**EEOB-563 Final Project**

**Introduction**

While animal mitochondrial DNA (mtDNA) has traditionally been depicted as a uniform, circular molecule encompassing a conserved set of genes (Gray et al. 1999, 2004; Gualberto et al. 2014; Smith and Keeling 2015), multiple exceptions to this general picture exist (Lavrov and Pett 2016). In particular, transitions from circular to linear mitochondrial genome architecture have occurred independently in at least three groups of animals, having been reported extensively among non-bilaterian lineages such as Medusozoa (Cnidaria) and Calcarea (Porifera) and within Bilateria in at least fifteen species of isopods (Bridge et al. 1992; Doublet et al. 2012; Lavrov et al. 2013).

Linear mtDNA architecture is ubiquitous among the medusozoan cnidarians, suggesting that it is an ancestral trait that first evolved in the group's common ancestor. Medusozoan mtDNA also includes an extra open reading frame (ORF) resembling DNA-polymerase, termed polB, and inverted terminal repeats (ITRs) that, in some hydrozoans, may contain a portion of *cox1* (Kayal et al. 2012). Within Calcarea, all mitogenomes appear linear and multipartite (Lavrov et al. 2013, 2016). There appear to be two unique cases of linear mtDNA within calcarea. Within *Clathrina* the mtDNA exists as a few chromosomes where a large proportion of it is repetitive DNA and the ends of the chromosomes are identical. While in others it appears to be composed of many chromosomes, with nearly 1 gene per chromosome, and many of these are “empty chromosomes” containing no identifiable coding sequences. Among isopod crustaceans, multiple members of the family Oniscidea have a mitogenome consisting of two molecules, a circular dimer formed from the fusion of two monomers in opposite polarities and a linear monomer (Doublet et al. 2012). The differences among linear mtDNA within metazoan suggests that perhaps there are more than one mechanism that can lead to linearization of mtDNA.

Outside Metazoa, linear mtDNA with ITRs has been reported in multiple lineages of fungi, plants, and other eukaryotes (Zardoya 2022). Within fungi, linear mtDNA arrangement is hypothesized to result from the integration of linear mitochondrial plasmids that encode DNA polymerase B (Mouhamadou et al. 2004). It has been hypothesized that among Medusozoa the transition to the linear mtDNA occurred via a similar process, as supported by the presence of a DNA-polymerase-B-like ORF (Kayal et al. 2012). However, the transition mechanism between circular and linear genome architecture in animals remains uncertain.

*Acanthella acuta* (Bubarida, Demospongiae) is a relatively common sponge species found in the Mediterranean Sea and the Atlantic Ocean and was one of the species sampled for our demosponge phylogeny project (Lavrov et al. 2019). Our previous analysis of the *A. acuta* mtDNA shows that is possesses a single, relatively large linear mtDNA molecule. The large size of its mt-genome is attributed to the presence of a large ORF resembling DNA-polymerase and RNA-polymerase, inverted terminal repeats, as well as 3 introns within the genes *cox1*, *cox2* and *rnl*. We believe the mechanism involved in the linearization of the *A. acuta* mtDNA is similar to that in medusozoans, as evidenced by the presence of a DNA-polymerase-like ORF and ITRs. Previous phylogenetic analysis of the ORF and introns suggests they might occur in the *A. acuta* mtDNA as a result of horizontal gene transfer. In order to better understand the linearization of mtDNA, it may be pertinent to learn when such HGT events would have occurred and how they impacted the *A. acuta* mtDNA architecture.

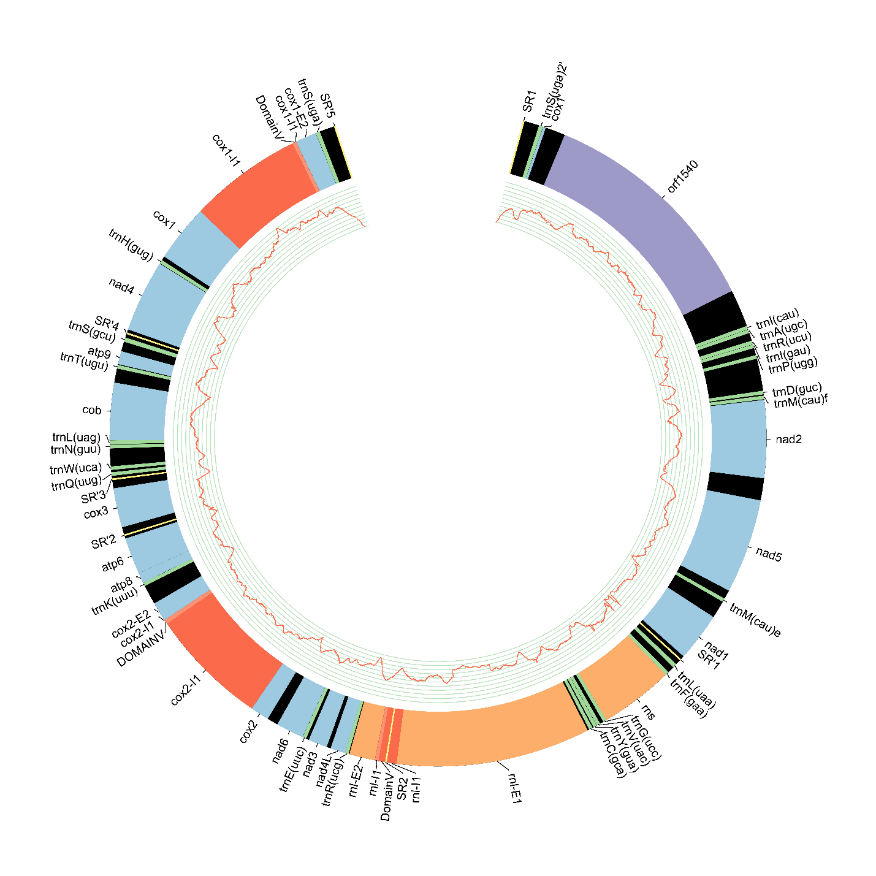


Figure 1 Acanthella acuta mtDNA

**Main questions**

* When did the *A. acuta* mtDNA acquire the DNA-polymerase like ORF?
* When did the *A. acuta* mtDNA acquire the introns within its genes?
* When did the *A. acuta* mtDNA acquire the ITR on its mt-chromosome?
* What can the order of acquisition of these genomic elements tell us about the process of linearization?

**Methods**

*Phylogenetic Tree Construction*

Phylogenetic analysis was conducted using sequences from the NCBI BLAST database for the introns (top 100 hits) and the UniProt BLAST database for *ORF1540* (top 50 hits), along with some sequences reported by Lavrov et al. (2019). Sequences were then filtered using CD-HIT with a similarity threshold of 95%. The filtered sequences were aligned using the MAFFT (v7.511) on auto settings. Aligned sequences were passed through GBlocks (-b5=a -p=n). Phylogenetic trees were constructed using RAxML-NG (v 1.0.2) using the WAG model, and bootstrap values were calculated using 200 replicates.

*Distance and Divergence Time Calculations*

Distance matrices for mitochondrial sequences in *cox1*, *cox2* and the parts of the *ORF1540* were calculated using the *A. acuta* sequences along with at least four of the most closely related sequences as inferred from the phylogenetic trees. The matrices were constructed with FastME, using the mtREV model of mitochondrial amino acid substitutions. FASTA files downloaded from the NCBI database were converted to NEXUS format using EMBOSS.

The divergence time between *A. acuta* and *P. ventilabrum* was reported by Lavrov et al. 2019 as 168MYA. Using the distance between *A. acuta* and *P. ventilabrium* and their divergence time, a rate of change for the *cox1* intron was calculated using the formula:

Assuming that the rate of change for introns within the *A. acuta* mt-genome remains constant, a divergence time was calculated for the *cox2* intron in *A. acuta* and the algae *Coleochaete scutata* and *Nitella hyaline*.

**Results and Discussion**

*Cox1 and cox2 introns*

While initial analysis of both introns suggested that they might have been horizontally transferred into the *A. acuta* mtDNA, Maximum Likelihood reconstruction, with the inclusion of data from Lavrov et al. 2019, of the *cox1* intron reveals that all demosponges included in this analysis form a monophyletic clade, suggesting that it is instead vertically inherited. The *cox2* intron clustered close to the green algae, *Coleochaete scutata* and *Nitella hyaline,* with no other reported similar sequences in any sponges. This supports the possibility of horizontal transfer; however, this could be a result of a lack of representation of Porifera mt-genomes on the NCBI database. The distance matrices are shown in table 1 and 2 for the *cox1* and *cox2* introns, respectively.

A picture containing graphical user interface

Description automatically generated

Figure 2. Maximum likelihood reconstruction of the intron within cox1 of the Acanthella acuta reveals it is most closely related to an intron found in cox1 of Phakellia ventilabrum. All demosponges included in this analysis form a monophyletic clade, suggesting that the intron was vertically inherited by A. acuta. Bootstrap values of 200 replicates are shown.

Diagram

Description automatically generated

Figure 3 Maximum likelihood reconstruction of the intron within cox of the Acanthella acuta reveals it is most closely related to an intron found in cox2 of the algae Coleachaete scutata and Nitella hyalina. The lack of representation of any closely related species to A. acuta could be an artifact of poor Poriferan mtDNA representation in the NCBI database or could suggest that the intron was acquired by A. acuta through horizontal gene transfer. Bootstrap values of 200 replicates are shown.

Table 1 Distance matrix for the cox1 intron in demosponge mitogenomes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ***Acanthella acuta*** | ***Phakellia ventilabrum*** | ***Thymosia sp*** | ***Svenzea flava*** |
| ***Acanthella acuta*** | 0 | 0.0696 | 0.2729 | 0.3950 |
| ***Phakellia ventilabrum*** | 0.0696 | 0 | 0.2992 | 0.4351 |
| ***Thymosia sp*** | 0.2728 | 0.2992 | 0 | 0.3231 |
| ***Svenzea flava*** | 0.3950 | 0.4351 | 0.3231 | 0 |

Table 2 Distance matrix showing distances between the most closely related sequences to the cox2 within A. acuta.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***Acanthella acuta*** | ***Coleochaete scutata*** | ***Nitella hyalina*** | ***Ulva compres*** | ***Ulva ohnoi*** | ***Ulva pertusa*** | ***Ulva torta*** |
| ***Acanthella acuta*** | 0 | 1.3974 | 1.5495 | 1.5911 | 1.7075 | 1.5348 | 1.5668 |
| ***Coleochaete scutata*** | 1.3974 | 0 | 0.3494 | 1.5431 | 1.6536 | 1.5681 | 1.7301 |
| ***Nitella hyalina*** | 1.5496 | 0.3494 | 0 | 1.7617 | 1.7217 | 1.813 | 1.9375 |
| ***Ulva compres*** | 1.5911 | 1.5431 | 1.7617 | 0 | 0.3973 | 0.3743 | 0.4020 |
| ***Ulva ohnoi*** | 1.7075 | 1.6536 | 1.7217 | 0.3973 | 0 | 0.4081 | 0.4972 |
| ***Ulva pertusa*** | 1.5348 | 1.5681 | 1.8130 | 0.3743 | 0.4081 | 0 | 0.3572 |
| ***Ulva torta*** | 1.5668 | 1.7301 | 1.9375 | 0.4020 | 0.4972 | 0.3572 | 0 |

The rate of change between *A. acuta* and *P. ventilabrum* was calculated to be 4.14553516x10-10 substitutions/year. Assuming that the rate of change is constant for all introns within the *A. acuta*, and the distance matrix in table 2, the divergence time between *A. acuta* and *C. scutata* was calculated to be 3.37 109 years.

This divergence time is too far in the distant past to be accurate. It is possible this large divergence time suggests that the assumption that rate of evolution for introns in the *A. acuta* mitogenome is not constant. It is also possible that the rate of evolution within *C. scutate* is accelerated. If the *cox2* intron is in fact not horizontally transferred, this could also lead to a miscalculation of divergence times. A more likely cause of this is that mtREV is based on vertebrate mitochondrial amino acids substitutions, and therefore its application on algae might result in an incorrect distance matrix.

Unfortunately, due to the inaccurate divergence time calculations, it is not possible to make any inferences about the timing of these insertions and their relationship with the linear mtDNA arrangement.

*Extra ORF and ITRs*

The extra ORF was composed of two parts, part 1 that resembled DNA-directed DNA-polymerase, and part 2 that resembled DNA-directed RNA-polymerase. Phylogenetic analysis of each part showed that part-1 is most closely related to the DNA-directed DNA-polymerase protein from the demosponge *Amphimedon queenslandica* (A0A1X7TG59), with which it shares 15.56% of amino-acid identity. Part-2 of the ORF grouped with probable mitochondrial DNA-directed RNA polymerases from *Podospora* and *Neurospora* fungi (Q01521, P33541) and shared with them 15.51% and 18.24% of amino acid identity, respectively. Such differentiated phylogenetic relationships for the two polymerase domains might suggest that *ORF1540* is the product of gene fusion. Due to a lack of well-dated samples, divergence times for these could not be calculated.