UNIVERSITY OF NORDLAND

MASTER THESIS

Molecular characterization of the deep-water sea anemone *Protanthea simplex* reveals true novel features in mitogenome organization

Dubin Arseny Vadimovich

BI309F MSc in Marine Ecology
Faculty of Biosciences and Aquaculture
May 2015

Index

Acknowledgments

The present study is the final part of two-year Master program at the Faculty of Biological Science and Aquaculture, University of Nordland, Norway.

I would like to thank the Norwegian Research Council for giving me an opportunity to take part in such an amazing project. The running cost to the experimental part of the sea anemone mitogenome sequencing was covered by RCN FriPro: Cold-Water Coral Genomics (Steinar D. Johansen, UiT).

I owe my deepest gratitude to my supervisor Professor Steinar Johansen who supported me throughout the entire way. His insights were invaluable when writing my thesis. Without him present thesis have not been possible. He was forgiving and merciful.

I would like to thank my co-supervisors Professor Truls Moum and Dr. Åse Emblem for their assistance till the end. I owe much of what I know about lab technique to Åse. Her enthusiasm and cheerfulness made my mistakes a little bit less painful.

I'm specially grateful to Erik Bergseth, Alica Kravtcova, Vladislav Vlaga for their company and support.

I would also like to note the help of Tor Erik Jørgensen with a genome assembly.

Abstract.
Resent advances in Next-Generation sequencing technology have completely changed our understanding of a natural world. Phylogenetic studies involving complete mitochondrial and nuclear genomes by using massive parallel sequencing have revealed previously unknown interaction between species and populations.
Next Generation sequencing in cooperation with bioinformatics could be a powerful combination, as the massive data generated using these methods are not possible to be analyzed by hand.
In this study compete mitogenome from cold-water sea anemone *Protanthea simplex* has been sequenced and annotated.
Results revealed unique mitogenome organization, completely different from what have been seen before.

1.0 Introduction.

Cnidarians represent very ancient and diverse taxa with a 500 million years of evolutionary history. According to the recent literature there are over 9000 cnidarian species that are grouped in 5 classes: the Anthozoa, the Hydrozoa, the Cubozoa, the Scyphozoa, and the Staurozoa (Osigus et al., 2013; Steele et al., 2011). Their life styles and strategies vary between the classes, but they all have several common features: lack of bilateral symmetry, sack-like body, single oral opening, only two tissue layers, simple nervous system, and the presence of specialized stinging cells - cnidocytes (Moen et al., 2004).

Cnidarians have received much attention in recent years due to their role in the ecosystem, ancient origins, and due to the recent discoveries of several bioactive compounds that can be used for treating serious diseases like AIDS or cancer (Fedorov et al., 2010; Miller and Ball, 2008; Miller et al., 2005; Plaisance et al., 2011; Sarma et al., 2009; Sunagawa et al., 2010). Most of these studies were possible only due to advances in the sequencing technologies.

1.1 Class Anthozoa.

Perhaps, one of the most notable members of the phylum are anthozoans. They occupy a wide variety of marine habitats, from shallow waters to the ocean depths. Many of them are considered a keystone, habitat-building species. Reef building corals can be used as example to show anthozoan role in the ecosystem. Tropical reefs estimated to harbour from quarter to one third of all marine species (Plaisance et al., 2011; Weis et al., 2008)

Class Anthozoa consists of two subclasses: subclass Hexacorallia and subclass Octocorallia.

The basis for division is the number of tentacles per individual animal, hence the names of the subclasses. Octocorals have eight tentacles, and hexacorals have multiple tentacles grouped in a cluster of 6.

Hexacorallia subclass divided into 6 orders: order Actinaria (sea anemones), order Zoanthidea (colonial anemones), order Ceriantharia (tube anemones), order Scleractinia (reef-building corals), order Corallimorpharia (mushroom corals), and order Antipatharia (black coralls) (Berntson et al., 1999).

Sea anemones play important role in the energy transfer between water column and benthic community. Sea anemones are also known for forming symbiotic relationships with other marine organisms, like hermit crabs, or clown fish (Daly et al., 2008).

Figure 1. *Protanthea simplex* on *Lophelia pertusa* reef. Photo by Åse Emblem.

P. simplex is a cold water sea anemone. Commonly found in boreal areas, P. simplex recently was located on the west coast of Scotland. The usual habitat for P. simplex are deep water (50 to 500 meters depth) Lophelia reefs. But it can also be found at much shallower waters, around 15 meters, attached to sea squirts or polychaete tubes.
It is a relatively small sea anemone, up to 20 mm high (not including tentacles) with 10 mm pedal disc. Body is salmon colored and has 24 distinct vertical furrows. Number of tentacles varies and can be up to 200. (Moen et.al, 2004).

Figure 2. Urticina eques. Photo by Steinar Johansen.

U. eques is a very common cold water sea anemone and can be found all along north-east Atlantic coasts. Pedal disc serves as a powerful sucker that attaches the animal to the solid surface. Can be seen from up to the low water line to 400 meters deep, but prefers deeper waters. Body is usually larger in diameter that in size, and is either white or pale orange in color with red-orange color spots. Tentacles are long and can be up to 300 mm. (Moen et.al, 2004)

1.2 Mitochondrial genome of the hexacorals and its applications

Like most of the organisms, hexacorals have have circular mitogenome organisation. Typical hexacoral mitogenome is 16-18 kb in length and include 13 or more protein coding genes, two ribosomal RNAs (small and large subunits), as well as two transfer RNAs (fMet and Trp) (Johansen et al., 2010).

Notable feature for all hexacorals, is that their mitogenomes contain group I introns that interrupt reading frames of essential protein coding genes. Cold water sea anemone M. senile, was the first animal described to have group I introns in mitochondrial DNA in 1996 by Beagly et al.

Intron that interrupts ND5 gene reading frame is found in every hexacoral sequenced to date. In contrast, intron that interrupts COI gene is considered optional in hexacorals, and found only in several species (Johansen et al., 2010).

Among sea anemones genome organisation is conserved. Despite of 3' heterogeneities in size of 5 protein coding genes, overall gene sequences appear to be conserved. The exception from this is *Nematostella sp*., which in phylogenetic studies positioned as the most basal of sea anemones (Emblem et al., 2014).

Mitochondrial DNA proven to be a useful marker in the variety of studies(De Paepe, 2012; Galtier et al., 2009; Sanders et al., 2014).

mtDNA gained popularity as a marker because it is easy to work with, it has reasonably conserved arrangements of genes, high mutation rate, low to none recombination, and mostly maternal inheritance. It is also a relatively cheap method to gather initial information about the new species(Galtier et al., 2009).

Development of Next Generation Sequencing (NGS) allowed researchers to do more comprehensive studies due to its relatively low cost, high coverage, high and consensus accuracy (de Magalhães et al., 2010). In 2010 there were 27 complete mitogenomes from five of the six known hexacoral orders available publicly in the NCBI database (Johansen et al., 2010). By now there are 63 complete hexacoral mitogenomes available and this number will only continue to grow.

Overall aim of this study is to sequence and characterize mitogenome of cold water sea anemone *Protanthea simplex* using various molecular biology tools NGS technology. The specific objective of this study are: to compare P. simplex mitogenome to the rest of the sea anemones.

2.0 Materials and methods

2.1 Sampling effort

Protanthea simplex was collected from *Lophelia pertusa* reef outside Nord-Leksa, Norway (63º36'N; 9º24'E) at 150-200 m depth, with a help of the research vessel «Gunnerus» and the Remotely Operated Vehicle «Minerva» (NTNU). Epidermal tissue samples from body wall and tentacles were immersed in absolute ethanol, frozen, and stored at minus 20 ºC with the intention to be used for further DNA isolation. In order to prevent RNA degradation, tissue samples collected for RNA isolation were placed in RNAlater® RNA Stabilization Solution (Life Technologies™), and kept at minus 20 ºC until they were shipped to laboratory facility at the University of Nordland.

Urticina eques was collected by scuba-diving at Mørkvedbukta Research station (67°16'N; 14°33'E), Bodø, Norway at 3-5 m depth. Animal specimens were placed in a plastic container, submerged in seawater and transported to laboratory facility at University of Nordland where they were immediately processed. Animals were alive until RNA extraction. Total RNA isolation was immediately performed from epidermal tissue samples. Remaining parts were preserved in ethanol and stored at minus 20 ºC.

2.2 DNA isolation

Tissue samples (approximately 2-10 mg) were thawed and air-dried for 5 minutes to evaporate remains of ethanol. Samples were then mechanically lysed in 2 ml MagNa Lyser Green Beads tube (Roche™) with 300 µl lysis buffer on a Precellys 24 homogenizer (Bertin technologies™) at 10,000 rpm until they were completely dissolved. Lysate was transferred to a 1.5 ml microcentrifuge tube that contained 2 µl proteinase K (50 µg/µl) and then incubated on a heat block at 55 °C for approximately 60 minutes with occasionally manual mixing of the tubes. Purification of total DNA was performed according to a standard phenol/chloroform protocol with ethanol precipitation. The DNA was diluted in 20 µl of water and stored in a freezer at minus 20 ºC.

After isolation of total DNA, one of the samples was split in two. Genomic DNA Clean & Concentrator™ kit (Zymo research™) was applied to one half of the sample in order to

eliminate possible polysaccharide contamination. The remaining half was left untouched in the event that purification procedures would affect DNA concentration in the sample.

Quantification and quality assessment of total DNA was performed on a Qubit® 2.0 Fluorometer (Invitrogen™) using Qubit® dsDNA BR Assay Kit (Invitrogen™), the Agilent 2200 TapeStation System (Agilent Technologies™) with Genomic DNA Screen tape and NanoDrop® ND-1000 (Thermo Fisher Scientific™) according to manufacturer's specifications.

Blank sample with ultrapure water was used as a negative control in all measurements. Qubit® 2.0 Fluorometer provides a fast and accurate measurement of the nucleic acid concentration, and the Agilent 2200 TapeStation System gives a visual picture of the nucleic acid size distribution in the sample. The NanoDrop® ND-1000 (Thermo Fisher Scientific™) gives easy and fast measurements, indicates the purity of the sample, but is not very accurate.

2.3 RNA isolation

Tissue samples were placed in a MagNa Lyser Green Beads tubes (Roche), suspended in 500 µl of TRIzol and lysed mechanically on a Precellys 24 homogenizer (Bertin technologies™) at 10,000 rpm. Total RNA precipitation and purification was done according to the standard TRIzol protocol (Chomczynski and Sacchi, 1987). All reactions were performed on ice and under the fume hood. In order to remove RNases from working surfaces RNaseZap® Solution (Life Technologies™) was used.

Quantitative and quality assessment of total RNA was performed on a Qubit® 2.0 Fluorometer (Invitrogen™) using Qubit® RNA HS Assay Kit (InvitrogenTM) and Agilent 2200 TapeStation System (Agilent Technologies™) with High Sensitivity RNA ScreenTape and Reagents and NanoDrop® ND-1000 (Thermo Fisher Scientific™) according to manufacturer's specifications.

In order to remove most of the ribosomal RNA, Low Input RiboMinus™ Eukaryote System v2 kit (AmbionTM) was applied to *P. simplex* total RNA sample. MicroPoly(A)Purist™ Kit (Ambion™) was used with *U. eques* RNA to remove all but poly-adenylated RNA.

2.4 Primer walking approach

In order to isolate and amplify the mitochondrial genome of *P. simplex* from total DNA, a primer walking approach was used. Primers that were used in this method were designed based on the published *U. eques* mitochondrial DNA (Emblem et al. 2014).

The core strategy is to use different combinations of primers in order to PCR amplify overlapping mitochondrial DNA fragments of various length. Primer combinations were applied such that the whole mitochondrial genome was covered by overlapping fragments. PCR amplifications were confirmed by gel electrophoresis.

2.5 PCR and gel electrophoresis

The TaKaRa PCR Amplification Kit with TaKaRa LaTaq (Takara Bio Company™) polymerase was used for amplification of mitochondrial DNA. All reactions were performed on ice.

The thermocycler was programmed for following conditions:

- Initial denaturation – 94 ºC (5 min.); 25 cycles:
- Denaturation – 94 ºC (30 sec.)
- Annealing – 55 ºC (30 sec.)
- Elongation – 72 ºC (4 min.)
- Final elongation - 72 ºC (4 min.); Hold at 4 ºC.

Annealing temperatures were determined according to the primer melt temperatures (Tm). To confirm the amplification, PCR products (5 µl product and 1 µl Bromophenol Blue loading buffer of each lane, with 1 lane per reaction) were run on a ~1% standard agarose gel with SYBR® Safe (Invitrogen™) nucleic acid stain to visualize dsDNA, and with 1 Kb Plus DNA Ladder (Invitrogen™) to determine size of amplicons. Gel Logic 200 Imaging system (Kodak™) and Safe Imager™ (Life Technologies™) was used for visualization.

2.6 Total DNA and RNA sequencing using Ion PGM™ System (Life Technologies™)

All library preparation and template preparation steps were performed according to the standard protocols. In short, total DNA library was fragmented chemically with Ion Shear™ Plus Reagents (Life Technologies™). Reaction volumes were scaled for 50–100 ng DNA input and were performed according to the protocol.

The whole transcriptome library was constructed by reverse transcription of fragmented total RNA with a reverse transcriptase mix (10X SuperScript® III Enzyme Mix).

When working with RNA, all safety procedures were considered. Working surfaces were treated with RNaseZap® Solution (Life Technologies™) and all reactions were performed on ice.

As recommended, the E-Gel® electrophoresis system with E-Gel® SizeSelect™ Agarose Gels (2%) (Life Technologies™) was used for size selection of DNA and RNA libraries.

Quality controls for both DNA and RNA libraries were performed on the Agilent 2200 TapeStation System (Agilent Technologies™) with High Sensitivity RNA and DNA ScreenTapes using appropriate reagents.

Prior to template preparation procedure, precise library concentrations were measured with quantitative real-time PCR (qPCR) using Ion Library Quantitation Kit and The StepOnePlus Real-Time PCR System (Life Technologies™). Template preparation was completed on Ion OneTouch™ 2 System with Ion PGM™ Template OT2 400 Kit (Life Technologies™).

Quality control of unenriched and enriched templates was performed on Qubit® 2.0 Fluorometer with Ion Sphere™ Quality Control Kit (Life Technologies™). Final preparation and sequencing were done by the use of Ion PGM™ Sequencing 400 Kit and according to the protocol. 316 v.2 chip was loaded using Simplified Ion PGM™ Chip Loading protocol with the Ion PGM™ Weighted Chip Bucket (Life Technologies™).

2.7 Bioinformatics and data analyses

Most of the data analyses were performed using the CLC Genomics Workbench (Qiagen™). These include: *De novo* assembly of *P. simplex* mitogenome, gene annotation, multiple sequence (both nucleic acid and protein) alignments, gene mapping, and transcriptome analyses of *P. simplex* and *U. eques*.

Basic Local Alignment Search Tool (BLAST) was also used for mitogenome annotation. For comparative purposes, mitogenomes of six sea anemones (Emblem et al. 2014) were acquired from National Center for Biotechnology Information (NCBI) genome database.

For assembly and annotation verification, MITObim script and MITOS web tool were used. MITObim (MITOchondrial Baiting and Iterative Mapping) is a Perl script that uses the MIRA assembler to reconstruct mitochondrial genomes of non-model organisms without the need for a reference genome.

Reconstruction of *P. simplex* mitogenome was completed using direct reconstruction without prior mapping with the "--quick" option and reconstruction mitogenomes from mitochondrial barcode seeds. As a seed for building mitogenome *P. simplex* COIII gene acquired from NCBI genomic database was used. MITOS is a web tool designed to help with de novo Metazoan mitochondrial genome annotation. MEGA v6 software was used to make concatenated gene sequences and build phylogenetic trees.

2.8 Amplicon sequencing and molecular cloning

To confirm the preliminary mitogenome assembly, to obtain better resolutions on sequenced regions, and to investigate problematic genome regions, re-sequencing attempt of the whole mitogenome was made. ND5 gene region, because of its unusual features, was cloned and send for Sanger sequencing to UiT/UNN. Primers that were used in this stage were designed based on Ion PGM sequencing results. Amplification of the mitochondrial DNA was done according to PCR conditions described previously.

Both the mitochondrial genome and the ND5 region were amplified with PCR and visualized on a ~1% standard agarose gel. To separate primers from amplicons, the entire sample was run on the same ~1% standard agarose gel and visualized on a Safe Imager™ (Life Technologies™). Bands that contained amplicons were cut out and collected in a Eppendorf LoBind® tubes (Eppendorf™).

To isolate amplified DNA from an agarose gel, the QIAEX II Gel Extraction Kit (Qiagen™) was used. All reactions were performed according to a manufacturer's agarose gel extraction protocol.

Molecular cloning of ND5 gene region was performed using TOPO® TA Cloning® Kit with One Shot® TOP10 Competent Cells (Life Technologies™). All reactions were performed according to the manufacturer's protocol.

To find the most optimal conditions for transformation, reaction pool was split in two groups: 2 µl of PCR product was added to group one, 4 µl of PCR product was added to group

two. For blue/white colony screening, LB agar plates were covered with 40 µl of IPTG (100mM) and 40 µl X-Gal were then germinated with three different concentrations of transformation culture for each group (2 µl and 4 µl groups): 50 µl, 100 µl and 150 µl. For each concentration separate plate was used. Kanamycin (50 mg/ml) was used as an antibiotic agent for both groups. For negative control plates, ampicillin was used as an antibiotic agent.

Selected single white colonies from each plate were transferred to 2 ml of LB medium with kanamycin using a pipet tip. Mixture was incubated overnight in a Multitron Standard Incubation Shaker (Infors HT™) at 150 rpm in 37 ºC. Overnight cultures were transferred to Eppendorf LoBind® tubes. To purify plasmids from a bacterial culture, PureLink® Quick Plasmid Miniprep Kit (Invitrogen™) was used. All procedures were performed according to the manufacturer's protocol.

Purified plasmid DNA was prepared for Sanger sequencing. Reaction mixture was prepared: 2 µl of purified plasmid DNA, 1 µl of BigDye 5 x Sequencing Buffer (Applied Biosystems™), 1 µl of BigDye 3.1 enzyme and nucleotide mix (Applied Biosystems™), 3 µl of primer, 3 µl of nuclease-free water. Reaction tubes were placed in a thermocycler. Cycler was programed for a following conditions: initial denaturation – 96ºC for 5 minutes; 25 cycles: 96 ºC for 10 seconds, 50 ºC for 5 seconds and 60 ºC for 4 minutes; hold at 4 ºC. Afterwards, 10 µl of RNase-free water was added to each sample, so the total reaction volume was equal to 20 µl. Samples were sent to the Sanger sequencing facility at UiT/UNN. Final preparation of the sequences and the actual sequencing was performed by the laboratory technician responsible for the sequencer platform. Sequencing was performed on the 3130xl Genetic Analyzer (Applied Biosystems™).

2.9 Confirming presence of COII gene in a total DNA sample

To test if COII gene is present in a mitochondrial genome of *P. simplex* (which may support a hypothesis that COII is localized at a mitochondrial mini-chromosome) an attempt was made to use specific PCR primers, amplify target sequence and visualize products on a gel. Six COII gene sequences from sea anemones were aligned in CLC Genomic Workbench (Qiagen™) to search for conserved regions. Four primers were then designed with intention not only to amplify not only whole COII gene but also mini chromosome that COII might be localized. Amplification was then confirmed on a ~1 % standard agarose gel.

PCR reactions and gel preparations were done according to the conditions described previously.

3.0 Results

3.1 Mitochondrial DNA and RNA isolation and sequencing results

Prior to template and library preparation, quality and quantity of total DNA and RNA were measured. Measurements showed 63 ng/µl of DNA in the sample. Possibly due to small amounts of RNA in the samples, the Agilent 2200 TapeStation (Agilent Technologies™) and Qubit® 2.0 Fluorometer (Invitrogen™) gave inconclusive measurements. Therefore only approximate concentrations were measured on NanoDrop® ND-1000 (Thermo Fisher Scientific™).

Attempts to isolate mitochondrial genome of *P. simplex* using the primer walking approach were apparently not successful. It seems that none of the primers were able to successfully bind to *P. simplex* mitochondrial DNA. Further, Genomic DNA Clean & Concentrator™ kit (Zymo research™) was applied to purify DNA as polysaccharide contaminants may affect the primer binding. Cleanup procedure resulted in significant loss of DNA in the sample and did not affect primer binding. Decision was made to proceed with sequencing of total DNA and RNA, and to mine out mitochondrial genome and transcriptome sequences from the data pool using bioinformatics approaches.

Total DNA was sequenced using 316 v2 chip. According to a final report:

- ISP loading was 82 %
- Total amount of reads was 3,114,365 bp and 60 % out of them were usable
- 33 % polyclonality was observed
- Mean read length size - was 349 bp

Total RNA was sequenced using 316 v2 chip. According to a final report:

- ISP loading for *P. simplex* run was 57 %, for *U. eques* 81 %
- Total amount of reads for *P. simplex* run was 982,890 and 31% out of them were usable. For *U. eques* total reads count was 3,963,997 with 77% usable reads
- Mean read length for *P. simplex* was 60 bp, for *U. eques* - 276 bp

- *U. eques* ISPs were 14 % polyclonal, *P. simplex* - 37 %

CLC genomic workbench *De novo* assembly tool was used in order to isolate mitochondrial reads from total read pool. The program looks for overlapping read pairs and attempts to create the longest scaffold possible. Out of 61,461 contigs, top 5 in length and in coverage were considered.

To find an approximate location and order of the genes BLAST search against NCBI genomic database was performed. The search revealed high match with mitochondrial genes of sea anemones.

In order to verify assembly, several methods were used: by hand assembly using COIII and ND5 genes as starting points, MITObim script with direct reconstruction without prior mapping using the "--quick" option and reconstruction from mitochondrial barcode seeds with COIII (KJ482993) as a barcode seed. Although genomes constructed with MITObim were similar to CLC assembled genome in size, gene content and organization, gene redundancies and errors in scaffolds showed that MITObim was ineffective at reconstructing *P. simplex* mitogenome.

3.2 Genome assembly verification.

Assembly revealed several interesting features that needed verification:

1. Unusual size of ND5-717 intron
2. Apparent lack of COII gene in the Ion PGM data pool

To get better resolution for ND5, primers covering this genome region were designed. PCR amplified products would then be sent for sequencing.

Despite 144x coverage of the genome, and high number of usable reads gathered from Ion PGM run, mapping of total reads pool on COII genes from other sea anemones gave no result. Primers based on the assembled genome were designed in order to verify length of the mitogenome, and confirm the absence of COII in the mitogenome. The goal was to PCR amplify overlapping fragments of 6-7kb length, covering the whole mitogenome, confirm amplification on an agarose gel and perform re-sequencing. Deviations from 6-7 kb length would be visible on a gel and could serve as an indication that genome assembly may be wrong.

Depending on the length of the fragments, it would have been possible to deduce if COII was present in the mitogenome. However, there was no deviation from intended length of the fragments when visualized on a gel (Figure 3).

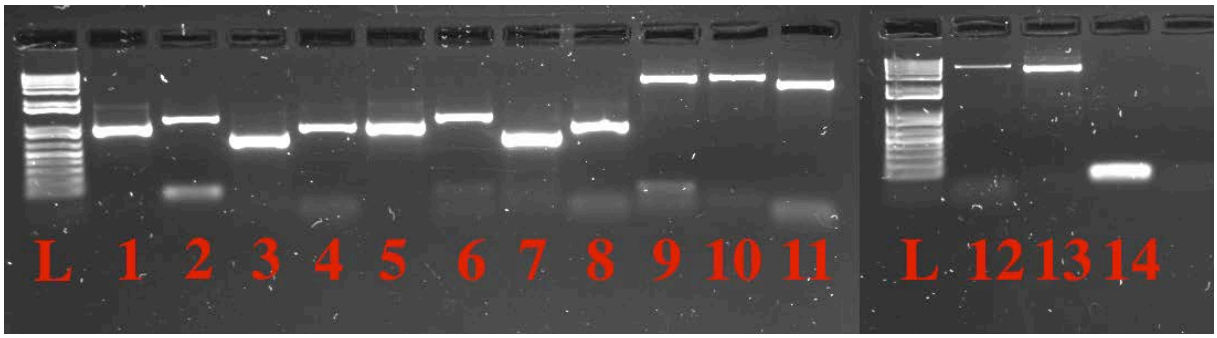


Figure 3. Agarose gel visualization. L, - 1kb+ ladder; lanes 1-8, amplified ND5 gene region; lanes 9-13, amplified genome fragments; lane 14, primer dimer

Fragments from 1 to 8, covering ND5 gene region, were selected. DNA from these fragments was isolated for subsequent molecular cloning experiments and Sanger sequencing (Figure 3).

As illustrated in Figure 1, genome fragments presented in lanes 9 to 13 were around 6-7 kb in length, which supports the assembly and possess the question about apparent lack of COII gene in the mitogenome.

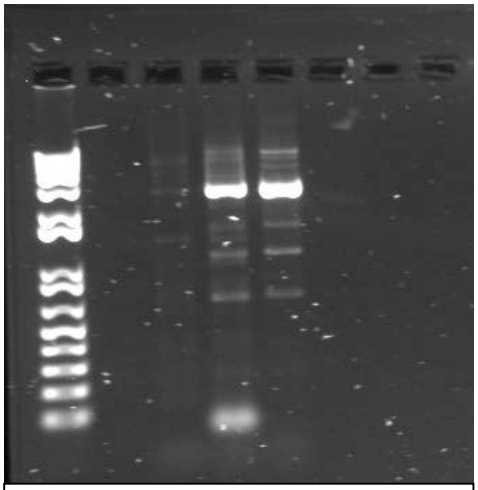


Figure 4. Agarose gel visualization.

L, – 1 kb+ ladder; lanes 1-2, failed primers; lanes 3-4, fragments that possibly contain COII (these have not yet been sequenced).

To exclude the possibility that library preparation has introduced bias in the data pool, and to detect the COII gene, primers based on aligned COII sequences from available sea anemone genomes were designed. Figure 4 presents distinct amplified fragments of approximately 5kb length in lanes 3-4, which might be the COII gene incorporated into a mini-chromosome. This issue will be addressed further in the Results section and in the Discussion. Due to time limitations of this thesis it has not been possible to provide a sequencing results that give a definite answer to these questions.

3.3 Protanthea simplex mitochondrial genome.

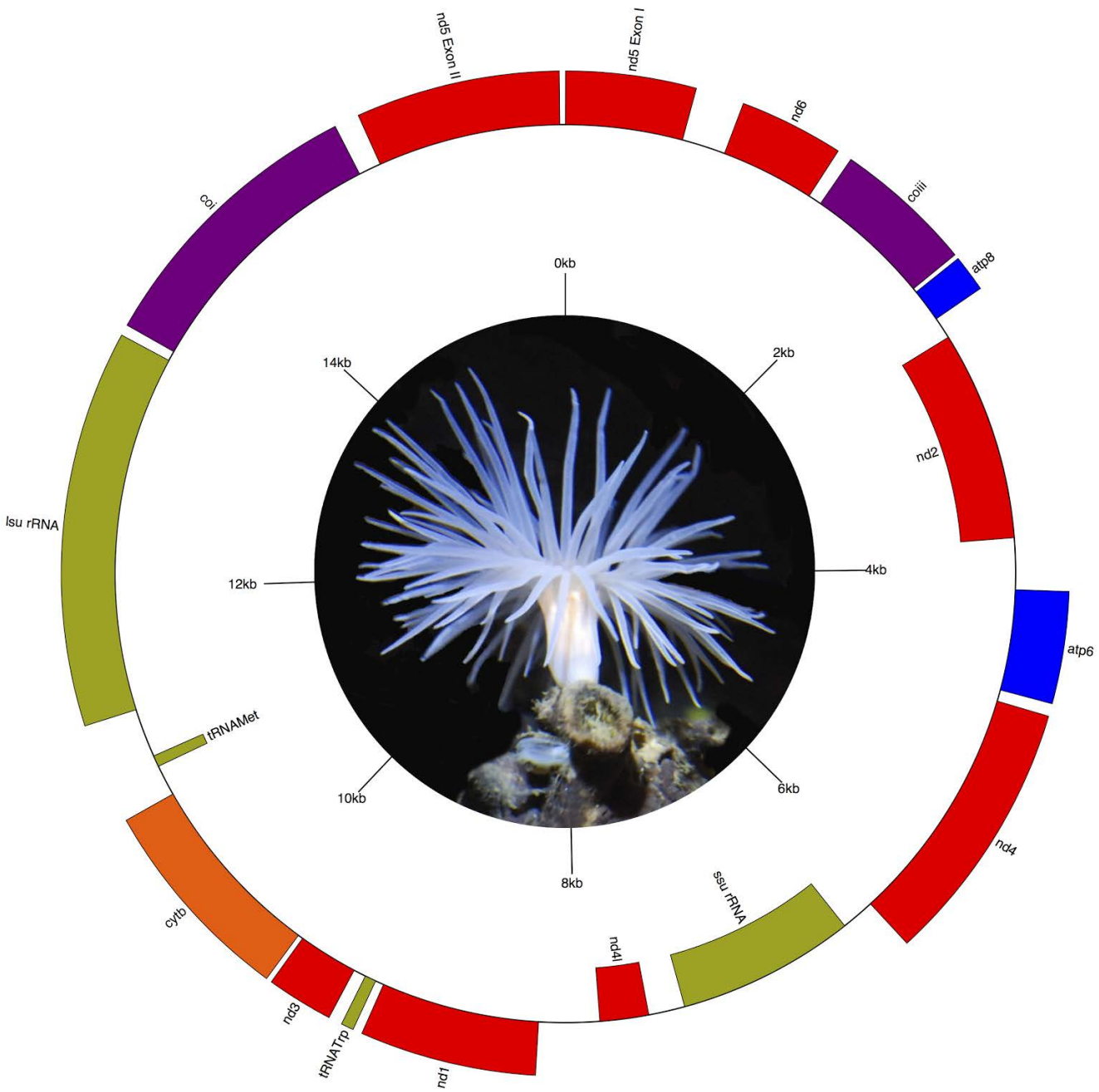


Figure 5. *Protanthea simplex* mitochondrial genome. Circular view. Red - Complex I genes, orange - Complex III genes, purple - Complex IV genes, blue - Complex V genes, green - RNA genes. Genes that are located on a different DNA strands are placed on outer or inner circle. P. simplex photo (Åse Emblem).

P. simplex mitogenome has circular organization and is approximately 17,123 bp in length. The mitogenome contained all mitochondrial genes common for sea anemones with the exception of COII: 12 protein coding genes: 7 ND genes, 3 CO genes, 2 ATPase genes, 2 genes coding for small (SSU) and large (LSU) ribosomal RNA subunits, and 2 tRNA genes (tRNA-Trp and tRNA-fMet). The average coverage was 141.47 times. Functional genes constitute approximately 85 % of the mitogenome and IGRs 15 %. The *P. simplex* mitogenome was compared to mitogenomes of other sea anemones that have been sequenced previously. All genomes were acquired from NCBI genomic database (Table 1).

Gene order in sea anemones.						
Protanthea simplex	*Urticina eques*	*Metridium senile*	*Hormathia digitata*	*Aiptasia pulchella*	*Bolocera tuediae*	*Nematostella vectensis*
ND5 Exon I	ND5 Exon I	ND5 Exon I	ND5 Exon I	ND5 Exon I	ND5 Exon I	ND5 Exon I
ND6	ND1	ND1	ND1	ND1	ND1	ND1
COIII	ND3	ND3	ND3	ND3	ND3	ND3
Atp8	ND5 Exon II	ND5 Exon II	ND5 Exon II	ND5 Exon II	ND5 Exon II	ND5 Exon II
ND2 (negative str.)	tRNA-Trp	tRNA-Trp	tRNA-Trp	tRNA-Trp	tRNA-Trp	tRNA-Trp
Atp6	ND2	ND2	ND2	ND2	ND2	ND2
ND4	SSU rRNA	SSU rRNA	SSU rRNA	SSU rRNA	SSU rRNA	SSU rRNA
SSU rRNA (negative str.)	COII	COII	COII	COII	COII	COII
ND4L (negative str.)	ND4	ND4	ND4	ND4	ND4	ND4
ND1	ND6	ND6	ND6	ND6	ND6	ND6
tRNA-Trp	CytB	CytB	CytB	CytB	CytB	CytB
ND3	tRNA-Met	tRNA-Met	tRNA-Met	tRNA-Met	tRNA-Met	tRNA-Met
CytB	LSU rRNA	LSU rRNA	LSU rRNA	LSU rRNA	LSU rRNA	LSU rRNA
tRNA-Met (negative str.)	COIII	COIII	COIII	COIII	COIII	COIII
LSU rRNA	COI Exon I	COI Exon I	COI Exon I	COI Exon I	COI Exon I	COI
COI	HEG	HEG	HEG	HEG	HEG	
	COI Exon II	COI Exon II	COI Exon II	COI Exon II	COI Exon II	
	ND4L	ND4L	ND4L	ND4L	ND4L	ND4L
ND5 Exon II	Atp8	Atp8	Atp8	Atp8	Atp8	Atp8
	Atp6	Atp6	Atp6	Atp6	Atp6	Atp6

Table 1. Summary of gene order features among completely sequenced sea anemones. The ND5-717 group I intron is obligatory to all sea anemones. The ND5 exons I and II are marked in red. The COI-884 group I intron is optional among sea anemones. The COI exons I and II are marked in purple. References of published sea anemones mitogenomes are: Medina et al. 2006; Emblem et al. 20014.

In comparison to the six other completely sequenced sea anemones mitogenomes, *P. simplex* has several notable features:

- The gene order is significantly different
- Most genes were of different length when compared to the other sea anemones
- The ND5-717 intron that has been found in all sea anemones to this moment, has an unusual large size and harbors all the mitochondrial genes (except ND5 and COII) between its 5' and 3' splice sites
- The lack of COI-884 intron that is present in most of the sea anemones, except the *Nematostella sp.*

- *P.simplex* mitogenome does not contain a homing endonuclease gene (HEG) like most of the sea anemones, and thus appears similar to the mitogenome of *Nematostella sp.*
- Despite the high coverage and numbers of reads acquired, the COII gene was apparently missing from mitogenome assembly and from total DNA reads pool
- All other hexacoral mitogenomes sequenced to date encode all the mitochondrial genes from the same strand. The genes coding for SSU rRNA, ND2, ND4L and tRNA-fMet were all located on the opposite strands

The following parts of the Results section will focus on the *P. simplex* mitogenome in more detail. Here, genes have been divided into 5 groups according to their mitochondrial function:

Protein coding genes:

- Complex I – 7ND genes
- Complex III – CytB gene
- Complex IV – COI and COIII genes
- Complex V – Atp6 and Atp8 genes

RNA genes

- RNA genes – SSU and LSU ribosomal RNA genes, and two tRNA genes

The *P. simplex* mitochondrial genome and corresponding proteome will be compared to the other sea anemones. Most of the genes show high heterogeneity both in length and in DNA sequence. Thus, in order to illustrate the difference between species, protein alignments will be used to high-light the mutations that affect protein sequences only. All protein and some nucleic acid alignments are presented in the appendix.

Complex I genes in sea anemones.														
Sea anemones	ND1		ND2		ND3		ND4		ND4L		ND5		ND6	
	Size (bp.)	Start/Stop codons	Size (bp.)	Start/Stop codons	Size (bp.)	Start/Stop codons	Size (bp.)	Start/Stop codons	Size (bp.)	Start/Stop codons	Size (bp.)	Start/Stop codons	Size (bp.)	Start/Stop codons
Protanthea simplex	972	ATG/TAA	X	X	369	ATG/TAA	X	X	300	GTG/TAA	1839	ATG/TAA	576	ATG/TAA
Urticina equcs	984	ATG/TAA	1164	ATG/TAG	357	ATG/TAA	1467	ATG/TAA	300	ATG/TAA	1833	ATG/TAG	609	ATG/TAA
Metridium senile	984	ATG/TAA	1158	ATG/TAG	357	ATG/TAG	1476	ATG/TAA	300	ATG/TAA	1833	ATG/TAA	609	ATG/TAA
Hormathia digitata	984	ATG/TAA	1377	ATG/TAA	357	ATG/TAG	1467	ATG/TAA	300	ATG/TAA	1833	ATG/TAA	609	ATG/TAA
Aiptasia pulchella	984	ATG/TAA	1158	ATG/TAA	357	ATG/TAA	1470	ATG/TAG	300	ATG/TAA	2172	ATG/TAA	609	ATG/TAA
Bolocera tuediae	984	ATG/TAA	1383	ATG/TAG	357	ATG/TAA	1476	ATG/TAA	300	ATG/TAA	1833	ATG/TAG	609	ATG/TAA
Nematostella vectensis	984	ATG/TAA	1110	ATG/TAA	357	ATG/TAA	1476	ATG/TAG	300	ATG/TAA	1818	ATG/TAA	600	ATG/TAA

Table 2. Summary of sizes and codon usage in Complex I genes among sea anemones.

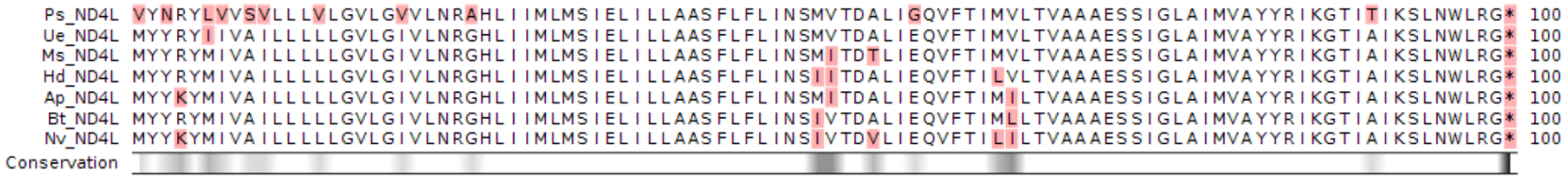


Figure 6. Protein alignment of the ND4L genes of 7 sea anemones. Amino acids deviating from the majority are marked in red.

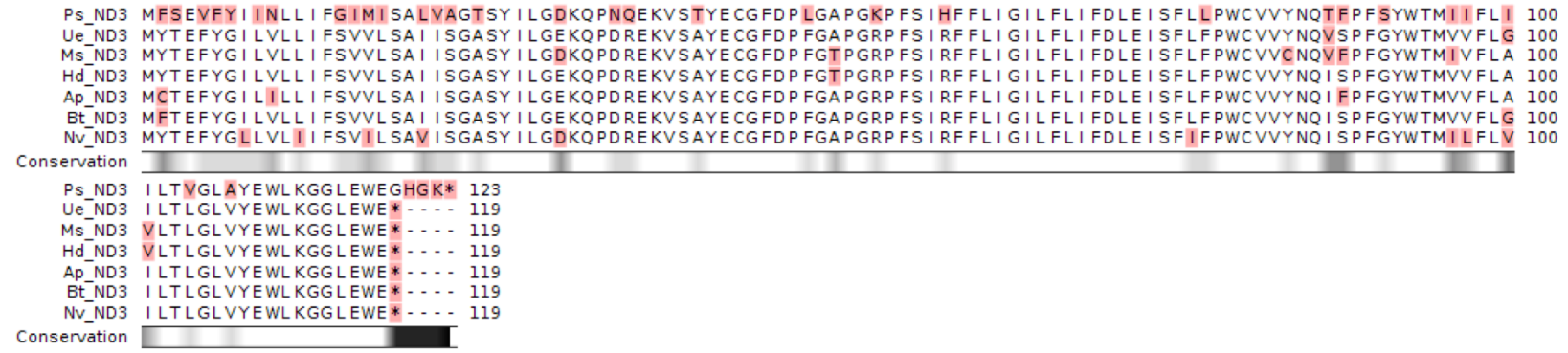


Figure 7. Protein alignment of the ND3 genes of 7 sea anemones. Amino acids deviating from the majority are marked in red.

3.4.1 DNA and amino acid conservation in P. simplex Complex I genes

In general, ND genes in P. simplex display high heterogeneity in size, in codon usage and protein sequence throughout reading frame. Several internal deleterious mutations, as well as deletions in the 5' and 3' ends appear to be the main cause of size variations among ND genes. Some of the genes are more conserved in size than others (Table 2). Differences in amino acid sequences for all genes can be studied in the appendix.

The longest gene is ND5 - 1839 bp (612 AA) and is also the most variable - 58 changes in the amino acid sequence. Two deletions and seven insertions resulted in the size deviations at the protein level in *P. simplex.* The ND5 protein sequence in *P. simplex* is longer than most sea anemones by 3 amino acids (Figure 8).

The shortest gene is ND4L - 300 bp (99 AA) and it is the most conserved one - 11 changes in amino acid sequence and no size variation (Figure 6).

Overall the ND4L, ND3, and ND1 appear to be the most conserved genes in size and in protein sequence among Complex I group.

To exclude the possibility of reading frame-shifts being the cause for 3' end size variation in protein coding genes, protein sequences were studied. As an example the ND3 size variation is illustrated in Figure 7. ND3 is longer by 4 amino acids in *P.simplex.* Frame-shift would have resulted in dramatic changes in sequence, which cannot be observed in the protein alignment.

3.4.2 Codon usage in ND genes

All ND genes except ND4L in *P. simplex* start with an ATG codon. ND4L is the only gene that starts with a GTG codon. However there are variations in stop codon usage. As can be seen from Table 2, ND2, ND3, ND4, and ND5 have variations in stop codons between species. Stop codons vary between TAA and TAG.

X in Table 2 indicates that it was not possible to identify reading frames for those genes.

Another unusual feature found in *P. simplex* mitogenome is that some of the genes, including ND2 and ND4L, are located on a different strand compared to the other ND genes. This has not been observed in sea anemones before.

3.4.3 ND5 gene

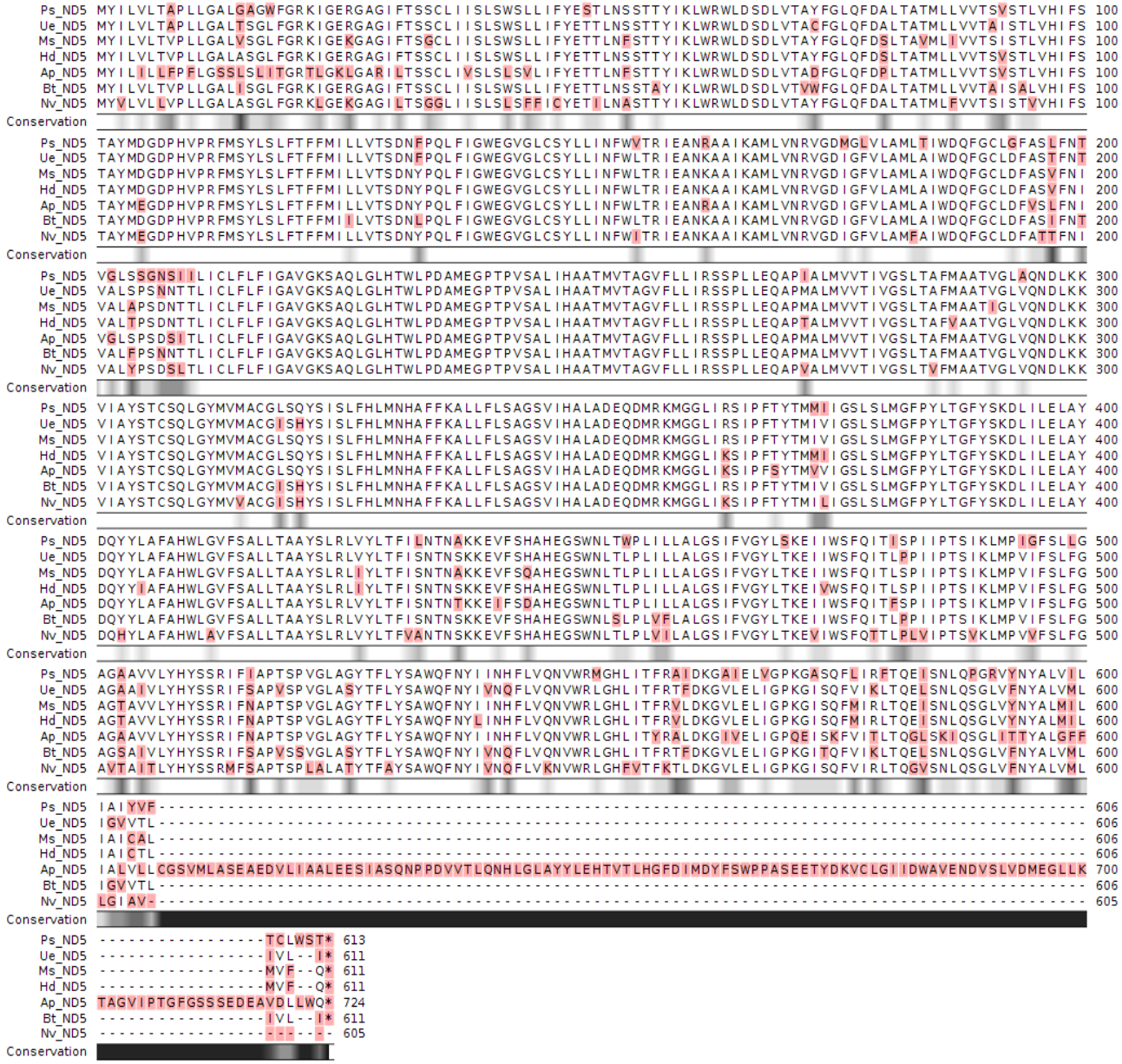


Figure 8. Protein alignment of the ND5 genes of 7 sea anemones. Amino acids deviating from majority are marked in red.

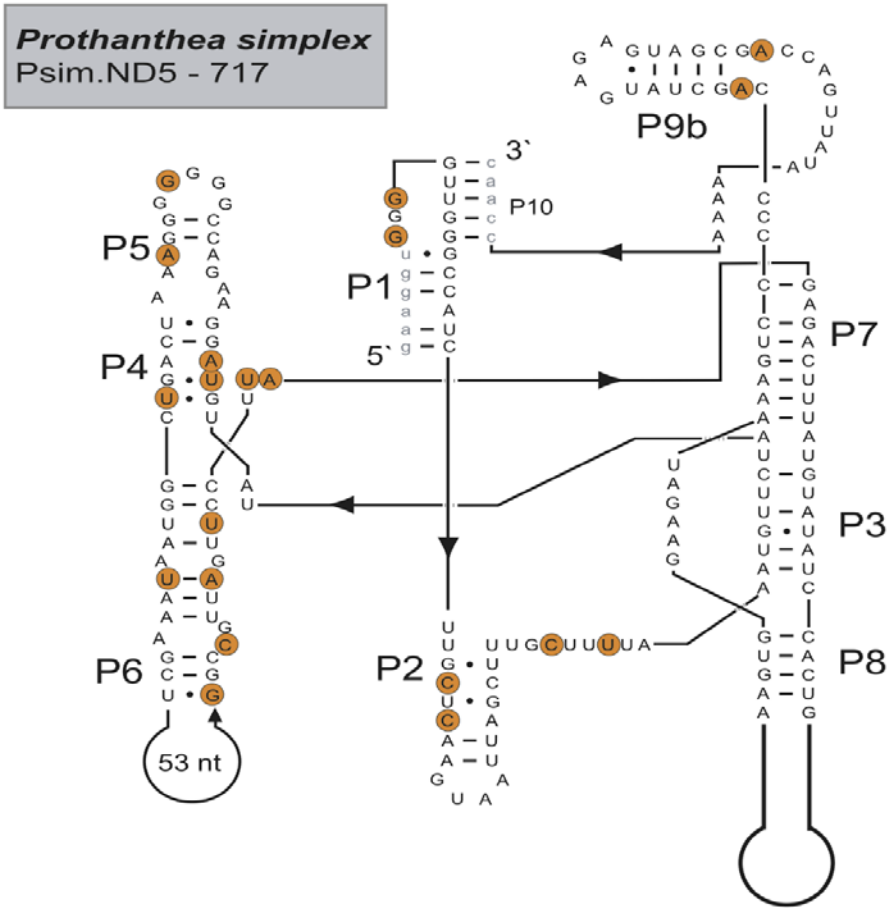
As mentioned previously, the ND5 is the longest and the most variable gene within the Complex I group. ND5 possesses high variability all across the reading frame in all sea anemones. 227 changes in *P. simplex* nucleotide sequence, resulted in the 58 amino acid substitutions. There is a notable variation at the C-terminal end of the protein with *A. pulchella* being 113 AA longer than most of the sea anemones, and *Nematostella sp.* 6 AA shorter. *P. simplex* is only 2 AA longer at the C-terminal end (Figure 8).

3.4.4 ND5-717 group I catalytic intron

Most splicing in eukaryotic cells occurs with the aid of the spliceosome complex. However, some of the introns are able to splice without the enzymatic aid by the spliceosome and they are considered to be self-splicing. Group I introns belong to this category. They have little conservation at the primary sequence level, but their secondary and tertiary structures appear highly conserved (Nielsen and Johansen, 2009).

All members of the Hexacorallia subclass studied to date have a group I intron in the mitochondrial ND5 gene. Furthermore, this ND5 intron harbors various mitochondrial genes, but the genes incorporated in the intron vary between orders (Emblem et al., 2014). The ND5-717 group I intron in *P. simplex* has the same location (nucleotide position 717 in ND5) as in other sea anemones, but it is considerably larger in size. The intron was found to be 15,251 bp, carrying all other mitochondrial genes. For comparison, the studied sea anemones have introns of approximately 1600 - 2000 bp, harboring only the ND1 and ND3 genes (Emblem et al., 2014).

A



B

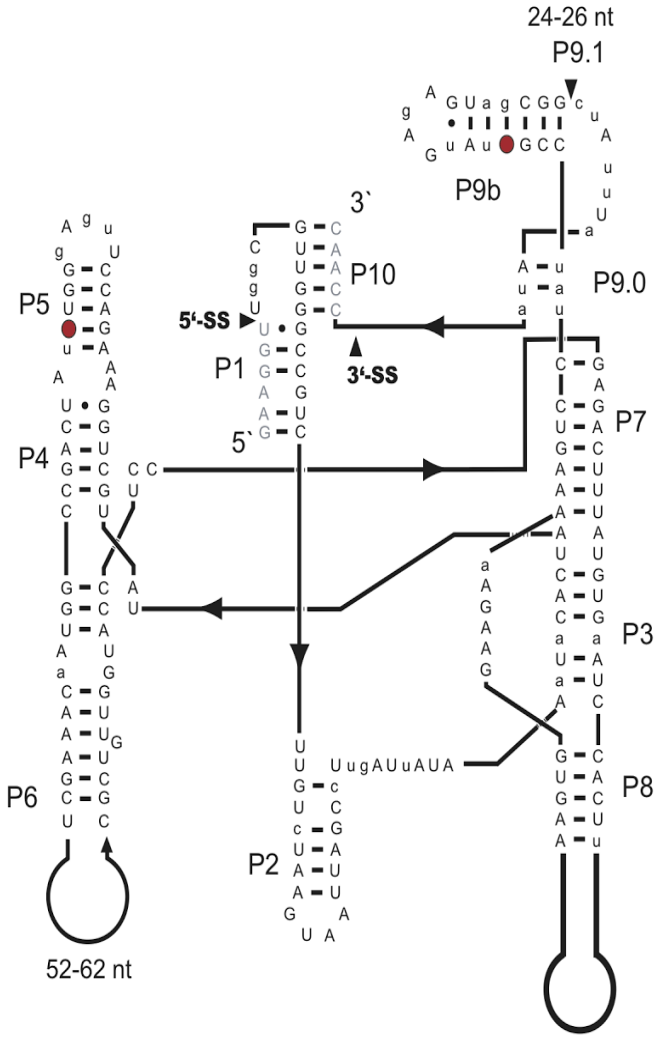


Figure 9.

A. Secondary RNA structure of P. simplex ND5-717 intron. P1 - P10, functional paired RNA segments. Orange dots indicate differences compared to conserved positions in the intron sequence of other sea anemones.

B. Consensus secondary structure of ND5-717 intron. Structure is based on intron sequences from: U. eques, M. senile, H. digitata, A. pulchella, B. tuediae, Nematostella sp. N, highly conserved positions; n, moderately conserved; red dot, varying positions. Figure 8B is from Emblem et al., 2014.

As illustrated in Figure 9, the ND5-717 intron in *P. simplex* contains nucleotide changes at the positions that are 100% conserved in other sea anemones. There are no changes in the conserved positions of P7, P3 and P8 of the intron catalytic domain. P7 stem is a binding site for free guanosine (guanosine co-factor) so conserved secondary structure in this region is expected.

3.5 Complex III genes

Complex III genes in sea anemones		
Sea anemones	CytB	
	Size (bp.)	Start/Stop codons
Protanthea simplex	1155	ATG/TAA
Urticina eques	1185	ATG/TAG
Metridium senile	1146	ATG/TAA
Hormathia digitata	1182	ATG/TAA
Aiptasia pulchella	1191	ATG/TAG
Bolocera tuediae	1152	ATG/TAG
Nematostella vectensis	1179	ATT/TAA

Table 3. Summary of sizes and codon usage in Complex III genes among 7 sea anemones.

The CytB gene in sea anemones shows considerable variations in size, nucleotide, amino acid sequence and in stop codon usage. (Table 3, Figure 10). As can be seen from Figure 10, most of the size variations are due to changes at the 5' and 3' ends of the sequence (N and C-terminal ends, respectively, in proteins). At the nucleotide level, when compared to other sea anemones, *P. simplex* has 128 substitutions, 28 deletions and 9 insertions. These changes are reflected at the protein level. *P. simplex* has 42 substitutions and 9 deletions in amino acid sequence. All deletions are located at the N-terminal end of the CytB protein sequence.

P. simplex CytB gene is one of the shortest among sea anemones - 1155 bp (384 AA) with only *M. senile* and *B. tuediae* being shorter. Start codons are mostly ATG with *Nematostella sp*. being an exception and use ATT as a start codon. Stop codons vary between TAA and TAG, with TAA being a majority.

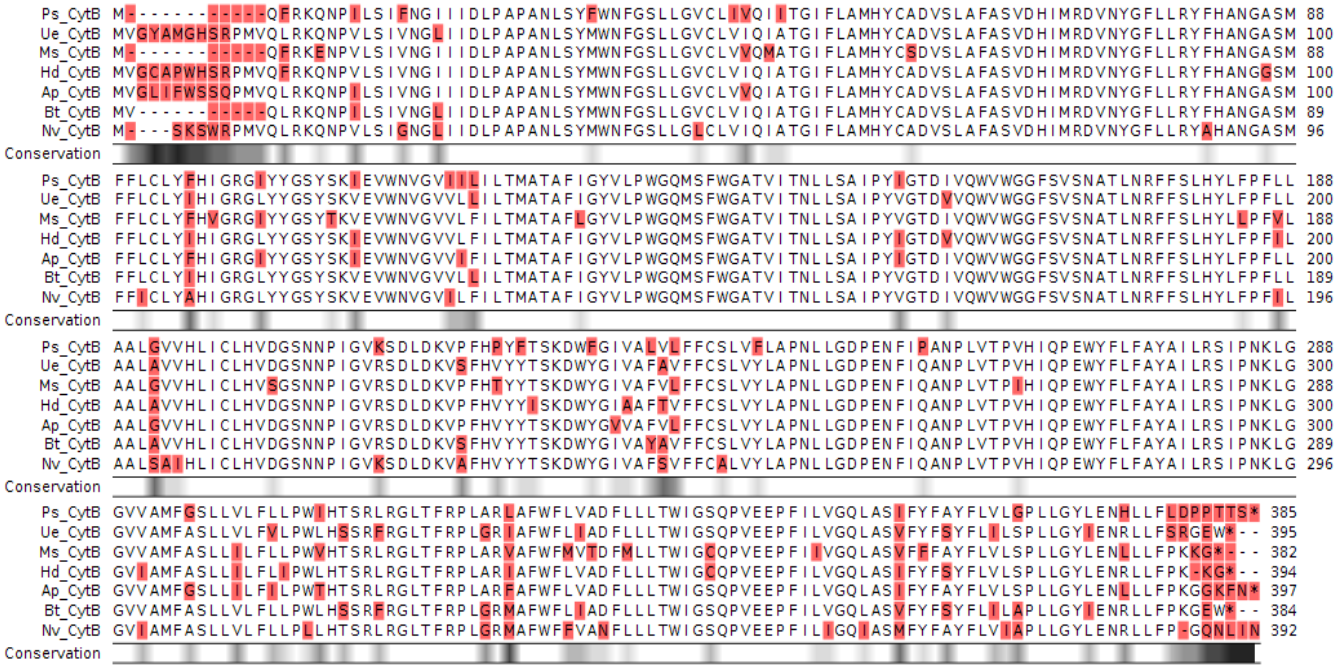


Figure 10. Protein alignment of CytB genes of 7 sea anemones. Amino acids deviating from the majority are marked in red.

3.6 Complex IV genes

Complex IV genes in sea anemones						
Sea anemones	COI		COII		COIII	
	Size (bp.)	Start/Stop codons	Size (bp.)	Start/Stop codons	Size (bp.)	Start/Stop codons
Protanthea simplex	1581	ATG/TAG	X	X	768	ATG/TAA
Urticina eques	1572	ATG/TAG	741	ATG/TAA	789	ATG/TAG
Metridium senile	1593	ATG/TAA	747	ATG/TAA	789	ATG/TAG
Hormathia digitata	1593	ATG/TAA	741	ATG/TAA	789	ATG/TAG
Aiptasia pulchella	1626	ATG/TAG	747	ATG/TAA	789	ATG/TAG
Bolocera tuediae	1572	ATG/TAA	747	ATG/TAA	789	ATG/TAA
Nematostella vectensis	1587	ATG/TAA	744	ATG/TAA	789	ATG/TAG

Table 4. Summary of sizes and codon usage in CO genes among 7 sea anemones.

Cytochrome oxidase (CO) subunits possess the least variations among protein coding genes. They show considerably low variations in amino acid sequence, length and start/stop codon usage.

3.6.1 COIII gene

At the DNA level the *P. simplex* COIII gene has 102 substitutions, 21 deletions when compared to the majority of the sea anemones. Deviations in the protein sequence (Figure 11) are mainly located at the internal sites, with 25 substitutions and 7 deletions (one at the N-terminal end of the protein and 6 at the internal sequence). There are small changes in the protein sequence throughout the reading frame. Among the sea anemones, *Nematostella sp.* shows highest heterogeneity throughout the reading frame.

Figure 11. Protein alignment of COIII in 7 sea anemones. Amino acids deviating from the majority are marked in red.

3.6.2 COI gene

All sea anemones, except *Nematostella sp.* and now *P. simplex*, that have been sequenced, have a group I intron with an internal HEG that interrupts the mitochondrial COI reading frame (Table 1). For comparative purposes, HEGs were removed from the protein alignment, and COI exon I and II were combined into a continuous reading frame.

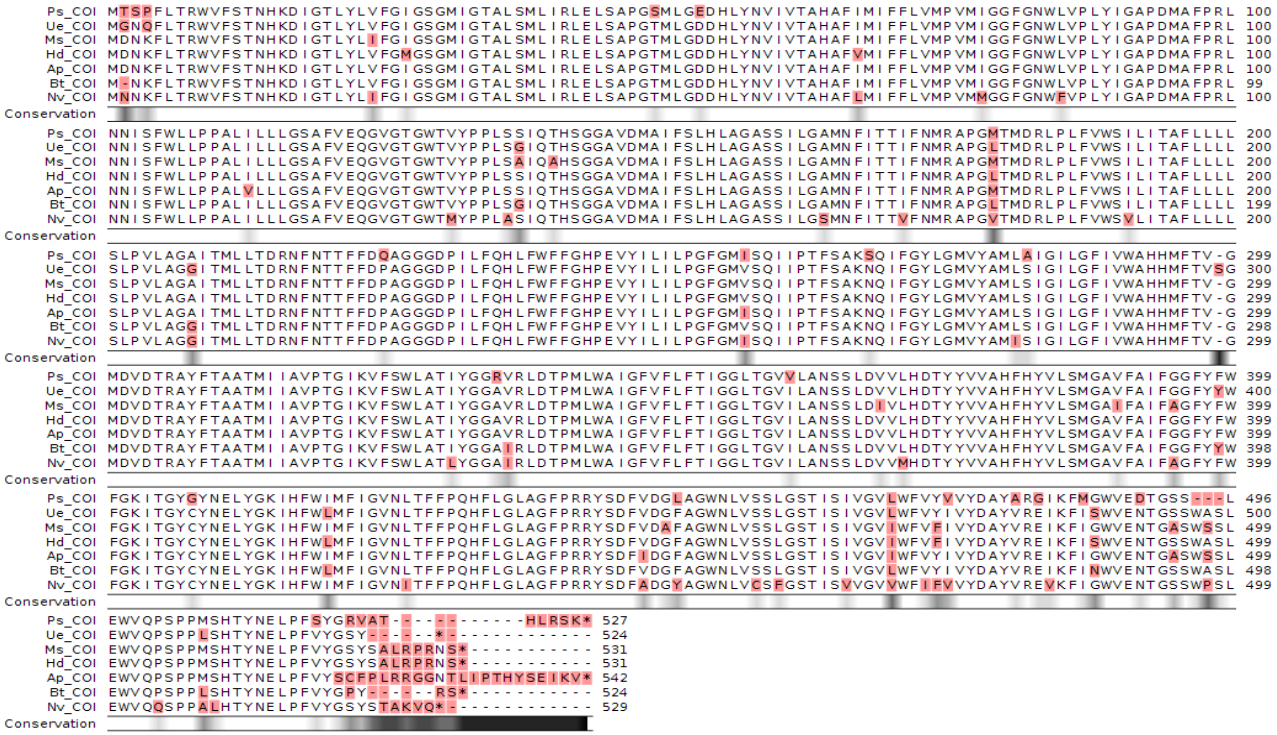


Figure 12. Protein alignment of COI in 7 sea anemones. Amino acids deviating from the majority are marked in red.

The COI amino acid sequence (Figure 12) is highly conserved in comparison to other protein coding genes. 194 nucleotide changes resulted in 33 substitutions in the protein sequence. Size variation is caused mainly by changes in a 3’ end, corresponding to a 3 AA internal gap in *P. simplex*.

3.6.3 COII gene

As mentioned previously, an attempt was made to identify the *P. simplex* COII gene with PCR using primers designed for conserved regions within COII gene (see primer key features in the appendix). As a result fragments around 5 kb were amplified. This correlates with a study made in lice (Shao et al. 2012; Jiang et al. 2013) that investigated mini-chromosomes in lice mitogenome. Researchers used a similar approach based on inverted PCR to characterize fragmented mitochondrial genomes and to identify circular mini-chromosomes, which was confirmed with an agarose gel run and subsequent sequencing.

Amplified fragments were from 3-5 kb in size, which correlates with our findings. The possibility of a mini-chromosome, it’s structure, and role in the mitochondria will be discussed in the next section.

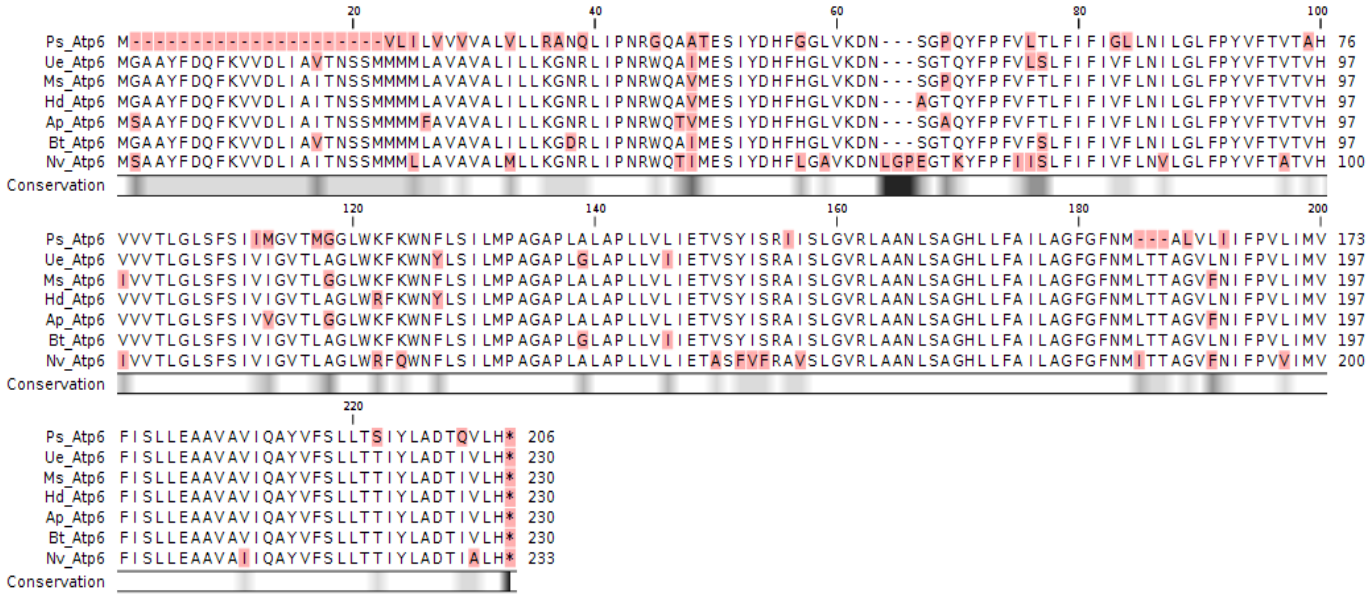
3.7 Complex V genes

Complex V genes in sea anemones				
Sea anemones	Atp6		Atp8	
	Size (bp.)	Star/Stop codons	Size (bp.)	Start/Stop codons
Protanthea simplex	618	ATG/TAA	204	ATG/TAA
Urticina eques	690	ATG/TAA	216	ATG/TAA
Metridium senile	690	ATG/TAA	216	ATG/TAA
Hormathia digitata	690	ATG/TAA	216	ATG/TAA
Aiptasia pulchella	690	ATG/TAA	216	ATG/TAA
Bolocera tuediae	690	ATG/TAA	216	ATG/TAA
Nematostella vectensis	699	ATG/TAA	231	ATG/TAG

Table 5. Summary of sizes and codon usage in ATPase genes among 7 sea anemones.

ATPase genes show moderate variation in protein sequence and in size (Table 5). *P. simplex* and *Nematostella sp.* are the most variable species in this genome region among sea anemones. For other sea anemones, the sizes are relatively conserved and amino acid sequences possess little variation.

Figure 13. Protein alignment of Atp6 in 7 sea anemones. Amino acids deviating from the majority are marked in red.



Size variations of Atp6 protein sequences in *P. simplex* and *Nematostella sp.* are caused by internal changes and changes in 5' end region (N-terminal end of protein). 85 substitutions, 72 deletions - 63 nucleotide gap at the 5' end and 9 nucleotide internal gap closer to the 3' end. As a result *P. simplex* has 27 AA substitutions and lacks 24 AA in it's sequence and *Nematostella sp.* has 27 AA substitutions and 3 additional amino acids. Other than size variation in 2 species, protein sequence is relatively conserved (Figure 13).

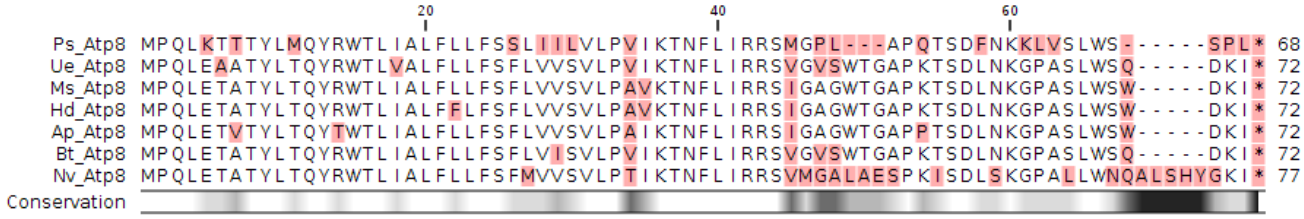


Figure 14. Protein alignment of Atp8 in 7 sea anemones. Amino acids deviating from the majority are marked in red.

As in Atp6, size variations in Atp8 are only found in *P. simplex* and *Nematostella sp.* These variations are caused by internal changes and changes in 3' region (C-terminal end of protein). While *P. simplex* lacks 4 AA, Nematostella sp. has 5 AA in addition. Like Atp6, the protein sequence of Atp8 is moderately conserved among sea anemones. *P. simplex* has 19 substitutions and 4 deletions in its amino acid sequence (Figure 14).

3.8 RNA genes

RNA genes in sea anemones				
Sea anemones	SSU rRNA Size (bp.)	LSU rRNA Size (bp.)	tRNA Trp Size (bp.)	tRNA Met Size (bp.)
Protanthea simplex	1093	2188	70	70
Urticina eques	1057	2214	70	71
Metridium senile	1082	2188	70	71
Hormathia digitata	1082	2189	70	71
Aiptasia pulchella	1066	2178	70	71
Bolocera tuediae	1082	2209	70	71
Nematostella vectensis	693	602	70	71

Table 6. Size variation in RNA genes among sea anemones.

Size of the rRNA genes was defined by, first aligning ribosomal subunits genes and second, by using transcriptomic data to define more precise borders of rRNA genes.

As illustrated in Table 6, sizes of ribosomal RNA subunits (SSU and LSU rRNAs) varies between species but within reasonable range with *Nematostella* sp. deviating from common picture (strongly suggesting that the *Nematostella* data need to be re-analysed).

P. simplex SSU rRNA sequence possess 247 changes: 192 substitutions, 19 deletions, and 36 insertions. LSU rRNA has 464 changes compared to majority of the sea anemones: 341 substitutions, 64 deletions, and 59 insertions.

tRNA genes are conserved (10 substitutions in tRNA-fMet and 8 substitutions in tRNA-Trp) among sea anemones with exception of *P. simplex* lacking in nucleotide in tRNA-fMet (Figure 15, Figure 16). Secondary structure is also conserved (Figure17)

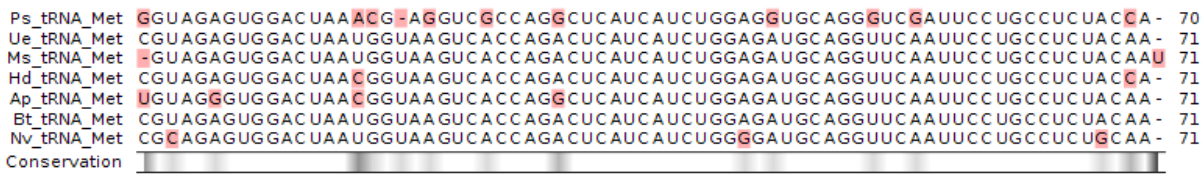


Figure 15. Alignment of tRNA-fMet of 7 sea anemones. Nucleotides deviating from the majority are marked in red.

Figure 16. Alignment of tRNA-Trp of 7 sea anemones. Nucleotides deviating from the majority are marked in red.

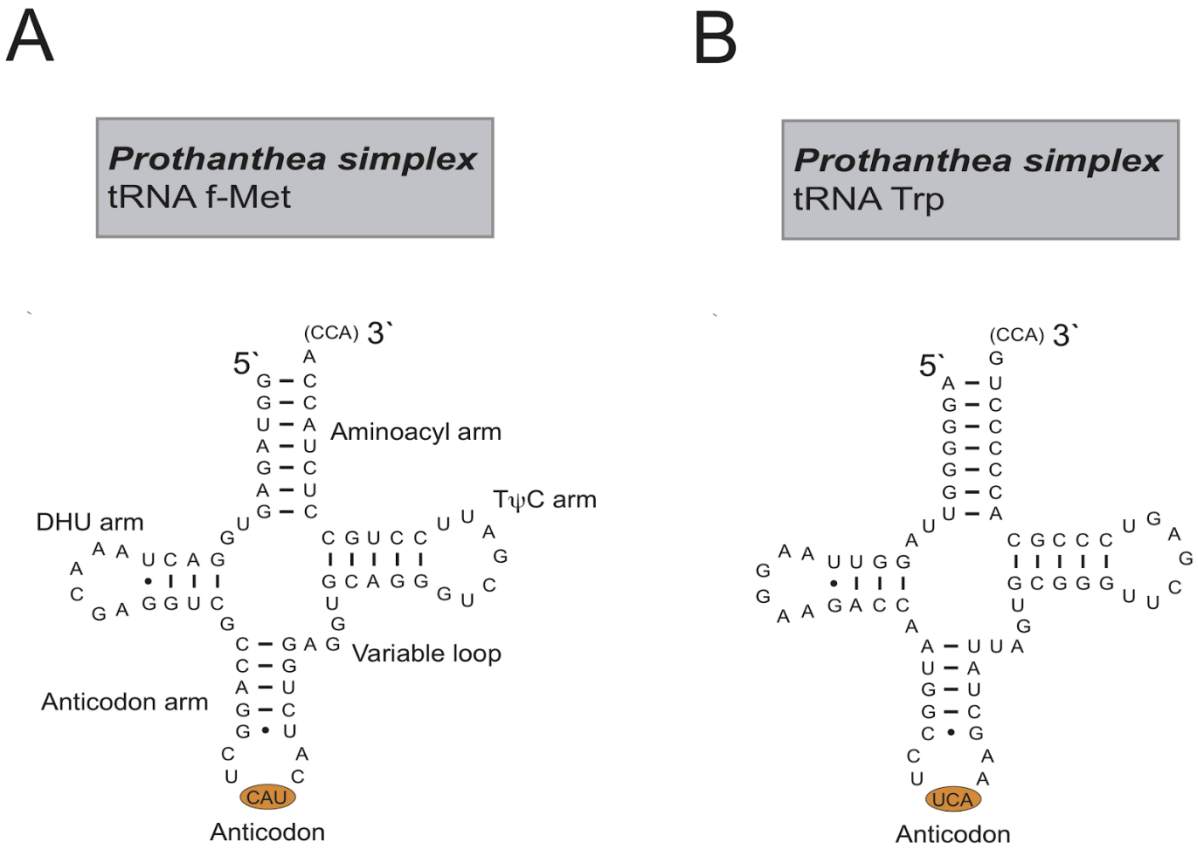
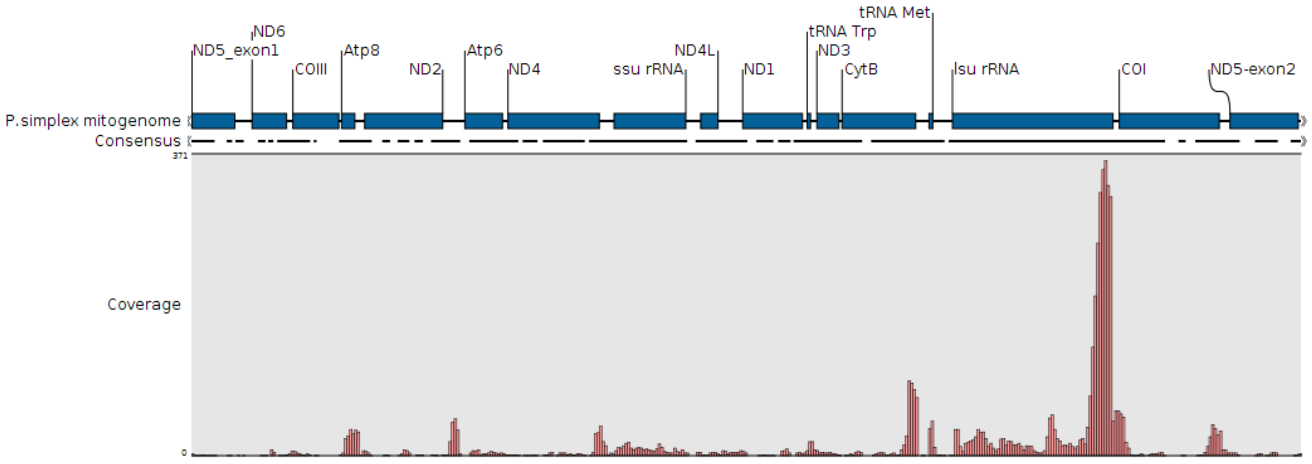


Figure 17. tRNA Secondary structure.

3.9 Total RNA sequencing data

Percentage of mapped reads for *P. simplex* mitogenome was 0.16 %, which indicates that most of the sequenced RNA was genomic. When mapped, total RNA reads pool showed highest coverage for LSU rRNA and CytB gene regions. In general, almost whole mitogenome was covered. Even though coverage for the most of the regions was low, sequencing results indicate that mitigenome was active.

It is important to mention that mapping total RNA seq. pool on COII genes from six sea anemones gave no result (Figure 18).



Figur 18. Mapping of total RNA reads on P. simplex mitogenome. Coverage of certain regions is illustrated as a bar plot with highest peak matching highest covered region. Genes are marked as blue boxes on the sequence.

Mapping *U. eques* total RNA reads pool on the mitogenome (acquired from NCBI), showed that only 0.06 % of all reads were mitochondrial. Compared to *P. simplex*, *U. eques* mitotranscriptome shows considerably lower coverage for most of the genome regions. LSU rRNA region, like in *P. simplex*, has the highest coverage, and SSU rRNA has the second highest (Figure 19).

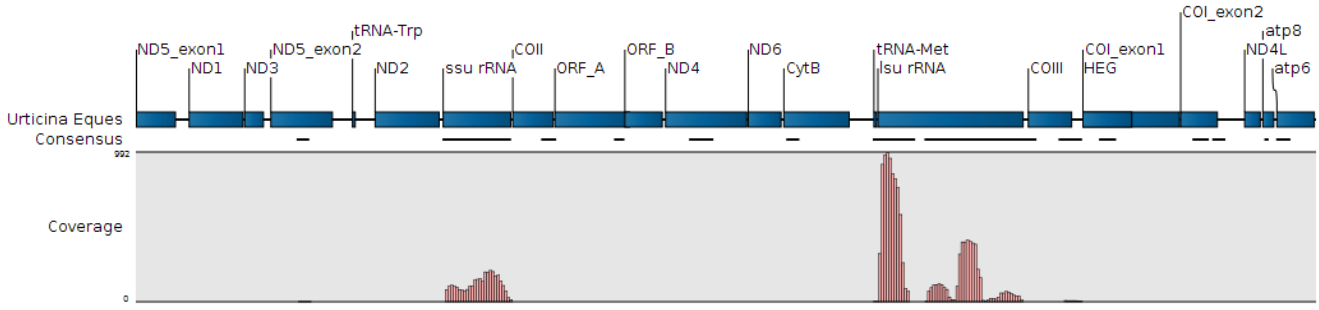


Figure 19. Mapping of total RNA reads on U. eques mitogenome. Coverage of certain regions is illustrated as a bar plot with highest peak matching highest covered region. Genes are marked as blue boxes on the sequence.

4.0 Discussion

Past studies have reported that mitochondrial genomes of the subclass Hexacorallia show low substitution rates in the mitochondrial protein coding genes despite an apparent high variation in the nuclear genome (Hellberg, 2006; Shearer et al., 2002). More recent studies, however, suggest that sea anemones mitogenomes might not be so inert as previously believed (Emblem et al., 2014).

In the present study we used NGS technology to obtain mitochondrial genome sequences and transcriptomic data of the cold-water sea anemone *Protanthea simplex,* as well as transcriptome of a second sea anemone species, *Urticina eques*.

Our study revealed that *P. simplex* deviates significantly from current "sea anemones" picture. The mitochondrial genome of *P. simplex* contains a novel genome organization and has a high heterogeneity in protein coding genes compared to the other studied sea anemones. These findings apparently support previous studies in hexacoral mitogenomics, suggesting that their mitogenomes are dynamic in structure and that group I introns could be involved in structural rearrangements (Emblem et al., 2012; Emblem et al., 2011; Emblem et al., 2014; Hedberg and Johansen, 2013; Lin et al., 2014).

Other notable findings include:

- Like *Nematostella sp*., *P. simplex* lacks the COI-884 group I intron.
- COII is missing from total DNA and total RNA data pools.
- Atypical size of ND5-717 group I intron, which is believed to be obligatory and essential for members of Hexacorallia subclass (Johansen et al., 2010). *P. simplex* ND5 intron includes all, except ND5 and COII, mitochondrial genes between it's 5' and 3' splice sites.
- Both strands are used for gene coding, which is unique among hexacorals.

The COI-884 group I intron is considered to be optional for hexacorals. In sea anemones, the COI intron harbors a gene coding for a homing endonucleases (HEGs) - a highly specific endonuclease that, affects intron mobility through the process of homing (Emblem et al., 2014; Haugen et al., 2005; Johansen et al., 2010). Goddard et al., 2006, suggested that slow substitution rates advance HEG invasion into Cnidarian mitogenomes. However, *P. simplex* lacks COI-884 group I intron and gene encoding for HE.

One explanation could be that the loss of a COI group I intron is a relatively recent evolutionary event in *P. simplex*.

It has been proposed that loss and gain of COI introns goes in cycles throughout evolutionary history (Goddard et al., 2006). Our findings corresponds to the model proposed by Emblem et al., that COI intron undergoes 5 stages from invasion into the host mitogenome, to complete loss of the COI-intron and HEG, through degradation (Emblem et al., 2014). *P. simplex* could be at the "loss stage" of this intron gain-and-loss cycle.

Despite multiple reports about stability and sequence conservation of Hexacorallia mitogenomes, *P. simplex* shows high heterogeneity in protein coding genes when compared to the previously sequenced sea anemone mitogenomes. A plausible interpretation might be that *P. simplex* represents a separate and distinct branch of hexacorals, and does not belong to the sea anemones clade. Sequencing errors as a reason for heterogeneity are, of course, a possibility. However, this explanation seems unlikely, when considering the high sequencing coverage. Furthermore sequencing errors would likely have resulted in multiple reading frame shifts, which cannot be observed when analyzing protein gene sequences.

The apparent lack of COII sequences in both, total DNA and the RNA data pools, raises a question about if this important and essential gene is present or not. Though it has not been possible to give a definite answer for this question due to the time limitations of this thesis, we propose a novel explanation for this phenomenon based on mitochondrial mini-chromosomes. This alternative will be discussed further below.

Finally it has not been possible to identify functional reading frames for ND2 and ND4 genes. Preliminary sequencing errors are likely explanations for unconfirmed reading frames, and re-sequencing of these regions have to be performed in order to eliminate errors.

4.1 Group I introns

Group I introns are self-catalytic genetic elements that interrupt functional genes. They have been observed in the nuclei of protists, in eukaryotic viruses and bacteriophages, in bacteria, in chloroplasts, and in mitochondria (Beagley et al., 1998; Haugen et al., 2005).

In sea anemones group I introns are found in mitochondria where they interrupt two essential protein coding gene sequences: the obligatory and vertically inherited ND5 intron, and the optional and mobile COI intron.

The ND5-717 (named according to the conserved ND5 gene position; Emblem et al., 2011) in sea anemones harbors two protein coding genes - ND1 and ND3 (Emblem 2014). *P. simplex*, in comparison, contains as many as 15 gene inside the ND5-717 intron. Similar intron sizes have been reported in Corallimorpharia (Medina et al., 2006) but our study reports the first example in sea anemones.

In all hexacoral mitogenomes sequenced to date, group I intron that interrupts the ND5 reading frame at position 717 position seems to be a highly conserved feature (Emblem et al, 2011). COI group I introns, on the other hand, have been found to be optional in hexacorals (Emblem et al., 2014; Goddard et al., 2006; Johansen et al., 2010). The genic position of COI introns vary among hexacorals, with at least three different insertion sites representing a specific variant of the ribozyme (Emblem et al., 2014; Goddard et al., 2006). However, all sea anemone COI introns are inserted after nucleotide position 884 in the COI gene (hence, the intron is named COI-884 according to – Emblem et al. 2011).

Group I introns have an ability to integrate themselves into the host genome and occasionally switch genic position. There are two mechanisms that assists integration and transposition of group I introns (Birgisdottir and Johansen, 2005b; Hedberg and Johansen, 2013):

1. Insertion into the homologous site of an intron-less allele (intron homing), initiated by a homing endonuclease that is encoded in some group I introns (like COI-884 group I intron).

2. Transposition into the new site of the same or different gene at the RNA level via mechanism called reverse splicing.

Reverse splicing depends on several complex processes: site specific recognition by internal guide sequence (IGS) at the RNA level and ribozyme-assisted insertion into specific site of the exon RNA, reverse transcription of the recombined RNA, and finally cDNA integration into the genome. It appears, that reverse splicing is only moderately site specific and that variations in the IGS may affect integration of the intron (Birgisdottir and Johansen, 2005a).

The role of ND5-717 introns in the mitogenome of hexacorals remain a mystery.

It has been shown that this intron has an origin within fungi mitogenomes and later became trapped inside the hexacoral mitogenome, probably due to inclusion of essential mitochondrial genes in its structure (Emblem et al., 2011). If ND5-717 has any biological role, or host benefitting function, remains to be seen.

4.2 The COII mitochondrial mini-chromosome hypothesis

We have made several attempts to assemble *P. simplex* mitochondrial genome. In all cases, genome scaffolds were lacking COII sequences. COII is the most conserved among the mitochondrial encoded genes in metazoans, and it is an important gene involved in the mitochondrial respiratory chain.

Our initial explanation for the absence of COII in the genome data was due to some kind of error during mitogenome assembly. An alternative explanation for missing COII sequences was due to a bias introduced by library preparation, and that COII have been unintentionally removed from the data pool.

As we previously presented in the Results section, two experiments using specific primers, PCR, and an agarose gel for visualization, were designed.

Gel visualization (Figure 3 and Figure 4) results supported the assembly, but also showed that total DNA probably contains COII gene sequences. This suggests that COII may have escaped during the library preparation process. How could an entire gene fall out of the genome on a stage of library preparation if the assembly was correct?

Several reports describing the mitochondrial genomes of blood-sucking mammalian lice, provided evidence that mitogenomes of these animals are fragmented into many functional mini-chromosomes (Jiang et al., 2013; Shao et al., 2012). Each mini-chromosome could contain from one, and up to eight mitochondrial genes, depending on the species. These mini-chromosomes were found to be 3-4 kb in size, organized as circular DNA and consisted of coding and noncoding regions. Noncoding regions contained tandem repeats of various size and copy number (12-75 bp repeats; 2-5 copies).

These studies provided new insight into what might have happened with COII gene in *P. simplex.*

Emblem et al., 2014, reported a “transposon-like” insertion element in the intergenic region (IGR) in some sea anemone species, including U. eques. In U. eques, this IGR islocated adjacent to COII gene and contained two open reading frames.

Sizes of the amplified fragments were around 5 kb in length, which is similar to what was found in the lice. Considering the length of the “transposon-like” element that was found in U. eques (2043 bp) and the size of COII gene in other sea anemones (around 740 bp) these 5 kb fragments may very well represent our hypothetical mini-chromosome.

Presence of a COII mini-chromosome could also explain how it escaped from the data pool. During the shearing stage, mini-chromosome may have been more affected by the enzyme.

Since shearing incubation time was calculated for large fragments of total DNA, considerably small mini-chromosome should be sheared into very DNA fragments.
Then, on the size selection stage, small fragments containing COII sequences could have been discarded. The benefits of having mitochondrial mini-chromosome remain to be determined.
This is a very intriguing hypothesis that suggests truly novel observations. However, still just little empirical evidence has been collected, and the results have to be verified by re-sequencing and detailed feature analysis.

4.3 Heterogeneity of the protein coding genes

Many researchers agree that the rate of molecular evolution of the sea anemones mitochondrial genome is unusually low (Daly et al., 2008; Hellberg, 2006; Miller and Ball, 2008; Shearer et al., 2002).
Sequencing of the *P. simplex* mitogenome revealed high heterogeneity of protein coding genes in size, and in DNA sequence when compared to the previously sequenced sea anemones.
After close examination of the protein sequences, transcriptomic data, and initial DNA pool, we concluded that sequencing errors could not have been a reason for gaps and deviations in the coding sequence. Reasonably high coverage could also serve as an insurance that sequencing errors are not responsible for the main heterogeneity in the mitogenome. Considering that rate of molecular evolution is low, a likely explanation might be that *P. simplex* could represent a separate branch of the sea anemones or, maybe, a separate branch of hexacorals.
During our comparative analyses, we have noted that P. simplex is not the only sea anemone that deviates from the rest. *Nematostella sp*., also possess considerably high heterogeneity in protein coding as well as rRNA genes, and does not have a group I intron that interrupts the COI reading frame.
Phylogenetic studies showed that *Nematostella sp*. appears as the most basal of the sea anemones (Daly et al., 2008; Emblem et al., 2014).

We assume that it is possible to expect a similar picture when the same analyses are applied to P. simplex mitogenome. However, it has not been possible to perform phylogenetic analysis

based on the complete mitogenome, since we still are unable to identify reading frames for two genes, and COII is missing from the assembly.

4.4 Transcriptomics data

In this study we generated mito-transcriptome data from two sea anemones species: *U. eques* and *P. simplex*. When compared, these transcriptomes reveal somewhat different expression levels in these species. Our results acquired from the Urticina transcriptome, compare well to previous findings (Emblem et al., 2014) and suggest universal but low level of expression for mitochondrial genes. The *P. simplex* mito-transcriptome, despite the low coverage, showed that most of the genes were transcribed and supported that the mitogenome is transcriptionally active.

For both species, rRNA genes showed the highest coverage, which was expected. Transcriptomics data supports P. simplex mitogenome assembly.

4.5 Suggestions and perspectives

Our results showed that the *P. simplex* mitogenome is highly unusual and has several surprising and true novel features. However, our results require further validation.

To validate the assembly and to remove any potential sequencing errors that could have been introduced by Ion Torrent, re-sequencing by a different technology platform should be performed. One possibility is to PCR amplify mitochondrial DNA using primers based on the acquired mitogenome, and Sanger sequence the amplicons. Another possibility is to use the SOLiD™ Next-Generation Sequencing System (Life Technologies™), or any other NGS platform. High number of reads coupled with a reasonable error rate, would ensure high coverage of the mitogenome and eliminate any uncertainties from previous assembly.

It is important to use physical shearing for the library preparation stage. Physical shearing would give more consistent and reliable results than enzymatic shearing. All protocols must be tested and optimized for coral tissue and DNA.

It could be useful to isolate pure mitochondria from coral tissues to ensure that obtained DNA and RNA are from the mitochondria and not from other parts of the cell. This might be especially beneficial for mitotranscriptome isolation. As we showed previously, different species have different expression levels. Some have low transcription levels, and thus low

amounts of extracted RNA in the sample. Mitochondria isolation should enrich samples with more copies of the transcriptome.

Larger tissue samples are recommended for this procedure and any reliable protocol could be used (Frezza et al., 2007). To remove the abundant rRNAs from the rest of the transcriptome RiboMinus™ (Life Technologies™) kit could be used. However, it is important to request the custom probes for rRNA depletion. Manufacturer recommends to include FASTA sequence file with a request, so the custom probes could be made specifically for desired rRNAs.

Mito-transcriptome analysis can be used not only to monitor see what genes have been transcribed, but it can also serve as a useful tool for assembly verification. Transcriptome data can also be used for finding the exact borders of the genes, but high quality transcriptome data are required for these applications.

Appendix A.

P. simplex mitogenome sequence

ATGTACATTTTAGTTTTAACTGCCCCCCTTTTGGGGGCCTTAGGAGCAGGTTGGTT
TGGTAGAAAAATAGGGGAAAGGGGCGCTGGAATTTTTACTTCCAGTTGTTTAATT
ATAAGTTTATCGTGGTCCCTTTTAATTTTTTATGAAAGCACATTAAACTCTTCAAC
TACATATATAAAACTATGGCGATGGTTGGATTCAGATTTAGTTACTGCCTATTTTG
GCTTGCAATTTGATGCTCTTACGGCCACTATGTTACTTGTGGTTACTAGTGTCTCC
ACCCTGGTACATATTTTCTCTACAGCCTACATGGACGGGGACCCCCACGTGCCTC
GATTTATGTCTTATTTATCATTATTTACTTTTTTTATGATTTTATTAGTCACTAGTG
ATAACTTCCCCCAACTTTTTATTGGCTGGGAGGGGGTGGGGCTATGTTCTTATTTA
TTAATAAATTTTTGGGTAACCCGGATTGAGGCTAATAGGGCCGCCATAAAAGCTA
TGCTTGTAAATCGGGTGGGGGATATGGGGCTTGTTCTAGCCATGTTAACAATTTG
GGACCAATTTGGGTGTTTGGGGTTTGCCTCTTTATTTAATACAGTTGGCCTTTCCT
CTGGGAATAGTATTATCTTGATATGTTTATTTTTATTCATCGGAGCAGTTGGAAAG
TCTGCGCAATTGGGATTGCACACTTGGTTACCGGATGCCATGGAAGGTGGGGTTG
GGCCATCTTGCTCAAGTAATTAGCTTTGCTTTTAAATGTTCTATATGTAGGAAGAC
CGGGGGGGGAAATCAGTCGGTAATAAAGCTGGGTGGAAAGGCCCCCTATAAGTGG
TTTGATGGGGTGGCGGCGGGCTATCTGGGGCCGTTAGTTCCTTAGAGACTTTATG
TATATCCACTGGGACTAGGGGGGCTAAGCCGATCACCGGGCCGGAACATGCTTA
AAAATTATAACAATACATAACATACCTGATCTAGTGTAGGATGATTACAGGTTTT
TACATCTTATCATTGGGGACTGTGGTTTCTGGAATAATGGTTATTTCTGCGCTTAA
TCCTGTTCATTCCGTTTTTTGGTTAGTGGTGGCCTTTATAAATTCAGCGGCCCTTTT
TATCTTATTGGGGGTTGATTTTATTGCCTTAATGTTGATAATAATTTATGTGGGTG
CTATAGCCATTTTATTTTTATTTGTAATAATGATGTTAAATTTAGCGGAGGGCGGG
GGAGAGTCCGACATGACAAATTATGCTCCAATTGGATTGGGGTTGGGCACCTTTT
TCTTGGGGGCTTTGGCCTCAAGGGGTGGGGGGCCCCCCACCCCCCTCCGGGGGGG
ATGAAATCTGACCCGCCCCTGGGTTTTACAAAAATCCCACAATATTGAAGCAATA
GGGAGGATACTTTATTCTGATTGTTATTACCTTTTTATTTTAGTTAGCTTTATATTG
TTAGTAGCCATGATTGGGGCCATAGTGCTCACCCAAGAAATTAGGAGCGAAATCG
GCCACCCTCCAAAGAAGCAAGATATTTGGGCGCAAACAAGCCGCTACATTAGAC
GATAATTCAATGGCCATTTTGAGAACCTAATGAGAGAGGAGACAAATCCAATCA
ACCAGAGAGCCATTTTATAATGGATTGTACTCATGAAGGCGGTTTATCATCCATA
CCATTTAGTAGACCCCAGCCCTTGGCCCTACGTGGGAGCCTGCGGAGCTCTTTTGT
TAACAGTAGGAAGTGTCGCCTATTTTCACTATAGTCAAACTTGAGTTTTATTGATG
GGGGCAATAGTCCTTAGTTTAACCATGATAGTTTGGTGGAGAGATGTAATAAGAG
AAGTGGCTTTCCAAGGTCTCCACACCATCATTGTAAAACAGGGTTTAAAATATGG
AATGCTCCTATTTATTCTTTCTGAAGTTTTGTTTTTTTTTCTTTCTAGTAGCCTGGC
CCCCAGTATAGAGTTAGGCGTGTGTTGGCCGCCTCAGGGAATAAACCCATTAAAT
CCTTTTTCAGTACCTCTTGTAAACACGGCCGTCCTTTTAAGTTCGGGGGCGACAGT
TACATGGGCGCACCATGCTTTAATTAGCGGAAAAAAAACTGAGGCCCGTAACGG
GTTGGCTGCTACTGTCATTCTGGGTGTTATATTTACAGGACTACAGGCAATGGAG
TATTACGAAGCCCCTTTTGCTATTTCAGACTCGGTCTATGGTTCTACTTTCTTCGTA
GCCACCGGGTTCCACGGGATTCATGTCATTATAGGAACAACGTTTTTAGCAGTCT
GTTTAAACAGGTTAGTCTATCATCAGCTTTCTCGCCACCACCACCTCGGCTTTGAA
GCTGCTAGCTGATATTGGCATTTTGTCGACGTAGTGTGGTTATTTTTATATGTTTG
TATTTATTGATGAGGGAGTTAAAAAATGATCTAGTGAGCTACAAATGATGCCTCA
ATTGAAAACTACTACTTATTTAATGCAATATAGATGGACTTTAATTGCTTTATTTT
TACTATTTTCTTCTTTAATCATTTTAGTATTGCCCGTTATTAAAACTAATTTTTTAA
TCCGAAGATCCATGGGGCCCCTGGCCCCCCAAACATCCGATTTTAATAAGAAGCT

GGTTTCGTTATGGAGTTCGCCCCTATAATCGGATATTAAATTCTTAATGATTTGAA
TCTGCGGAGCCTAGTGAGCAAGTCCCTAGGGGCCGCAGATAATTTAACAAAGCA
AACCCCGAAAAACAGAAAAGGGGGTGCACCTCTTTTATGTCCGCGAAAATTTAG
AACATCCCCACTGCGGATCCGTGGGCTAATTGAAATAATAGATTGGGGCAAATAA
TACCCAGCCCCAAAAGGTAAAAACAGGCCCCAATTAACAGGGCCCTCCCCAAAT
TAATTCTTTGGGGGCCCCTTAAAGCTTTTAGCCCCACCCCTCAAAAAGCATGAGTT
TGAAAAAATATGATTTTAACTAACCTAACATAATAAACCCCCGCAACTACCGAAC
AAACCACCGCGGTAATTGAAACTAAATAATATCCATAATTTATCCCGGCCAATAA
AATTCACCATTTCCCCAAAAAGCCCAATAAGGGAGGGACCCCCGCAATAGAAAA
AAAAGCTAACCCCAAAGTTATAGCCAAAATGGGCTCTTTTCTTGAAAGGCCCCCC
AATTCAACAATTAGATTTTGGCCCCCCCAAACCACCACTATAGCAAAGGCACAAA
TAGCCATAATAACATATAAGACCATATATATTAAACTGGCTTGAACACTTTCAAG
TGATCCCATTTCTATGCCCCAAAGAACAAAACCCATGTGAACTATCCCACTGTAG
GCCAAAAGCCGTTTAATTTTTATTTGGTTTAAAGCGCCCCCGGCACCTACCACTAA
GGAGAATAAAACGGCAATCAAAAGGGCTTGAGCCACGGGGCCAAGGGCAACCA
AAATGGAAAAAACTCCCACCTTGGGCACCGTGGCCAATAAAGCGGTTGTGGCAG
TGGGGGCCCCTTCGTAAACATCGGGAGCCCACATGTGGAAGGGGGCGGCGGATA
ATTTGAATAAAAGGGCCCCCGTTATTAAAAGGCCACCCCCCCCCAGAAATTCCCCC
CCGGGAGACCCCAAGTTCTTCTCCCGGAAAACCGGGGCCTCCTGAGAATCCACAT
AAGAGGCCACAGCCAAATAAAAATAACCCCGAGCTAACGCCCCTAATACAAAAT
ATTTTAACCCCGCTTCGGCGCTTAACCCCGACCCCCCCTTTTGGGACACCAAAATA
AAAAGAGAAAGGTGGGCAATTCCATAGCCAAATAAAGAGATATATAATTACTTG
CCGAGACTAAAAGGGTGCTCCCCAAGGCCACTAATAAAACTAAAACGGGTGTGG
CCCCCCCCCCTAGAGGGGCCAATCAAAATAAAAGGGGATAGGGCCCCAACAAGAA
TCACCATTTTGGTGCTTACAGCCCAACTATTAATTGTTAAAAACCCATTTTGCCAA
GGGAGGAGAACGTCAAGCCACGCTCAATACGTTACACACCCTAAAACTAAAAGA
GATATTTTTAAAGTAGAATTCTTTACACTATAAATAAGTAAAAAAAGTATGGACA
TGGTCAAAAAAAGTAAAAACATCCCCAACGTACATATTCAACTCCCCCAAAAAAT
AAGGAATGGGTTAAAGTAGCCCAGATAAAAAAATGATGTCTCTACCAATAATAG
CCATTAGTTTTTCAGTAAGTATAATAGTGATAAAATGTTGGGGCAGCAGGGGGGC
TGTTGCTCCTATAGGTGAGATTTCTATTCCTGCCCCTAGTGAGATTGTTGGCAGCG
GGGGCTCTAATAATGCTAATTTTTTTGATCAATTTCAACTCCTGAATTGAGGACAC
TTAACTAACTTGGTTATGGTTCTTATATTGGTGGTGGTTGTAGCTCTAGTGTTATT
AAGGGCTAATCAGTTAATTCCTAACCGAGGGCAGGCCGCCACGGAGTCGATTTAT
GATCATTTTGGGGGGCTAGTCAAAGATAATTCGGGGCCACAATACTTCCCCTTTG
TTCTTACCCTCTTTATTTTTATAGGGCTTTTAAATATTTTAGGGTTATTTCCTTATG
TTTTTACGGTAACAGCTCATGTGGTGGTGACGTTGGGCTTGTCATTTTCAATTATA
ATGGGGGTAACTATGGGGGGGCTTTGGAAGTTTAAATGGAACTTTTTAAGTATTC
TTATGCCAGCCGGAGCCCCTTTGGCCCTGGCCCCCCTTTTGGTATTAATTGAAACA
GTAAGTTATATATCCAGGATTATCTCTTTAGGGGTCCGTCTCGCCGCCAATTTATC
TGCGGGGCATTTATTATTTGCTATTTTGGCTGGCTTTGGCTTTAATATGGCCTTAG
TTCTTATTATTTTTCCTGTTTTGATTATGGTCTTTATAAGTTTACTAGAGGCCGCGG
TAGCGGTTATTCAGGCCTACGTATTTTCTTTGCTTACAAGTATTTACTTAGCTGAT
ACTCAAGTTTTACACTAAGTTAGTCGGTTAGTGAGCAAGTTAAGGTTATCGGACC
CTGATTGGGCGGGGTGTTGCGACCCTAAAAAGATGGCCTGCGGCGGATTGGGTTG
AATTTTCCTTTTACTTTTCTTGGGGATGTTAAATGTAATGGGAATTCCCAGAAACA
ATTATGGTAAATTAAAAAGAGCCGCGTTGAGATGGACTTTGGCAACCCTGGCTGG
CACTCTAATTTTATGGGGGGCTTGGGGGTTGGAAAGTAATTTTCAATATCCCTATC
AAATGGGGGGAATCGCCCCCACATCCTTAAGTTTTAAATGGGGCCCCCTTTTTTTA
GCCATTGACGGAATTTCTTTGTTTTTTTTATTTTAACTGCGCTGTTGACACCCATAT

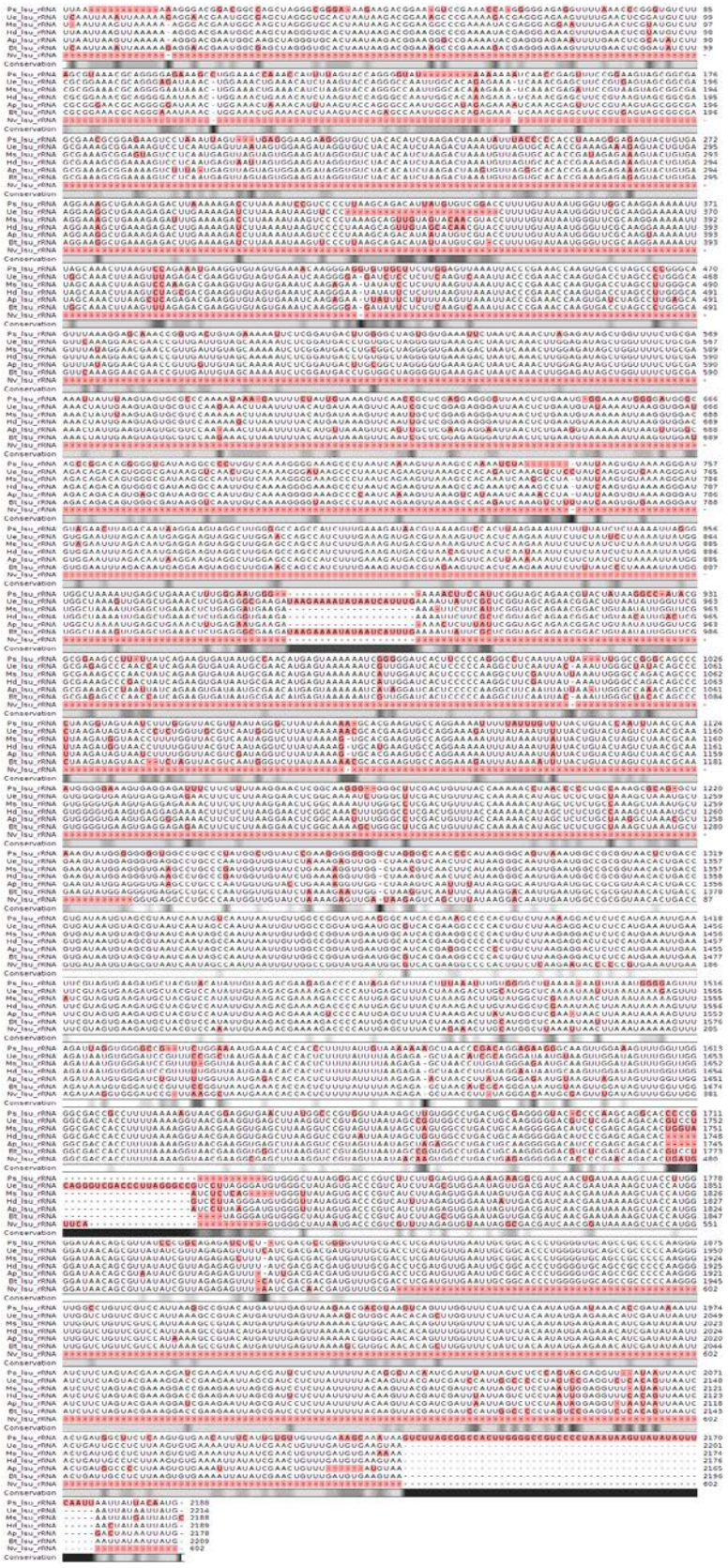
GCATTTTAATAAGTTGAAACTCCATAAAATTTTTATTAAAAGAGTTTTTGATGTGT
CTTTTATTTTTGGAAGTTTTATTAATGGGGGTCTTTTCCGCCCTAGACCTTTTATTG
TTTTATATTTTATTCGAGGGGGTTTTAATCCCCATGTTTCTTTTAATTGGGATTTGG
GGCTCTAGAGAAGAAAAAGTCCGGGCTTCTTTTTATTTCTTCTTTTACACTTTTAT
TGGCTCAGTGTTTATGCTTTTGGGGCTATTTCAATTATATAGTGTGACCGGGACCA
CTAACTACCTGGCCCTATTAAATTGGGGGTTGCCCCCCTCCACCCAAAAATGGGT
TTTTGTGGGCTTCTTTTTAAGTTTAGCCGTGAAAATTCCCCTGGTTCCTTTTCATAT
TTGGCTCCCCCAGGCTCATGTGGAGGCGCCAGTTTCGGGTTCGGTAATTTTGGCC
GGAGTTTTATTAAAACTTGGGGGGTATGGTTTTTTAAGGTTTACATGGCCCCTTTT
GCCCGGGGCCTCGGAATATTTTACCCCCTTTATAATAATGTTAAGTGTTTTAGCAA
TAATCTATGGAAGCTTGACCACTTGCCGCCAGGTGGATTTAAAGAGGCTCGTTGC
GTATTCTTCGGTGGCCCACATGGGGCTGGTGACTTTGGCAATATTTACCCATACA
ATTGAGGGCATGGGGGCGGCCCTTTTAATGATGTTGGCCCATGGGCTTGTAAGCT
CGGCGCTTTTTATTGCTGTGACTTATTTATATGAGCGCCACCACACCCGTCTTATT
AAATACTTCCGGGGTGTGGCTTTTTCTATGCCCCTATTTGTGTTTGTTTTTGGGGTT
TTAACATTAACTAACATGGCTATCCCCCCAAGTTTCAACTTTGTGGCCGAATTTCT
CTCTTTGCTGGCGGCCTTTCAATATAATTTGGTCATGGGGATATTGGCAGCCACGG
GGATGGTATGGGCTGCGGCCTATTCTCTTTTTTTATATAATCGGGTTGGTTTTGGG
GTCCCCTCAAGCCATCTTCTTTTTACAAGAGACTTAAATAGGCGGGAACTTTTGGC
TCTTTTCCCTCTGGTTTGACCCGTCTTTATTTTGGGGGTATTTCCTTCCCCCATTTT
GGACCCCCTGAAAAAGGCTTTGGTAATTAACCCTGGGGGTGCCTAATAAGACAA
GTTTCAAATGGGAGTGGTTTTAGTTATTCTTGGAATGCAAATTTTAGGCGCCCCAA
TCTAGTGAGCAAGTAAGTAGGGTTTACCCAAATTAATATGTGCCAATGGCCAGTA
AAAAATAAAGACCATGTTACGACTTTCAACTAGCCGCCTCATTTTAATTCCCCAG
TCTAGGAAGGCTGAAGCCAATAATCACCGGCAAGGGGAGGTTCCCCTCCCCTCAC
CATGTTACGACTTTCTCTACCTCATAGATCTTTGTTGCCACTTTAAGCGTCCAAAT
AATCTTTTTAAGTAAAGGCGACGGGCGATTTGTACCAACACTGATAGTAATTCAC
CGGGACGCTCTACTTCCCGATTACTATTAAATCCAACTTCGGGGGGCATTCCAGC
CCCTTTTCGGGCCCCTACTTTGGGGATTTGCCAGAGCATCCCATTATATAAGAAA
ATGTAGCGCATAAATCCCCCAACATGAGGACCATGAGACCTGTCTTTCTCCGAGC
TAAGGCGCGGATTGCGTACCCTTTTGTTAAACACGCGGCCTTAAGACGGCCATGC
AGTACCTGTTTTGTACATGGGCTTAAGGCGGCCATGCAATACCTGTTTTGTTTCAA
GCTGGGGTAAGGTTACACGCGGACCATCGAATTAAACAACACGCTCCTCTAATTT
AAGAAGTGAACAGCCAAACTTTTTGAATTTTAATTTTGCAATCGTACTACTCAGG
CGGAAGATTTTCCTTTAGGGTCTGCCCCGCATAATCTTCATAGTTCCCGGGGTGGA
CTACCAGGGTCTCTAATCCTGTTTGCTTCCCACCCTCTCATGTTTCAGCCTTTTTTA
GAGTAATTGGTTTTAACCTTGGAGGTTCTTTTTTCTATTTACACATTCTACCGCTAC
AGAAAAATTCCCTTTACCTTATCTTATAATTTTGGCCGACACATCCTTTACGCCTT
TTTGGTTATAGAGACCAGCCCTTAAGTTTTACCGCGTCTGCTGGCACTTAAGTTGG
ACAGACTGGGTACGTCCTCTCTGTGTCACGCTTTCGCTCATTGCACAAGATTCTCT
ACTGCGGTCTCCGGGTGAGCCTCCCCCTTGTTCCAGGGAAAGTGGGGCTCAAAGT
AAGCTTCCCATCTTAAGCAGGGTGGTCTTTTACACCCCCCTCCACCTAATAAGGTC
CCAGCCCAATCCTTTGGGTCCGGACTTACTCCCCTGTTTGCGGGGTTCTCCTAAAT
TCTTAGGCCCCACTTGCATGTATTAAAAGGGCAGCCTGCCTTCTATATCCCCCAAA
ATCTAAGGACTAATTCTCACTTTATTTGTTTTTAACAAATTTTACTATAACTTGTCT
CCCTATTAGAGCGTAAAATCCTAAAAGTAATACCCGGTAAAGGAAAATTAATGG
GTGCCCAGGATATTAATCTTTTTCCCTAAGGCCCCTTCTCCTGATACCCTATTAAT
TACCCCCGAGCAGACACGCCCCGGGGGATGTGGGTTATAGTGAGAGTTAAACCC
ATTGCAGTAATTTAACCCCTGAGCCAATTAAGAGATTTGATAGTGATTGTTCCCTT
TATTCGATAATAAGCCACCATAATGGCCAAACCAATAGAGGACTCTGCTGCGGCT

ACCGTTAATACCATAATAGTAAAAACCTGCCCGATTAACGCATCCGTTACCATAG
AATTTATTAAAAACAAAAAAGAGGCCGCCAATAAAATTAGTTCTATTGACATTAA
CATAATTATCAGATGCGCCCTATTCAGCACCACTCCTAATACCCCCAAGACTAAC
AATAAGACTGAAACCACCAGGTACCTGTTATACACCTCTTAAACGTTCCCACTAA
CATAGCTAGACCCAGATTCCCCACCGGAAGAATTCGAAAGGGGTACCCCCAATAT
TATTTTATCATTAACCATCAGCTGTAAATTATTAGATTTTATTAATCGGGCAAAAA
TAAAAGCTGCCTGGGGCGCTACCCCTCTTAAACGTTCCTAGAACCGTGCAAATAA
ATTTTAGTCAGAAGCAAGAAATTTTTAAAATTTCCCTGATGGCATAGGAATTTAC
TTATTTCAGTGTCAAATGGGTGAAATATTTTGTAAGGTAGTCATCAAGGCCTCGTT
TTTAAAATTTGCATTGCGAGTCGTGGTACTCAATCTCAAGAGCTTCCTGAAATGA
GATTGATTATGAGGTTAGATGATTATGGAGATTCTAAAAATTTTGCTAATAGTTAT
TCCGTTACTGGTAGCGGTGGCTTATTTAACTTTGGCTGAACGGAAGGTTTTGGGGT
ACACGCAAGCTAGGAAGGGACCAAATGTAGTGGGTGTTTATGGGTTATTCCAACC
TCTGGCCGATGGGATAAAACTATTTTCTAAAGAATTAGTTGTTCCCCACTTTGCTA
ATTTGTTTATTTATGTGGCAGCCCCCGTTTTTTCATTCACATTGGCTTTATTAGCGT
GGGCCATCATCCCCTATGAGAAAGGGGCTGTAATAAGTGATTTGAAAATTGGAGT
TTTGTTTACTTTAGCCGTGTCTTCTATAAGTGTCTATGCCATTTTGATGTCCGGGTG
GGCCAGCCAATCCAAATATGCGTTCCTGGGGGCCATAAGAGCTGCTGCCCAAATG
ATAAGTTATGAAGTTTCTATTGGATTAATAATAATAACAGTTATTCTGTGTGTGGG
CTCTTTGAATATTACTGAAGTGGTTTTGGGGCAAAGTTGGGGCACTTGGTTTTTTT
TCCCCCTCTTCCCGGCGGCAATGATGTTTTTGGCTTCGGCCCTGGCCGAAACCAAT
CGAGCCCCCTTTGATTTAACGGAGGGGGAGTCCGAGCTGGTGTCGGGCTACAATG
TAGAGTACTCCTCTATGTCTTTTGCTTTATTTTTTCTTGCAGAATATGCCCATATAA
TTTTGATGAGTTGTTTAACTACAATTTTATTTTTTGGGGGGTGACTTTCCCCGGGG
GGTTTCCTAAGGGGGGGGGCAGTTTGGTTGGGTTTTAAGGCCTCTTTTATAATTTT
TCTTTTTGTTTGGGTAAGAGCCTCTTTTCCCAGGGTGCGATATGACCAATTAATGG
CCTTATTATGGAAATCTTACCTGCCCCTAAGCTTGGGCTTTGTAATTTGGGTGGCC
AGTCTCTTATGAGGGGGAGGCGGGGCGCCCCCCTACTAAGAGGAGACAATAGTA
TTTGTTTTTATGCCCCCGGGGGCTTGTTATCTTACTAGGGGGTTAGGTTAAGGAAG
ACCAATGGCCTTCAAAGCTATTAGTGCGGGTTCGAGTCCCGCACCCCCTGGCATT
AAATATCTGGGGCCCCGGCATAGCATAAACCATGAGCTGGCGGTATTAAAACCTT
CCCCATAAACTTGAGATGTTCAGTGAAGTTTTTTATATTATTAATTTACTAATCTT
TGGAATTATGATCTCTGCTCTTGTGGCGGGAACTTCTTATATCTTGGGGGACAAAC
AACCTAACCAAG
AGAAGGTTTCTACTTACGAATGTGGGTTTGACCCCTTGGGAGCCCCGGGCAAGCC
TTTCTCCATTCATTTTTTTCTTATAGGAATTCTTTTCCTTATTTTTGATTTAGAAATT
TCTTTTTCTTTTACCTTGGTGCGTCGTCTATAATCAAACTTTTCCTTTTAGTTATTGA
ACCATGATCATTTTCCTAATCATTCTGACTGTGGGGTTAGCCTATGAGTGGTTAAA
GGGGGGCTTAGAGTGAGAAGGCCATGGAAAATAAGAAGGCTCAGATCCACCAGG
AGTCTTTAAGATGCAATTTAGAAAACAAAATCCTATCTTATCTATTTTTAATGGAA
TAATCATTGATTTACCGGCCCCCGCCAATCTTAGTTATTTTTGGAACTTTGGTTCTT
TATTGGGAGTGTGTTTAATAGTACAAATCATTACGGGCATATTTTTGGCGATGCAT
TATTGTGCGGACGTAAGCTTAGCTTTTGCTTCTGTAGACCATATTATGCGGGATGT
TAATTATGGGTTTTTACTAAGGTACTTTCATGCTAATGGGGCCTCCATGTTTTTCC
TATGTCTTTATTTTCATATCGGCCGAGGAATTTATTATGGAAGTTATTCTAAAATT
GAAGTTTGGAATGTCGGCGTTATAATATTAATATTGACAATGGCGACGGCCTTTA
TTGGATATGTTTTACCTTGGGGGCAAATGTCTTTTTGAGGCGCTACAGTTATTACC
AACTTATTGTCTGCTATTCCTTATATAGGAACTGATATTGTTCAATGAGTATGAGG
AGGGTTTAGCGTATCTAATGCTACACTCAATCGTTTTTTTAGCCTTCATTATTTATT
TCCTTTTCTTTTAGCAGCTTTAGGGGTTGTTCATTTAATTTGCTTACATGTAGATGG

TTCTAATAATCCTATCGGAGTAAAATCTGACCTAGATAAAGTTCCCTTTCACCCCT
ACTTTACTTCAAAGGATTGGTTTGGTATAGTGGCATTAGTGCTATTTTTTTGCTCC
TTAGTTTTCTTAGCCCCCAATTTATTAGGGGACCCTGAGAATTTTATTCCAGCTAA
CCCCTTAGTGACTCCAGTACATATTCAACCAGAGTGGTATTTTTTATTTGCCTATG
CCATTTTACGTTCTATCCCCAATAAATTAGGAGGGGTGGTGGCTATGTTTGGTAGT
CTTCTTGTATTATTTTTACTTCCTTGGATTCATACTAGTCGATTAAGGGGATTAACT
TTTCGACCATTAGCTCGATTGGCTTTTTGGTTTTTAGTAGCGGACTTTTTACTTCTT
ACTTGGATAGGATCTCAGCCTGTTGAAGAGCCCTTTATTTTAGTAGGACAATTAG
CCTCTATCTTTTATTTTGCTTACTTTTTAGTATTAGGTCCCCTGTTAGGTTATCTTG
AAAATCATCTTTTATTTCTAGACCCCCCAACAACTTCATAATATTCAGTGAGCAAG
ACCCCAACACAAGATATTGGATTTTTTATCGGGAATGTTTAGAGATTCCAGTCCCT
AATGACAGCTATTAGAGATGGTCTCTAGTGGTCGTAGATAAAATTAAAATCCACC
ATTTTGTAAAGCAGGATTAACATGATCCGGCCAAGAAGGATGTCCCTAATGACAC
CCTGATTTTGGTAGAGGCAGGAATCGACCCTGCACCTCCAGATGATGAGCCTGGC
GACCTCGTTTAGTCCACTCTACCAGATCTTTATTTGTTTTTAAAATTTGGGGCTGTT
GCTATAACTAGTAATAAAATAGTTATTCTATTGGTGTGAAACCATTTTGTCCAGCT
GGGCTCCCTTTAGTGGCTACATTCGTCCTTCTTTCTAGGGGGGACATACCTATAAT
TTTCGTTGGCCTCCCCCTATGAGTGTGGGTTTGATTCTGTGGGAAGATTTCCGATG
CGAGTTTTTAAAAACAAATAAAGTGGAGGCTAAAAAATGCTCCTGACTTGATGTT
GTTACAAGGCACAAATAGACATAACCTCTACCAAAACCAGGTTTAAAAAGGGAC
GGACGGCCAGCTAGGGCGGGAAAGAAGACGGAAGTCCGAAACCAGGGGGAGAG
GTTTTAAACCCGGGTGTCTTAGCGTAAACGCAGGGGAGAAAGCCTGGAAACCAA
ACCATTTTAGTACCAGGGGTATAAAAAAATCAACCGAGTTTCCGGAAGTAGCGGC
GAGCGAACGCGGAGAAGTCCTAAATTAGTTGAGGGAAGAAGGGTGTCTACACAT
CTAAGACTAAATATTTACCCCCACCTAAAGGGAGAGTACTGTGAAGGAAAGCTG
AAAGAGACTTAAAAAGACCTTAAAATCCGTCCCCTTAAGCAGACATTATGTGTCG
GACCTTTTGTATAATGGGTTGGCAAGGAAAAATTTAGCAAACTTAAGTCCAGAAA
TGAAGGTGTAGTGAAAACAAGGGAGGTGTTGCTTCTTGGAGTTAAATTACCCGAA
ACCAAGTGACCTAGCCCCGGGCAGTTTAAAGGAGCAAACCGGTGACTGTAGAAA
AATTCTCGGATGACTTGGGGCTAGTGGTGAAATTCTAATCAAACTTAGAGATAGC
TGGTTTTCTGCGAAAATTATTTAAGTAGTGCGCCCAAAATAAAGATTTTCTATTGT
AAAGTTCAACCGCTCGAGGAGGGGTTAACTCTGAATGGGAAAATGGGGATGGGC
AGCCGGACAGGGGGTGATAAGGCCCCTTGTCAAAAGGGGAAAGCCCTAATCAAA
AGTTAAAGCCAAAATCTATATTAAGTGTAAAAGGGATGTAGAACTTAGACAATA
AGGAAGTAGGCTTGGGGCCAGCCATCTTTGAAAGATAACGTAAAAGTCCACTTA
AGAAATTCTTTTATCTCTAAAATTAGGGTGGCTAAAATTGAGCTGAAACTTTGGG
AATGGGAAAACTTCCATTCGGTAGCAGAACGTACTATAAGGCCATACGGCGGAA
GCCTTTTATCAGAAGTGATAATGCCAACATGAGTAAAAAATCGGGGGATCACTTC
CCCAGGGCCTCAATTATTATTGGCCGGGCAGCCCCTAAGGTAGTAATCTTTGGGT
TACGTTAATAGGGCTTATAAAAAAGCACGAAGTGCCAGGAAAATTTTTATTTGTT
TTACTGTACCAATTTAACGCAAATGGGGGAAAGTGAGGAGATTTCTTCTTTTAAG
GAACTCGGCAAGGGGGGCTTCGACTGTTTACCAAAAACCTAACCCCCTGCCAAAG
CGCAGGCTAAAGTATGGGGGGGGTGGCCTGCCCTATGGCTGTATCCGAAGGGGG
GGGGCTAGGGCCAACCCCATAAGGGCAGTTAAATGGCCGCGGTAACTCTGACCG
TGATAATGTAGCGTAATCAATAGTCAATTAATTGTTGGCCGGTATGAAGGGCATC
ACGAAAGCCCCACTGTCTTAAAAGGACTCTCCATGAAATTGAATTCGTAGTGAAG
ATGCTACGTACATATTGTAAGACGAAGAGACCCCATAGAGCTTTACTTTAAATTT
GTGGGGCTTAAAAAATTTAAATGGGGAGTTTAGATTAGGTGGGGCCGTTCTGGAA
AATGAAACACCACCCTTTTATTGTAAAAAAAGCTAACCCGACAGGAGAAGGGCA
AGTTGGAAAGTTTGGTTGGGGCGACCGCCTTTTAAAAAGTAACGGAGGTGAACTT

ATGGCCCGTGGTTAATAGCTTGTGGCCTGACTGCGAGGGGTACCCCCAAGCAGGC
ACCCCCGGTGGGCTATAGGGACCCGTCTTCTTGGAGTGGAAAAGAAGGCGATCA
ACTGATAAAAGCTACCTTGGGGATAACAGCGTTATCCCGGCAGAGATCTCTTCGA
CGCCGGGGTTTGCGACCTCGATGTTGAATTGCGGCACCCTGGGGTGCAGCCGCCC
CCAAGGGTTGGCCTGTTCGTCCATTAAGGCCGTACATGATTTGAGTTAAGAACGA
CGTAAGTCAGTTTGGTTTCTATCTACAATATGAATAAACACCGATAAAATTATCTT
CTAGTACGAAAGGATCGAAGAATTAGCGATTCTCTTATTTTTACAGGGTACAATC
GATTTATTAGTCTCCCAGTAGGAGGTTATAATTAATCACTGATGGCTTCTCAAGTG
TGAACATTTCATTGTGTTGTTTGAAAGCAAATAAGTCTTAGCGGCCACTTGGGGC
CGTCCCCTAAATAAGTTATATATTTCAATTAATTATTACAATGTTACTAAAAAGTT
AGAACGTTGCTAACAAAGGAGGGGCCGTCCTTGGAATATTTTTTAAGATGACAAG
CCCATTTTTAACTCGTTGAGTATTTTCTACTAATCATAAAGATATTGGAACATTAT
ATTTAGTCTTTGGAATAGGATCCGGTATGATAGGCACAGCGTTAAGCATGTTAAT
AAGATTGGAATTATCCGCCCCCGGTTCTATGTTGGGGGAGGACCATCTTTATAAT
GTCATAGTAACAGCCCATGCTTTTATTATGATTTTCTTTTTAGTAATGCCCGTAAT
GATAGGTGGGTTTGGAAATTGGTTAGTCCCCCTATATATTGGGGCTCCCGACATG
GCTTTCCCTCGATTAAACAATATTAGTTTTTGGTTGCTTCCGCCTGCGCTTATATTA
TTATTGGGCTCGGCCTTTGTGGAACAAGGAGTGGGGACAGGGTGAACAGTCTATC
CCCCCTTGTCCAGCATCCAAACACACTCGGGGGGGGCGGTTGACATGGCCATTTT
TAGTCTTCATTTGGCGGGTGCTTCTTCTATTTTAGGGGCCATGAATTTTATAACCA
CCATATTTAATATGAGGGCCCCAGGAATGACTATGGACAGACTCCCCTTATTTGT
GTGGTCCATTTTAATCACAGCCTTTTTATTGTTACTCTCCCTACCAGTTTTAGCAG
GGGCCATTACGATGTTATTGACTGATAGGAATTTTAATACAACTTTCTTTGATCAG
GCCGGGGGCGGAGACCCTATTTTATTCCAACATTTATTTTGGTTCTTTGGGCACCC
CGAGGTTTATATTTTAATTCTCCCAGGGTTTGGCATGATTTCTCAAATCATACCCA
CTTTTTCTGCTAAAAGTCAAATTTTTGGGTATTTGGGCATGGTCTATGCCATGTTA
GCTATTGGAATACTGGGCTTTATTGTGTGGGCACATCACATGTTTACGGTTGGCAT
GGATGTTGACACAAGAGCTTATTTCACTGCCGCCACAATGATTATTGCTGTCCCC
ACTGGAATCAAGGTATTTAGTTGGTTGGCCACTATTTATGGTGGTAGGGTAAGAC
TAGACACCCCCATGCTTTGAGCCATTGGGTTTGTCTTCCTATTCACAATAGGGGGC
CTCACCGGGGTTGTTTTAGCTAACAGTTCTTTAGATGTAGTTTTACATGATACTTA
TTATGTAGTGGCCCATTTCCACTACGTCCTATCAATGGGGGCTGTTTTTGCCATCT
TTGGGGGGTTCTATTTTTGGTTTGGGAAAATTACGGGCTATGGCTACAATGAACT
CTATGGGAAAATCCATTTTTGGATAATGTTTATCGGGGTTAATTTAACATTTTTCC
CTCAGCATTTCTTAGGTCTAGCGGGCTTTCCTAGACGATACTCTGATTTTGTTGAT
GGTTTGGCGGGTTGGAATTTAGTTAGCTCCTTAGGTTCCACCATATCAATTGTGGG
TGTCCTTTGGTTTGTCTATGTAGTATATGATGCCTATGCTCGAGGAATAAAATTTA
TGGGCTGGGTGGAGGACACGGGGTCTTCTTTAGAGTGGGTACAACCCTCCCCCCC
TATGTCTCATACCTACAATGAATTACCCTTCTCTTATGGAAGAGTGGCCACTCACT
TACGGAGTAAATAGAAAATAGTGAGCAAGTATAAAGACCTTGCGCATCAGAAAG
ATATTCTTACGATAATTAATTGTTGCCCCTCGAGAGCAAGTAAAGAAGTGGAAGA
TAAAGTCCCCCCAGCTATGAGAGTAGCGACCAGTTATAAAAACCAACCCCAGTGT
CTGCTTTAATACATGCTGCCACCATGGTTACAGCCGGCGTTTTTTTATTAATTAGA
TCGTCCCCCCTATTAGAGCAGGCGCCCATAGCTTTAATGGTTGTCACCATTGTGGG
GTCACTTACGGCATTTATGGCCGCCACCGTTGGCCTGGCACAAAATGATTTAAAG
AAAGTAATAGCTTATTCAACTTGTAGTCAATTGGGGTACATGGTGATGGCTTGTG
GGCTTTCTCAATATTCTATAAGCCTATTTCATTTAATGAACCATGCTTTTTTTAAA
GCTTTATTATTCTTAAGCGCGGGGTCCGTGATTCATGCCTTGGCCGATGAACAAG
ATATGAGAAAAATGGGAGGGCTTATAAGATCGATCCCCTTTACTTATACTATGAT
GATTATTGGCTCTCTGTCTTTAATGGGTTTTCCTTATTTGACAGGGTTTTATTCTAA

AGATTTAATTTTGGAATTAGCCTACGATCAATATTATTTAGCCTTTGCTCATTGGC
TGGGGGTTTTTTCGGCCCTATTAACAGCCGCCTATTCCCTTCGATTAGTGTACCTT
ACTTTCATATTAAATACTAACGCTAAAAAGGAGGTTTTTTCCCATGCCCATGAGG
GCTCTTGAAATTTAACTTGGCCTCTAATTTTATTGGCCCTGGGGAGCATTTTTGTG
GGTTATTTATCTAAAGAGATTATTTGATCATTTCAAATCACTATCTCTCCTATTATT
CCCACCTCTATTAAGTTAATGCCTATAGGCTTTAGTCTTTTGGGGGCGGGGGCGG
CGGTAGTTCTTTATCATTATTCTTCACGAATCTTTATTGCACCAACTTCGCCCGTTG
GCCTGGCTGGTTACACCTTTTTATACTCTGCTTGGCAGTTTAATTATATTATTAATC
ATTTTTTGGTCCAGAATGTGTGAAGAATGGGCCATCTTATTACATTCCGTGCCATT
GATAAAGGTGCCATTGAGCTCGTTGGGCCCAAAGGAGCCTCTCAATTCCTGATCC
GATTCACCCAAGAAATTAGTAACCTACAGCCTGGTAGAGTATATAATTACGCGTT
GGTTATTCTAATCGCTATTTATGTGTTTACATGTTTATGGTCAACTTAAAAAGTAT
CGAAAGTCATATTGAAGAATAAATGG

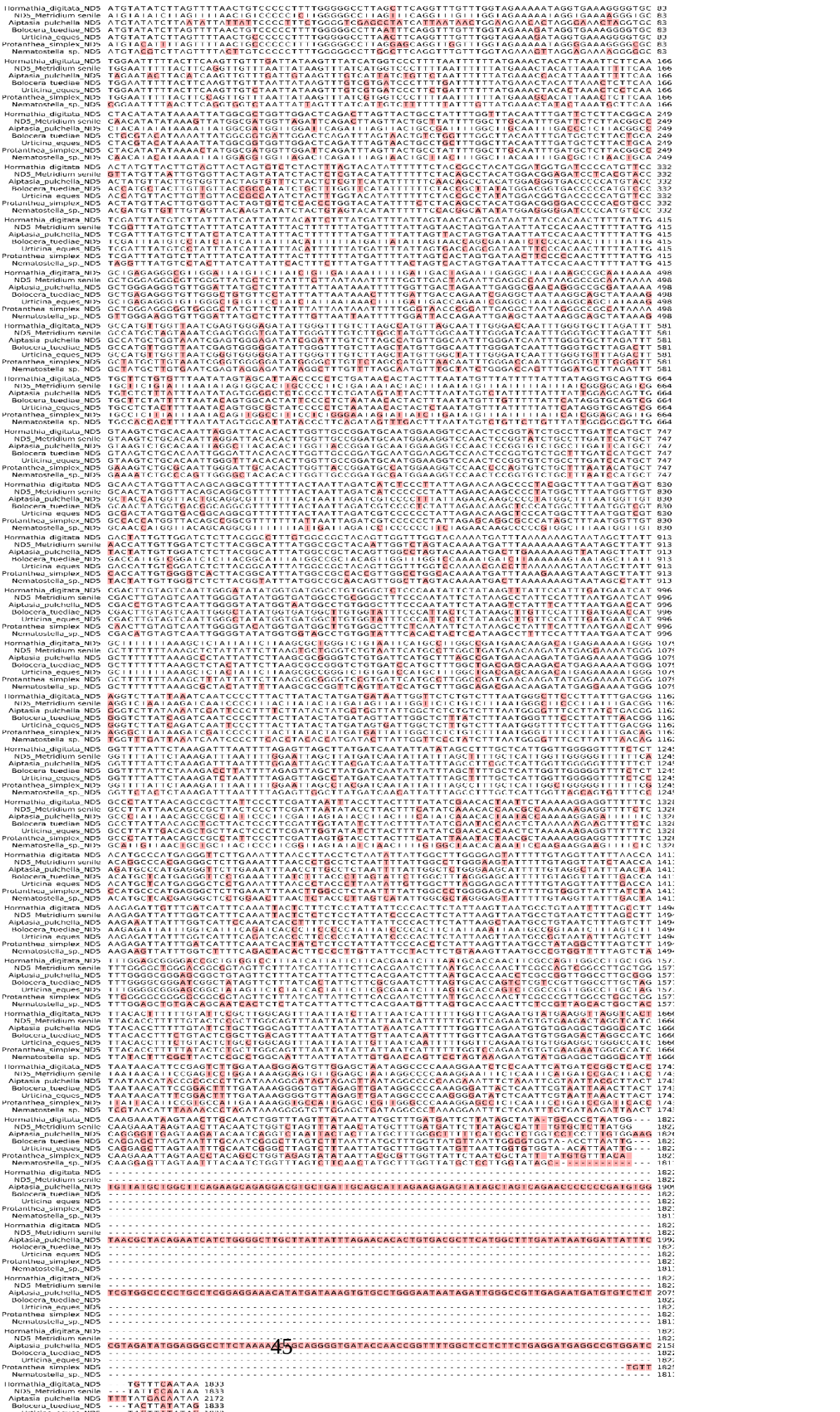


Appendix B

Multiple sequence alignment of protein coding and rRNA genes

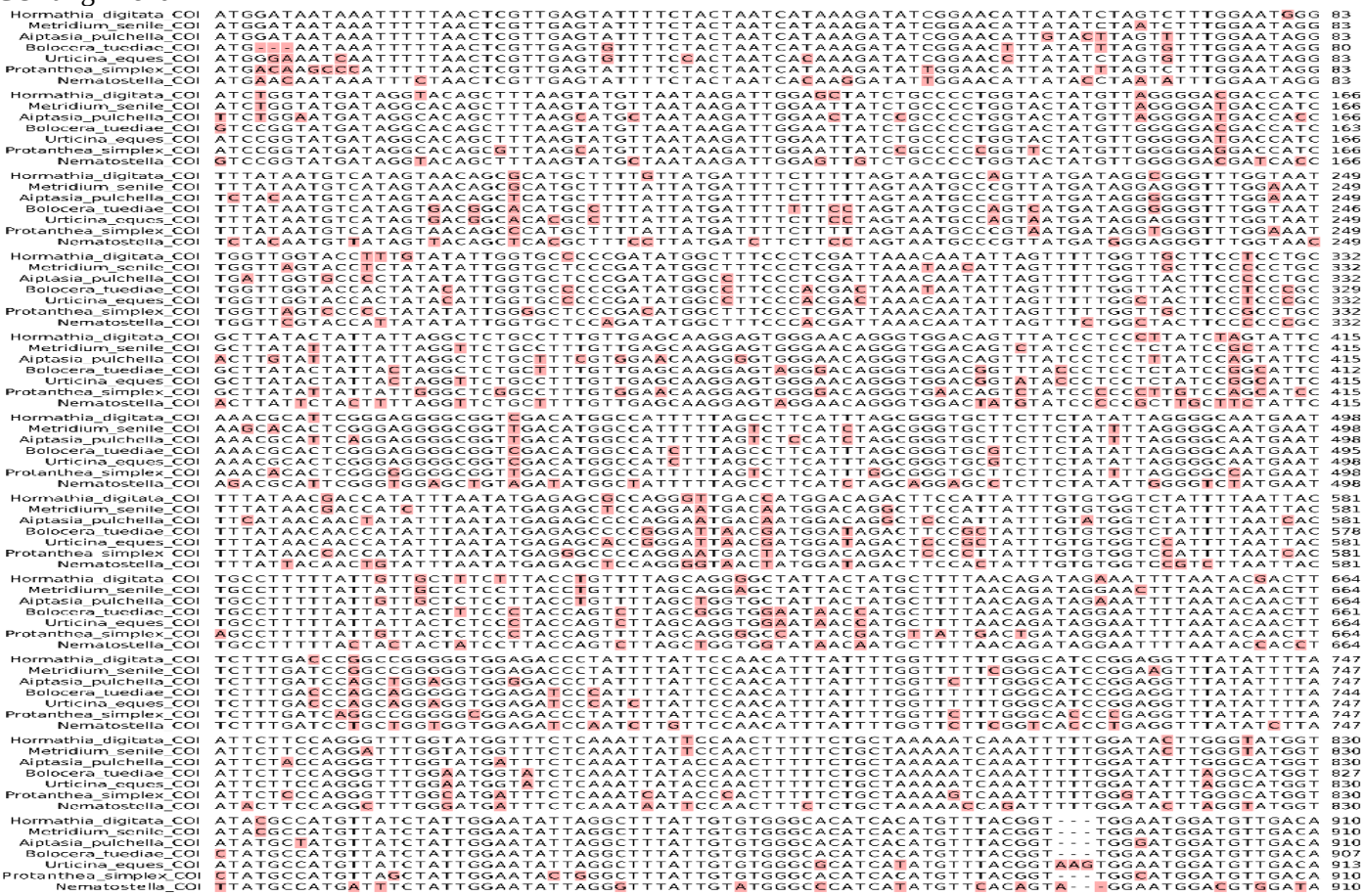
LSU rRNA

ND5



COII primer design

COI alignmend



References/Bibliography

Beagley, C.T., Okimoto, R., and Wolstenholme, D.R. (1998). The mitochondrial genome of the sea anemone Metridium senile (Cnidaria): introns, a paucity of tRNA genes, and a near-standard genetic code. Genetics *148,* 1091-1108.

Berntson, E.A., France, S.C., and Mullineaux, L.S. (1999). Phylogenetic relationships within the class Anthozoa (phylum Cnidaria) based on nuclear 18S rDNA sequences. Mol Phylogenet Evol *13,* 417-433.

Birgisdottir, A.B., and Johansen, S.D. (2005a). Reverse splicing of a mobile twin-ribozyme group I intron into the natural small subunit rRNA insertion site. Biochemical Society transactions *33,* 482-484.

Birgisdottir, Å.B., and Johansen, S. (2005b). Site-specific reverse splicing of a HEG-containing group I intron in ribosomal RNA. Nucleic Acids Research *33,* 2042-2051.
Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Analytical biochemistry *162,* 156-159.

Daly, M., Chaudhuri, A., Gusmão, L., and Rodríguez, E. (2008). Phylogenetic relationships among sea anemones (Cnidaria: Anthozoa: Actiniaria). Molecular Phylogenetics and Evolution *48,* 292-301.

de Magalhães, J.P., Finch, C.E., and Janssens, G. (2010). Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. Ageing research reviews *9,* 315-323.
De Paepe, B. (2012). Mitochondrial Markers for Cancer: Relevance to Diagnosis, Therapy, and Prognosis and General Understanding of Malignant Disease Mechanisms. ISRN Pathology *2012,* 15.

Emblem, A., Karlsen, B.O., Evertsen, J., Miller, D.J., Moum, T., and Johansen, S.D. (2012). Mitogenome polymorphism in a single branch sample revealed by SOLiD deep sequencing of the Lophelia pertusa coral genome. Gene *506,* 344-349.

Emblem, Å., Karlsen, B.O., Evertsen, J., and Johansen, S.D. (2011). Mitogenome rearrangement in the cold-water scleractinian coral Lophelia pertusa (Cnidaria, Anthozoa) involves a long-term evolving group I intron. Molecular Phylogenetics and Evolution *61*, 495-503.

Emblem, Å., Okkenhaug, S., Weiss, E.S., Denver, D.R., Karlsen, B.O., Moum, T., and Johansen, S.D. (2014). Sea anemones possess dynamic mitogenome structures. Molecular Phylogenetics and Evolution *75*, 184-193.

Fedorov, S., Dyshlovoy, S., Monastyrnaya, M., Shubina, L., Leychenko, E., Kozlovskaya, E., Jin, J.-O., Kwak, J.-Y., Bode, A.M., Dong, Z., *et al.* (2010). The anticancer effects of actinoporin RTX-A from the sea anemone Heteractis crispa (=Radianthus macrodactylus). Toxicon : official journal of the International Society on Toxinology *55*, 811-817.

Frezza, C., Cipolat, S., and Scorrano, L. (2007). Organelle isolation: functional mitochondria from mouse liver, muscle and cultured filroblasts. Nat Protocols *2*, 287-295.

Galtier, N., Nabholz, B., Glemin, S., and Hurst, G.D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Molecular ecology *18*, 4541-4550.

Goddard, M.R., Leigh, J., Roger, A.J., and Pemberton, A.J. (2006). Invasion and Persistence of a Selfish Gene in the Cnidaria. PLoS ONE *1*, e3.

Haugen, P., Simon, D.M., and Bhattacharya, D. (2005). The natural history of group I introns. Trends in Genetics *21*, 111-119.

Hedberg, A., and Johansen, S.D. (2013). Nuclear group I introns in self-splicing and beyond. Mobile DNA *4*, 17.

Hellberg, M.E. (2006). No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. BMC Evolutionary Biology *6*, 24-24.

Jiang, H., Barker, S.C., and Shao, R. (2013). Substantial variation in the extent of mitochondrial genome fragmentation among blood-sucking lice of mammals. Genome biology and evolution *5*, 1298-1308.

Johansen, S.D., Emblem, A., Karlsen, B.O., Okkenhaug, S., Hansen, H., Moum, T., Coucheron, D.H., and Seternes, O.M. (2010). Approaching marine bioprospecting in hexacorals by RNA deep sequencing. New Biotechnology *27*, 267-275.

Lin, M.-F., Kitahara, M.V., Luo, H., Tracey, D., Geller, J., Fukami, H., Miller, D.J., and Chen, C.A. (2014). Mitochondrial Genome Rearrangements in the Scleractinia/Corallimorpharia Complex: Implications for Coral Phylogeny. Genome biology and evolution *6*, 1086-1095.

Medina, M., Collins, A.G., Takaoka, T.L., Kuehl, J.V., and Boore, J.L. (2006). Naked corals: Skeleton loss in Scleractinia. Proceedings of the National Academy of Sciences of the United States of America *103*, 9096-9100.

Miller, D.J., and Ball, E.E. (2008). Cryptic complexity captured: the Nematostella genome reveals its secrets. Trends Genet *24*, 1-4.

Miller, D.J., Ball, E.E., and Technau, U. (2005). Cnidarians and ancestral genetic complexity in the animal kingdom. Trends in Genetics *21*, 536-539.

Moen, F. E., Cochrane, S., Pleijel, F. & Svensen, E. 2004. *Marine fish & invertebrates of Northern Europe*, Kristiansund, KOM.

Osigus, H.-J., Eitel, M., Bernt, M., Donath, A., and Schierwater, B. (2013). Mitogenomics at the base of Metazoa. Molecular Phylogenetics and Evolution *69*, 339-351.

Plaisance, L., Caley, M.J., Brainard, R.E., and Knowlton, N. (2011). The Diversity of Coral Reefs: What Are We Missing? PLoS ONE *6*, e25026.

Sanders, L.H., McCoy, J., Hu, X., Mastroberardino, P.G., Dickinson, B.C., Chang, C.J., Chu, C.T., Van Houten, B., and Greenamyre, J.T. (2014). Mitochondrial DNA damage: molecular

marker of vulnerable nigral neurons in Parkinson's disease. Neurobiology of disease *70*, 214-223.

Sarma, N.S., Krishna, M.S., Pasha, S.G., Rao, T.S., Venkateswarlu, Y., and Parameswaran, P.S. (2009). Marine metabolites: the sterols of soft coral. Chemical reviews *109*, 2803-2828.

Shao, R., Zhu, X.Q., Barker, S.C., and Herd, K. (2012). Evolution of extensively fragmented mitochondrial genomes in the lice of humans. Genome biology and evolution *4*, 1088-1101.

Shearer, T.L., Van Oppen, M.J., Romano, S.L., and Worheide, G. (2002). Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). Molecular ecology *11*, 2475-2487.

Steele, R.E., David, C.N., and Technau, U. (2011). A genomic view of 500 million years of cnidarian evolution. Trends in genetics : TIG *27*, 7-13.

Sunagawa, S., Woodley, C.M., and Medina, M. (2010). Threatened corals provide underexplored microbial habitats. PLoS One *5*, e9554.

Weis, V.M., Davy, S.K., Hoegh-Guldberg, O., Rodriguez-Lanetty, M., and Pringle, J.R. (2008). Cell biology in model systems as the key to understanding corals. Trends in ecology & evolution *23*, 369-376.