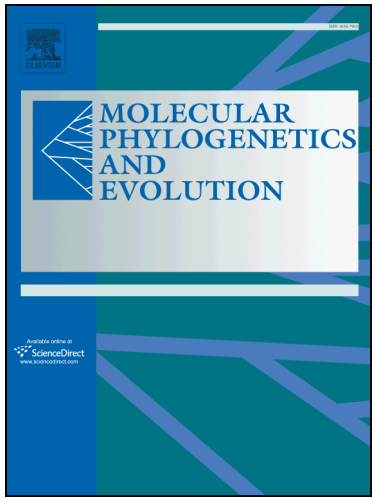


Short Communication

Complete mitochondrial genome and evolutionary analysis of *Turritopsis dohrnii*, the "immortal" jellyfish with a reversible life-cycle

A.A. Lisenkova, A.P. Grigorenko, T.V. Tyajelova, T.V. Andreeva, F.E. Gusev, A.D. Manakhov, A.Yu. Goltsov, S. Piraino, M.P. Miglietta, E.I. Rogaev

PII: S1055-7903(16)30335-9
DOI: http://dx.doi.org/10.1016/j.ympev.2016.11.007
Reference: YMPEV 5666

To appear in: *Molecular Phylogenetics and Evolution*

Received Date: 5 April 2016
Revised Date: 10 October 2016
Accepted Date: 10 November 2016

Please cite this article as: Lisenkova, A.A., Grigorenko, A.P., Tyajelova, T.V., Andreeva, T.V., Gusev, F.E., Manakhov, A.D., Goltsov, A.Yu., Piraino, S., Miglietta, M.P., Rogaev, E.I., Complete mitochondrial genome and evolutionary analysis of *Turritopsis dohrnii*, the "immortal" jellyfish with a reversible life-cycle, *Molecular Phylogenetics and Evolution* (2016), doi: http://dx.doi.org/10.1016/j.ympev.2016.11.007

Complete mitochondrial genome and evolutionary analysis of *Turritopsis dohrnii,* the "immortal" jellyfish with a reversible life-cycle

Lisenkova A.A.[a], Grigorenko A.P.[abc], Tyajelova T.V.[a], Andreeva T.V.[ac], Gusev F.E.[ab], Manakhov A.D.[ad], Goltsov A.Yu.[a], Piraino S.[e], Miglietta M.P[f] and Rogaev E.I.[abcd]

[a]Department of Genomics and Human Genetics, Laboratory of Evolutionary Genomics, Vavilov Institute of General Genetics, Russian Academy of Sciences, Gubkina 3, Moscow, 119991 Russia; [b]Brudnick Neuropsychiatric Research Institute, University of Massachusetts Medical School, 303 Belmont Street, Worcester, MA 01604, USA; [c]Center for Brain Neurobiology and Neurogenetics, Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia; [d]Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow, 119991 Russian Federation; [e]Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, Università del Salento, I-73100 Lecce, Italy; [f]Texas A&M University at Galveston, Dept. of Marine Biology, OCSB, Galveston, TX, 77553.

Abstract

Turritopsis dohrnii (Cnidaria, Hydrozoa, Hydroidolina, Anthoathecata) is the only known metazoan that is capable of reversing its life cycle via morph rejuvenation from the adult medusa stage to the juvenile polyp stage. Here, we present a complete mitochondrial (mt) genome sequence of *T. dohrnii*, which harbors genes for 13 proteins, two transfer RNAs, and two ribosomal RNAs. The *T. dohrnii* mt genome is characterized by typical features of species in the Hydroidolina subclass, such as a high A+T content (71.5%), reversed transcriptional orientation for the large rRNA subunit gene, and paucity of CGN codons. An incomplete complementary duplicate of the *cox1* gene was found at the 5' end of the *T. dohrnii* mt chromosome, as were variable repeat regions flanking the chromosome. We identified species-specific variations (*nad5*, *nad6*, *cob*, and *cox1* genes) and putative selective constraints (*atp8*, *nad1*, *nad2*, and *nad5* genes) in the mt genes of *T. dohrnii*, and predicted alterations in tertiary structures of respiratory chain proteins (NADH4, NADH5, and COX1 proteins) of *T. dohrnii*. Based on comparative analyses of available hydrozoan mt genomes, we also determined the taxonomic relationships of *T. dohrnii,* recovering Filifera IV as a paraphyletic taxon, and assessed intraspecific diversity of various Hydrozoa species.

Keywords

Hydrozoa; *Turritopsis dorhnii*; linear mitochondrial DNA; phylomitogenomics; mitochondrial genome structure; intraspecific diversity.

1. Introduction

Tissue and cell differentiation are key morphogenetic processes of metazoans and are usually coupled with irreversible cell cycle arrest and final cellular commitment. The jellyfish *Turritopsis dohrnii* (Weismann, 1883) (Hydrozoa, Oceaniidae) was the first metazoan described

with the ability to break this biological dogma by reversing its ontogeny under stressful conditions (Piraino et al., 2004; Schmich et al., 2007), which is manifested by back transformation of either the newly liberated, immature medusa, or adult specimen with ripened gonads (Piraino et al., 1996) into the preceding stage, the polyp (Bavestrello et al. 1992). This remarkable feature and its high reproducibility has generated broad interest in this species as a potential biological model for research on ageing and the molecular mechanisms of cell differentiation and transdifferentiation (Bosch et al., 2014; Petralia et al., 2014; Sanchez-Alvarado and Yamanaka, 2014).

Although several genes have been sequenced in *T. dohrnii* for phylogenetic reconstruction (Miglietta et al., 2007; Miglietta and Lessios, 2009; Quiquand et al., 2009, Devarapalli et al., 2014; Zheng et al., 2014), the complete mitochondrial genome has not yet been reported. Here we present the complete sequence of the mitochondrial genome of *T. dohrnii*, including analysis of its structure and organization. We also explore the phylogenetic relationship between *T. dohrnii* and other hydrozoan species for which complete or nearly complete mitochondrial genomes are available, and their intraspecific diversity.

2. Materials and Methods

2.1. *DNA extraction, PCR amplification and sequencing*

Colonies of *Turritopsis dohrnii* were collected in shallow waters (2 m depth) on vertical rocky cliffs at two different locations (Italy and Turkey): the T10 population was sampled near Punta Palacia, Otranto (Strait of Otranto, Italy), and the BP population was sampled at Kuş Burnu (Bird Point) on the northern side of the large Turkish island of Gokceada.

Batches of ten hydranths (T10) or 15 hydranths and 9 medusas (BP) isolated from the same parental colony were pooled in separate Eppendorf tubes. Total DNA was extracted using QIAGEN Mini Spin Columns according to manufacturer's protocols (QIAGEN). For preparation of Illumina PE genomic libraries for *T. dohrnii* BP specimens, gDNA was fragmented on a Covaris S220 Focused ultrasonicator. Indexed paired-end (PE) genomic libraries were constructed using the NEBNext® Ultra™ DNA Library Prep Kit (NEB). For preparation of libraries from T10 specimens, 1 µg of genomic DNA was sheared on a Covaris S220 Focused ultrasonicator, and PE genomic libraries were constructed using the End-It™ DNA End-Repair Kit, Exo-Minus Klenow DNA Polymerase for A-Tailing and the Fast-Link- DNA Ligation Kit (Epicentre, Illumina) with PE adapters (Illumina). PfuUltra II Fusion HS DNA Polymerase (Agilent Technologies) and PCR primers PE 2.0 and PE 1.0 were used to amplify size-selected fractions of T10 specimen libraries.

Sequencing was performed on a HiSeq 2000 sequencer (Illumina) at the Vavilov Institute of General Genetics RAS (Moscow, Russia) and at UMASS Medical School (Worcester, USA).

Paired-end reads from the 100 base pair (bp) HiSeq 2000 lane for *T. dohrnii* BP specimens were assembled *de-novo* using Minia 2.0.2 (Salikhov et al., 2013). Long mitochondrial contigs were identified through BLAST searches against closely related hydrozoan species and were used to obtain a draft mt genome sequence of *T. dohrnii*. Based on this preliminary reconstruction, over thirty PCR oligonucleotide primers were designed for additional amplification and Sanger sequencing to ensure the correct gene order and boundaries (Supplementary Table 1). The incorporation of Sanger sequencing results into the draft mt genome sequence was accompanied by additional alignments of HiSeq 2000 paired-end reads using bowtie2 v.2.2.2 (Langmead and Salzberg, 2012). Flanking sequences that contained duplicated portions of the *cox1* gene were difficult to align using deep sequencing data. These sequences were recovered by PCR amplification using specific primers (Tur_L0_R, Tur_L1_R, Nst6; Supplementary Table 1). PCR was conducted using PicoMaxx High-Fidelity PCR Master Mix (Agilent Technologies) on *T. dohrnii* total DNA. Long-range PCR was performed using GoTaq Long PCR Master Mix (Promega). All PCR products were Sanger sequenced on a 3730*xl* DNA Analyzer (Applied Biosystems).

Authenticity of the complete mt genome was ensured by checking sequencing quality and coverage for each individual nucleotide position. Every position that contained more than 3% of aligned reads with non-reference alleles was checked against the Sanger sequencing data to exclude possible nuclear DNA reads.

Coverage of mt genome for the BP and T10 sequences (excluding the duplicated portion of the *cox1* gene from each end of the chromosome) was calculated using GATK v3.3-0 (DePristo et al., 2011; Van der Auwera et al., 2013). Average coverages for the BP and T10 mt genomes were ~900-fold and ~130-fold, respectively. The complete mt sequence was reconstructed for BP specimen and near complete sequence was determined for T10 specimen, except for the latter *rrnL* gene and both copies of the *cox1* gene were partially recovered.

2.2. *Sequence annotation and analysis*

Gene annotations for the assembled complete mtDNA sequence were made using MITOS (Bernt et al., 2013) and then checked against GenBank and sequences of other hydrozoans using BLAST (Madden, 2003). ORFs for protein-coding genes were specified and adjusted using ExPASy (Gasteiger, 2003) and ORF Finder (Wheeler, 2003). In all analyses, the Coelenterate Mitochondrial Code was used for translation.

Predictions of tRNA genes and their secondary structures were performed using the ARWEN (Laslett and Canbäck, 2008) and tRNAscan-SE (Schattner et al., 2005) services along with comparisons to other hydrozoan species. The boundaries of rRNA genes were predicted in the locARNA software using the locARNA-P probabilistic mode (Will et al., 2007; Smith et al., 2010; Will et al., 2012) and were then checked against closely related hydrozoans.

Tertiary structures of *Turritopsis dohrnii* mt proteins were modeled using the I-TASSER server (Zhang, 2008; Roy et al., 2010; Yang et al., 2015) for protein prediction on the default settings and were visualized in Pymol (https://www.pymol.org).

All structural and sequence analyses were performed using the complete mt genomic sequence of a *T. dohrnii* BP specimen. Analyses of nucleotide composition of the mt genomes was conducted using MEGA6 (Tamura et al., 2013). Percentage similarity between sequences was assessed using MUSCLE (Edgar, 2004). Codon usage for protein-coding sequences was estimated using the CUSP program in EMBOSS (Rice et al., 2000) on the concatenated set of all genes of each *T. dohrnii* specimen (excluding the *cox1* duplicate at the 5'-end of the mt chromosome).

2.3. *Phylogenetic analysis*

The phylogenetic position of *Turritopsis dohrnii* was assessed on a mtDNA data set including most Hydroidolina species with complete or nearly complete mitochondrial genomes and the subclass Trachylina as the outgroup (Supplementary Table 2) (Cartwright et al., 2008). Both BP and T10 specimens of *T. dohrnii* were used in phylogenetic reconstructions.

Multiple alignments of nucleotide and amino acid sequences were individually created for all protein-coding genes using MUSCLE as implemented in MEGA v.6.06. The duplicate *cox1* gene was never used in phylogenetic analysis. Alignments were then checked and corrected using Gblocks (Talavera and Castresana, 2007).

We then proceeded to data partitioning based on the codon positions of each gene individually using PartitionFinder v.1.1.1 (Lanfear et al., 2014). Subsets and respective best-fitting models were chosen using the Bayesian Information Criterion (BIC) for Maximum Likelihood (ML) and Bayesian Inference (BI) analysis. Subsets containing third codon positions were then excluded from all further analyses.

Multiple alignments of the complete *rrnS* sequences along with consensus secondary structure were created using the PicXAA-R server with default settings (Sahraeian & Yoon 2011).

The final concatenated dataset included 13 protein-coding genes and the *rrnS* gene and was 12,478 bp in length. The datasets comprising *rrnS* sequences and consensus secondary structures for their multiple alignment, as well as 13 concatenated protein-coding gene sequences were analyzed separately.

ML analyses for these datasets were conducted using RAxML (Stamatakis, 2014) with the GTR+G model. For *rrnS* dataset, consensus secondary structure was taken into account using the S16 model. For the combined dataset, rRNA sequences were analyzed in one separate partition, along with partitions for the protein-coding genes. Node support was evaluated by bootstrapping with 2,000 replicates.

BI phylogenetic analyses were performed in MrBayes (Ronquist et al., 2011) using the Metropolis-coupled Markov chain Monte Carlo (MCMCMC) algorithm with 3 heated and 1 cold chains (heating coefficient = 0.2), 2,000,000 generations and a burn-in of 25% from the cold chain. Sampling occurred every 500 generations. The *rrnS* sequence alignment, in both the standalone and combined datasets, was manually divided into two partitions based on the elements of its consensus secondary structure; nucleotide pairs belonging to stems were analyzed under the doublet model and unpaired nucleotides from loops were analyzed under 4by4 model. Convergence of the parameters in use was inferred using Tracer v1.6 (Rambaut et al., 2014). Bayesian Posterior Probabilities (PP) were calculated as a measure of node support.

Estimates of evolutionary and selective constraints were made using branch-site model A in PAML v.4.8a (Yang, 2007). Estimates were performed on a concatenated set of protein-coding genes only for the same set of species. For Model A, the Bayes Empirical Bayes (BEB) statistical test was used to calculate the posterior probabilities. The FMutSel-F model of codon substitution (CodonFreq = 7; estFreq = 0) (Yang and Nielsen, 2008) was used for selective constraint estimation.

In order to infer intraspecific nucleotide diversity within Hydrozoa, we examined the number of nucleotide differences per site (π) in species with at least two complete or partial mitochondrial genomes available from databases (Supplementary Table 2). The number of segregating (polymorphic) sites (S) and π were calculated using DnaSP v.5.10.1 (Librado and Rozas, 2009; Rozas, 2009). Interspecific comparison of π values was conducted using Student's t-test in STATISTICA v.10 (StatSoft, Inc., 2011).

3. Results and Discussion

3.1. *Organization, structure and nucleotide composition of Turritopsis dohrnii mitochondrial genome*

We obtained two sequences of the *Turritopsis dohrnii* mitochondrial genome, complete sequence for BP specimen (GenBank submission #KT020766) and incomplete for T10 specimen (#KT899097). Mitochondrial genome of *T. dohrnii* was organized as a single 15,425 bp linear chromosome with 13 protein-coding genes, two tRNAs ($tRNA^{Met(cau)}$, *trnM*; and $tRNA^{Trp(uca)}$, *trnW*) and two rRNA subunits (Table 1). The boundaries for most genes were separated by intergenic regions with an overall length of 150 bp. Several neighboring genes overlapped by 1, 2, 10 or 47 nucleotides (Supplementary Fig. 1).

Gene arrangement in the *T. dohrnii* mt chromosome was found to be most similar to that of non-aplanulatan Hydroidolina species (Kayal et al., 2015). We identified a partial duplicate of the *cox1* gene at the 5'-end of the *T. dohrnii* mitochondrial genome that corresponds to the 3' part of the *cox1* gene, overlaps with the end of the 16S rRNA gene by 47 bp, and has a reversed nucleotide sequence orientation. This feature is also found in Aplanulata species, members of Capitata, Leptothecata, and Filifera III. A highly polymorphic region is located at the 5' end of the mt chromosome downstream of the *cox1* duplicate. It consists of multiple variations of nucleotide sequence motif (gggggggg), variable in content and length. For the $(G)_8$ variant of this repeat region, an inverted copy was found at the 3' end of the *T. dohrnii* chromosome, with no variation in contents. Although similar (but longer) repeats occur in various members of Hydroidolina, *T. dohrnii* is the only species for which they are known to be highly polymorphic at the 5' end of the mt chromosome. The approximated percentage of heteroplasmy for the polymorphic repeat region at the 5'-end of the mt chromosome was 67%. The degree of heteroplasmy observed in this region was the following: position 1 G/A, with allele A reaching 13% at 237-fold coverage; position 2 G/GGG, where (GG) insertions had 27% frequency at 315-fold coverage; position 9 A/G, with allele A reaching 60% and allele G 40% frequency, respectively (at 576-fold coverage).

According to analyses of nucleotide composition of the complete mt genome, *T. dohrnii* skewed strongly towards high A+T (71.5%) content, which is typical for the class Hydrozoa (Kayal et al., 2012).

The analysis of intraspecific diversity revealed relatively high diversity of mt genome sequences between the two *T. dohrnii* colonies (T10 and BP) from the Strait of Otranto, Italy and Bird Point, Turkey (Fig. 1; Table 2). For different protein-coding mt genes, the average number of nucleotide differences per site (π) in *T. dohrnii* varied from 1.7% to 4.2%. The variations at

synonymous sites, $\pi_{synonymous}$, were as high as 12.8% in the *atp8* gene. In order to determine if these high values are unique for *T. dohrnii*, a comparative analysis of intraspecific mtDNA sequence diversity between five other Hydrozoa species with at least two published complete or partial mitochondrial genomes was performed (Fig. 1; Table 2). Two of the other species analyzed, *Alatina moseri* (Smith et al., 2011) and *Craspedacusta sowerbyi*, exhibited similarly high intraspecific diversity at synonymous sites (Fig. 1, A; Table 2). No statistically significant differences were found between $\pi_{synonymous}$ of the protein-coding mt genes in pairwise analysis of these three species (p=0.32; 0.33; and 0.93, t-test). *Physalia physalis* and *Ectopleura larynx* demonstrated significantly lower values of $\pi_{synonymous}$ for the protein-coding mt genes (however, we could not confirm the genetic isolation for sampled specimens of these two species). No statistical differences in $\pi_{synonymous}$ (p=0.16, t-test) were found between these two species. The three species with high (*T. dohrnii*, *A. moseri,* and *C. sowerbyi*) and two species with low intraspecific diversity (*P. physalis* and *E. larynx)* differed from each other significantly in pairwise comparison. The $\pi_{nonsynonymous}$ varied greatly between species and genes (Fig. 1, B); however, no statistical differences were found among species in the pairwise comparison using t-test.

These results suggest that relatively high genetic heterogeneity exists in natural marine populations of *T. dohrnii*, *A. moseri* and *C. sowerbyi*, and may be indirect evidence of large effective population sizes for these species (Lynch and Conery, 2003). High intraspecific diversity does not appear to be exclusive to T. dohrnii and is observed in other species of Hydrozoa. The factors contributing to putatively high intraspecies diversity in one species but not in others will require larger studies at the population level.

3.2. *Properties of protein-coding genes*

Turritopsis dohrnii had a complete set of 13 mt protein coding genes that are involved in oxidative phosphorylation and respiration. Among them, we found the lowest A+T percentage (61.6%) for *cox1*, whereas *nad4L* and *nad6* had the highest (~78%) (Table 1). Within the three codon positions of these genes, the 3rd position had the highest A+T content (82.4%), which is typical for most of the hydrozoan species within our dataset of 33 species.

Initiation and termination codons for *T. dohrnii* energy pathway protein-coding genes mostly did not differ from its close relatives, except for a few genes: TAG termination codon for *nad6;* GTG start-codon for *nad2* and *nad5* genes, although *nad5* had a putative ATG codon 153 bp upstream of the predicted start of translation with a continuous ORF.

T. dohrnii mtDNA genome incorporated low numbers of CGN family codons (used only one to six times), although they were still found more frequently than in other hydrozoan species (Kayal and Lavrov 2008; Zou et al., 2012; Kayal et al., 2015).

Several proteins of *T. dohrnii* exhibited differences in their tertiary structures compared to those found in its closest relatives, *N. bachei* and *Rathkea octopunctata* (Supplementary Fig. 2). The COX1 protein in *T. dohrnii* was 52 aa longer than in other hydrozoans (Supplementary Fig. 2, A). The 3' portion of the functional *cox1* sequence that was duplicated at the 5' end of the mt chromosome contained 11 of the 15 CGN codons used in the *T. dohrnii* protein coding sequences.

Structural differences were also detected in NADH4 and NADH5 proteins. In the tertiary structure of NADH5, two β-strands at the same position were significantly elongated in both *R. octopunctata* and *T. dohrnii* compared to *N. bachei* (Supplementary Fig. 2, B), and in NADH4, one additional short α-helix and two β-strands are present between the second and third transmembrane helices, where neither *N. bachei* nor *R. octopunctata* have such structural elements (Supplementary Fig. 2, C).

It is of interest that compared to its closest relatives, *T. dohrnii* exhibits changes in tertiary structure and sequence of mt proteins comprising proton-pumping modules of mt NAD:ubiquinone oxidoreductase complex 1 (Zickermann et al., 2015). These primary data must be investigated further.

An amino acid substitution in an evolutionary conserved site inside a putative proton-channel (Met^{384} in *T. dohrnii*, Ile^{384} in other hydrozoans, except Val^{384} in *Plotocnide borealis*) was also detected in the *nad4* gene. Additionally, species-specific amino acid substitutions were found in *nad5* (Ser^{137} in *T. dohrnii* BP sequence), *nad6* (GTG start codons), and *cob* (Tyr^{317}) genes of *T. dohrnii*.

3.3. *Phylogenetic analysis*

Phylogenetic reconstructions using either rRNA or protein-coding sequences alone or combined (sequences of rRNA + protein-coding genes) can yield different results. The results may also differ slightly depending on the method used. All phylogenetic trees, except those reconstructed using only *rrnS*, reveal paraphyly of the Filifera IV group. In reconstructions based on the combined dataset (Fig. 2 and Supplementary File 1, C), *T. dohrnii* and *R. octopunctata* formed a group distinct from the rest of the [Filifera III + Filifera IV] clade, with relatively high support (ML, 69%; BI, 0.997 PP). Additionally, many of the groupings recovered by our combined dataset, such as [Aplanulata + Filifera I + Capitata] (ML, 84%; BI, 0.9 PP), [Leptothecata + Filifera III/IV] (ML, 74%; BI, 1 PP), and Siphonophorae being the earliest

branching taxon within the Hydroidolina were consistent with the findings presented in Kayal et al. 2015. Reconstructions based on either combined rRNA and protein-coding sequences (Fig. 2 and Supplementary File 1, C) or protein-coding data alone (Supplementary File 1, D) were consistent with each other as well as with other reported studies utilizing these markers (Zou et al., 2012; Kayal et al., 2013; Kayal et al., 2015) (Fig. 2), whereas the trees based on *rrnS* sequences and their consensus secondary structure (Supplementary File 1, A and B) were inconsistent in clade formation and had extremely low support values. Additionally, when the combined dataset was used, the inclusion of *rrnS* as an additional phylogenetic marker improved the support values for deeper nodes.

Branch lengths of *Laomedea flexuosa* and *Obelia longissima* were nearly equal, as were branch lengths of *Hydractinia polyclina* and *H. symbiolongicarpus*, which is consistent with previous analysis (Zou et al., 2012). Branch lengths of BP and T10 *T. dohrnii* specimens, however, had greater differences then these species. To assess these results and compare intraspecific diversity of *T. dohrnii* to interspecific nucleotide differences in these four species, nucleotide diversity of their mt genes was compared to π values of other analyzed species (Fig. 1; Table 2). We found that the intraspecific differences of protein-coding genes between *L. flexuosa and O. longissima* were the lowest (π=0.161%) among all analyzed Hydrozoa species. For two *Hydractinia* species, the nucleotide diversity was π=6.18%. For comparison, two *Aurelia aurita* specimens (NC_008446 and HQ694729), which possibly belong to different cryptic subspecies (Dawson and Jacobs, 2001; Park et al., 2012), exhibited significant diversity ranging from 49.2% for *cox3* to 69% for *nad4L* (π=18.69%) (data for genes not shown; total π shown in Table 2). This discrepancy between morphological data, which identifies *L. flexuosa* and *O. longissima* as two different species, and results of the genetic analysis requires further independent studies of other specimens of these species and application of nuclear genomic markers.

Despite relatively high nucleotide diversity of *T. dohrnii* specimens, it was lower than in both *Hydractinia polyclina/H. symbiolongicarpus* (different species) and *A. aurita* (different subspecies) samples. Therefore BP and T10 specimens of *T. dohrnii* most likely represent different populations rather than different subspecies.

We assessed the evolution of *T. dohrnii* mt genes using PAML. Branch-site tests for the *T. dohrnii* crown branch identified 9 amino acid sites with significant BEB statistical test results of at least 0.95 PP (Supplementary Table 3). This finding was statistically significant according to a chi-square test ($p<<0.0001$). These sites were identified for *T. dohrnii* in the *atp8*, *nad1*, *nad2,* and *nad5* genes.

Acknowledgments

This work was supported by Russian Science Foundation Grant № 14-50-00029 to E.I.R., and NSF DEB grant № DEB-1456501 to M.P.M.

The authors would like to thank Vladimir Lebedev for help with phylogenetic analysis verification and advice, and Ellen Kittler for sequencing assistance.

References cited:

1. Bavestrello, G., Sommer, C., Sarà, M., Hughes, R.G., 1992. Bi-directional conversion in *Turritopsis nutricula*. In: Bouilllon, J., Boero, F., Cicogna, F., Gili, J.M., Hughes R.G. (Eds.). Aspects of Hydrozoan Biology Sci. Mar. 56, 137–140.

2. Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., and Stadler, P.F., 2013. MITOS: Improved *de novo* metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69, 313–319, doi:10.1016/j.ympev.2012.08.023.

3. Bosch, T. C. G., Adamska, M., Augustin, R., Domazet-Loso, T., Foret, S., Fraune, S., Funayama, N., et al. (2014). How do environmental factors influence life cycles and development? An experimental framework for early-diverging metazoans. BioEssays, 36, 1185-1194, doi:10.1002/bies.201400065.

4. Cartwright, P., Evans, N.M., Dunn, C.W., Marques, A.C., Miglietta, M.P., Schuchert, P., and Collins, A.G., 2008. Phylogenetics of Hydroidolina (Hydrozoa: Cnidaria). JMBA 88, 1663, doi:10.1017/S0025315408002257.

5. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498, doi:10.1038/ng.806.

6. Devarapalli, P., Kumavath, R.N., Barh, D., Azevedo, V., 2014. The conserved mitochondrial gene distribution in relatives of *Turritopsis nutricula*, an immortal jellyfish. Bioinformation 10, 586-591, doi:10.6026/97320630010586.

7. Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797, doi:10.1093/nar/gkh340.

8. Gasteiger, E., 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31, 3784–3788, doi:10.1093/nar/gkg563.

9. Kayal, E., Bentlage, B., Cartwright, P., Yanagihara, A. A., Lindsay, D. J., Hopcroft, R. R., & Collins, A. G., 2015. Phylogenetic analysis of higher-level relationships within Hydroidolina (Cnidaria: Hydrozoa) using mitochondrial genome data and insight into their mitochondrial transcription. PeerJ 3, e1403, doi:10.7717/peerj.1403.

10. Kayal, E., Bentlage, B., Collins, A.G., Kayal, M., Pirro, S., and Lavrov, D.V., 2012. Evolution of Linear Mitochondrial Genomes in Medusozoan Cnidarians. Genome Biol Evol 4, 1–12, doi:10.1093/gbe/evr123.

11. Kayal, E., and Lavrov, D.V., 2008. The mitochondrial genome of *Hydra oligactis* (Cnidaria, Hydrozoa) sheds new light on animal mtDNA evolution and cnidarian phylogeny. Gene 410, 177–186, doi:10.1016/j.gene.2007.12.002.

12. Kayal, E., Roure, B., Philippe, H., Collins, A.G., and Lavrov, D.V., 2013. Cnidarian phylogenetic relationships as revealed by mitogenomics. BMC Evol. Biol. 13, 5, doi:10.1186/1471-2148-13-5.

13. Lanfear, R., Calcott, B., Kainer, D., Mayer, C., and Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14, 82, doi:10.1186/1471-2148-14-82.

14. Langmead, B., and Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359, doi:10.1038/nmeth.1923.

15. Laslett, D., Canbäck B., 2008. ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24, 172-175, doi:10.1093/bioinformatics/btm573.

16. Librado, P. and Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452, doi:10.1093/bioinformatics/btp187.

17. Lynch, M. and Conery, S., 2003. The Origins of Genome Complexity. Science 302, 1401–1404, doi:10.1126/science.1089370.

18. Madden T. The BLAST Sequence Analysis Tool. 2002 Oct 9 [Updated 2003 Aug 13]. In: McEntyre, J., Ostell, J., editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2002-. Chapter 16. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21097/

19. Miglietta, M.P., and Lessios, H.A., 2009. A silent invasion. Biol. Invasions 11, 825–834, doi:10.1007/s10530-008-9296-0.

20. Miglietta, M.P., Piraino, S., Kubota, S., and Schuchert, P., 2007. Species in the genus Turritopsis (Cnidaria, Hydrozoa): a molecular evaluation. J. Zoolog. Syst. Evol. Res. 45, 11–19, doi:10.1111/j.1439-0469.2006.00379.x.

21. Petralia, R. S., Mattson, M. P., & Yao, P. J. (2014). Aging and longevity in the simplest animals and the quest for immortality. Ageing Res. Rev., 16, 66-82, doi:10.1016/j.arr.2014.05.003.

22. Piraino, S., Boero, F., Aeschbach, B., Schmid, V., 1996. Reversing the life cycle: medusae transforming into polyps and cell transdifferentiation in *Turritopsis nutricula* (Cnidaria, Hydrozoa). Biol. Bull. 190, 302–312.

23. Piraino, S., De Vito, D., Schmich, J., Bouillon, J., Boero F., 2004. Reverse development in Cnidaria. Can. J. Zool. 82, 1748–1754, doi:10.1139/Z04-174.

24. Quiquand, M., Yanze, N., Schmich, J., Schmid, V., Galliot, B., & Piraino, S., 2009. More constraint on ParaHox than Hox gene families in early metazoan evolution. Dev. Biol. 328, 173-187, doi:10.1016/j.ydbio.2009.01.022.

25. Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer.

26. Rice, P. Longden, I., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet 16, 276-277.

27. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst. Biol. 61, 539–542, doi:10.1093/sysbio/sys029.

28. Roy, A., Kucukural, A., & Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc., 5, 725-738, doi:10.1038/nprot.2010.5.

29. Rozas, J., 2009. DNA Sequence Polymorphism Analysis using DnaSP, in: Posada, D. (Ed.) Bioinformatics for DNA Sequence Analysis; Methods in Molecular Biology Series Vol. 537. Humana Press, NJ, USA, pp. 337-350.

30. Sahraeian, S. M. E., & Yoon, B. J., 2011. PicXAA-R: efficient structural alignment of multiple RNA sequences using a greedy approach. BMC Bioinform. 12, S38, doi:10.1186/1471-2105-12-S1-S38.

31. Salikhov, K., Sacomoto, G. and Kucherov, G., 2014. Using cascading Bloom filters to improve the memory usage for de Brujin graphs, Algorithms Mol Biol. 9, 1, doi:10.1186/1748-7188-9-2.

32. Sánchez Alvarado, A., & Yamanaka, S., 2015. Rethinking Differentiation: Stem Cells, Regeneration, and Plasticity. Cell, 157, 110-119, doi:10.1016/j.cell.2014.02.041.

33. Schattner, P., Brooks, A.N., and Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33, W686–W689, doi:10.1093/nar/gki366.

34. Schmich, J., Kraus, Y., De Vito, D., Graziussi, D., Boero, F., and Piraino, S., 2007. Induction of reverse development in two marine Hydrozoans. Int. J. Dev. Biol. 51, 45–56, doi:10.1387/ijdb.062152js.

35. Smith, C., Heyne, S., Richter, A.S., Will, S., and Backofen, R., 2010. Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LOCARNA. Nucleic Acids Res. 38, W373–W377, doi:10.1093/nar/gkq316.

36. Smith, D.R., Kayal, E., Yanagihara, A.A., Collins, A.G., Pirro, S., and Keeling, P.J., 2012. First Complete Mitochondrial Genome Sequence from a Box Jellyfish Reveals a Highly Fragmented Linear Architecture and Insights into Telomere Evolution. Genome Biol. Evol. 4, 52–58, doi:10.1093/gbe/evr127.

37. Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313, doi:10.1093/bioinformatics/btu033.

38. StatSoft, Inc., 2011. STATISTICA (data analysis software system), version 10. www.statsoft.com.

39. Talavera, G., and Castresana, J., 2007. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. Syst. Biol. 56, 564–577, doi:10.1080/10635150701472164.

40. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725–2729, doi:10.1093/molbev/mst197.

41. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al., 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline: The Genome Analysis Toolkit Best Practices Pipeline, in: Bateman, A., Pearson, W.R., Stein, L.D., Stormo, G.D., Yates, J.R. (Eds), Current Protocols in Bioinformatics, (Hoboken, NJ, USA: John Wiley & Sons, Inc.), pp. 11.10.1–11.10.33.

42. Wheeler, D.L., 2003. Database resources of the National Center for Biotechnology. Nucleic Acids Res. 31, 28–33, doi:10.1093/nar/gkg033.

43. Will, S., Joshi, T., Hofacker, I.L., Stadler, P.F., and Backofen, R., 2012. LocARNA-P: Accurate boundary prediction and improved detection of structural RNAs. RNA 18, 900–914, doi:10.1261/rna.029041.111.

44. Will, S., Reiche, K., Hofacker, I.L., Stadler, P.F., and Backofen, R., 2007. Inferring Noncoding RNA Families and Classes by Means of Genome-Scale Structure-Based Clustering. PLoS Comput. Biol. 3, e65, doi:10.1371/journal.pcbi.0030065.

45. Yang, Z., 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24, 1586-1591, doi:10.1093/molbev/msm088.

46. Yang, Z., & Nielsen, R., 2008. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. Mol. Biol. Evol. 25, 568-579, doi:10.1093/molbev/msm284.

47. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. Nat. Methods, 12, 7-8, doi:10.1038/nmeth.3213.

48. Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinform. 9, 40, doi:10.1186/1471-2105-9-40.

49. Zheng, L., He, J., Lin, Y., Cao, W., and Zhang, W., 2014. 16S rRNA is a better choice than COX1 for DNA barcoding hydrozoans in the coastal waters of China. Acta Oceanol. Sin. 33, 55–76, doi:10.1007/s13131-014-0415-8.

50. Zickermann, V., Wirth, C., Nasiri, H., Siegmund, K., Schwalbe, H., Hunte, C., and Brandt, U., 2015. Mechanistic insight from the crystal structure of mitochondrial complex I. Science 347, 44–49, doi:10.1126/science.1259859.

51. Zou, H., Zhang, J., Li, W., Wu, S., and Wang, G., 2012. Mitochondrial Genome of the Freshwater Jellyfish *Craspedacusta sowerbyi* and Phylogenetics of Medusozoa. PLoS ONE 7, e51465, doi:10.1371/journal.pone.0051465.

52. Dawson, M. N., and Jacobs, D. K., 2001. Molecular evidence for cryptic species of *Aurelia aurita* (Cnidaria, Scyphozoa). The Biological Bulletin 200, 92-96.

53. Park, E., Hwang, D. S., Lee, J. S., Song, J. I., Seo, T. K., and Won, Y. J., 2012. Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. Molecular phylogenetics and evolution 62, 329-345, doi:10.1016/j.ympev.2011.10.008.

Tables

Table 1. Nucleotide composition of *T. dohrnii* mitochondrial genes and comparison of the gene properties to other closely related species (*C. multicornis*, *N. bachei, R. octopunctata*). Sizes of genes are listed including termination codons. Asterisks for *N. bachei* and *R. octopunctata* indicate gene lengths that may be biased due to the partial genome reconstructions. Double asterisks indicate start and stop codons for *T. dohrnii* that differ from at least one of the compared species.

Mitochondrial genes	**Nucleotide composition of *T. dohrnii* (%)**					**Number of encoded nucleotides**				**Start codon**	**Stop codon**
	T	***C***	***A***	***G***	***A+T***	***T. dohrnii***	***C. multicornis***	***N. bachei***	***R. octopunctata***		
cox1	33.2	20.9	28.4	17.5	61.6	1758	1566	1300**	>710**	ATG	TAA**
cox2	40	11.1	29	19.9	69.0	738	738	738	738	ATG	TAA
cox3	44.4	13.9	25.3	16.4	69.7	786	786	786	786	ATG	TAA
nad1	41.6	13.5	30.4	14.5	72.0	987	990	990	987	ATG	TAA
nad2	46.1	9.4	31.4	13	77.5	1359	1362	1353	1344	GTG**	TAA
nad3	43.4	12	31.4	13.2	74.8	357	357	357	357	ATG	TAA
nad4	42.1	13.3	31.2	13.4	73.3	1458	1458	1458	1458	ATG	TAA**
nad4L	44.8	9.4	33.3	12.5	78.1	297	300	294	297	ATG	TAA
nad5	43.9	10.6	30.9	14.6	74.8	1830	1833	1833	1833	GTG**	TAA**
nad6	43.9	10.8	34.9	10.4	78.8	558	564	561	558	ATG	TAG
atp6	46.7	10.9	29.5	12.9	76.2	705	705	705	705	ATG	TAA
atp8	45.1	9.8	34.8	10.3	79.9	204	204	204	204	ATG	TAA
cob	39.4	15.3	31.1	14.3	70.5	1140	1143	1143	1146	ATG	TAA
trnM	36.2	13	36.2	14.5	72.4	71	71	69	69	-	-
trnW	34.3	14.3	31.4	20	65.7	70	70	70	70	-	-
rrnL	45.5	14.5	28.2	11.8	73.7	1766	1749	1507*	>1580*	-	-
rrnS	35.2	10.8	37.8	16.2	73.0	907	924	918	902	-	-

Notes: * Gene lengths may be biased due to *N. bachei* and R. octopunctata genome reconstructed partially.
** Start and stop codons of *T. dohrnii* genes different from at least one of compared species.

Table 2. Intra- and interspecific nucleotide diversity (π) of mitochondrial protein-coding and rRNA genes in various hydrozoan species.

Species	Protein-coding genes			rRNA genes
	π, %	$\pi_{synonymous}$, %	$\pi_{nonsynonymous}$, %	π, %
*Turritopsis dohrnii**	2.801	10.659	0.517	1.358
*Alatina moseri**	2.715	10.111	0.302	0.974
Craspedacusta sowerbyi	2.783	9.789	0.54	0.91
Physalia physalis	0.653	1.953	0.255	0.456
Ectopleura larynx	0.594	1.553	0.313	0.14
*Aurelia aurita**	18.691	61.251	5.185	9.386
Laomedea flexuosa/Obelia longissima	0.161	0.637	0.023	7.362*
Hydractinia polyclina/H. symbiolongicarpus	6.18	22.977	1.326	1.82

Notes: * Specimens confirmed as genetic isolates; π, average number of nucleotide differences per site. For intraspecific analysis, species with at least two partial mitochondrial genomes available were chosen.

Figure captions

Fig. 1. Nucleotide diversity (π, %) of protein-coding and rRNA genes of six hydrozoan species at synonymous (A) and nonsynonymous sites (B). *Aurelia aurita* was not included due to its exceptionally high intraspecies diversity.

Fig. 2. Phylogenetic analysis of Hydrozoa class based on nucleotide sequences of *rrnS* + protein-coding sequences (combined dataset) by Maximum likelihood (topology shown as a bootstrap consensus tree). Numbers at the nodes indicate bootstrap support values and Bayesian posterior probabilities respectively. Support values greater than 99%/0.999 PP are not shown. Scale bar indicates number of substitutions per site.

Fig. 1

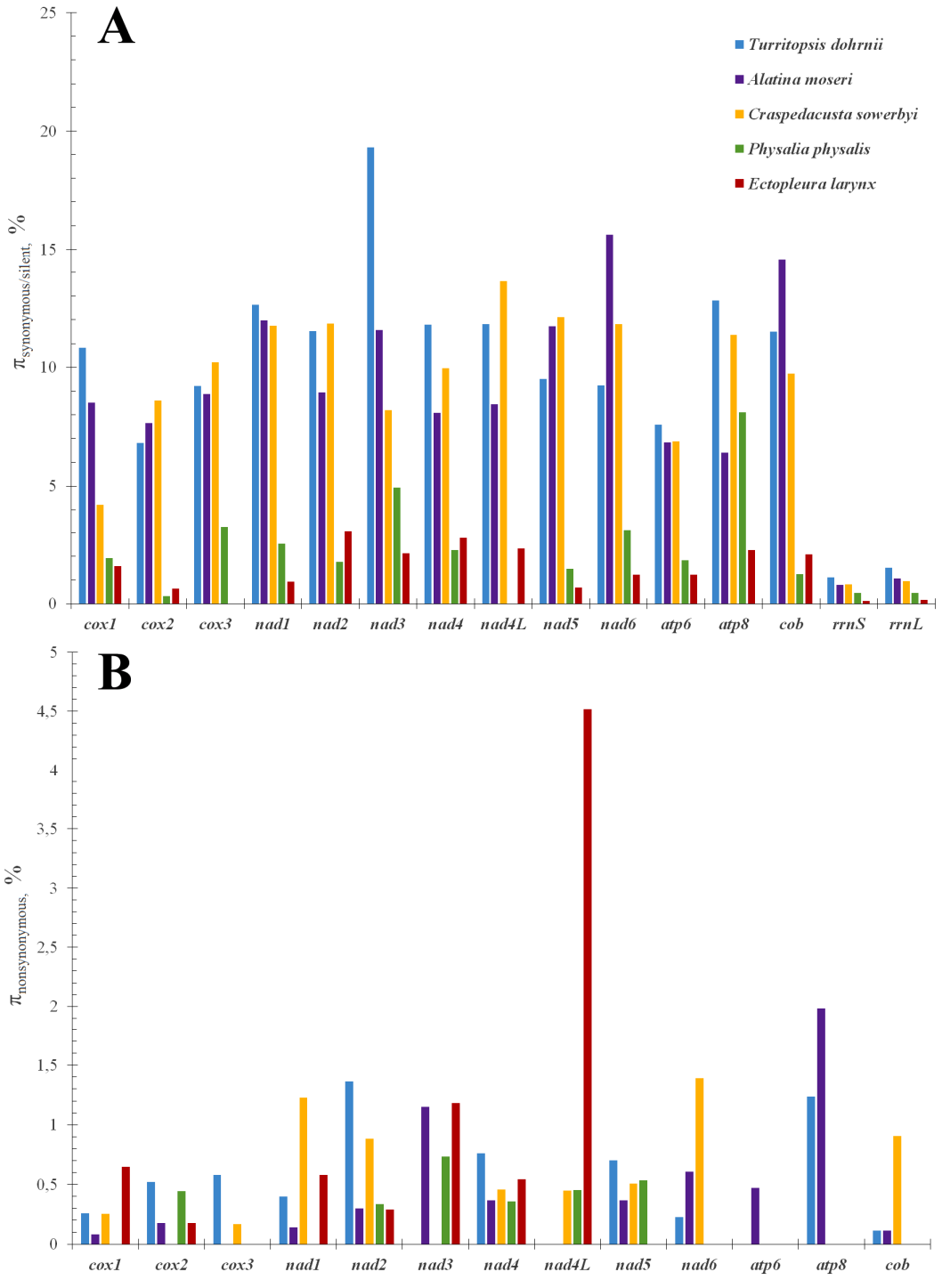
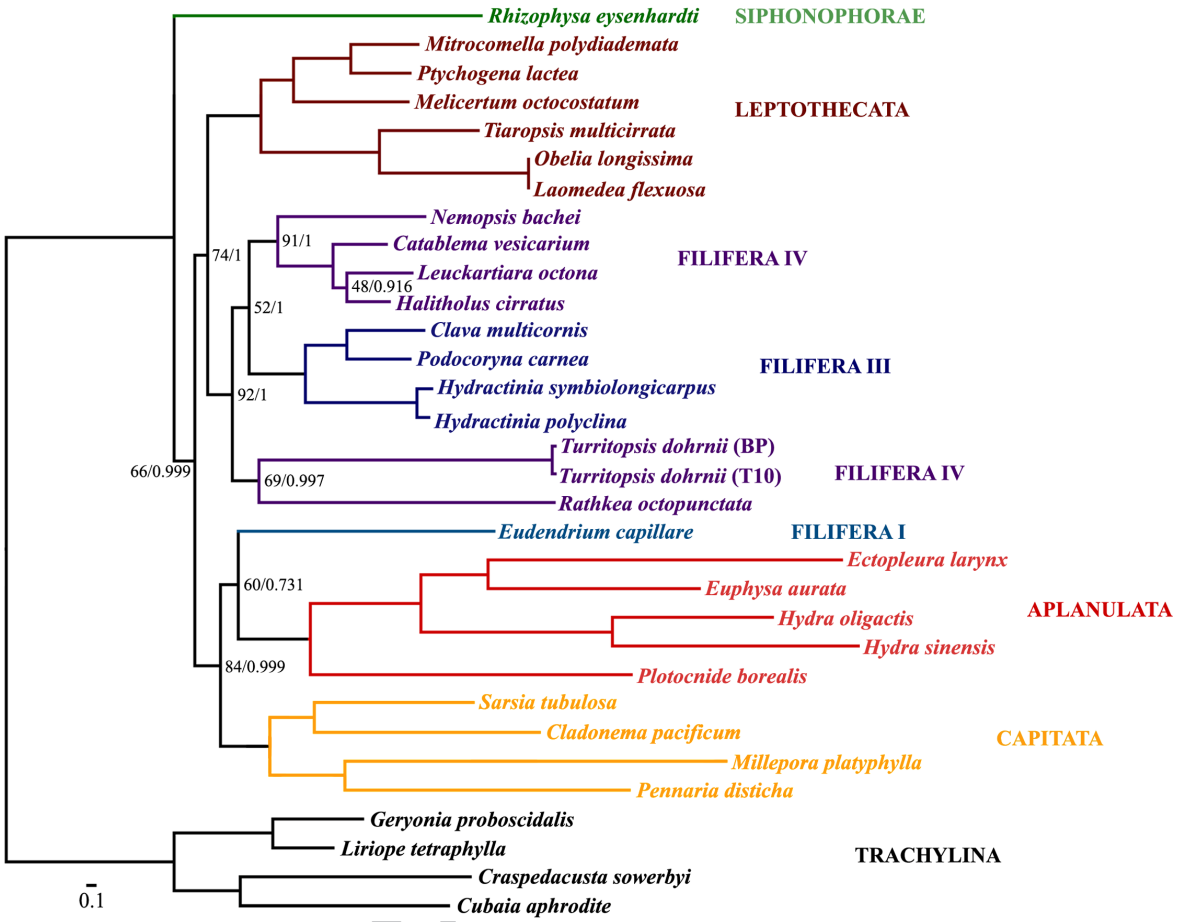
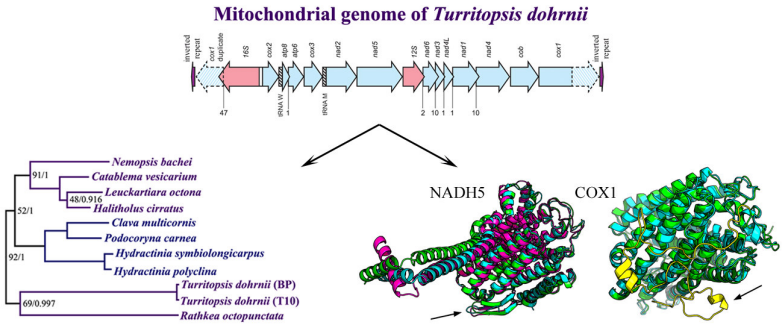


Fig. 2



Graphical abstract



Highlights

- We sequenced complete mitochondrial genome of *Turritopsis dohrnii*.
- Duplicated *cox1* gene and polymorphic regions were found in the linear mt chromosome.
- *T. dohrnii* has unique changes in the tertiary structures of proton-pumping proteins.
- Intra- and interspecific diversity was compared among various cnidarian species.
- Filifera IV was recovered as a polyphyletic group using mixed phylogenetic markers.