, Jacobs DK, Greenspan RJ. The genome of the jellyfish Aurelia and the evolution of animal complexity. Nat Ecol Evol. 2019;3(1):96–104.
9. Leclère L, Horin C, Chevalier S, Lapébie P, Dru P, Peron S, Jager M, Condamine T, Pottin K, Romano S, et al. The genome of the jellyfish *Clytia hemisphaerica* and the evolution of the cnidarian life-cycle. bioRxiv. 2018. https://doi.org/10.1101/369959.
10. 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(7288):592–6.
11. 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(7360):320.
12. 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(5834):86–94.
13. Galliot B, Quiquand M, Ghila L, de Rosa R, Miljkovic-Licina M, Chera S. Origins of neurogenesis, a cnidarian view. Dev Biol. 2009;332(1):2–24.
14. Shen W, Wang D, Ye B, Shi M, Ma L, Zhang Y, Zhao Z. GC3-biased gene domains in mammalian genomes. Bioinformatics. 2015;31(19):3081–4.
15. Wan X-F, Xu D, Kleinhofs A, Zhou J. Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. BMC Evol Biol. 2004;4(1):19.
16. Tatarinova TV, Alexandrov NN, Bouck JB, Feldmann KA. GC3 biology in corn, rice, sorghum and other grasses. BMC Genomics. 2010;11:308.
17. 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(7406):231–4.
18. Seipel K, Schmid V. Evolution of striated muscle: jellyfish and the origin of triploblasty. Dev Biol. 2005;282(1):14–26.
19. Cegolon L, Heymann WC, Lange JH, Mastrangelo G. Jellyfish stings and their management: a review. Mar Drugs. 2013;11(2):523–50.
20. Leclere L, Rottinger E. Diversity of cnidarian muscles: function, anatomy, development and Regeneration. Front Cell Dev Biol. 2016;4:157.
21. Watanabe H, Fujisawa T, Holstein TW. Cnidarians and the evolutionary origin of the nervous system. Develop Growth Differ. 2009;51(2):167–83.
22. 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(9):920–6.
23. Chourrout D, Delsuc F, Chourrout P, Edvardsen RB, Rentzsch F, Renfer E, Jensen MF, Zhu B, de Jong P, Steele RE, et al. Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. Nature. 2006; 442(7103):684–7.
24. 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(5675):1335–7.
25. Kusserow A, Pang K, Sturm C, Hrouda M, Lentfer J, Schmidt HA, Technau U, von Haeseler A, Hobmayer B, Martindale MQ, et al. Unexpected complexity of the Wnt gene family in a sea anemone. Nature. 2005;433(7022):156–60.
26. Nusse R. An ancient cluster of Wnt paralogues. Trends Genet. 2001;17(8):443.
27. Han C-H, Uye SI. Combined effects of food supply and temperature on asexual reproduction and somatic growth of polyps of the common jellyfish Aurelia aurita sl. Plankton Benthos Res. 2010;5(3):98–105.
28. Maiorova TD, Kosevich IA, Melekhova OP. On some features of embryonic development and metamorphosis of Aurelia aurita (Cindaria, Scyphozoa). Ontogenez. 2012;43(5):333–49.
29. Müller WA, Leitz T. Metamorphosis in the Cnidaria. Can J Zool. 2002;80(10): 1755–71.
30. Schmich J, Trepel S, Leitz T. The role of GLWamides in metamorphosis of Hydractinia echinata. Dev Genes Evol. 1998;208(5):267–73.
31. 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(3):263–73.
32. Cunningham TJ, Duester G. Mechanisms of retinoic acid signalling and its roles in organ and limb development. Nat Rev Mol Cell Biol. 2015;16(2):110–23.
33. So EN, Crowe DL. Characterization of a retinoic acid responsive element in the human ets-1 promoter. IUBMB Life. 2000;50(6):365–70.
34. 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(5):721–9.
35. 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(4):551–7.
36. Rawlings ND, Barrett AJ. Evolutionary families of metallopeptidases. Methods Enzymol. 1995;248:183–228.
37. 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 Proteome. 2014;106:17–29.
38. 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.
39. Jouiaei M, Yanagihara AA, Madio B, Nevalainen TJ, Alewood PF, Fry BG. Ancient venom systems: a review on Cnidaria toxins. Toxins (Basel). 2015;7(6):2251–71.
40. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013. https://arxiv.org/abs/1303.3997.
41. 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(16):2078–9.
42. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.
43. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.
44. 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(6):563–9.
45. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2010;27(4):578–9.

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;1(1):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(Web Server issue):W435–9.

48. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):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, Brajkovic J, Tautz D. A phylostratigraphy approach to uncover the genomic 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 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;2:e1079.

53. Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13(9):2178–89.

54. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22(21):2688–90.

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 using CAFE 3. Mol Biol Evol. 2013;30(8):1987–97.

56. Zdobnov EM, Apweiler R. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics. 2001;17(9):847–8.

57. 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(8):1494–512.

58. 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(3):562–78.

59. 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(1):136–8.

60. 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(6):835–41.

61. 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(1):e4231.

62. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490.

63. 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(1):186–99.

64. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972–3.

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

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

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

68. 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.

69. McCoy RC, Taylor RW, Blauwkamp TA, Kelley JL, Kertesz M, Pushkarev D, Petrov DA, Fiston-Lavier AS. Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. PLoS One. 2014;9(9):e106689.

70. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

71. She R, Chu JS, Wang K, Pei J, Chen N. GenBlastA: enabling BLAST to identify homologous gene sequences. Genome Res. 2009;19(1):143–9.

72. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 2005;6:31.

73. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11.

74. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110(1–4):462–7.

75. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005;21(Suppl 1):i351–8.

76. Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics. 2006;22(23):2971–2.

77. Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. The evolution of mammalian gene families. PLoS One. 2006;1:e85.

78. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4(1):44–57.

79. Bond JS, Beynon RJ. The astacin family of metalloendopeptidases. Protein Sci. 1995;4(7):1247–61.

80. Bork P, Beckmann G. The CUB domain. A widespread module in developmentally regulated proteins. J Mol Biol. 1993;231(2):539–45.

81. Rachamim T, Morgenstern D, Aharonovich D, Brekhman V, Lotan T, Sher D. The dynamically evolving nematocyst content of an anthozoan, a scyphozoan, and a hydrozoan. Mol Biol Evol. 2015;32(3):740–53.

82. Brekhman V, Malik A, Haas B, Sher N, Lotan T. Transcriptome profiling of the dynamic life cycle of the scypohozoan jellyfish Aurelia aurita. BMC Genomics. 2015;16:74.

83. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect Math Life Sci. 1986;17(2):57–86.

84. Herbert A, Rich A. RNA processing and the evolution of eukaryotes. Nat Genet. 1999;21(3):265–9.

85. Seidel U, Bober E, Winter B, Lenz S, Lohse P, Goedde HW, Grzeschik KH, Arnold H. Alkali myosin light chains in man are encoded by a multigene family that includes the adult skeletal muscle, the embryonic or atrial, and nonsarcomeric isoforms. Gene. 1988;66(1):135–46.

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

87. Lesh-Laurie GE, Suchy PE. Cnidaria: scyphozoa and cubozoa. Microsc Anat Invertebrates. 1991;2:185–266.

88. Nakanishi N, Yuan D, Hartenstein V, Jacobs DK. Evolutionary origin of rhopalia: insights from cellular-level analyses of Otx and POU expression patterns in the developing rhopalial nervous system. Evol Dev. 2010;12(4): 404–15.

89. Mazza ME, Pang K, Martindale MQ, Finnerty JR. Genomic organization, gene structure, and developmental expression of three clustered otx genes in the sea anemone Nematostella vectensis. J Exp Zool B Mol Dev Evol. 2007; 308(4):494–506.

90. Epstein RJ, Lin K, Tan TW. A functional significance for codon third bases. Gene. 2000;245(2):291–8.

91. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25(8):1091–3.

92. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11): 2498–504.

93. Campbell NA, Reece JB, Taylor MR, Simon EJ, Dickey J. Biology: concepts & connections, vol. 3: Pearson/Benjamin Cummings; 2009.

94. Dillies M-A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform. 2013;14(6):671–83.

95. Holland PW. Evolution of homeobox genes. Wiley Interdiscip Rev Dev Biol. 2013;2(1):31–45.

96. Park TY, Woo J, Lee DJ, Lee DC, Lee SB, Han Z, Chough SK, Choi DK. A stem-group cnidarian described from the mid-Cambrian of China and its significance for cnidarian evolution. Nat Commun. 2011;2:442.

97. Collins AG, Schuchert P, Marques AC, Jankowski T, Medina M, Schierwater B. Medusozoan phylogeny and character evolution clarified by new large and small subunit rDNA data and an assessment of the utility of phylogenetic mixture models. Syst Biol. 2006;55(1):97–115.

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, Laudet 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, Piatigorsky 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. Chambon 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, Anno YN, Chatagnon A, Samarut E, Poch O, Laudet V, Benoit G, Lecompte O, 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. Mariottini GL, Pane L. Cytotoxic and cytolytic cnidarian venoms. A review on health implications and possible therapeutic applications. Toxins (Basel). 2013;6(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, Benzie 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 carciniopados. Comp Biochem Physiol B Biochem Mol Biol. 2002;132(3):571–8.

112. Johnson LS, Eddy SR, Portugaly 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