nature ecology & evolution

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41559-019-0853-y

In the format provided by the authors and unedited.

Medusozoan genomes inform the evolution of the jellyfish body plan

Konstantin Khalturin[1]*, Chuya Shinzato[1,6], Maria Khalturina[1], Mayuko Hamada[1,7], Manabu Fujie[2], Ryo Koyanagi[2], Miyuki Kanda[2], Hiroki Goto[2], Friederike Anton-Erxleben[3], Masaya Toyokawa[4], Sho Toshino[5,8] and Noriyuki Satoh[1]

[1]Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan. [2]DNA Sequencing Section, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan. [3]Zoological Institute, Christian-Albrechts-Universität, Kiel, Germany. [4]Seikai National Fisheries Research Institute, Nagasaki, Japan. [5]Tropical Biosphere Research Center, University of the Ryukyus, Sesoko, Okinawa, Japan. [6]Present address: Atmosphere and Ocean Research Institute, University of Tokyo, Kashiwa, Chiba, Japan. [7]Present address: Ushimado Marine Institute, Okayama University, Setouchi, Okayama, Japan. [8]Present address: Kuroshio Biological Research Foundation, Otsuki, Kochi, Japan.
*e-mail: konstantin.khalturin@oist.jp

Supplementary Information

Medusozoan genomes inform the evolution of the jellyfish body plan

Konstantin Khalturin[1,*], Chuya Shinzato[1,§], Maria Khalturina[1], Mayuko Hamada[1,#], Manabu Fujie[2], Ryo Koyanagi[2], Miyuki Kanda[2], Hiroki Goto[2], Friederike Anton-Erxleben[3], Masaya Toyokawa[4], Sho Toshino[5,±], Noriyuki Satoh[1]

[1] Marine Genomics Unit, OIST, 1919-1 Tancha, Onna-son, Kunigami-gun, Okinawa, 904-0495 Japan;
[2] DNA Sequencing Section, OIST, 1919-1 Tancha, Onna-son, Kunigami-gun, Okinawa, 904-0495 Japan;
[3] Zoological Institute, Christian-Albrechts-Universität, Am Botanischen Garten 1-9, 24118 Kiel, Germany
[4] Seikai National Fisheries Research Institute, 1551-8 Taira-machi, Nagasaki, 851-2213 Japan;
[5] Tropical Biosphere Research Center, University of the Ryukyus 3422 Sesoko, Motobu, Okinawa 905-0227, Japan

Current address:
[§] Atmosphere and Ocean Research Institute, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, 277-8564 Japan
[#] Ushimado Marine Institute, Okayama University, 130-17 Kashino, Ushimado,Setouchi, Okayama, 701-4303 Japan
[±] Kuroshio Biological Research Foundation, 560 Nishidomari, Otsuki, Hata, Kochi, 788-0333 Japan

* - Corresponding author (konstantin.khalturin@oist.jp)

This file includes:

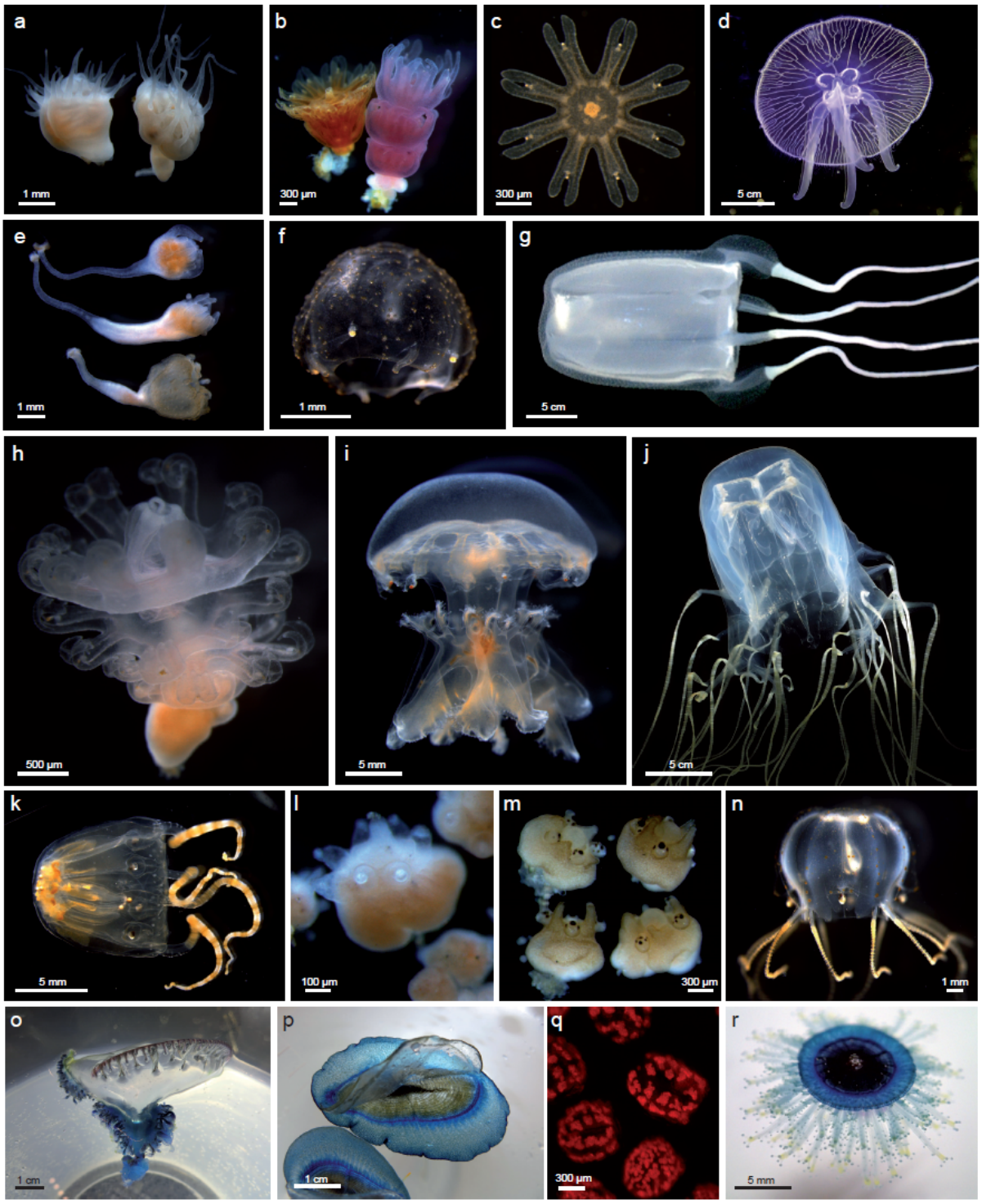


Figure S1: Cnidarian species used for genome and transcriptome sequencing. **(a)** *Aurelia aurita* polyps from Atlantic (left) and Pacific (right) strains. **(b)** Strobilating polyps of Pacific (left) and Atlantic (right) strains of *Aurelia*. **(c)** Ephyra of *Aurelia* (Pacific strain) **(d)** Adult jellyfish of *Aurelia aurita* (Pacific strain). **(e)** Metamorphosis of *Morbakka* polyps resembles monodiscoid strobilation in Scyphozoa. **(f)** *Morbakka* jellyfish 72 hours after the completion of metamorphosis. **(g)** Adult jellyfish of *Morbakka virulenta*. **(h)** Strobila of *Nemopilema nomurai*. **(i)** Jellyfish of *Nemopilema* in laboratory culture four weeks after strobilation. **(j)** Adult jellyfish of *Chironex yamaguchi*. **(k)** Adult male jellyfish of *Copula*. **(l)** Polyp of *Tripedalia cystophora*. **(m)** Initial phase of metamorphosis in *Tripedalia*. Tentacles are transforming into ropalia. **(n)** Adult jellyfish of *Tripedalia*. **(o)** Floating *Physalia* colony. **(p)** Floating *Velella* (sail by the wind) colony **(q)** Symbiotic algae inside the jellyfish of *Velella* (fluorescence microscopy). **(r)** Floating colony of *Porpita*.

Figure S2: Jellyfish as art objects. Depiction of a "fire jellyfish" *Morbakka virulenta* **(a)** and **(b)** *Aurelia* from the book of Tanshu Kurimoto, Japan, late 18th - early 19th century. Images from the Digital Collection of the National Diet Library of Japan (国立国会図書館), DOI 10.11501/1286924. **(c)** *Aurelia aurita* as shown on Panel 98 in the book "Kunstformen der Natur", Ernst Haeckel (Verlag des Bibliographischen Instituts, 1904).

Links:
http://dl.ndl.go.jp/info:ndljp/pid/1286924/4 (*Morbakka* image)
http://dl.ndl.go.jp/info:ndljp/pid/1286924/6 (*Aurelia* image)

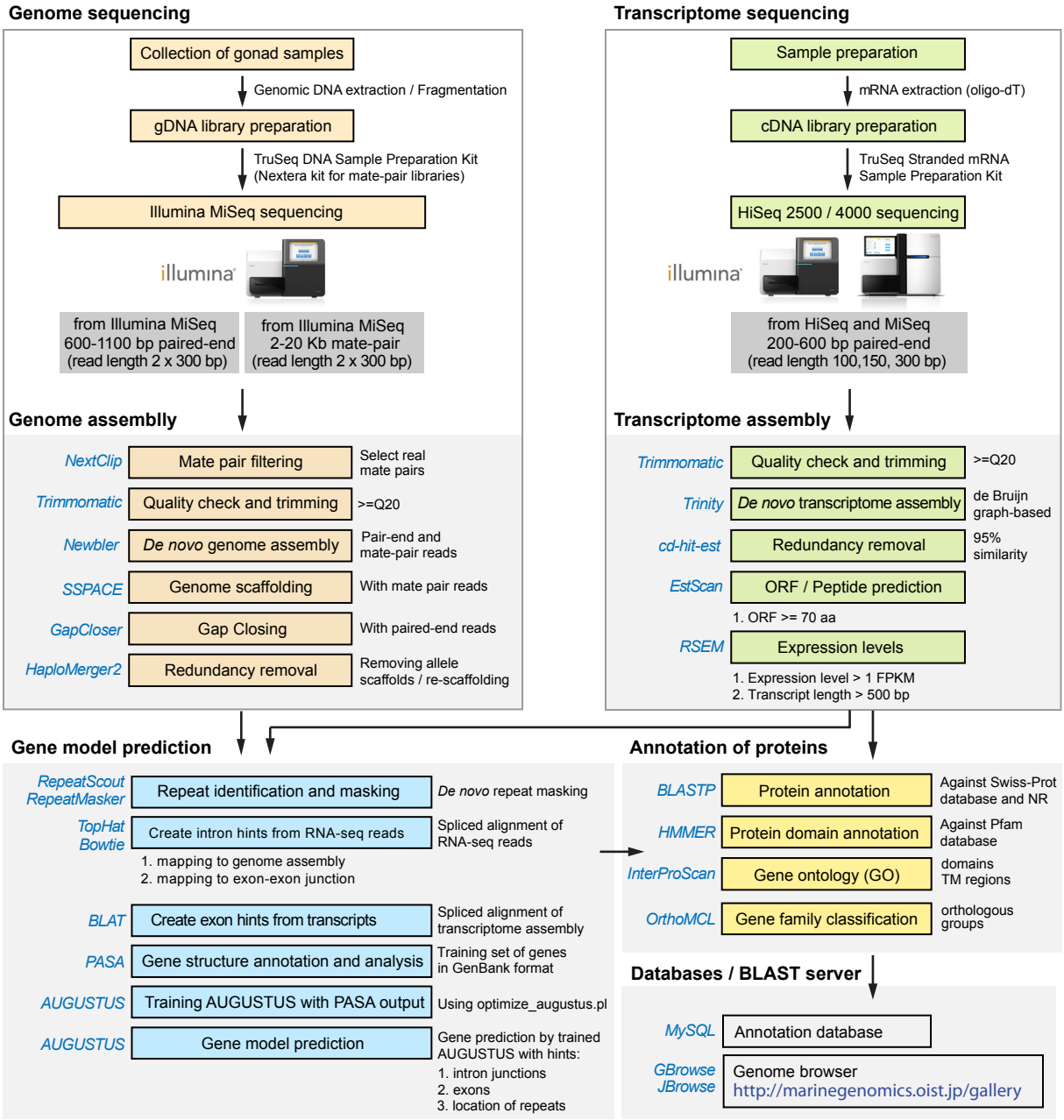


Figure S3: Genome assembly and annotation workflow for *Aurelia* and *Morbakka*.

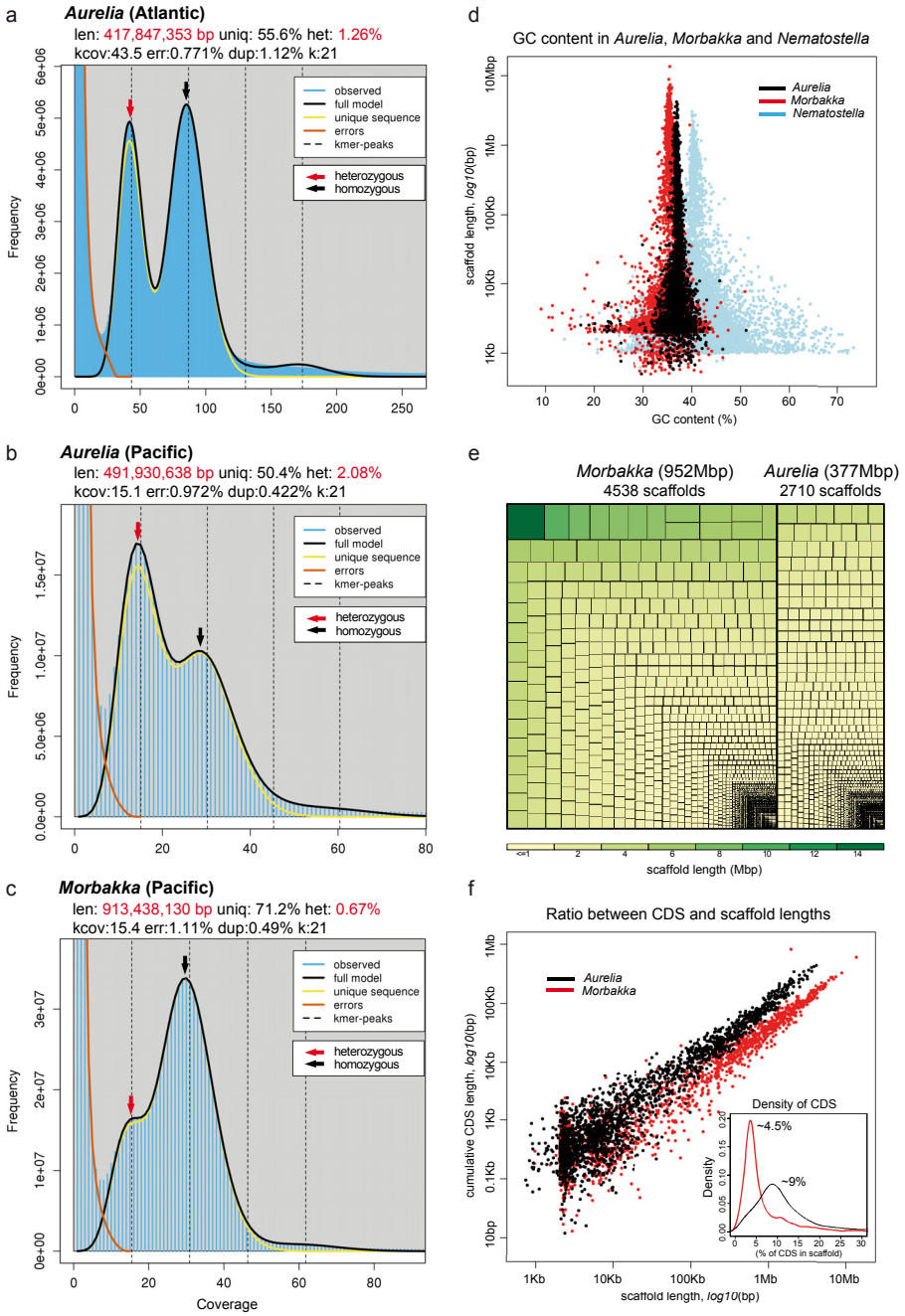


Figure S4: k-mer frequency distribution plots, GC content, scaffold size distribution and coding region densities. **(a)** *Aurelia* strain from the Atlantic Ocean shows nearly identical heights of hetero- and homozygous peaks. Heterozygosity level is estimated to be 1.26%, estimated genome size is 417 Mbp. **(b)** *Aurelia* strain from the Pacific Ocean ("Roscoff" laboratory culture). Estimated heterozygosity level is 2.08% with estimated genome size of 491 Mbp. **(c)** *Morbakka* k-mer frequency plot. A dominant homozygous peak is present. The heterozygosity level is relatively low, at 0.67%. Estimated genome size is 913 Mbp. **(d)** GC content per scaffold in *Aurelia* (black, ABSv1 assembly), *Morbakka* (red) and *Nematostella* (blue). The majority of the scaffolds align close to the mean GC% typical for each species. Scaffolds with GC<30% are relatively short (<10Kb) and contain AT-rich repeats. Scaffolds with GC>60% usually represent bacterial sequences. Dispersal of GC content values is considerably higher in *Nematostella* than in *Aurelia* and *Morbakka*. Stretches of NNNN were removed from all the genome assemblies prior to GC content calculation **(e)** Scaffold length distribution in the genome assemblies of *Morbakka* and *Aurelia*. Each rectangle represents a scaffold in the genome assembly with the area proportional to the scaffold length. The genome of *Morbakka* is ~2.5x larger than that of *Aurelia* and scaffolds are on average 2-fold longer. **(f)** Density of coding regions in the genomes of *Aurelia* and *Morbakka*. Ratio between cumulative length of all coding exons (CDS) and scaffold length is shown for each scaffold, *Aurelia* (black), *Morbakka* (red). About 9% and 4.5% of the genome encodes proteins in *Aurelia* and *Morbakka*, respectively. Coding density is a species-specific constant (see diagonal line in **Fig.S4f**) and can be used for effective contamination removal, especially in combination with GC content plots shown in **Fig.S4d**. In our case the genome assemblies were essentially free of any contaminations (except for *Aurelia* Roscoff strain). Plots in **(a)**, **(b)** and **(c)** were generated using GenomeScope software by k-mer based statistical approach (http://qb.cshl.edu/genomescope/).

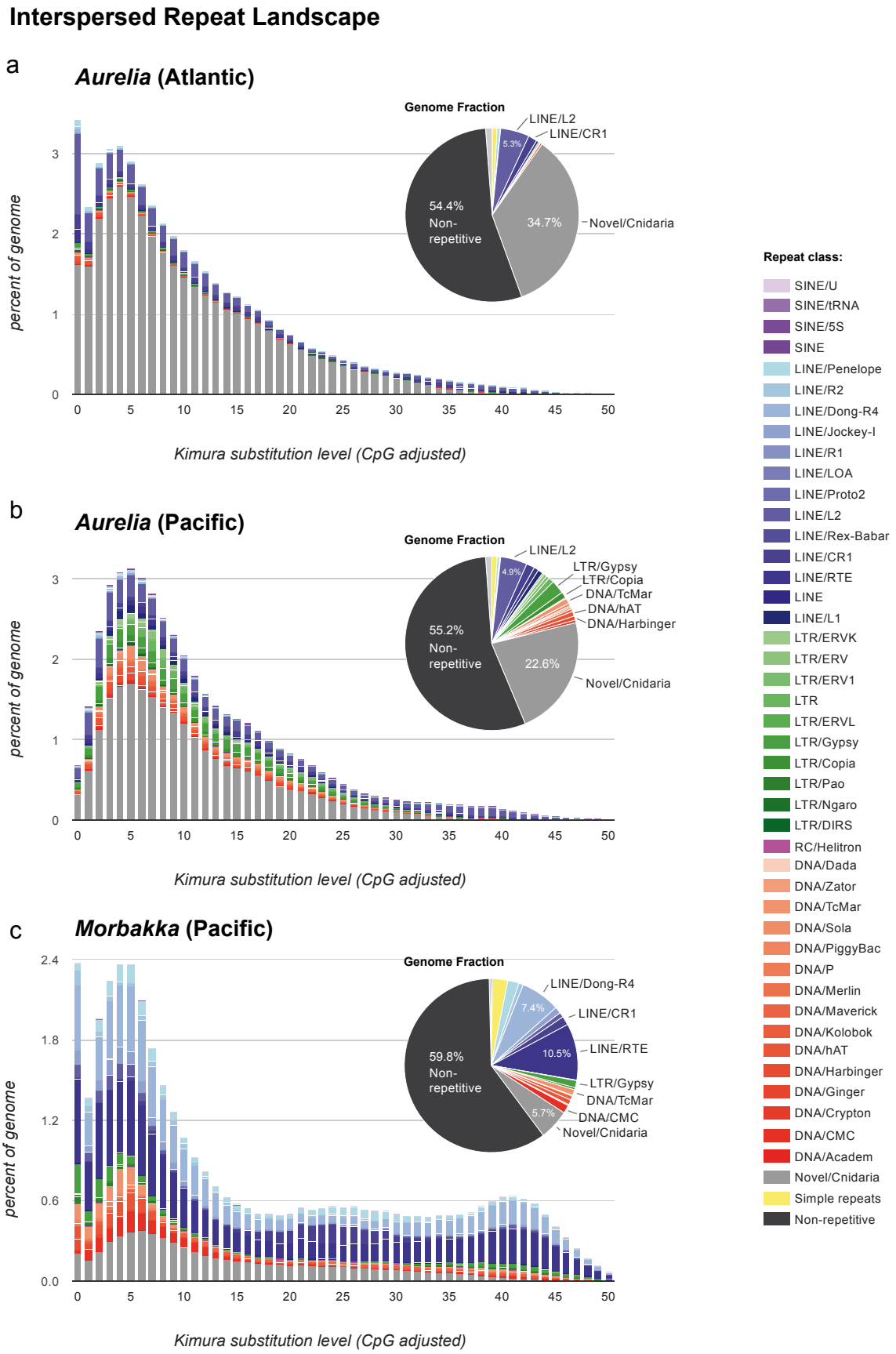


Figure S5: Analysis of repetitive elements. Repeat landscapes and proportions of various repeat classes **(a)** *Aurelia* from the Atlantic Ocean. **(b)** *Aurelia* from the Pacific Ocean. **(c)** The *Morbakka* genome. Different repeat classes are color-coded as shown on the right side of the figure. The non-repetitive fraction of the genome is shown in black. The "Novel/Cnidaria" repeat class represents the elements without any BLASTN or BLASTX hits in the repeat databases used.

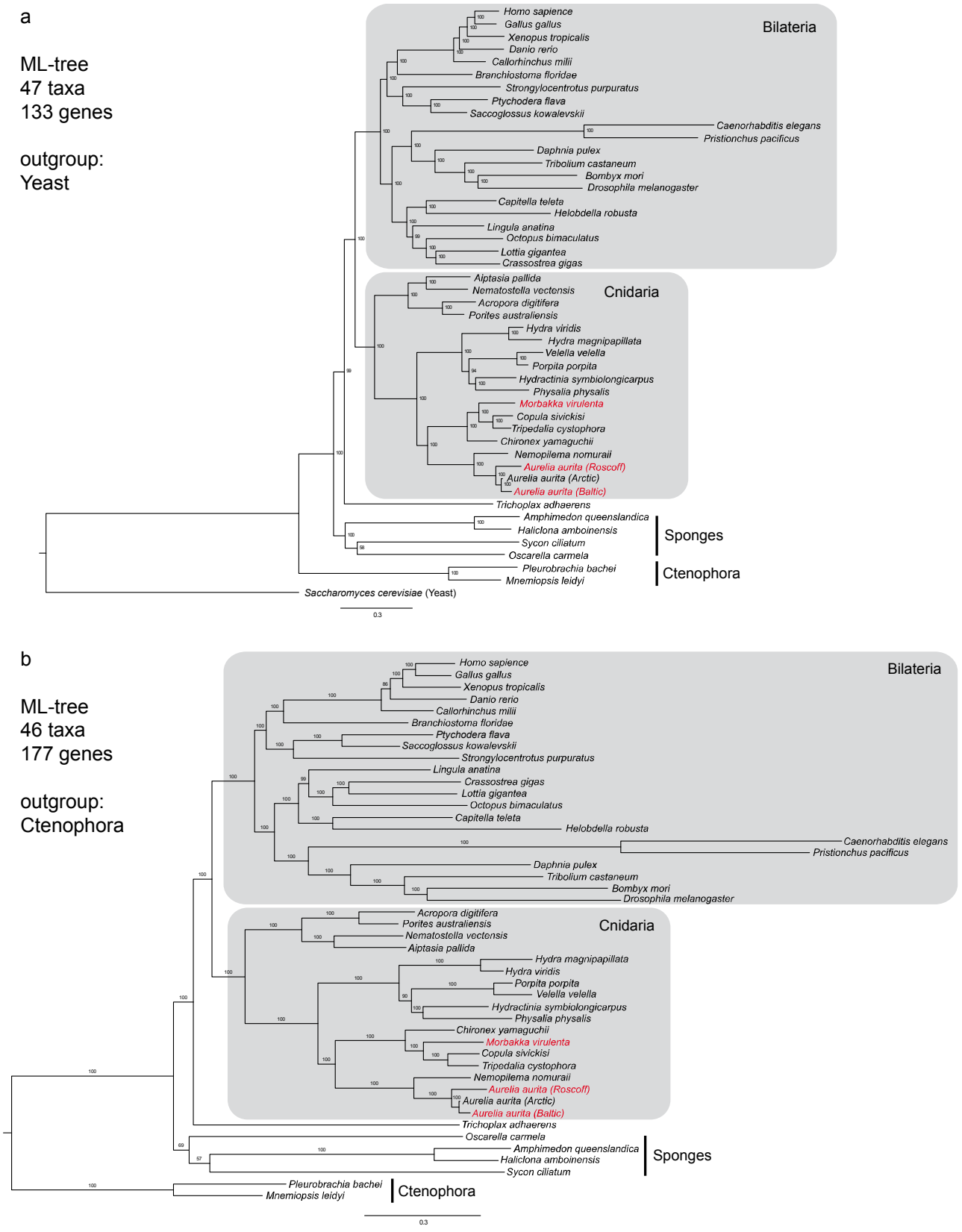


Figure S6: Phylogenetic relations within the Cnidaria. **(a)** A maximum likelihood tree based on 133 one-to-one orthologs from 47 taxa, using the LG substitution model with 100 bootstrap replicates. Yeast was used as an outgroup. **(b)** Maximum likelihood tree based on 177 one-to-one orthologs from 46 taxa, LG substitution model with 100 bootstrap replicates. Two species of Ctenophora were used as an outgroup. Three-letter species code and proteome sources are shown in **Table S5**. apa - *Aiptasia*, nve - *Nematostella*, adi - *Acropora*, Pau - *Poritas*, hvi - *Hydra viridis*, hma - *Hydra magnipapillata*, vel - *Velella*, Por - *Porpita*, hsy - *Hydractinia*, phy - *Physalia*, mvi - *Morbakka*, cop - *Copula*, tri - *Tripedalia*, chi - *Chironex*, nem - *Nemopilema*, ars - *Aurelia* (Pacific, Roscoff strain), aws - *Aurelia* (Arctic, White sea), abs - *Aurelia* (Atlantic, Baltic sea).

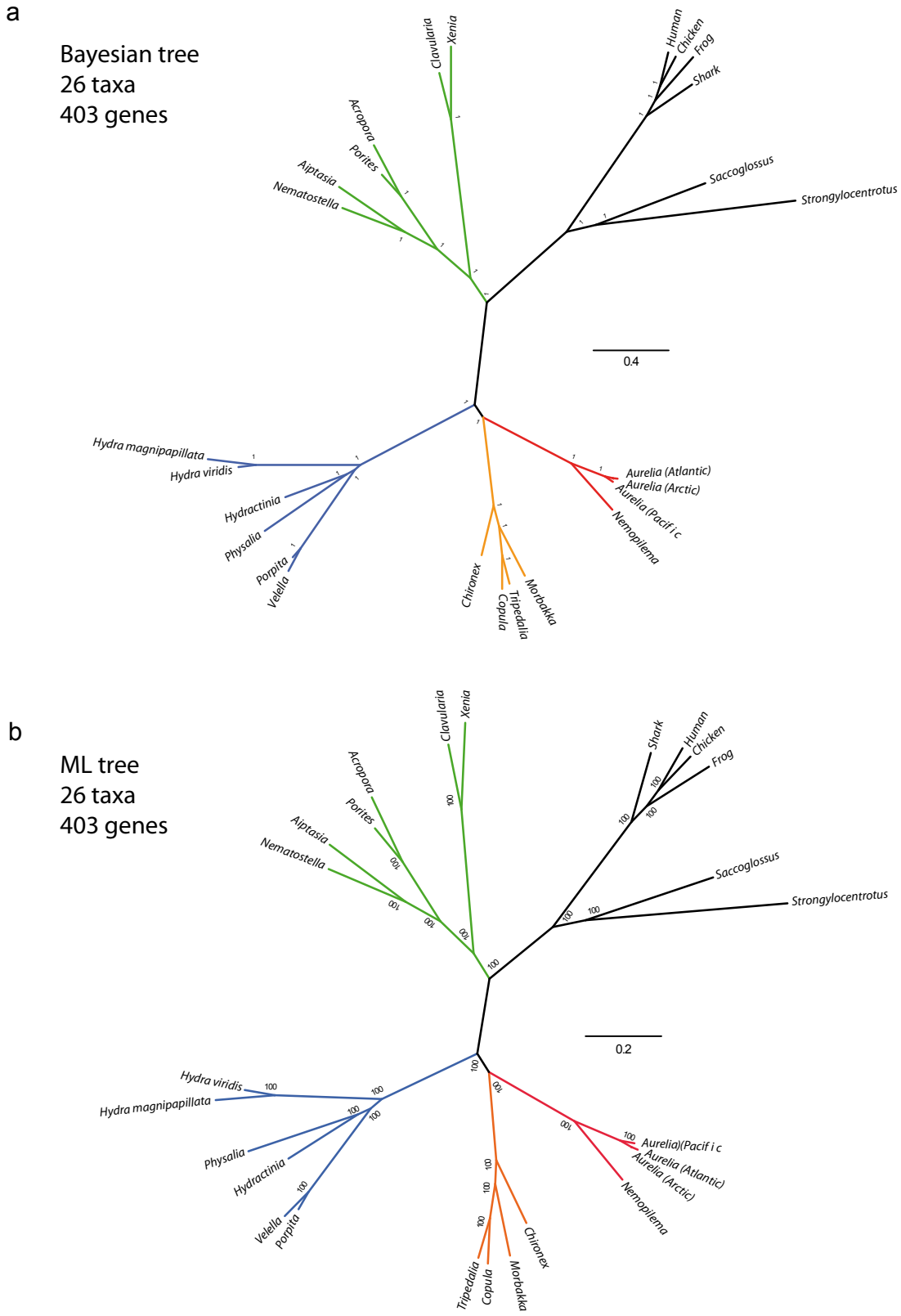


Figure S7: Phylogenetic relationships of 26 species representing the Deuterostomia, Anthozoa, and Medusozoa, based on Bayesian and Maximum likelihood approaches. Cnidarian clades are color-coded - Anthozoa (green), Scyphozoa (red), Cubozoa (orange), Hydrozoa (blue). **(a)** Bayesian analysis was conducted on a supermatrix of 403 orthologs using Phylobayes v1.16j software with global exchange rates inferred from the data (CAT-GTR model). A consensus tree with posterior probabilities at nodes is shown. **(b)** A maximum likelihood tree based on the same set of 403 orthologs, using the LG substitution model. Bootstrap support values are shown at nodes.

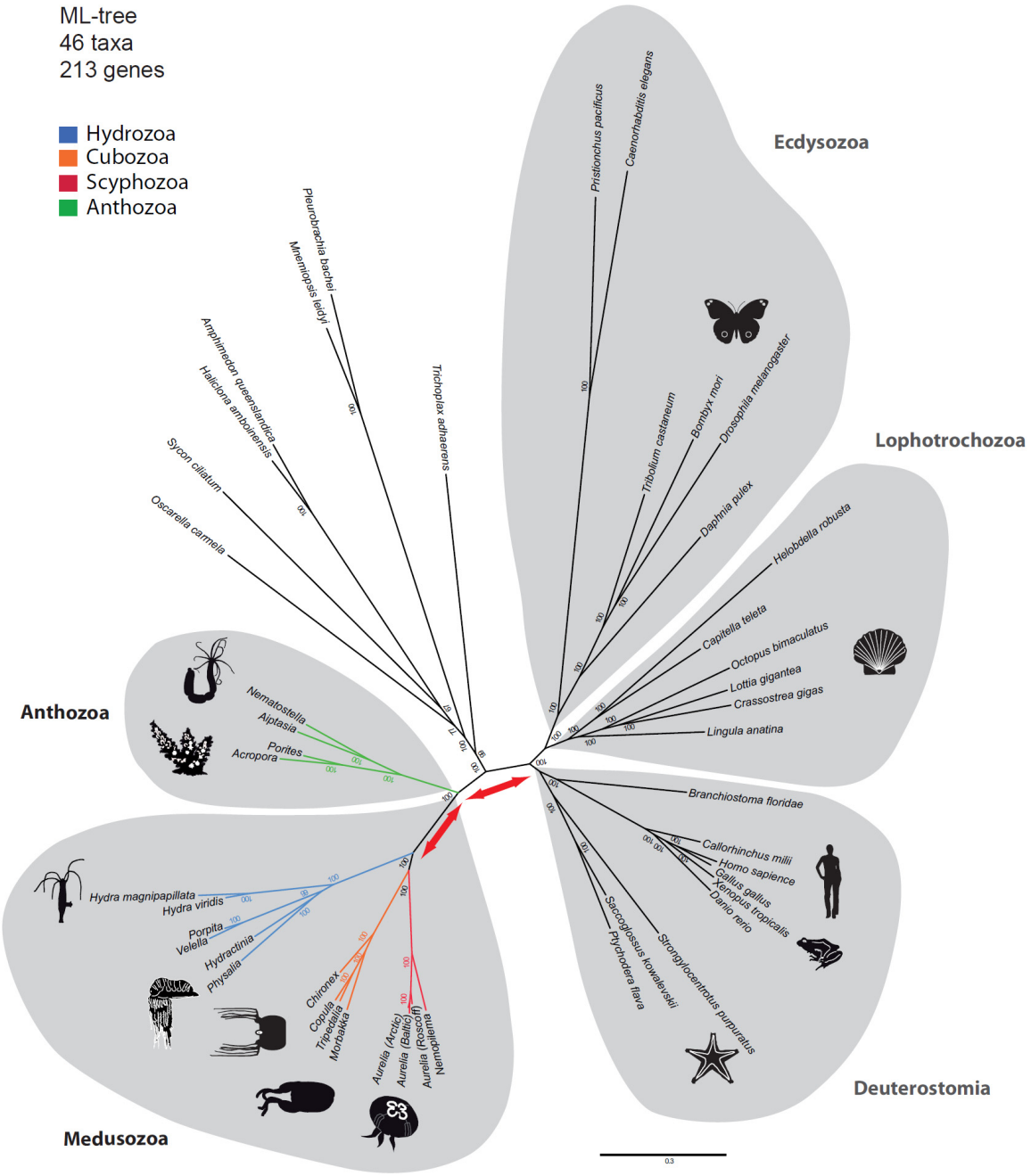
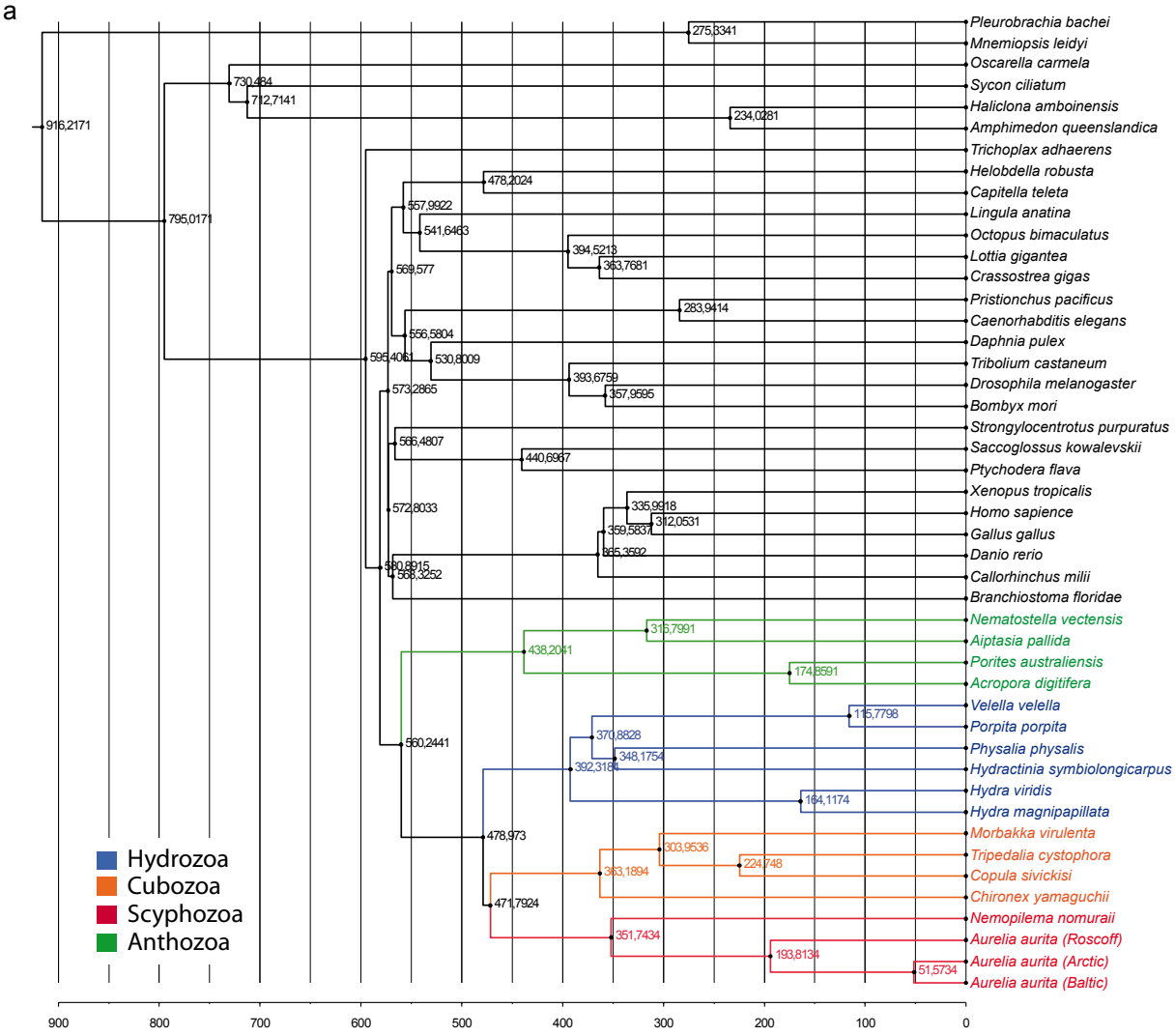


Figure S8: Phylogenetic relationships of 46 taxa representing the Cnidaria, Bilateria, Placozoa, Sponges and Ctenophores inferred by Maximum likelihood approach. LG substitution model with 100 bootstrap replicates was used. Cnidarian clades are color-coded - Anthozoa (green), Scyphozoa (red), Cubozoa (orange), Hydrozoa (blue). Anthozoa and Medusozoa are widely diverged in the cnidarian phylogeny with genetic differences among Hydrozoa, Scyphozoa and Cubozoa being equivalent to the difference among deuterostomian groups. Red arrows show genetic distances between Anthozoa and Medusozoa as well as between Anthozoa and Bilateria.

Molecular dating (complete chromogram of the sub-tree shown in Figure 2A)

a



b

List of fossil calibration constrains used (in Phylobayes format)

```
calibration_46_taxa_984-581_mod02.conf:
************************************
13
hsa gga 330 312
hsa xtr 350 330
hsa bfl 581 518
hsa spu 581 518
cte dme 581 531
cte lgi 581 531
cte hro 581 305
dpu dme 543 510
tca dme 414 307
hma nve 581 520
nve hsa 581 531
mle tad 984 581
adi pau 240 -1

*****************
```

Figure S9: Molecular dating results and fossil calibration settings. **(a)** Complete chronogram of the sub-tree shown in **Fig.2a**. Cnidarian clades are color coded - Anthozoa (green), Scyphozoa (red), Cubozoa (orange), Hydrozoa (blue). Estimated divergence times (in Mya) are shown at each node of the tree. The tree in NEXUS format is available in the Supplementary file, "46taxa_chronogram.nwk". **(b)** A fossil calibration table in Phylobayes format. 3-letter species codes as in **Table S5**.

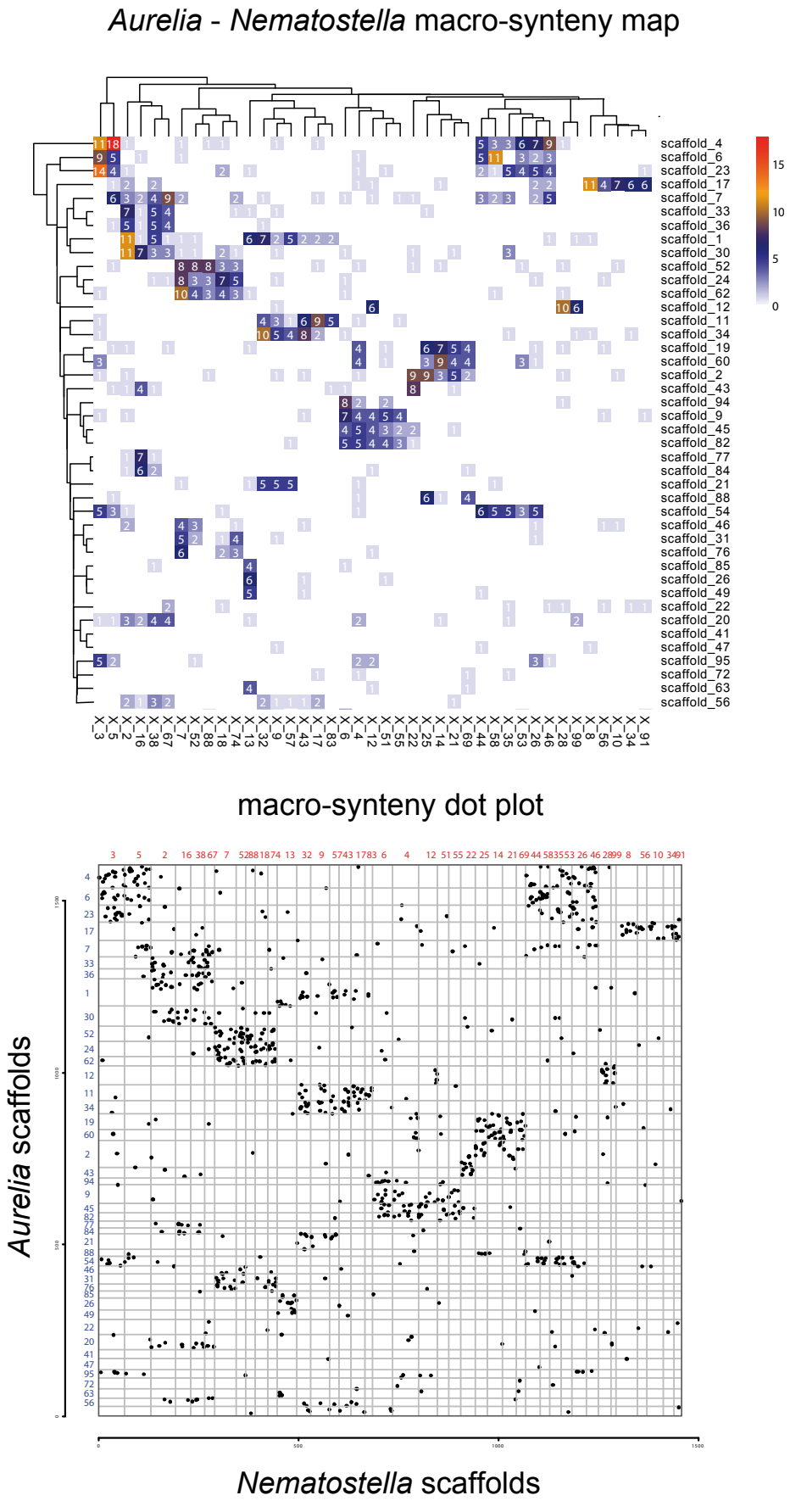


Figure S10: *Aurelia* - *Nematostella* macro-synteny maps. Clustered Oxford grid and dot-plot representations. Numbers of shared orthologs per scaffold pair are shown in the cells of the Oxford grid. For example, *Aurelia* scaffold 4 and *Nematostella* scaffold 3 share 11 orthologs. Grid cells are coloured according to the number of shared orthologs.

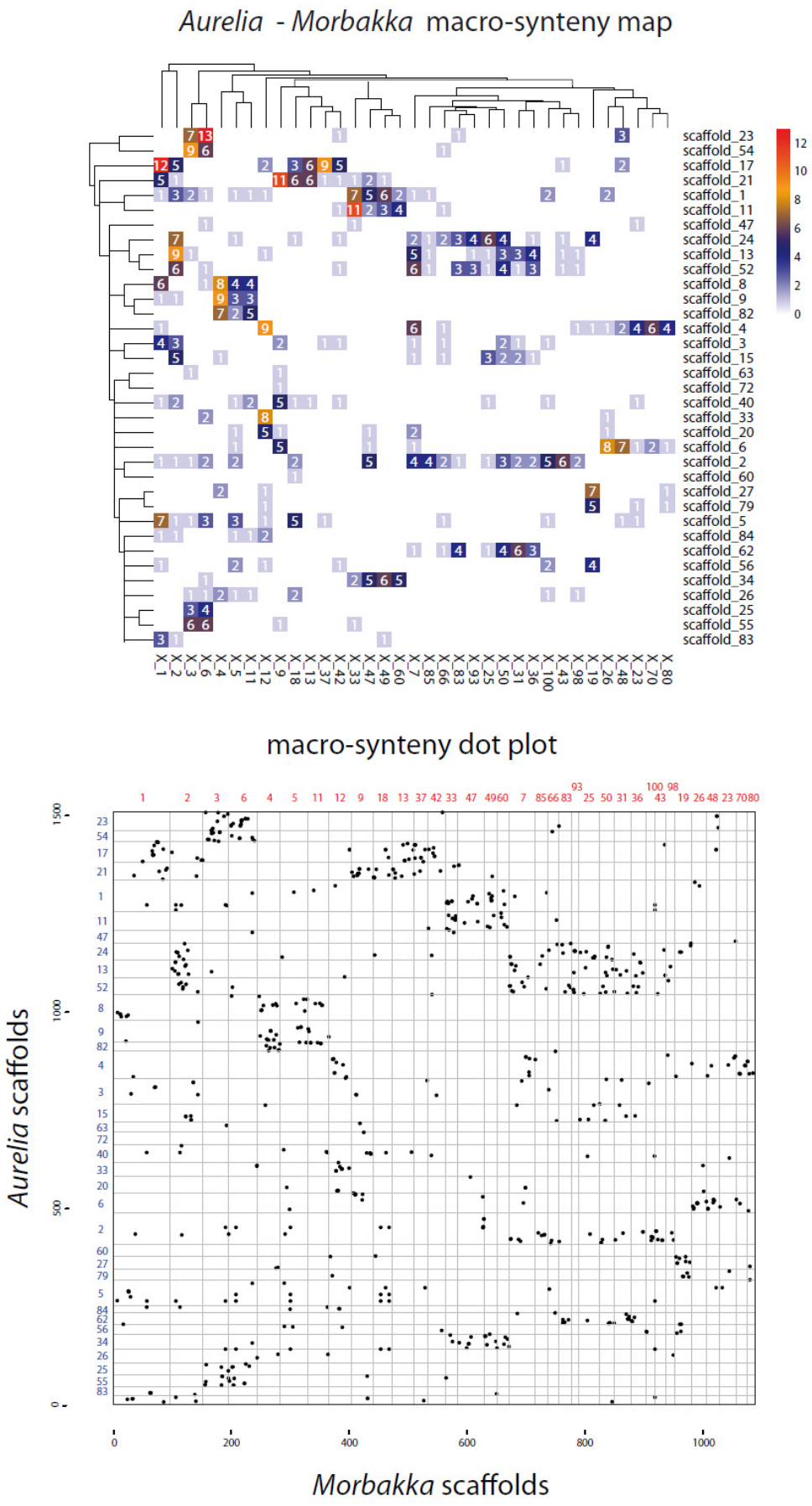


Figure S11: *Aurelia* - *Morbakka* macro-synteny maps.

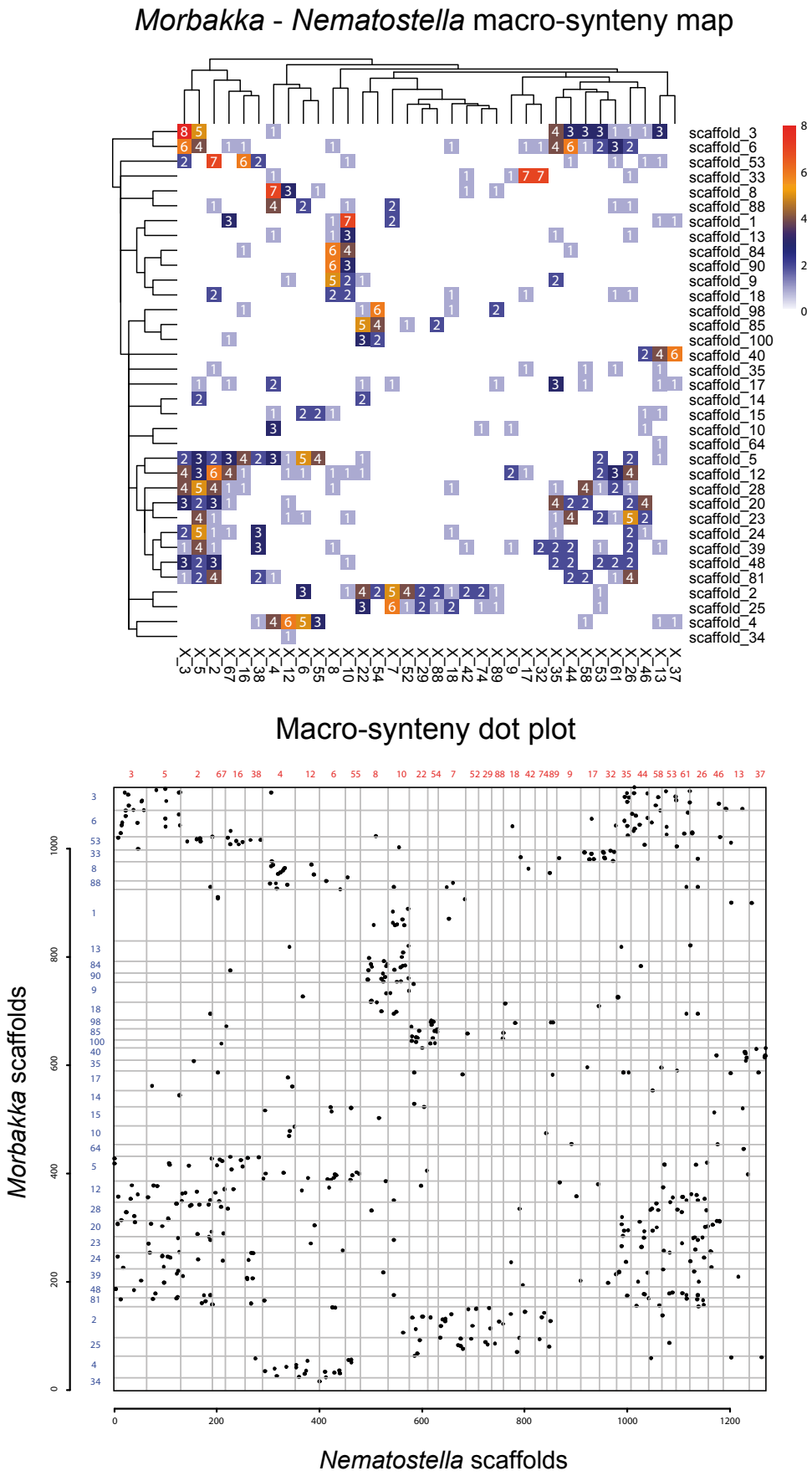


Figure S12: *Morbakka* - *Nematostella* macro-synteny maps.

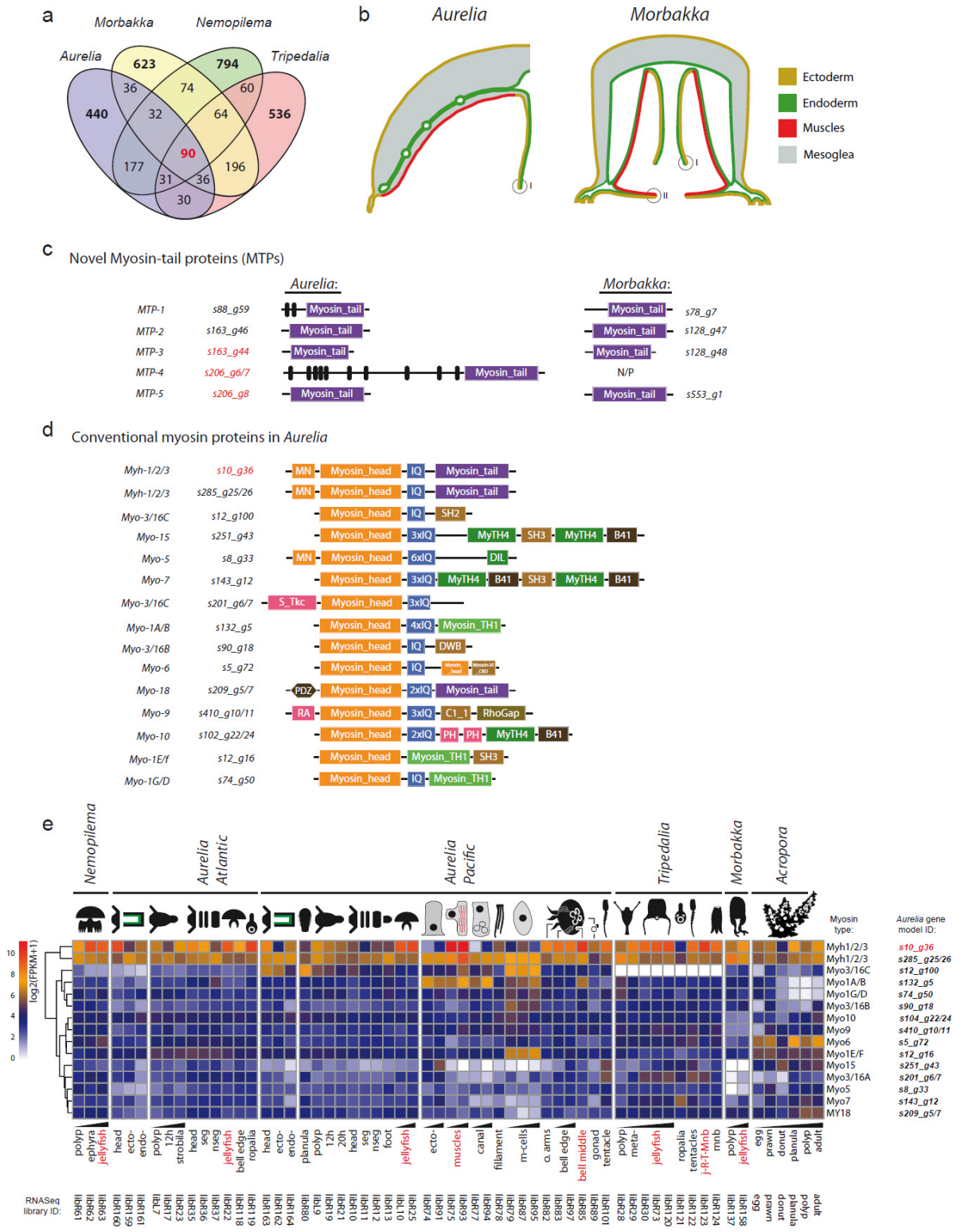


Figure S13: Stage-specific genes**. (a)** A Venn diagram represents the number of shared jellyfish-specific orthologous groups in *Aurelia*, *Morbakka*, *Nemopilema*, and *Tripedalia*. Genes were considered jellyfish-specific if the expression ratio between the jellyfish and polyp stages was >=4, or if genes were exclusively expressed in a jellyfish stage. Expression levels were estimated by RNASeq approach. **(b)** Schematic representation of tissue layers in *Aurelia* and *Morbakka* jellyfish. Layers of striated swimming muscles are shown in red. **(c)** Muscle proteins of *Aurelia* and *Morbakka* with novel domain composition, MTP - myosin tail protein. Gene model IDs are given as ***s88_g59***. Gene IDs with jellyfish-specific expression are shown in red. (**d)** Complement of conventional myosins in *Aurelia* and their domain organization. Names of corresponding bilaterian orthologs are shown on the left side (Myo-1/2/3, Myh-1/2/3 and others). **(e)** Expression of myosin genes in various life stages and tissues of *Nemopilema*, *Aurelia*, *Tripedalia*, *Morbakka*, and *Acropora*. Black-filled wedges indicate that the sampling points shown in the hit map follow each other within the life cycle (e.g. polyp > ephyra > jellyfish or young and adult jellyfishes).

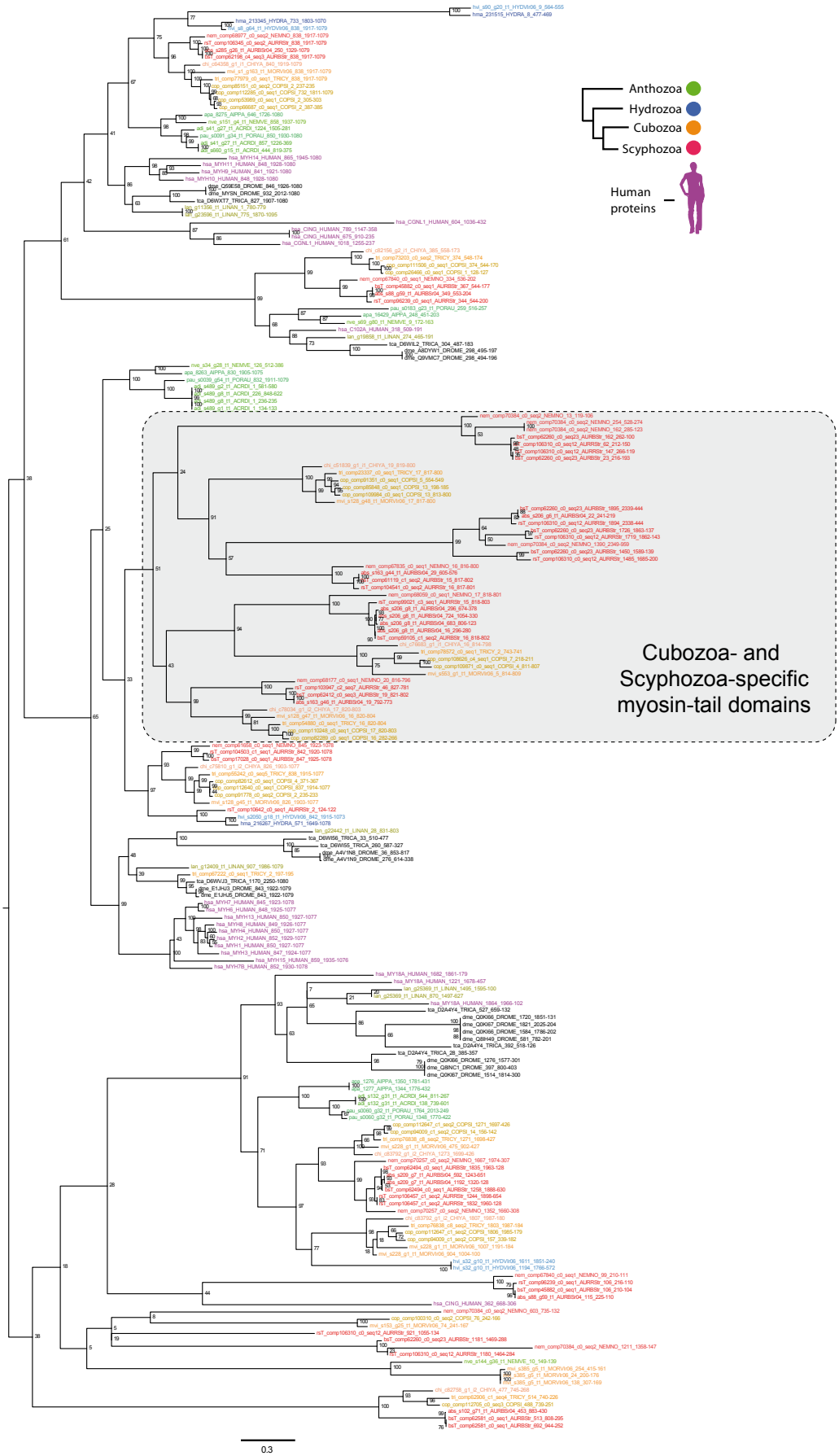


Figure S14: Maximum likelihood tree of Myosin_tail_1 domains (Pfam PF01576.16). Cubozoan and scyphozoan sequences are descendents from the conventional myosins (MYH1, MYH2, MYH3, etc.), but were independently diversified within Medusozoa (grey area). LG substitution model with 100 bootstrap replicates was used.

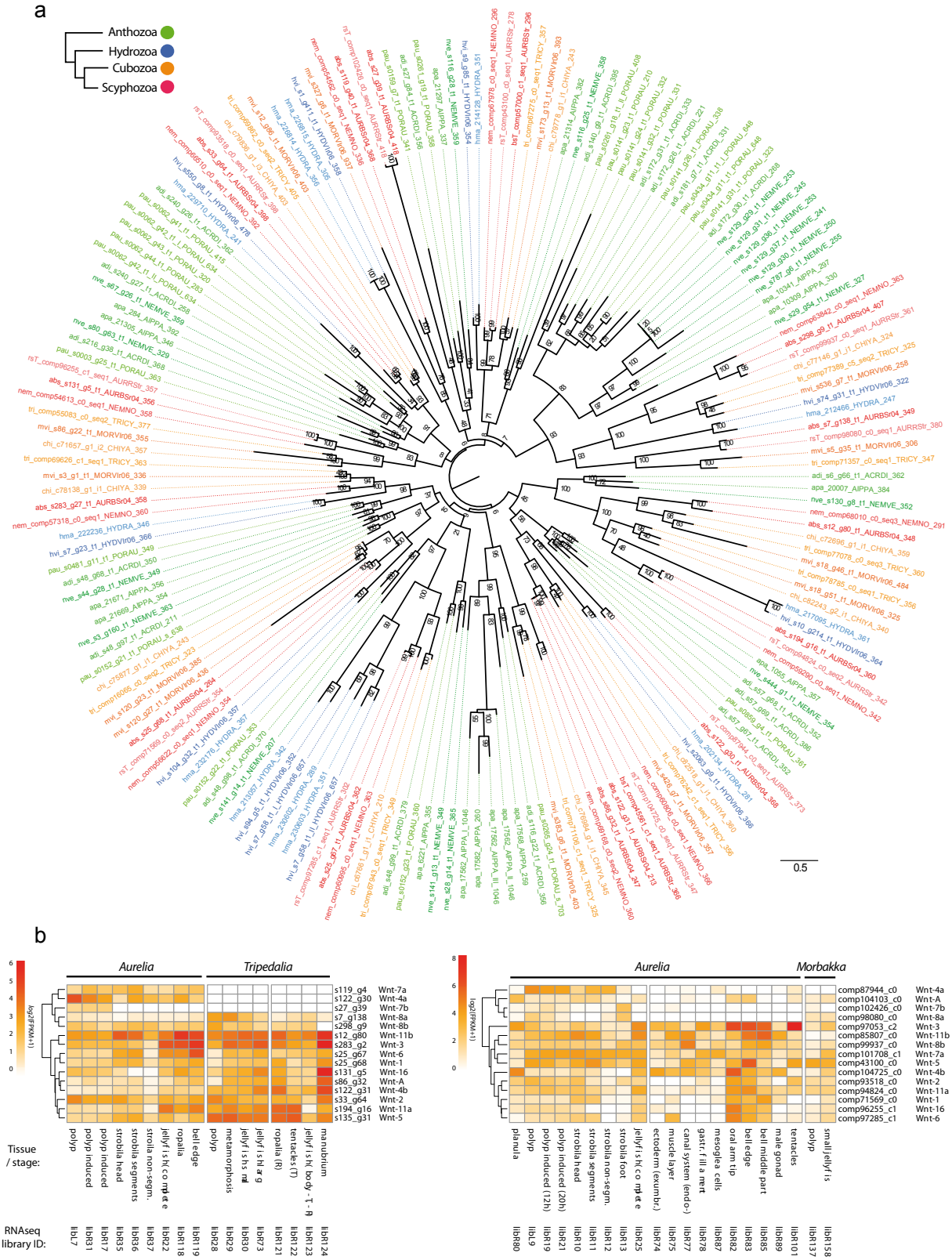


Figure S15: Classification of *Wnt* genes and their expression in *Aurelia*, *Tripedalia* and *Morbakka*. **(a)** Relations among cnidarian *Wnt* genes based on the maximum likelihood approach. Complete sequences of 172 Wnt proteins from the Anthozoa and Medusozoa were used. LG substitution model with 100 bootstrap replicates. Sequence identifiers (IDs) are color-coded (Anthozoa - green, Hydrozoa - blue, Cubozoa - orange, Scyphozoa - red). Each ID includes a 3-letter species code. For example, **'abs_s27_g39_t1_AURBSr04_418'** is an *Aurelia* (Baltic sea) gene located in scaffold 27, gene 39, transcript variant 1, assembly AURBSr04, protein length is 418 amino acids. Species 3-letter codes are listed in the **Table S5**. **(b)** Expression of *Wnt* genes in various life cycle stages and tissues of *Aurelia*, *Tripedalia*, and *Morbakka*.

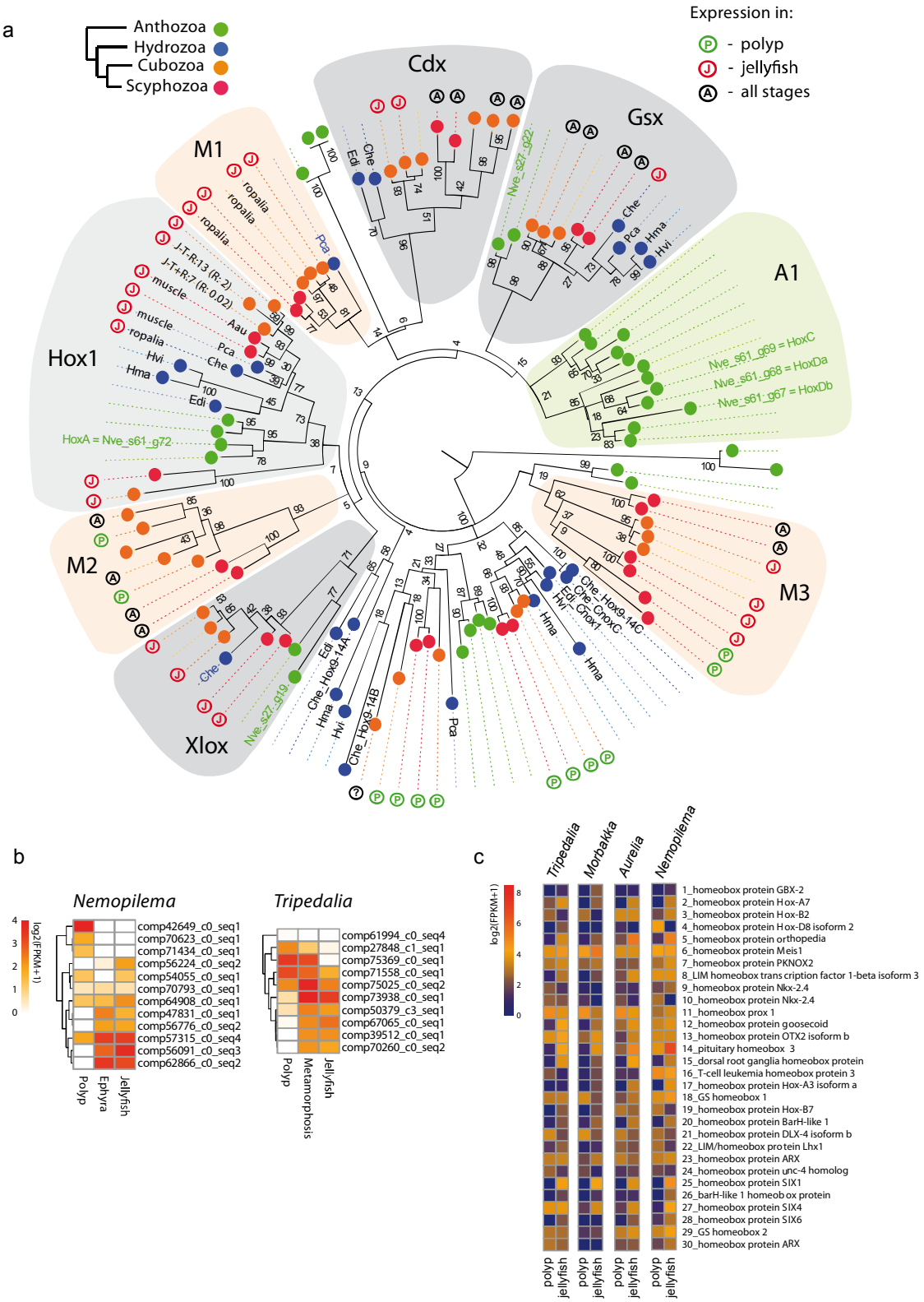


Figure S16: Relations of homeobox genes and their expression in *Nemopilema*, *Tripedalia*, *Morbakka*, and *Aurelia*. **(a)** Phylogenetic relations of Hox and Para-Hox genes of the Medusozoa and Anthozoa. Three groups (M1,M2,M3) of *Hox* genes are Medusozoa-specific. Many of these genes are expressed in jellyfish-specific organs and tissues such as striated muscles and ropalia. Proteins are color-coded: Anthozoa (green), Hydrozoa (blue), Cubozoa (orange), and Scyphozoa (red). An ML tree with the LG substitution model. **(b)** Expression of Hox-like genes during the polyp-to-jellyfish transition in *Nemopilema* and *Tripedalia.* **(c)** Expression of 30 homeobox genes in polyp (P) and jellyfish (J) stages of *Tripedalia*, *Morbakka*, *Aurelia*, and *Nemopilema*. Genes are named according to their best BLASTP hits in the human proteome. The majority of homeobox genes exhibit similar expression dynamics in all four jellyfish species.

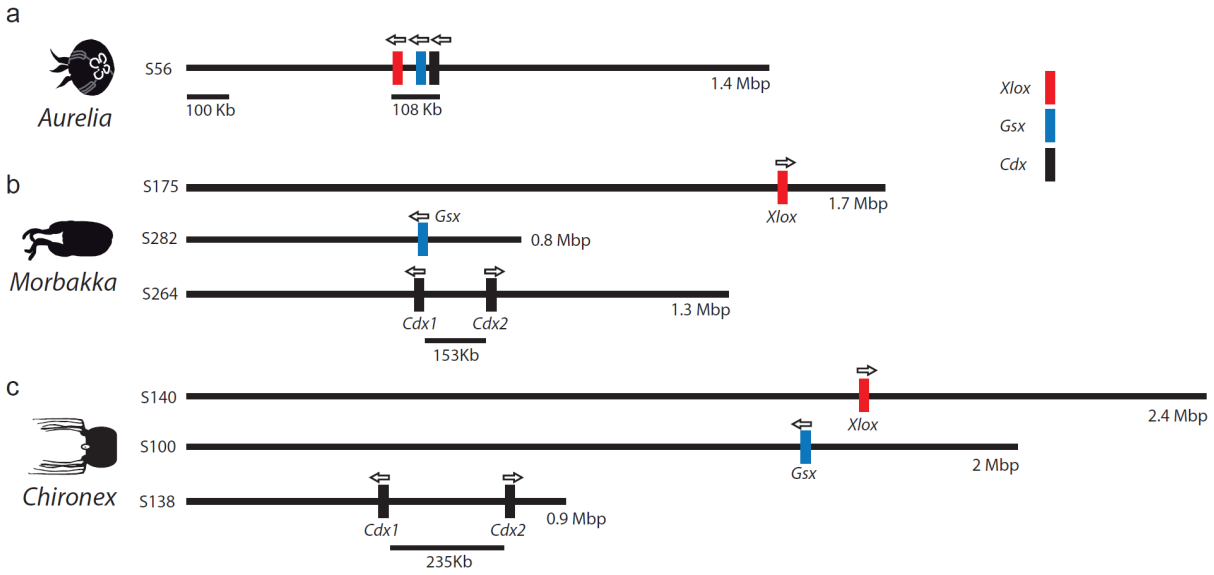


Figure S17: Location of Para-Hox genes in the genomes of *Aurelia aurita*, *Morbakka* virulenta and *Chironex yamaguchii*. **(a)** In the *Aurelia* geneome Cdx>Gsx>Xlox are located as a cluster within 108 Kb in the scaffold 56 (Assembly ABSv1, REGM00000000). *Xlox* is strongly upregulated in a jellyfish stage. **(b)** In *Morbakka*, *Xlox*, *Gsc* and two *Cdx* paralogs are located on three different scaffolds with the flanking regions of at least ~250Kb. Scaffolds from the hybrid assembly "MOR05_r06_PacBio_SS-gc" (combination of Illumina and long PacBio reads) are shown. Assembly is accessible at http://203.181.243.155/aurelia/blast.html. **(c)** In *Chironex*, similar to that in *Morbakka* compact cluster of *Xlox*, *Gsx* and *Cdx* has not been detected. Scaffolds from the draft NanoPore assembly "dbgen_Chironex_Para-Hox_cluster_scaffolds " are shown (http://203.181.243.155/aurelia/).

Supplementary Notes - Materials and Methods

Supplementary Note 1: Background information, material and methods

1.1 Background information on *Aurelia* and *Morbakka*

Aurelia

The moon jellyfish, *Aurelia aurita* (Linnaeus 1758), is one of the most recognizable scyphozoans (Fig.1b,c; Fig.S1a-d; Fig.S2b,c). Inhabiting coastal waters throughout the world, it has been initially considered to be a single, cosmopolitan species with a high level of morphological variability. Later, based on anatomical differences, several additional species with distinct geographic distributions were proposed (*A. limbata*, *A. labiata*, *A. coeurulea*). Finally, phylogenetic studies based on molecular markers revealed a complex of cryptic species consisting of 7 to 9 major clades, which by themselves are composed of numerous local strains[1-3]. Sequence divergence of nuclear DNA marker (ITS-1/5.8S region) among *Aurelia* clades vary from 5% to 32%[2]. If the evolutionary rate of the ITS-1/5.8S locus is roughly assumed to be 1.2% per million years[4], then *Aurelia* strains diverged 4-29 million years ago. However, estimations of ITS divergence times in Schlötterer et al. were calculated for *Drosophila* species and their direct extrapolation to cnidarians is difficult due to the differences in generation times and evolutionary rates.

Current geographic distributions of *Aurelia* strains (and their genetic identities) do not necessarily correspond to their places of origin. Polyps and jellyfish have been extensively transported across oceans as a result of human maritime activities. One such example is the laboratory strain called "Roscoff", which despite of its name originates from the western part of the Pacific Ocean. In general, four clades of *Aurelia*, well distinguishable by genetic markers, roughly correspond to the results of speciation that has been occurring in the Atlantic and Pacific Oceans, as well as in the polar and tropical regions.

In our study, we used three "strains" of *Aurelia aurita* - Baltic sea (ABS), White sea (AWS), and Roscoff (ARS). Two of them (Baltic Sea and White Sea) originate from the Northern Atlantic and were collected in the Bay of Kiel (Kieler Förde, Baltic sea, Germany) and Chupa Bay of the White sea (Керетский рейд, Чупинская губа, Russia), respectively (see Table S1). The "Roscoff" strain was kindly provided by Dr. Gerhard Jarms (University of Hamburg). According to 16S and ITS1/5.8S sequence analysis, the "Roscoff" strain is most related to "ubiquitous" *Aurelia* (see details in Fuchs et al.[5]) and is genetically identical to the animals that inhabit Tokyo Bay and the Sea of Japan near Sasebo and Nagasaki. Baltic Sea and White Sea strains are most related to the "boreal" group according the definition by Schroth et al.[2].

According to our k-mer based estimation[6], the genomes sizes of Baltic and Roscoff strains of *Aurelia* are 417Mbp and 491Mbp, respectively (Fig.S4a,b). Genome size measurements for other strains of *Aurelia* based on cytofluorometry and re-association kinetics gave higher values ranging from 0.71pg (~695±9.8 Mbp)[7] to 0.73pg[8]. Interestingly, the number of chromosomes in *Aurelia* from the Japanese sea (2n=34) and White sea (2n=38) is also different[9].

The Baltic Sea strain (ABS) has ~1.5-fold lower heterozygosity than the "Roscoff" animals (see Fig.S4a,b). Probably the population of *Aurelia* in the relatively narrow Bay of Kiel, which is >5 km long, is to some extent reproductively isolated from the rest of the Baltic sea. Due to the lower level of genetic diversity, which greatly facilitates genome assembly, the Baltic sea strain was selected as a primary target for genome sequencing.

Morbakka

Morbakka virulenta (Kishinouye, 1910) is one of the largest representatives of about 50 Cubozoa species known worldwide (see Fig.1e; Fig.S1g)[10,11]. This species has a relatively small areal and is usually reported along the coasts of Kyushu and Shikoku Islands in Japan, as well as in the Seto Inland Sea. With a bell length of 25 cm and extended tentacles of about 3 metres it is well known to local fisherman as a dangerous species. Due to its painful fiery stings it is called "fire jellyfish" in Japanese (火水母 - "hi kurage") and is well known to the public. For example, vivid natural history illustration of *Morbakka* produced by the Japanese painter and medical doctor Tanshu Kurimoto (栗本丹洲) dates back to the end of 18th century (Fig.S2a).

Morbakka is one of the few cubozoans where the complete life cycle is known and stable polyp culture is available[10]. This species has several morphological and physiological traits that clearly distinguish it from other cubozoans. In contrast to other cubozoans, which preferentially occur in warm tropical waters, adult *Morbakka* medusae appear in fall or at the beginning of winter when the water temperature is around +21^{0}C at the surface. Polyps of *Morbakka* are the largest of all known cubozoan polyps and their shape and size are surprisingly similar to the polyps of scyphozoan jellyfish (Fig.1d; Fig.S1e). While asexual buds of other cubozoans develop into tiny, creeping, worm-like polyps, those of *Morbakka* are free swimming and have a triangular shape with two tentacles. In this respect, they are unique. During metamorphosis, a *Morbakka* polyp undergoes transverse fission and produces one jellyfish in a process that greatly resembles monodiscoid strobilation in Scyphozoa[11] (Fig.S1e). After jellyfish release, the remaining part of the polyp regenerates tentacles and is able to iterate jellyfish production.

Such a combination of attributes, which strongly resemble that of scyphozoans, indicates that *Morbakka* may have retained ancestral traits of the last common ancestor of the Scyphozoa and Cubozoa. That makes this species especially interesting for the investigation of a jellyfish body plan evolution within Medusozoa.

Genomes of cubozoans are large, often exceeding 1 Gbp, which is ~2-3 times larger than the genomes of anthozoans and similar to that in *Hydra magnipapillata*[7,12]. Taking into consideration high level of heterozygosity, which is typical for invertebrates, cubozoan genome sequencing is a challenging task. Fortunately, the population of *Morbakka* from the Seto Inland Sea, along the coast of Okayama Prefecture, seems to be reproductively isolated and has a heterozygosity level of ~0.67% which is unprecedentedly low for any wild-type marine invertebrate. Taken together, the presence of the putative ancestral traits, availability of a polyp culture and a low genetic diversity make *Morbakka* one of the most attractive candidate for the cubozoan genome sequencing (Fig.S4c).

1.2 Biological materials and sampling

1.2.1 *Aurelia*

Extraction of genomic DNA and total RNA is notoriously difficult from Scyphozoa. In *Aurelia*, the standard phenol-chloroform extraction does not allow isolation of high-molecular-weight DNA of sequencing grade (>50-100 Kbp) from somatic tissues. This may be due to the presence of powerful DNAses that are not effectively inhibited by EDTA, but remain in the water phase during chloroform extraction and completely digest genomic DNA before EtOH precipitation. Commercial column-based kits for DNA isolation (Qiagen) also

do not provide satisfactory results. Only DNA isolated from jellyfish spermatozoa is sufficiently intact to allow generation of long mate-pair libraries spanning more than 10 Kb. Thus, high-molecular-weight DNA for genome sequencing was isolated from spermatozoa of a single *Aurelia* jellyfish of the Baltic sea strain, collected in the Bay of Kiel (see Table S1).

Genomic DNA was also extracted from purified mesoglea cells of the Roscoff strain jellyfish because sexually mature male medusae were not available. Mesoglea cells seem to be the only tissue in *Aurelia* where the activity of endogenous DNAses is relatively low. Polyps from the laboratory culture of Roscoff strain were induced to metamorphose using 5-methoxy-2-methyl-indole[5]. Ephyra and medusae were further cultured in the Marine Genomics Unit aquarium facility at OIST. Animals were fed 5 times per week with freshly hatched *Artemia* nauplii. Jellyfish with a bell diameter of 5-7 cm were dissected. Blocks of mesoglea (pure ECM with mesoglea cells without any traces of ectodermal cells or gastrovascular system) were digested with *Clostridium* collagenase (Sigma C0130-100MG, 0.1 mg/ml dissolved in filtered sea water) and mesoglea cells were collected by centrifugation (5 min at 500g). After three rounds of washing in filtered sea water (FSW), mesoglea cells were pelleted, lysed in DNA extraction buffer (10mM Tris-HCl, pH 8.0, 75mM EDTA, 1% N-lauroylsarcosine) and DNA was extracted using the standard phenol-chloroform method. This approach allowed isolation of genomic DNA from Roscoff strain that was sufficiently long to allow generation of ~7Kb mate-pair libraries using a Nextera Kit (Illumina).

1.2.2 *Morbakka*

Morbakka virulenta medusae were collected in Ondo Fishing Port, Hiroshima Prefecture, Japan (N 34.193240, E 132.535295, Table S1). Gonads of three male specimens were dissected and deep frozen. 500 mg of gonad tissue were grinded up in liquid nitrogen, mixed with 5 ml of extraction buffer (10mM Tris-HCl, pH 8.0, 75mM EDTA, 1% N-lauroylsarcosine) and standard phenol-chloroform DNA extraction procedure was performed. Polyps of *Morbakka* were maintained at Sesoko Marine station (Okinawa) and in the aquarium facility of Marine Genomics Unit (MGU) at OIST. Polyp-to-jellyfish transition was induced by 5-methoxy-2-methyl-indole similar to that in *Aurelia*. Polyps were fed three times per week with freshly hatched *Artemia* nauplii. Several hours after feeding water was exchanged. Polyps and juvenile jellyfishes used for RNAseq were starved for 48 hours prior to mRNA extraction.

1.2.3 RNAseq samples

Eleven cnidarian species, representing the Anthozoa, Hydrozoa, Scyphozoa, and Cubozoa, were used as a source of comparative transcriptomic data (Table S2). In all species, mRNA was extracted by oligo-dT affinity chromatography. Complete animals at various stages of the life cycle, their tissues, isolated cells, or body parts were used (Table S2). All sequencing libraries were produced using Illumina TruSeq Stranded mRNA Sample Prep kit, quantified by Real-Time PCR (StepOnePlus, Applied Biosystems) and quality controlled using capillary electrophoresis on Bioanalyzer. Libraries were sequenced on MiSeq and HiSeq2500 instrument using 600-cycle or 2x100-150-cycle chemistry, respectively. In total, 16 reference transcriptomes from 11 species were generated (Table S2). *Aurelia* of "Roscoff" strain and animals from Kujukushima were genetically identical and were considered as belonging to the same species and strain. Polyp cultures of *Aurelia*, *Tripedalia* and *Morbakka* used for transcriptomic sequencing were generated from the single polyps propagated by asexual reproduction. Hence, all individual polyps in the cultures used for RNAseq were genetically identical clones. Usually we used 5-10 polyps for each mRNA extraction in order to average

possible physiological differences in gene expression among the polyps within the culture. Transcriptomes and RNAseq raw reads of *Aurelia* and *Morbakka* were used as hints for gene model prediction (PASA and AUGUSTUS steps in Fig.S3).

1.3 Genome sequencing, assembly, gene prediction and annotation

High-molecular-weight DNA was quality checked using agarose gel electrophoresis, and for paired-end library preparations, it was fragmented by sonication (Covaris M220). Size selection was done by electrophoresis on the BluePippin system (Sage Science). For paired-end libraries, the DNA fraction with a mean fragment size of 600bp was used. Mate-pair libraries of various fragment sizes ranging from 1 Kb to 20 Kb were constructed using Nextera Mate Pair Sample Prep kit (Illimina). Following tranposase-mediated fragmentation, DNA fractions of the desired sizes were selected by electrophoresis on a BluePippin system. Resulting paired-end and mate-pair libraries were sequenced on a MiSeq system with 600-cycle chemistry (2x300 reads). Read lengths of 300bp are preferable for assembly of highly repetitive genomes than shorter reads produced by HiSeq2500/4000 instruments. Longer MiSeq reads allow extraction of larger amounts of valuable sequence information per read from mate-pair libraries. This peculiarity of MiSeq data is highly important and increases the performance of the scaffolding phase.

Raw Illumina reads were quality filtered (Q20, 99% accuracy) and 5 bp were trimmed on both ends to remove sequencing bias and low-quality bases using Trimmomatic (v0.30)[13]. At this step, raw reads were also screened against the collection of Illumina adaptor sequences as well as mitochondrial sequences of *Aurelia* and *Alatina*[14,15]. Raw mate-pair reads were additionally filtered and reverse-complemented with NextClip (v0.8)[16]. Genome sizes were estimated from Illumina raw reads by k-mer distribution analysis (Table 1; GenomeScope at http://qb.cshl.edu/genomescope/, Fig.S4a-c). Jellyfish v2.0.0 was used to count k-mers and their frequencies[17].

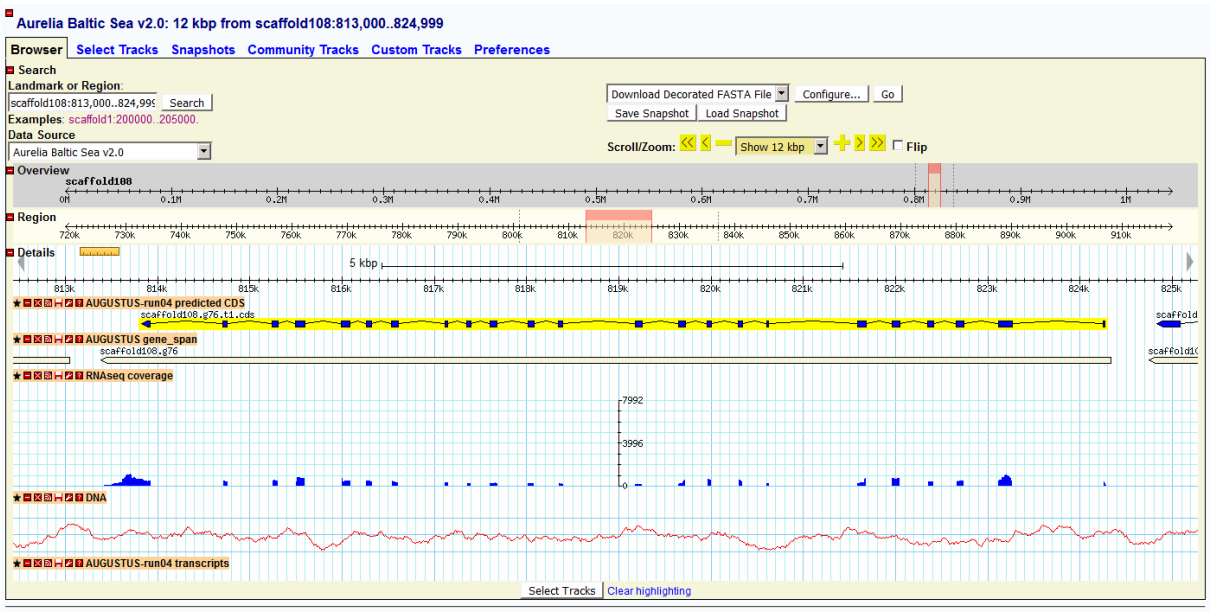
Genome assembly was conducted using Newbler v2.9 software and 37 Gbp (~90x) and 34 Gbp (~30x) Illumina reads for *Aurelia* and *Morbakka*, respectively (see flowchart in Fig.S3). Dedicated Linux servers in the OIST SQC with 64 cores and 750Gb of RAM were used. Assembly times using 48 cores was ~120 hours. Several assemblies with different parameters and amounts of data were performed and their results were compared. The best assemblies were selected for scaffolding. In case of *Morbakka* the assembly with 50x coverage had worse quality metrics and the assembly with 30x coverage was used for scaffolding and further analysis. Assembly with 50x coverage can be accessed at OIST BLAST server (http://203.181.243.155/aurelia/blast) under the name: "dbgen_Morbakka_genome_assembly_MOR06".

Scaffolding step was performed with SSPACE v3.0[18] and mate-pair reads ranging from 1-20 Kbp. GapCloser v1.12 was used for filling gaps in scaffolds. Next, one round of the Haplomerger2 processing pipeline[19] was applied to eliminate redundancy in scaffolds and to merge haplotypes.

Final assemblies contain 377 Mbp for *Aurelia* Baltic sea strain (ABSv1, PRJNA494057, REGM00000000), 429 Mbp for *Aurelia* Roscoff strain (ARSv1, PRJNA494062, REGL00000000) and 952 Mbp for *Morbakka* (MVIv1, PRJNA494059, RDPX00000000). Genome assembly of Baltic sea *Aurelia* is of much higher continuity than that of the Roscoff strain (see Table 1). The major reason for the difference is the origin of the starting genomic DNA. In the Baltic sea strain, genomic DNA was isolated from spermatozoa and the resulting

DNA fragments were larger than 150Kb in size. For the Roscoff strain, sexually mature male jellyfish were not available and DNA was isolated from mesoglea cells. As a result, the maximum length of the genomic DNA was ~20 Kb which precluded generation of mate-pair libraries with insert sizes larger than 7Kb.

The quality and completeness of the genome assemblies were checked by searching for the set of 248 highly conserved eukaryotic genes using CEGMA v2.4[20] and BUSCO 3.0.2[21] (see Table S3, S4). CEGMA values were mostly utilized in order to compare assembly runs performed with different parameters. The major complication with CEGMA and BUSCO usage is that the consensus sequences used by both programs include cnidarian sequences only from *Hydra* and *Nematostella*. As a result, due to large sequence differences, at least 10% of genes are completely missed or referred to as fragmented by both programs (Table S3). Sub-optimal gene detection performance in medusouzoans is obvious from the comparison of BUSCO values for *Aurelia* genome sequence (79.8% complete), predicted protein from the same genome (82.9% complete) and mRNAs (86%). Most problematic for the detections are those *Aurelia* and *Morbakka* genes that contain a lot of short exons. For example, as shown below, gene ***s108_g76*** (ortholog of EOG091G05AH, "WD40 repeat" protein) consists of 23 exons and is referred to as "missing" by BUSCO program.



In reality, as shown above, this gene is present in *Aurelia* assembly ABSv1 and its' structure is correctly predicted. In case of *Morbakka* the discrepancy between the "complete" "genomic" and "mRNA" BUSCO values is even larger, being 81.5% and 90.0%, respectively.

BUSCO analysis is a convenient method, but the results should be taken with care especially in the cases when no proteins from any closely related species have been used for the construction of consensus sequences utilized in TBLASTN and HMM searches (in the phylogenetic tree *Hydra* and *Nematostella* are very far away from *Aurelia* and *Morbakka*, see Fig.S7,S8). Probably in order to get the proper impression about the completeness one should consider BUSCO values for "predicted mRNAs" in C+F (complete + fragmented) column in the Table S3. There *Aurelia* genome assembly has 94.3%, *Morbakka* 95.7%, *Nematostella* 98.5%. Probably these values are the most reliable and close to reality.

It is important to mention that 14 proteins from the "metazoa_odb9" set were reported as missing by BUSCO in ***all*** 5 newly sequenced transcriptomes of the Cubozoa and Scyphozoa (*Aurelia*, *Chironex*, *Copula*, *Nemopilema* and *Tripedalia*). Taking into consideration that the "complete" BUSCO score of these transcriptome assemblies ranges between 95.2% - 96.5%, the genes shown below most probably have been lost or extremely derived in these lineages (see the list of "missing" BUSCOs in Table S4).

metazoa_odb9 ID:	**Gene name:**
EOG091G1757	"chromosome 12 open reading frame 57"
EOG091G0XKP	"phosphorylated adaptor for RNA export"
EOG091G0M09	"Ribosome-binding factor A"
EOG091G0LMQ	"deleted in primary ciliary dyskinesia homolog (mouse)"
EOG091G0IHL	"mitochondrial ribosomal protein S34"
EOG091G0GA7	"N-6 adenine-specific DNA methyltransferase 2 (putative)"
EOG091G0G1G	"BCL2-associated athanogene 2"
EOG091G0BKX	"tRNA 5-methylaminomethyl-2-thiouridylate methyltransferase"
EOG091G0AX1	"NAD kinase"
EOG091G07PR	"chromosome 16 open reading frame 70"
EOG091G07PH	"DTW domain containing 1"
EOG091G054P	"hdc homolog, cell cycle regulator"
EOG091G03JG	"PDX1 C-terminal inhibiting factor 1"
EOG091G01TX	"sterile alpha and TIR motif containing 1"

We estimate that as a consequence of aforementioned gene losses and sequence divergence the maximal theoretically achievable BUSCO score (complete+partial) for *Aurelia* and *Morbakka* can not exceed 97.8. We also predict that even with the 100% of the *Aurelia* genome being assembled the "complete" category will never exceed 87-89%. Differences in protein sequences between *Aurelia* and current "metazoa_odb9" set will not allow detection of short exons - these exons will always fall below cut-off threshold in TBALSTN search and, consequently, the corresponding genes will be reported as missing or fragmented.

Taken together we estimate that the real completeness level of *Aurelia* and *Morbakka* genome assemblies are ~2-3% worse than that of *Nematostella* and *Aiptasia* and better than *Acropora*, *Hydra* and *Stylophora* (Table S3). Continuity of scaffolds (N50 value) is better than in all previously published cnidarian genomes, even that of *Nematostella* (see scatter-plot in Fig.S4d and relative scaffold lengths in Fig.S4e). No assembly is perfect, but we are confident that our data is within the same quality range as previously published cnidarian genomes. Due to their continuity *Aurelia* and *Morbakka* assemblies represent valuable information for large-scale genome architecture comparisons with other cnidarians and representatives of Deuterostomia (Supplementary Note 3.1 "Macrosynteny analysis").

Heterozygosity rates in *Aurelia* Baltic strain (1.26%), *Aurelia* Roscoff strain (2.08%) and *Morbakka* (0.67%) were estimated by calculation of ratios between homo- and heterozygous peaks in k-mer frequency distribution plots (GenomeScope, Fig.S4a-c). Repetitive regions in the genomes were identified with RepeatScout v1.0.5[22] and soft masked with RepeatMasker v4.0.3[23] (Fig.S5).

Gene models were predicted using AUGUSTUS v3.0.2[24]. RNASeq transcripts were mapped to the genome assembly using PASA v2.0.1 pipeline[25]. Resulting transcript models (*.gff3 file) were converted by "gff2gbSmallDNA.pl" into GenBank format and used as a training set

for AUGUSTUS (autoAugTrain.pl). Exon and intron hints for gene predictions were generated by mapping raw RNAseq reads and cDNAs from transcriptome assemblies to the genome sequence with BLAT v.34. Gene models, transcript models, and peptide predictions from gene models are stored in GFF3 and FASTA format (see the folder /gallery on the MGU server at http://marinegenomics.oist.jp). Genome sequences were deposited at NCBI under project accession numbers REGM00000000, REGL00000000 and RDPX00000000.

Genome browsers for two strains of *Aurelia* and *Morbakka* are accessible via the web page of Marine Genomics Unit at OIST (http://marinegenomics.oist.jp/gallery). Protein sets from *Aurelia* and *Morbakka*, OrthoMCL groups and all transcriptomic assemblies are also accessible via OIST BLAST server at http://203.181.243.155/aurelia/blast.html

Links to data files (*.gff, *.fasta, *.sql) and project related web pages are listed in Tables S1, S2, S3. Please address Konstantin Khalturin (konstantin.khalturin@oist.jp) in case of any data related questions.

1.4 Transcriptome sequencing, assembly and annotation

Libraries for transcriptome sequencing were prepared from mRNA as a starting material with TruSeq Stranded mRNA Sample Preparation kits and sequenced on MiSeq and HiSeq2500/4000 instruments (Illumina). Raw reads were quality filtered (Q20) and trimmed at both ends to remove low-quality regions with Trimmomatic v0.30[13]. Transcripts were assembled *de novo* with Trinity (r20140717, v2.0.6, v2.3.2)[26]. Peptides encoded by transcripts were predicted with ESTScan 3.0.3[27] or TransDecoder[28]. Only peptides with more than 70 amino acids were retained and used in further analyses. Resulting peptides were searched against the non-redundant NCBI peptide database and several selected protein sets (human, *Nematostella*, *Acropora*, *Hydra*) using BLASTP. Redundancy in protein sets was removed by CD-HIT v4.6.1[29] with 95% similarity cut-off value. Protein domains and their coordinates were identified by search with HMMER v3.1b2 using release 29 of the Pfam-A database (ftp://ftp.ebi.ac.uk/pub/databases/Pfam/releases/Pfam29.0/).

Transcriptome assemblies and raw reads were used as hints for gene model prediction (see flow diagram in Fig.S3) and for analysis of gene expression dynamics at various stages of the life cycle and in different tissues (Figures 3-5). In order to generate expression tables, quality filtered reads from all RNAseq libraries available for a given species were mapped back to the reference transcripts with Bowtie2[30] and transcript abundance was estimated with RSEM v1.2.5[31]. Finally, transcript sequences, peptide predictions, BLAST search results and expression values were imported and integrated into relational database (MySQL 5.6.15). The expression database is accessible via web browser at the OIST BLAST server http://203.181.243.155/aurelia/

1.5 Repetitive elements

Repetitive elements (RE) in the draft genome assemblies of *Aurelia* and *Morbakka* were identified *de novo* with RepeatScout v.1.0.5[22] and RepeatMasker v.4.0.6[23]. RE were filtered by their lengths and occurrences so that only sequences longer than 50 bp and present more than 10 times in the genome were retained (see instructions in the README file of RepeatScout package). Resulting sets of RE were annotated by BLASTN and BLASTX searches against RepeatMasker.lib (35996 nucleotide sequences) and RepeatPeps.lib (10544 peptides) bundled with RepeatMasker v.4.0.6. Results of both searches were combined and

BLASTX result was given a priority in cases where both BLASTN and BLASTX searches gave hits. Original identifiers produced by RepeatScout (i.e., >R=0,1,2,3, etc.) were exchanged for the results of BLAST searches according to the RepeatMasker format of sequence headers (i.e. >Chapaev-16_HM#DNA/CMC-Chapaev). For putative novel repeats, one additional class "Novel/Cnidaria" was introduced and those sequences were assigned headers like >R=0#Novel/Cnidaria. Next, RepeatMasker was run with annotated species-specific collections of repeats. To visualize the expansion history of repetitive elements, we calculated Kimura substitution levels between the repeat consensus to its copies in the genome using two Perl scripts from RepeatMasker package ("calcDivergenceFromAlign.pl" and "createRepeatLandscape.pl"). Resulting stack bar graphs of the repeat landscapes are shown in Figure S5. An original script "createRepeatLandscape.pl" was modified in order to include a "Novel/Cnidaria" repeat class into the input and output.

In *Aurelia,* the most abundant classes of non-LTR retrotransposons are L2 (5.3%) and CR1 (1.6%) (Fig.S5a). However, from the total of 19704 repeat types identified by RepeatScout only 3521 types could be assigned by BLASTN and BLASTX searches to already known sequences present in repeat databases. Thus, novel and putatively cnidarian-specific repeats are the most abundant, occupying 34.7% of the genome. Interestingly, RE composition is drastically different in Atlantic and Pacific strains of *Aurelia*. LTR transposons (Gipsy and Tc/Mariner) together with DNA transposons (hAT and Harbinger) occupy ~12% and novel RE occupy 22.6% of the Pacific *Aurelia* genome (Fig.S5b). At the same time, the proportion of LINEs is similar to that in the Atlantic strain, with L2 (4.9%) and CR1 (1.6%) being the most abundant representatives (Fig.S5b).

In *Morbakka*, non-LTR retrotransposons, such as Dong-R4 (7.4%) and RTE (10.5%), are the major repetitive elements (Fig.S5c). Novel repeats occupy only 5.7% of the genome which is ~6-fold less than in *Aurelia*. The proportions of the genomes occupied by RE in *Aurelia* (44.7%) and *Morbakka* (37.4%) are quite similar. Drastic genome size differences between these two species, therefore, are not due to the accumulation of mobile elements in *Morbakka* genome.

Annotated repeats of *Aurelia* and *Morbakka* were added to OIST BLAST server as combined database "Repeats_in_ABSv1_ARSv1_MVIv1_genomes". They are also stored in the "Downloads" section of OIST Genome browser (http://marinegenomics.oist.jp/gallery/).
Files "AUR21_r04_250316_repeats.fa.gz" and "MOR05_r06_genome_repeats.fa.gz" include 19,704 (82.1% novel) and 13,698 (49.7% novel) distinct repetitive elements, respectively. Repeat information was also added as "Repeat" tracks to the genome browser of each species.

Supplementary Note 2: Molecular phylogeny and dating

2.1 Background information on phylogeny of Cnidaria and datasets used

The position of the Cnidaria as a sister group to the Bilateria has been well established by phylogenetic reconstructions based on 18S and 28S ribosomal genes, as well as by phylogenetic analysis done within the genomes projects of *Nematostella*, *Trichoplax*, and *Amphimedon*[32,33,34]. Within Cnidaria, four classes (Anthozoa, Hydrozoa, Scyphozoa, and Cubozoa) are usually recognized, but the Staurozoa (stalked jellyfishes) are sometimes also referred to as a separate class within the Medusozoa (see discussion in Collins et al.[35]). Previously in the absence of genome sequences from Scyphozoa and Cubozoa, phylogenetic reconstructions among cnidarians were usually performed based on ribosomal markers[35]. Relations between *Acropora*, *Porites*, *Montastraea*, *Nematostella*, *Aurelia*, and *Hydra* were inferred for the first time based on a supermatrix of single-copy nuclear genes in Simakov et al.[36].

Here, we analyzed phylogenetic relations with maximum likelihood and Bayesian approaches using datasets derived from the newly sequenced genomes of *Aurelia* and *Morbakka*, as well as from transcriptomic data of species belonging to Hydrozoa, Cubozoa, and Scyphozoa. A complete list of species used for phylogenetic reconstructions is shown in Table S5. Our major aim was to estimate relationships among Cnidarian groups as well as the scale of genetic differences between cnidarians and bilaterians. Species were selected so that they represent the major clades within the animal kingdom. Another important requirement was to include representatives of animal groups whose time of origin had been verified based on paleontological records.

2.2 Molecular phylogeny

To identify reliable phylogenetic markers, we used OrthoMCL v.2.0.9[37] to cluster orthologous genes from 47 selected proteomes (see Table S5). In this taxonomic set, we included yeast as an outgroup and 46 metazoan proteomes (two Ctenophores, four sponges, *Trichoplax*, 18 cnidarians, and 22 bilaterians representing Protostomia and Deuterostomia). OrthoMCL results were imported into a custom-made MySQL database that contained protein sequences and information about their placement into orthologous groups (OGs). In this database, we identified 133 marker genes (OGs) represented by a single copy in the majority of selected species. For each of 133 OGs in 47 species, protein sequences were selected from the database and aligned with MAFFT v.7.130b (with --maxiterate 1000 --localpair options). Poorly aligned regions were trimmed with TrimAl v.1.2rev.59 (-gappyout option). After trimming, aligned blocks were concatenated with a Perl script "FASconCAT-G_v1.02.pl" (http://www.zfmk.de/bioinformatics/FASconCAT_v1.0.zip). Phylogenetic trees were constructed using RAxML software v. 8.2.7[38]. The rate heterogenity model GAMMA with the LG substitution matrix was used.

In addition to the initial protein set representing 133 OGs from 47 proteomes, we also used several additional sets with reduced taxonomical sampling, but increased numbers of genes (set A, set B, set C, set D, set E). Additional marker sets included more OGs due to the exclusion of the most ancient and (or) derived organisms (yeast, ctenophores, and sponges). Alternative protein sets included 213 OGs from 46 taxa (setA); 394 OGs from 32 taxa (setB); 483 OGs from 27 taxa (setC) and 403 OGs from 26 taxa (setE). Independent of the marker set used, tree topologies describing relations among cnidarian groups remained identical and correspond to the tree shown in Fig.1f. Phylogenetic trees obtained with sets of 47 and 46 taxa

by ML approach are shown in Fig.S6. Phylogenetic tree obtained with the set of 46 taxa and 213 genes by ML approach is shown in Fig.S8.

Bayesian analysis was conducted in order to evaluate and compare topologies obtained via the ML method. Several aforementioned protein sets were analyzed using PhyloBayes v.1.6j with -cat -gtr model[39]. Using "Sango" HPC (OIST) ~4000 trees were usually generated in each chain before maxdiff < 0.1 was reached. Convergence analysis was performed with 500 initial trees discarded as burn-in phase and sampling of every fifth tree from the remaining set. Topologies of consensus trees derived from ML and Bayesian approaches were identical (see Fig.S7a,b based on protein set E with 403 orthologs).

OrthoMCL clustering results used for phylogenetic reconstruction are stored in the "Datasets" section of OIST BLAST server (http://203.181.243.155/aurelia/datasets.html) in the file "OrthoMCL_data_run07.tar". This file is a *.tar archive (533Mb) which contains all proteins from 48 datasets (see Table S5, in total 1.1897.20 peptides), pairs of orthologs with 40% and 50% length overlap cut-off during all-against-all BLASTP search and grouping of proteins by MCL into OGs with two inflation rates I=1.2 and I=1.5:

Content of "OrthoMCL_data_run07.tar":

Size:	File Name:
281M	goodProteins_r07.fasta.gz <== 1.189.720 proteins
19M	groups_r07_1.5
19M	groups_r07-40pct_1.2
111M	orthomcl_pairs_run07-40pct.tar.gz
105M	orthomcl_pairs_run07-50pct.tar.gz

Subset of the cnidarian sequences (490,385 peptides) used in OrthMCL clustering is stored in the file "**peptides_omcl07.fasta.gz**". This combined set of cnidarian proteins can be searched in the BLAST section. BLAST database name is "Cnidarian_protein_sets_OrthoMCL_run07"

2.3 Molecular dating

For molecular dating, we used tree topologies obtained by ML approach with 133 OGs (47 taxa with yeast as outgroup) or 233 OGs (46 taxa with two Ctenophora species as outgroups). PhyloBayes v.4.1c was used with two different root prior settings according to the outgroup organisms used (-ln -bd -sb -rp 2000 1000 or -ln -bd -sb -rp 984 500). Fossil calibrations were similar to those used by Simakov et al.[36]. In addition, we set a 240 Mya lower constraint on the *Acropora* - *Nematostella* split (the first scleractinian corals appeared in fossil record after 240 Mya). Fossil calibrations used are shown in Fig.S9b. Major parameters that influence the time scale are the root prior and the oldest time point among all fossil calibrations used. We do not have fossil calibrations for yeast - Ctenophora, Ctenophora - Cnidaria or Anthozoa - Medusozoa splits. Thus, the root prior remains the major source of variation in our molecular dating. There are several Pre-Cambrian and Middle Cambrian fossils referred to as jellyfish[40,41], but we can not attribute them to any extant cnidarian class with confidence. The majority of fossil calibrations are from bilaterian data where paleontological records are available[42,43]. Branch lengths in phylogenetic trees are similar in cnidarian and bilaterian clades; therefore, we assume that substitution rates in the Cnidaria and Bilateria are similar. Based on that, we can calculate divergence times, which are just the best possible estimates in the absence of phylogenetically reliable fossils of medusozoans.

Supplementary Note 3: Characterization of genomes

3.1 Macro-synteny analysis

Comprehensive comparative analysis of genomic organization within the Cnidaria has not been performed previously, but drastic variations in relative positions of Hox genes have been reported in *Nematostella*, *Hydra*, *Eleuteria* and *Aiptasia*[32,44,45,46]. This finding indicates that cnidarian genomic architecture might be more flexible than that of deuterostomes. Here we compare the overall genomic architecture of *Aurelia* and *Morbakka* with that of *Nematostella* with a major emphasis on identification of features which may be related to the origin of a jellyfish stage. In order to check for the presence of conserved macro-synteny blocks, we calculated the number of shared orthologous genes for all pairwise combinations of genomic scaffolds in *Aurelia*, *Morbakka*, and *Nematostella.* Resulting values were tabulated in the form of Oxford grids and represented as synteny dot plots, as reported previously[32].

To examine the conservation of large-scale synteny we followed the approach utilized previously for *Nematostella* - human and *Amphioxus* - *Saccoglossus* comparisons[32,36]. In brief, all predicted genes in the genomes of *Nematostella* (nve), *Aurelia* **(**abs**)** and *Morbakka* (mvi) were assigned identifiers (3-letter code *nve*, *abs*, *mvi*), scaffold number and the position of a gene within the scaffold (e.g. abs|s9_g95 is *Aurelia* gene in scaffold number 9, 95-th gene in the scaffold and corresponds to the gene nve|s94_g37 in *Nematostella*). The first gene in a scaffold obtained index "g1" and genes g1, g2, g3, g_n follow each other). Thus, gene order is stored in a human-readable format that is easy to handle in any downstream analysis. Scaffolds of *Aurelia* and *Morbakka* were initially ordered according to their lengths; therefore, scaffold_1 is longer than scaffold_2 and so on. Longer scaffolds usually contain more gene models.

As a next step, orthologous genes in *Nematostella* (nve), *Aurelia* **(**abs**)** and *Morbakka* (mvi) were identified with OrthoMCL using standard settings (minimum 50% alignment overlap) and MCL inflation rate set to 1.5. Resulting OrthoMCL output files were used as an input for synteny analysis (an example of an *Aurelia* - *Nematostella* comparison is shown below):

```
abs|s1000_g1_t1_AURBSr04        nve|s39_g9_t1_NEMVE        0.881
abs|s1002_g1_t1_AURBSr04        nve|s135_g26_t1_NEMVE      0.643
abs|s1005_g2_t1_AURBSr04        nve|s5_g28_t1_NEMVE        0.146
..
..
abs|s9_g95_t1_AURBSr04          nve|s94_g37_t1_NEMVE       0.83
abs|s9_g98_t1_AURBSr04          nve|s55_g27_t1_NEMVE       0.543
abs|s9_g9_t1_AURBSr04           nve|s1_g182_t1_NEMVE       0.964
```

The table with orthologs was re-formatted and genes were sorted according to their order in the *Aurelia* genome:

```
abs  1     1  s1_g1_t1_AURBSr04      nve  125  31  s125_g31_t1_NEMVE  1.527
abs  1     3  s1_g3_t1_AURBSr04      nve  131  20  s131_g20_t1_NEMVE  0.623
..
abs  2681  1  s2681_g1_t1_AURBSr04   nve  62   12  s62_g12_t1_NEMVE   0.1
abs  2693  1  s2693_g1_t1_AURBSr04   nve  4627 3   s4627_g3_t1_NEMVE  0.088
```

From columns #2 and #6 of such a table (e.g. "orthologs_abs_nve.tab ") the number of shared orthologous genes was calculated for all pairs of scaffolds with the following command:

[user@sango]$ awk -F"\t" '{printf "%s\t%s\n", $2,$6}' orthologs_abs_nve.tab | sort -k1,1n -k2,2n - | uniq -c - | sort -k2,2n - > orthologs_abs_nve.shared_per_scaffold.sorted

The resulting table was converted into the Oxford grid matrix, which contains information about the number of shared orthologous genes for all pairwise combinations of scaffolds in the pair of species studied. A perl script used for this purpose together with example data files (see "oxford_grid.tar.gz") is available at http://203.181.243.155/aurelia/datasets.html. Only scaffolds that contained 10 genes or more were used in the final grid. Additionally, only scaffolds that had at least 10 hits to orthologous genes in the scaffolds of another species were retained (i.e. cumulative number of hits in any row or column of the Oxford grid must be at least 10). These two filtering steps allowed a drastic reduction of matrix complexity by removing scaffolds where just a few orthologs or hits were present. For example, the final Oxford grid for the *Aurelia* - *Nematostella* comparison has 225 rows and 201 columns (grid with 45225 cells) while the total numbers of scaffolds with predicted genes in *Aurelia* and *Nematostella* are 2023 and 4407, respectively (This would result in a grid with 8,915,361 cells). According to our screening criteria ~10% of scaffolds in *Aurelia* and 5% in *Nematostella* had detectable levels of synteny conservation.

Scaffolds contained in the Oxford grid were further clustered in order to better visualize putative ancestral linkage groups (PALs) shared by two species. During the clustering procedure, the number of shared orthologous genes per scaffold pair was used as a similarity metric. Hierarchical clustering and visualization were performed in R using the *pheatmap* function with the "complete" linkage option as a distance measure. As a result of clustering, scaffolds in the Oxford grid were reordered, bringing together scaffolds with similar gene content, thereby revealing conserved linkage groups which were probably present in the last common ancestor of the compared species. Dot plots, which represent positions of individual orthologous genes within respective scaffolds, were also generated in R (Fig.2b,d). Sizes of cells along x- and y-axes are proportional to the total number of genes in the compared scaffolds. Distributions of genes on dot plots makes it obvious that the order of genes within the scaffolds was strongly scrambled (not conserved) (see also Fig.S10-12).

Subsets of clustered Oxford grids for *Aurelia-Nematostella* and *Aurelia-Morbakka* are shown in Fig.2b,d. *Morbakka-Nematostella* pair is shown in Fig.S12. Complete grids and dot plots are shown in Fig.S10-12.

Seven *Aurelia* scaffolds (4, 6, 23 | 52, 24, 62 | 17) correspond to 14 *Nematostella* scaffolds that had been assigned to three ancestral linkage groups A, B, and C[32]. In our analysis, scaffold 61 of *Nematostella*, belonging to PAL A, where HoxA, C, Da, Db genes are located, did not qualify as a syntenic block due to insufficient number of shared orthologs. In *Aurelia, Hox1* gene (a clear ortholog of *HoxA* of *Nematostella*) is located on scaffold 6.

Aurelia scaffolds with the highest level of linkage conservation (s4 [277 genes], s6 [200 genes], s23 [195 genes]) have remarkable content. Clusters of *minicollagens* and *Dickkopf* genes (*Dkk*) are located on scaffold 4. Scaffold 6 contains *Hox1* (*HoxA* ortholog) and the set of surrounding genes is conserved throughout the Medusozoa. Scaffold 23 contains clusters of *Nk2* and *Otx* genes involved in the formation of the jellyfish sensory system. Thus, the regions corresponding to *Nematostella* PAL A[32] have obvious functional importance. Reasons for high linkage conservation in *Aurelia* groups of scaffolds 52, 24, 64 and scaffold 17 (see Fig.2b) are not immediately clear from their gene content and require additional analysis in the future.

The *Aurelia* - *Nematostella* comparison suggests that despite of the deep evolutionary split between the Medusozoa and Anthozoa (Fig.S7-8), there must have been very strong selective pressure that prevented inter-chromosomal translocations in several regions of *Aurelia* genome. Thus, remnants of ancient chromosomal segments, which existed before separation of Cnidaria and Bilateria, were also retained throughout the evolutionary history of the Medusozoa and remain unchanged in three clades that independently evolved for more than 500 million years.

Patterns observed in *Morbakka* - *Aurelia* and *Morbakka* - *Nematostella* comparisons (Fig.2d, Fig.S11, Fig.S12) imply a greater degree of structural changes in the genomes of *Morbakka* ancestors, and corroborate the notion that gene order may vary considerably among cnidarian lineages[45]. It is not clear why ancient linkage groups of the eumetazoan ancestor were retained in the genome of Aurelia and dissolved to a large extent in *Morbakka* (compare the patterns in Fig.2b and 2d). The overall genomic architecture of *Aurelia* is obviously closer to that of *Nematostella*; therefore, it has been subject to fewer changes. It might be premature to generalize based on just two medusozoan species, but our data suggest that Scyphozoa may have retained more ancestral traits in their genome architecture than Cubozoa.

In the future, with more genomes to compare, we may be able to reconstruct the gene sets of at least some chromosomes of the common eumetazoan ancestor. *Chironex* and *Nemopilema*, which are closer to the Scyphozoa-Cubozoa split (see Fig.1f and Fig.2a) are the most promising candidates for comparative studies that may shed more light on the trends of genome architecture changes within Medusozoa.

3.2 Gene sets of Cnidaria

In order to compare gene sets among Cnidaria (presence / absence of orthologs) we used the results of OrthoMCL clustering which have been utilized previously for our phylogenetic analysis (Fig.1f). From the complete matrix (Table S6) that shows the number of shared orthologs in all pairwise combinations of 48 proteomes we selected a subset - 14 cnidarian species and five representatives of Deuterostomia as an out-group. Resulting matrix with 19 species was clustered in order to bring together the species with the most similar sets of genes. Result of the analysis are shown as heatmap in Fig.3a with the colour of each cell being proportional to the number of shared orthologous genes between species.

Matrix in Table S6 can be also utilized for measuring the completeness of genomic and transcriptomic assemblies. For example, *Saccharomyces* is a good empirical reference to estimate the assembly completeness (similar to CEGMA or BUSCO approaches). Yeast has 2988 orthologous groups that represent the core eukaryotic protein set and all high quality proteomes (derived from genomic or transcriptomic data) usually share >2000 orthologous groups with yeast (for example, 2353 OGs are shared between yeast and human, 2366 OGs are shared with *Nematostella*). Comparison with the set of yeast proteins makes it obvious that the assembly quality differs considerably between *Aurelia* from the Baltic sea (2195 OGs) and *Aurelia* Roscoff strain (only 1784 OGs shared with yeast). At the same time the transcriptomic assembly of *Aurelia* Roscoff (rsT) is very good (2146 OGs shared with yeast). From that we can conclude that Roscoff assembly misses a lot of genes and it is better to use Baltic sea *Aurelia* genome as a reference. *Morbakka* assembly shares 2250 OGs with yeast

This approach can not be used as alternative for CEGMA or BUSCO as it requires much more computation. However, it is very informative and we used it in addition to the aforementioned programs for quality assessment of our assembly data.

3.3 Stage-specific gene sets

It is a matter of debate whether anthozoan polyps or medusozoan jellyfish have higher morphological complexity. However, among medusozoans, the anatomical complexity of a jellyfish relative to that of a polyp is obvious. Therefore, in the life cycle of medusozoans, the ultimate complexity of the adult organism is achieved not as a result of planula-to-polyp transition, as in the Anthozoa, but requires one additional metamorphosis step, which occurs long after embryonic development and the planula-to-polyp transition (Fig.1a). Development of new structures during life cycle progression (such as eyes, statocysts, swimming muscles) requires structural genes and transcription factors with jellyfish-specific expression. That implies the presence of regulatory factors and structural genes that are active only during the polyp-to-jellyfish transition and in the adult state. To what extent are they novel or shared with genes already present in the Anthozoa? What proportion of medusozoan genomes is devoted to production of stage-specific structures? In order to answer these questions, we identified stage-specific genes and analyzed their phylogenetic distribution within the Cnidaria. Differential gene expression was analyzed in *Aurelia*, *Morbakka*, *Tripedalia*, and *Nemopilema*. *Nematostella* and *Acropora* genomes were used as anthozoan references.

Stage-specific genes were identified by mapping quality-trimmed RNAseq reads (Illumina HiSeq or MiSeq) against transcript models (mRNA sequences including 5' and 3' UTR). PE reads were size-trimmed to maximum length of 100bp for better mapping with bowtie2. Transcript models for *Aurelia* and *Morbakka* were extracted from genome sequences and annotation files (*.gff) using the Perl script "gffGetmRNA.pl" from AUGUSTUS v3.2.1 package. The longest transcript representing a given gene was used for mapping. In the case of *Tripedalia* and *Nemopilema*, their transcriptome assemblies were used as references for RNAseq read mapping. RSEM and edgeR were utilized to convert mapped read counts obtained with bowtie2 into TMM-normalized FPKM values. Tables with expression results were imported into a relational database and differentially expressed genes were identified with SQL queries as described previously[5].

All expression values shown in Supplementary Tables 7,8,11-15 are TMM normalized FPKM values. Heatmaps of expression dynamics shown in Figs. 2-4 were generated in R using the "*pheatmap*" function (R libraries gplots, ggplot2, pheatmap, RColorBrewer have been used). In the majority of cases, TMM-normalized FPKM values were log2-transformed. Z-scores were not used except in Fig.3b as they do not allow direct visual comparisons of expression levels between the genes.

Genes that are expressed exclusively in polyps or jellyfish or that have >4x higher expression in one stage were considered stage-specific. Such genes were identified and categorized (see Table S7, S8). Interestingly, there is no significant difference in the proportion of Medusozoa-specific genes among the genes with jellyfish-specific or polyp-specific expression (Fig.3b). This observation probably indicates that selective pressure affects all stages of the life cycle to a similar degree and novel genes are necessary for the function of *Aurelia* polyp as well. When compared to the gene set without stage-specific expression, both stage-specific sets are significantly enriched for the genes that are not present in the genomes of *Acropora* and *Nematostella* ($p<0.001$, $\chi 2$-test). The data (from Fig.3b) are tabulated as follows:

	no blast hits	blast hits	Marginal Row Totals
gene is stage-specific	1340*	1960**	3300
gene is NOT stage specific	6193	15700	21893
Marginal Column Totals	7533	17660	25193

$\chi 2 = 207.613$, df = 1,$\chi 2/df = 207.61$,$P(\chi 2 > 207.613) < 0.00001$
* - 1340 = 400+940 | ** - 1960 = 726+1234 (jellyfish-specific + polyp-specific genes)

Combined transcriptomic analysis of *Aurelia*, *Morbakka*, *Nemopilema* and *Tripedalia* showed that genes belonging to 90 orthologous groups had jellyfish-specific expression in all four species (Fig.S13a, Tables S12-15). In case of *Aurelia*, that corresponds to 137 genes (only ~10% of the total number of jellyfish-specific genes in that species). Within each class (*Aurelia-Nemopilema* and *Tripedalia-Morbakka* pairs) the number of shared jellyfish-specific genes doubles (177 and 196 OGs), but still the proportion of jellyfish-specific orthologous groups that are species-specific remains high (43% in *Aurelia*, 30% in *Morbakka*, 37% in *Nemopilema*, and 23% in *Tripedalia*). Unexpectedly large differences in jellyfish-specific gene sets and small overlapping set might be explained by long periods of independent evolution of scyphozoans and cubozoans or independent origins of a jellyfish stage in the different medusozoan lineages.

3.4 Muscle-specific genes

Genes specific for striated muscles in *Aurelia* were identified based on the comparison of gene expression profiles between ectodermal epithelial cells of the exumbrella and striated muscle cells lining the subumbrella (shown as "Ectoderm" and "Muscles," respectively, in Fig.S13b). These two cell types share a common origin from ectodermal epithelial cells of a polyp, but the latter develops striated sarcomeres during the polyp-to-jellyfish transition; therefore, they should express structural and regulatory genes responsible for jellyfish propulsion.

Myosins and myosin tail proteins were identified by HMMER search with Pfam models PF00063.18 (myosin head) and PF01576.16 (Myosin_tail_1). Orthologs of *Aurelia* muscle-specific genes from *Nemopilema*, *Tripedalia*, *Morbakka*, and *Chironex* were identified by OrthoMCL clustering and reciprocal BLASTP searches. The position of myosin tail domains in cnidarian proteins were identified by HHMER. These regions were selected from a relational database using SQL query, aligned by MAFFT and the phylogenetic trees shown in Fig.3i and Fig.S14 were reconstructed by RAxML.

Scyphozoan and cubozoan propulsion relies exclusively on swimming musculature located in the subumbrella (Fig.3d; Fig.S13b). This specialized layer of epitheliomuscular cells develops during the polyp-to-jellyfish transition and is absent in the polyp stage. Jellyfish swimming muscles are striated, but in contrast to mesodermal muscles of deuterostomes, they are of ectodermal origin and develop during metamorphosis, rather than during embryogenesis. During metamorphosis of *Aurelia,* ectodermal epithelial cells of a polyp trans-differentiate either into upper epithelium of the bell or into muscle cells with striated myofibers lining the inner side of the bell (Fig.3d-g).

In *Aurelia*, the first signs of striated musculature appear during the late stages of metamorphosis, and in total, 88 genes are expressed differentially between oral and aboral ectodermal layers of a jellyfish (4x-fold difference, see examples in Fig.3E-G). Among these genes, 65 are also present in anthozoans and 23 are putative taxonomically restricted genes - TRGs (for example, gene ***s226_g16*** in Fig.3f).

The most prominent muscle-specific jellyfish motor-protein is a homolog of the conventional bilaterian myosin Myh1/2/3 (gene ***s10_g36*** in Fig.3h; Fig.S13d-e). Dynamics of Myh1/2/3 expression during metamorphosis are similar across various species. In *Aurelia*, *Nemopilema*, *Tripedalia*, and *Morbakka*, Myh1/2/3 is strongly up-regulated in the jellyfish stage (Fig.3h); therefore, it might be one of the most important proteins for jellyfish propulsion. Second, Myh1/2/3 paralog (***s285_g25/26***) shows higher expression in the polyp stage. The other 13 myosins do not exhibit any clear life-stage specificity and seem to be expressed in patterns typical for housekeeping genes (Fig.S13e).

In addition to conventional motor-proteins, striated muscles of scyphozoans and cubozoans contain proteins with unusual domain architecture (Fig.3j; Fig.S13c). In contrast to conventional myosins, these proteins lack Myosin_head and IQ domains and contain only myosin_tail_1 domains (PFAM PF01576.16). In two *Aurelia* proteins (***s206_g6/7*** and ***s88_g59***), the myosin_tail_1 domain is linked to a long repetitive N-terminal region (Fig.3j; Fig.S13c). Interestingly, this protein type is absent in *Hydra* and anthozoans, while it is expanded into five genes in scyphozoans and four genes in cubozoans (Fig.3i). In the genomes of *Aurelia* and *Morbakka*, these novel "myosin tail proteins" (*MTPs*) are organized in mini-clusters composed of two genes (Fig.3j). In *Morbakka*, *MTP-2* and *MTP-3* are co-localized with one copy of conventional myosin (***mvi_s128_g45*** in Fig.3j). Based on phylogenetic reconstruction, these novel Myosin_tail proteins probably appeared before the separation of the Cubozoa and Scyphozoa (Fig.3k).

Another prominent group of genes specifically expressed in striated muscle cells encodes proteins typically involved in regulation of smooth muscle contraction. These are myosin light chain (*MYLC6B* - ***s9_g140***), myosin light chain kinase (***s187_g36***), three calmodulin-like proteins, and myosin phosphatase subunits. According to the transcriptomic analysis, some of these genes are exclusively expressed in striated swimming muscles and are not detected in the polyp stage (see Tables S7 and S11).

From the evolutionary perspective, jellyfish propulsion probably emerged based on smooth muscle cells and required "polymerization" of myofilaments into regular, high-order repetitive units analogous to striated muscles of bilaterians[47]. Thus, additional specialized proteins for interconnections of sarcomeres, analogous to Z-disks and M-lines, were necessary. In bilaterians, these functions are carried out by *titin* and *nebulin*.

In accordance with an earlier report[47], homologs of giant Z-disk and M-line proteins, such as titin, nebulin, and myomesin, defining the structure and spatial arrangement of thin and thick filaments within myofibrils, are absent in the representatives of Scyphozoa and Cubozoa. Hence, sarcomeric structures resembling M- and Z-lines, which are observed with electron and light microscopy in *Aurelia* and *Tripedalia*[48,49] must be composed of other proteins. Titin is the largest protein in the genomes of higher animals, being ~30,000 amino acids long in humans; therefore, it should be readily identifiable in proteomes. In *Aurelia* and *Morbakka*, two genes encoding the largest proteins are ~10,000 amino acids long and also include domains typical of titin. Their expression, however, is not consistent with the function in Z-

disks, as they are not up-regulated in swimming muscles and are ubiquitously expressed in all stages of the life cycle. We propose that the role analogous to titin in *Aurelia* and *Morbakka* may be played by lineage-specific myosin tail proteins (MTPs). We do not have functional data to support our hypothesis yet, but expression of these proteins is specifically restricted to the jellyfish stage and swimming muscles (Fig.3g,h).

Swimming muscles of jellyfish are strikingly similar to bilaterian striated muscles on the electron or light microscopy level. Structurally they are organized analogously to striated myofibrils of bilaterians; nevertheless, their contraction cycle must be regulated purely by machinery typical of smooth muscles (the calmodulin-caldesmon complex and phosphorylation of regulatory myosin light chain 6B by myosin light chain kinase)[50]. As mentioned earlier, all these regulatory genes are highly up-regulated in jellyfish muscles while genes encoding the troponin complex (central regulatory components of bilaterian striated muscles) are absent in the genomes of *Aurelia* and *Morbakka*, as in other cnidarians (see discussion in Steinmetz et al.[47]).

Apparently, in parallel to the Bilateria, ancestors of Cubozoa and Scyphozoa "upgraded" their smooth muscles by independent invention of striation. By using a set of novel proteins the Medusozoa developed their taxon-specific way to arrange the myosin and actin filaments into iterative longitudinal modules (sarcomeres). However, our data as well as the recent results of Tanaka et al.[50] indicate that in order to regulate contraction of their novel striated muscles, the jellyfish are still using “archaic” smooth muscle regulatory machinery, based on myosin light chain phosphorylation. Interestingly, novel muscle-specific proteins differ in representatives of Scyphozoa and Cubozoa (Fig.3i,j and Fig.S13c). Thus, not only the muscles of Cnidaria and Bilateria appear to differ in origin, but the same may be true for Scyphozoa and Cubozoa.

3.5 Wnt genes

Sequences of Wnt proteins were identified in 11 cnidarian proteomes using BLASTP searches (genomic and transcriptomic data were from Atlantic and Pacific strains of *Aurelia*, *Morbakka*, *Nemopilema*, *Tripedalia*, *Chironex*, *Aiptasia*, *Nematostella*, *Pories*, *Acropora*, *Hydra viridis* and *Hydra magnipapillata*). In *Nematostella*, Wnt protein sequences from gene models (*.gff) were utilized[32] as this set contains more non-redundant Wnt proteins than the set published in Kusserow et al.[51]. All sequences were manually curated to exclude possible artefacts of automated peptide predictions. Duplicated identical sequences in the peptide sets of *Hydra* and *Aiptasia* were removed. In case of contradictions between protein sequences derived from automatic gene prediction and transcriptomes, the latter were given priority. The resulting set of 172 Wnt proteins was aligned with MAFFT and a phylogenetic tree was generated with RAxML using the LG substitution model (Fig.4a; Fig.S15a). Groups of Wnt proteins in Fig.4a were named according to the assignments made by Kusserow et al.[51]. Protein sequences used in Fig.4a are stored in Supplementary file "Wnt_peptide_dataset_172_genes.fasta". Corresponding phylogenetic tree in NEXUS format is stored in the file "Wnt_peptide_dataset_172_genes.nwk.figtree" (FigTree v1.4.2 compatible format).

Wnt genes of anthozoans and medusozoans usually belong to distinct clades. Especially profound is the case of a large *Wnt-8* group expansion in the anthozoans. Lineage-specific duplications also occurred in *Wnt-4,-7,-8,-11* groups in *Aurelia* and *Wnt-1* and *Wnt-11* in *Morbakka* (Fig.4b). Variation in the set of *Wnt* genes also occurred in the cubozoans - *Wnt-4a*,

-7a, *-7b* were not present in the transcriptomes of *Tripedalia* and *Copula*, while in *Chironex* and *Morbakka* both *Wnt-7a* and *Wnt-7b* were present and only *Wnt-4a* was missing.

Wnt genes are dynamically expressed throughout the life cycle of a jellyfish. For example, in *Aurelia*, *Wnt-4a* is predominantly expressed in the polyp stage and is down-regulated during the polyp-to-jellyfish transition, while *Wnt-4b* is up-regulated in jellyfish. Interestingly, *Wnt-4* genes are adjacent in the genome with opposing transcriptional directions. In *Aurelia*, *Wnt-7a, -8b* are predominantly polyp-specific while *Wnt-3, Wnt-11a*, and *Wnt-11b* are more strongly expressed in jellyfish. Similar dynamic expression is typical for *Tripedalia* (Fig.S15b).

For chemical interference with *Wnt* signalling we used 1-Azakenpaullone (SIGMA A3734-1MG) which is a cell permeable selective inhibitor of glycogen synthase kinase 3β (GSK3β). Stock solution of 1-Azakenpaullone (AzP) with 2 mM concentration was prepared in DMSO. Working concentration of AzP was 2μM in all the experiments. Chemical treatments were performed in triplicates with 20 polyps per experiment. Animals were kept in plastic Petri dishes with 5 ml filtered sea water (FSW 25 ‰). In the first experimental setup, *Aurelia* polyps were incubated in FSW containing 2μM AzP and 1μM 5-methoxy-2-methyl-indole (metamorphosis inducer, see Fuchs et al. for details[5]). In that case multiple strobila segments were formed, but the borders between the segments were not developed properly (representative polyp after this treatment is shown in Fig.4t). In the second experimental setup *Aurelia* polyps were pre-incubated for 48 hours in FSW containing 2μM AzP and after that 5-methoxy-2-methyl-indole was added to the culturing medium (final concentration was 1μM). By this treatment the segmentation process was completely blocked (representative polyp is shown in Fig.4u). Control animals were treated with the corresponding dilutions of the solvents (DMSO and methanol). 72 hours after the induction of metamorphosis, the experimental and control animals from aforementioned incubation setups were transferred into Petri dishes with FSW and photographed side-by-side (Fig.4t,u).

3.6 Hox / Para-Hox genes

In the majority of publications devoted to the early stages of Hox gene evolution, phylogenetic trees are usually based on alignments of homeobox domains[45,52,53]. This is the most reasonable option for comparison of cnidarian and bilaterian Hox genes, because the areas outside homeodomain are highly variable between distant taxonomic groups. Here our primary intent was to study relations of Hox genes within the Cnidaria, where the scale of variation outside the homeodomain region is much lower. Thus, in order to increase the number of useful informative sites, we utilized complete protein sequences of cnidarian Hox genes. In the proteomes of *Hydra magnipapillata*, *Acropora digitifera*, *Aiptasia*, *Chironex*, *Hydra viridis*, *Nemopilema*, *Nematostella*, *Porites*, *Tripedalia*, *Morbakka*, *Aurelia* Hox genes were identified with a combination of BLASTP and HMMER searches. In addition, previously published Hox genes from *Clytia*, *Podocoryna*, and *Eleutheria* were downloaded from NCBI and included into the analysis. In total, 111 Hox / Para-Hox genes were subjected to ML analysis using RAxML with the LG substitution model (Fig.S16a). Protein sequences used in Fig.S16a are stored in Supplementary file "Hox_peptide_dataset_111_genes.fasta". The corresponding phylogenetic tree in NEXUS format is stored in the file "Hox_peptide_dataset_111_genes.nwk.figtree" (FigTree v1.4.2 compatible format).

Orthologs of bilaterian *Hox1* (*HoxA*) and Para-Hox genes form distinct groups with high bootstrap support (*Hox1*, *Xlox*, *Cdx* and *Gsx* groups in Fig.S16a). Anthozoan homeobox genes form a distinct group to which *Nematostella HoxC*, *Da*, *Db* belong (A1 group in Fig.S16a)

while *HoxA* of *Nematostella* belongs to *Hox1* group. Three clades designated M1, M2, and M3 include exclusively medusozoan sequences representing clear lineage-specific expansions. Groups M2 and M3 include only scyphozoan and cubozoan sequences while M1 also contains one gene from a hydrozoan, *Podocoryne*[54]. Based on transcriptomic analysis, *Hox* genes of *Aurelia* and *Morbakka* were classified into three categories according to the life cycle stage with preferential expression - Polyp-specific, Jellyfish-specific and All-stages (Fig.S16a). Interestingly, all orthologs of *Hox1* have medusa-specific expression in all jellyfish studied so far [55]. Members of M1 groups are also expressed exclusively in a jellyfish stage (Fig.S16a).

The absence of "bilaterian-like" Hox cluster and considerable variation in relative locations of *Hox* genes have been reported previously in *Nematostella*, *Aiptasia*, *Hydra*, and *Eleutheria*[45]. In *Aurelia* and *Morbakka*, the *Hox* cluster is also absent. Any *Hox* cluster, if it ever existed in ancestral cnidarians, was scrambled by inter- and intra-chromosomal rearrangements and a new lineage-specific cluster has not appeared, since there may have been no functional need for it. However, the set of *Nematostella* genes in scaffold 61 might be viewed as an anthozoan *Hox* cluster "in development." Three genes, ***s61_g67***, ***g68***, ***g69*** are clearly anthozoan-specific. Most probably they appeared as a result of local gene duplication and may have already acquired as yet unknown taxon-specific functions important for *Nematostella* development

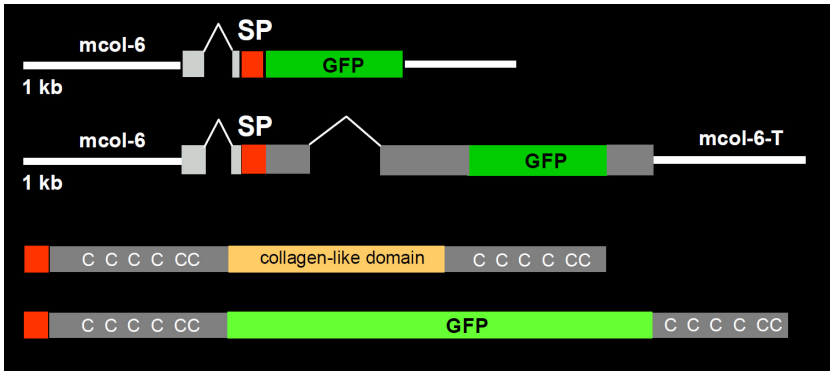
As shown on Fig.5c in *Aurelia*, five Hox genes are jellyfish-specific (*s6_g77*, *s7_g66*, *s56_g35*, *s191_g41* and *s376_g7*) and three genes are polyp-specific (*s7_g67*, *s45_g38, s45_g45*). A similar situation is observed in *Morbakka* where five genes are jellyfish-specific (*s6_181*, *s27_g7*, *s70_g42*, *s189_g37*, *s321_g6*) and four are polyp-specific (*s7_g177*, *s45_g78*, *s70_g44*, *s216_g16*). In most cases, orthologous *Hox* genes in *Aurelia* and *Morbakka* exhibit similar stage-specificity, except for *Cdx* of *Morbakka* which has been duplicated (Fig.5d). Such conserved expression dynamics may indicate that a subset of medusozoan homeobox genes function as molecular "switches," regulating the transition from a polyp to a jellyfish. If this is the case, *Hox* genes acquired different functions in bilaterian and cnidarian lineages. While *Para-Hox* genes are still involved in patterning and development of the digestive system (Fig.5e), *bona fide* Hox genes (together with medusozoan-specific *Hox-like* genes) might be responsible for tissue specificities throughout the life cycle.

Pairs of *Hox* genes with mutually exclusive expression in polyp and jellyfish stages are genomically linked in *Aurelia* and *Morbakka* (Fig.5b). Although genomic localization of their orthologs in *Tripedalia* and *Nemopilema* is not known, their expression dynamics is identical to that in *Aurelia* and *Morbakka*, which strongly supports the generality of this phenomenon (Fig.S16b). Taken together, in the Medusozoa, the molecular switch between life stages might be based at least in part on the activity of these homeobox genes.

3.7 Minicollagens and nematocyte-specific genes

From the evolutionary perspective, nematocytes are considered strongly derived nerve cells. Nematocysts differ across the Cnidaria and their molecular composition and development has been extensively studied in *Hydra* and to lesser extent in other species[56]. Nematocytes develop in specialized areas of the cnidarian body. Major structural components of capsules are minicollagens, which constitute the capsule wall and a thread (Fig.6b-d).

Packaging of a thread inside of the nematocysts and their wall structure (Fig.6b-d) were visualized by the reporter construct where EGFP was fused to N-terminal and C-terminal cystein rich domains of *minicollagen-6* (see below).



SP- signal peptide; GFP - green fluorescent protein; C-C-C-C-CC - CRD domain

Expression of the reporter construct was driven by 1-kb of the 5' upstream region of *Hydra minicollagen-6* gene. All types of the nematocytes were visualized by this approach. Transgenic *Hydra* lines were produced by embryonic microinjection as described previously[57].

In *Hydra,* 410 structural and soluble proteins making up the nematocysts have been identified and several of them have been studied functionally [58]. To study the phylogenetic distribution of nematocyte-specific genes within the Cnidaria, we constructed a database of 410 *Hydra magnipapillata* peptides based on the nematocyst proteome published by Balasubramanian et al.[58]. Gene models that correspond to 21 *Hydra* minicollagens were manually searched in the *H. magnipapillata* genome. Gene models for *mcol-4,7,9,16,17* were not identified in the current version of *Hydra* genome assembly[44]. Next, we clustered all proteins in 12 cnidarian species into orthologous groups using OrthoMCL with relaxed MCL inflation rate (-I 1.2) and 40% sequence identity cut-off. Many nematocyte proteins are composed of repetitive stretches; therefore, "relaxed" stringency of the clustering procedure increased the chances to identify orthologous genes. In total, orthologs of 334 genes were identified in at least one cnidarian species other than *Hydra magnipalillata* (see complete set of hits in Table S17). The resulting matrix shown in Fig.5k represents distribution and copy number of nematocyte-specific genes within the Cnidaria. This figure clearly shows three classes of genes with low (Group I), intermediate (Group II), and high conservation (Group III) among cnidarian species.

The locations of minicollagen genes in the genomes of *Aurelia*, *Hydra viridis*, *Morbakka*, and *Nematostella* were determined using a pattern matching SQL script and BLASTP with an *Aurelia minicollagen* sequence (***s4_g186***) as a query. A general feature of all minicollagens is the presence of two cystein-rich domains (CRDs) with consensus sequences CxxxCxxxCxxxCxxxCC and poly-proline stretches. Hence, the following SQL command effectively selects all peptides with above mentioned features from a peptide database table named "au3.omcl_run07_pep":

```
select * from au3.omcl_run07_pep
where pep_seq like '%C___C___C___C___CC%' and pep_seq like '%PPP%';
```

Peptides predicted from gene models have identifiers that specify their locations within the genome (e.g., *abs_s93_g53*, *abs_s93_g54*, *abs_s93_g55*, *abs_s93_g56*, *abs_s93_g59* - cluster of *Aurelia* minicollagens in scaffold 93). Thus, ordering the results of pattern matches (or

BLAST searches) according to gene identifiers allows easy identification of gene groups located within one scaffold and direct assessment of their proximity. By this approach, clusters of minicollagen genes were identified in four cnidarian genomes (Fig.6l). Transcription orientation of the genes was obtained from corresponding genome annotation files (*.gff). In addition to conventional minicollagens (Fig.6l), a number of novel proteins with CRDs typical of minicollagens were identified (for example, ***s4_g179***, ***s4_g183*** and ***s4_g195*** in *Aurelia*). One possible scenario is that these genes represent ongoing gene duplication events within clusters of minicollagens. In *Aurelia*, based on RNAseq data, four out of five novel minicollagen-like genes are co-expressed with the conventional minicollagens.

It cnidarians, similar to that in Bilateria the principles of gene clustering and collinearity are also utilized, but are applied to the development of cnidarian-specific cell type. Presence of a phylotypic gene cluster harbouring minicollagens underscores the generality of evolutionary processes and is a good example of parallelisms in animal evolution.

Supplementary Tables (stored in **the** Excel file **- "Suppl_Tables_S1-17.xls"**)

Table S1. Sampling locations.

Table S2. Summary of transcriptome assemblies and corresponding RNA-seq libraries.

Table S3. Comparison of BUSCO values for the assemblies of genomes and transcriptomes.

Table S4. Results of TBLASTN search for 141 "missing" BUSCOs among predicted mRNAs of Aurelia (Baltic sea genome ABSv1) and in the transcriptomes of Aurelia, Chironex, Copula, *Nemopilema* and *Tripedalia* (cut-off value E-value = 1e-5).

Table S5. Datasets (proteomes and genomes) used in comparative analyses (phylogeny reconstruction, molecular dating and expression profiles).

Table S6. Numbers of shared orthologs between pairs of species (48 proteomes). Subset of this matrix is shown in Fig.3a.

Table S7. Genes with jellyfish-specific expression in *Aurelia* (expression values, best BLASTP hits and peptide sequences). Dataset used in Fig.3b.

Table S8. Genes with polyp-specific expression in *Aurelia* (expression values, best BLASTP hits and peptide sequences). Dataset used in Fig.3b.

Table S9. InterProscan annotation of *Aurelia* genes with jellyfish-specific expresison.

Table S10. InterProscan annotation of *Aurelia* genes with polyp-specific expresison.

Table S11. Orthologs of *Aurelia* and *Morbakka* with jellyfish-specific expression - shared set of 97 genes shown in Fig.3c.

Table S12. *Aurelia* genes (OG list) with jellyfish-specific expression (dataset used in Fig.S13a)

Table S13. *Morbakka* genes (OG list) with jellyfish-specific expression (dataset used in Fig.S13a)

Table S14. *Nemopilema* genes (OG list) with jellyfish-specific expression (dataset used in Fig.S13a)

Table S15. *Tripedalia* genes (OG list) with jellyfish-specific expression (dataset used in Fig.S13a)

Table S16. Number of genes with transcription factor domains

Table S17. Conservation of the nematocyte-specific proteins among the representatives of the Anthozoa and the Medusozoa.

Supplementary References

1. Dawson MN, Jacobs DK. (2001) Molecular evidence for cryptic species of *Aurelia aurita* (Cnidaria, Scyphozoa). Biol Bull. 200(1):92-6.

2. Schroth W, Jarms G, Streit B, Schierwater B. (2002) Speciation and phylogeography in the cosmopolitan marine moon jelly, *Aurelia sp*. BMC Evol Biol. 2:1.

3. Hubot N, Lucas CH, Piraino S. (2017) Environmental control of asexual reproduction and somatic growth of *Aurelia spp.* (Cnidaria, Scyphozoa) polyps from the Adriatic Sea. PLoS One. 2017 Jun 14;12(6):e0178482. doi: 10.1371/journal.pone.0178482.

4. Schlötterer C, Hauser MT, von Haeseler A, Tautz D. (1994) Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. Mol Biol Evol. 11(3):513-22.

5. Fuchs B, Wang W, Graspeuntner S, Li Y, Insua S, Herbst EM, Dirksen P, Böhm AM, Hemmrich G, Sommer F, Domazet-Lošo T, Klostermeier UC, Anton-Erxleben F, Rosenstiel P, Bosch TC, Khalturin K. (2014) Regulation of polyp-to-jellyfish transition in *Aurelia aurita*. *Curr Biol.* ***24(3)***, 263-73.

6. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. (2017) GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 33(14):2202-2204. doi: 10.1093/bioinformatics/btx153.

7. Adachi K, Miyake H, Kuramochi T, Mizusawa K, Okumura S. (2016). Genome size distribution in phylum Cnidaria. Fisheries Science. 10.1007/s12562-016-1050-4.

8. Goldberg RB, Crain WR, Ruderman JV, Moore GP, Barnett TR, Higgins RC, Gelfand RA, Galau GA, Britten RJ, Davidson EH.. (1975) DNA sequence organization in the genomes of five marine invertebrates. Chromosoma 51(3):225–251.

9. Kotova AV, Shaposhnikova TG, Podgornaya OI, Adonin LS (2014) Karyological and molecular phylogenetic analysis of the three populations of *Aurelia aurita* (White, Black and Japanese seas. Abstract from 49th European Marine Biology Symposium, Saint-Petersburg, Russia, September 8-12.

10. Toshino S, Miyake H, Ohtsuka S, Okuizumi K, Adachi A, Hamatsu Y, Urata M, Nakaguchi K, Yamaguchi S. (2013). Development and polyp formation of the giant box jellyfish *Morbakka virulenta* (Kishinouye, 1910) (Cnidaria: Cubozoa) collected from the Seto Inland Sea, Western Japan. Plankton and Benthos Research. 8. 1-8. 10.3800/pbr.8.1.

11. Toshino S, Miyake H, Ohtsuka S, Adachi A, Kondo Y, Okada S, Hirabayashi T, Hiratsuka T. (2015) Monodisc strobilation in Japanese giant box jellyfish *Morbakka virulenta* (Kishinouye, 1910): a strong implication of phylogenetic similarity between Cubozoa and Scyphozoa. Evol Dev. 17(4):231-9. doi: 10.1111/ede.12127.

12. Zacharias H, Anokhin B, Khalturin K, Bosch TC. (2004) Genome sizes and chromosomes in the basal metazoan *Hydra*. Zoology (Jena). 107(3):219-27.

13. Bolger AM, Lohse M, Usadel B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114-20. doi: 10.1093/bioinformatics/btu170.

14. Shao Z, Graf S, Chaga OY, Lavrov DV. (2006) Mitochondrial genome of the moon jelly *Aurelia aurita* (Cnidaria, Scyphozoa): A linear DNA molecule encoding a putative DNA-dependent DNA polymerase. Gene. 2006 Oct 15;381:92-101.

15. Smith DR, Kayal E, Yanagihara AA, Collins AG, Pirro S, Keeling PJ. (2012) First complete mitochondrial genome sequence from a box jellyfish reveals a highly fragmented linear architecture and insights into telomere evolution. Genome Biol Evol. 2012;4(1):52-8. doi: 10.1093/gbe/evr127.

16. Leggett RM, Clavijo BJ, Clissold L, Clark MD, Caccamo M. (2014) NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries. Bioinformatics. 30(4):566-8. doi: 10.1093/bioinformatics/btt702.

17. Marcais G and Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27(6): 764-770

18. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. (2011) Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 27(4):578-9. doi: 10.1093/bioinformatics/btq683.

19. Huang S, Kang M, Xu A. (2017) HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. Bioinformatics. 33(16):2577-2579.

20. Parra G, Bradnam K, Korf I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 23(9):1061-7.

21. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31(19):3210-2. doi: 10.1093/bioinformatics/btv351.

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

23. Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-4.0*. http://www.repeatmasker.org

24. Stanke M, O. Schöffmann , B. Morgenstern, S. Waack (2006) Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7, 62.

25. Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith Jr, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D. et al. (2003) Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res., 31, 5654-5666.

26. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. (2011) Full-length

transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol. 29(7):644-52. doi: 10.1038/nbt.1883.

27. Iseli C, Jongeneel CV, Bucher P. (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol. 1999:138-48.

28. Haas & Papanicolaou, http://transdecoder.github.io

29. Li W, Godzik A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22(13):1658-9.

30. Langmead B, Salzberg SL. (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods. 9(4):357-9. doi: 10.1038/nmeth.1923.

31. Li B, Dewey CN. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12:323. doi: 10.1186/1471-2105-12-323.

32. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, Terry A, Shapiro H, Lindquist E, Kapitonov VV et al. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* ***317(5834)***, 86-94.

33. Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, Kuo A, Mitros T, Salamov A, Carpenter ML et al. (2008) The *Trichoplax* genome and the nature of placozoans. 454(7207):955-60.

34. Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, Mitros T, Richards GS, Conaco C, Dacre M, Hellsten U et al. (2010) The *Amphimedon queenslandica* genome and the evolution of animal complexity. 466(7307):720-6.

35. Collins AG, Schuchert P, Marques AC, Jankowski T, Medina M, Schierwater B. (2006) 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. 55(1), 97-115.

36. Simakov O, Kawashima T, Marlétaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, Bredeson J, Shoguchi E, Gyoja F et al. (2015) Hemichordate genomes and deuterostome origins. Nature *527(7579)*, 459-65.

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

38. Stamatakis A (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics 30 (9): 1312-1313.

39. Lartillot N, Rodrigue N, Stubbs D, Richer J. (2013) PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst Biol. 62(4):611-5. doi: 10.1093/sysbio/syt022.

40. Cartwright P, Halgedahl SL, Hendricks JR, Jarrard RD, Marques AC, Collins AG, Lieberman BS. (2007) Exceptionally preserved jellyfishes from the Middle Cambrian. PLoS One. 2(10):e1121.

41. Liu AG, Matthews JJ, Menon LR, McIlroy D, Brasier MD. (2014) *Haootia quadriformis* n. gen., n. sp., interpreted as a muscular cnidarian impression from the Late Ediacaran period (approx. 560 Ma). Proc Biol Sci. 281(1793). pii: 20141202. doi: 10.1098/rspb.2014.1202.

42. Benton, M.J., Donoghue, P.C.J., and Asher, R.A. 2009. Calibrating and constraining molecular clocks, p. 35-86. In Hedges, S.B. and Kumar, S. (eds.), The Timetree of Life. Oxford University Press, Oxford.

43. Benton, Michael J., Donoghue, Philip C.J., Asher, Robert J., Friedman, Matt, Near, Thomas J., and Vinther, Jakob. (2015) Constraints on the timescale of animal evolutionary history. Palaeontologia Electronica 18.1.1FC; 1-106; palaeo-electronica.org/content/fc-1

44. Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, Rattei T, Balasubramanian PG, Borman J, Busam D. et al. (2010). The dynamic genome of Hydra. *Nature* ***464(7288)***, 592-6.

45. Kamm K, Schierwater B, Jakob W, Dellaporta SL, Miller DJ. (2006) Axial patterning and diversification in the cnidaria predate the Hox system. *Curr Biol.* ***16(9)***, 920-6.

46. Baumgarten S, Simakov O, Esherick LY, Liew YJ, Lehnert EM, Michell CT, Li Y, Hambleton EA, Guse A, Oates ME, Gough J, Weis VM, Aranda M, Pringle JR, Voolstra CR. (2015) The genome of *Aiptasia*, a sea anemone model for coral symbiosis. *PNAS*. ***112(38)***, 11893-8.

47. Steinmetz PR, Kraus JE, Larroux C, Hammel JU, Amon-Hassenzahl A, Houliston E, Wörheide G, Nickel M, Degnan BM, Technau U. (2012) Independent evolution of striated muscles in cnidarians and bilaterians. *Nature*. ***487(7406)***, 231-4.

48. Chapman DM (1978) Microanatomy of the cubopolyp, *Tripedalia cystophora* (Class Cubozoa). Helgoländer wiss. Meeresunters. 31, 128-168.

49. Matsuno A. & Hisamatsu T. (1982) Electron microscopy on striated muscles in the ephyra of *Aurelia aurita*. Mem.Fac.Sci., Shimane Univ., 16, 61-67.

50. Tanaka H, Ishimaru S, Nagatsuka Y, Ohashi K. (2018) Smooth muscle-like Ca2+-regulation of actin-myosin interaction in adult jellyfish striated muscle. Sci. Rep. 8(1):7776.

51. Kusserow A, Pang K, Sturm C, Hrouda M, Lentfer J, Schmidt HA, Technau U, von Haeseler A, Hobmayer B, Martindale MQ, Holstein TW. (2005) Unexpected complexity of the Wnt gene family in a sea anemone. *Nature*. ***433(7022)***, 156-60.

52. Ryan JF, Burton PM, Mazza ME, Kwong GK, Mullikin JC, Finnerty JR. (2006) The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: evidence from the starlet sea anemone, *Nematostella vectensis*. Genome Biol. 7(7):R64.

53. Ryan JF, Mazza ME, Pang K, Matus DQ, Baxevanis AD, Martindale MQ, Finnerty JR. (2007) Pre-bilaterian origins of the Hox cluster and the Hox code: evidence from the sea anemone, *Nematostella vectensis*. PLoS One. 2(1):e153.

54. Masuda-Nakagawa LM, Gröer H, Aerne BL, Schmid V. (2000) The HOX-like gene Cnox2-Pc is expressed at the anterior region in all life cycle stages of the jellyfish *Podocoryne carnea*. Dev Genes Evol. 210(3):151-6.

55. Chiori R, Jager M, Denker E, Wincker P, Da Silva C, Le Guyader H, Manuel M, Quéinnec E. (2009) Are Hox genes ancestrally involved in axial patterning? Evidence from the hydrozoan Clytia hemisphaerica (Cnidaria). *PLoS One **4(1)***:e4231. doi: 10.1371/journal.pone.0004231.

56. David CN, Özbek S, Adamczyk P, Meier S, Pauly B, Chapman J, Hwang JS, Gojobori T, Holstein TW. (2009) Evolution of complex structures: minicollagens shape the cnidarian nematocyst. *Trends Genet. **24(9)***, 431-8.

57. Wittlieb J, Khalturin K, Lohmann JU, Anton-Erxleben F, Bosch TC. (2006) Transgenic Hydra allow in vivo tracking of individual stem cells during morphogenesis. Proc Natl Acad Sci U S A. 103(16):6208-11.

58. Balasubramanian PG, Beckmann A, Warnken U, Schnölzer M, Schüler A, Bornberg-Bauer E, Holstein TW, Ozbek S. (2012) Proteome of Hydra nematocyst. *J Biol Chem. **287(13)***, 9672-81.