Analysis of intron conservation in diatoms, including newly assembled diatom *Psammoneis japonica*

Marissa Ashner¹, Matthew Parks², Matthew Johnson², Norm Wickett²

¹Illinois Institute of Technology, Chicago, IL 60616; email: mashner@hawk.iit.edu; ²Chicago Botanic Garden, Glencoe, IL 60022

**Introduction**

Diatoms are a group of heterokont algae thought to be responsible for approximately 20% of the world’s primary productivity. Diatoms are very diverse, and there are estimated to be over 100,000 species. To date, only two diatom genomes have been annotated and published (Thalassiosira pseudonana and Phaeodactylum tricornutum). These genomes represent the two major orders of diatoms, pennate (lateralisymmetric) and centric (radially symmetric). Nonetheless, a broader diversity of sequenced and annotated diatom genomes would greatly contribute to understanding evolutionary patterns among this immensely diverse group. In these efforts, sequencing both coding and non-coding regions of the genome can reveal variable yet complementary information about the evolutionary history of organisms. This project focuses specifically on introns, a major class of non-coding DNA in the genome.

Introns are non-coding regions of genes that are removed by RNA spliceosomes prior to translation. While the frequency of introns varies in every eukaryotic genome, a commonly noted between intron-rich and intron-poor genomes is that a significant number of intron positions are conserved from distant eukaryotic relatives. This implies that a proportion of introns are retained from earlier stages of eukaryotic evolution. Previous work has demonstrated that the diatom Thalassiosira pseudonana, although intron rich, does not share a multitude of intron positions with ancient relatives. The genome of the pennate diatom Phaeodactylum tricornutum has been assembled at the Chicago Botanic Garden, and the first round of annotations have been made. The focus of this project is to use annotations from Psammoneis, Thalassiosira, Phaeodactylum, and a non-diatom heterokont out-group (Nannochloropsis gaditana) to quantify intron conservation across distantly related heterokonts and other eukaryotic lineages.

**Objectives/Hypotheses**

**Big Picture:** Investigate evolutionary relationships between diverse diatom species

**Smaller Focus:** Sequence and annotate nuclear genomes of more diatom species

**My Objective:** Compare patterns of intron conservation between three diatom species and more distant eukaryotic relatives

**Heterokont Genome Sequencing**

Hypothesis 1: Intron density and conservation in Psammoneis will be more similar to Phaeodactylum because they are both pennate diatoms, as opposed to Thalassiosira, a centric diatom.

Hypothesis 2: Intron density and conservation in diatoms will be more similar to each other as opposed to more distant eukaryotes.

**Methods**

**Data Collection:**

Genome annotations and nucleotide/protein FASTA files for the newly sequenced Psammoneis japonica, published diatoms Phaeodactylum tricornutum and a heterokont out-group (Nannochloropsis gaditana) produced in-house or downloaded from NCBI genome database.

**Finding and Aligning Orthologous Genes between Species for Comparison:**

OrthFinder (v) software used to find single copy ortholog clusters between diatoms Thalassiosira, Phaeodactylum, and Psammoneis, and the out-group Nannochloropsis. MAFFT[7] used to align orthologous nucleotide sequences including introns.

**Intron Position Conservation Analysis:**

Custom Python scripts created to analyze intron position conservation in aligned single copy orthologs. Conservation was defined as two species having the same intron position in regions where 75 nucleotides flanking the intron on each side have at least 32% nucleotide identity.

**Results**

**Figure 1:** Intron density for all genes of diatom species. Green vertical lines represent the average intron density across the genome (introns / gene) for each diatom species. For comparison, average intron density for Nannochloropsis (black), Drosophila (orange), Arabidopsis (purple), and Mus (pink) are also shown as vertical lines.

**Figure 2:** The central Venn Diagram represents a four-way comparison of intron position conservation between all species in conserved regions of single copy orthologs (Psammoneis– gold, Thalassiosira– blue, Phaeodactylum– black, Nannochloropsis– red). Other chord diagrams are 3-way species comparisons, with each line representing an intron conserved between the indicated species (color key same as Venn Diagram) and the species to which the line is drawn. Each outer segment of the chord diagrams represents an artificial concatenation of conserved single-copy orthologous genes.

**Conclusions/Discussion**

**Hypothesis 1:** It was hypothesized that, because Psammoneis is in the same diatom order as Phaeodactylum (pennate), the introns in this species would show similarities to the Phaeodactylum genome. However, while Phaeodactylum has an average intron density of <1 intron per gene, Psammoneis has almost 2 introns per gene, which is higher than the intron conservation from Thalassiosira (Figure 1). This implies that the intron patterns of diatoms in general may not be distinguishable simply by the split between the two major orders of diatoms. The between-diatom Israel intron conservation comparisons show Phaeodactylum shares more of its introns with Psammoneis (23%) than Thalassiosira (16%), which was expected (Figure 2). However, both Psammoneis and Thalassiosira share more of their intron positions with each other (13% and 17% respectively) than with Phaeodactylum (10% and 9% respectively), which supports the results shown in Figure 1. Additionally, Phaeodactylum has a higher frequency of unique introns which might imply that if the lack of introns in the diatom came from an intron loss event, these lost introns are likely to have had unique positions.

**Hypothesis 2:** The average intron densities of all three diatoms are < 2 introns per gene, 0.2-0.5 > that of more distant eukaryotic ancestors Arabidopsis, Mus, and Drosophila (Figure 1). The intron density of Nannochloropsis, also a photosynthetic heterokont, is similar to the diatoms; however, Nannochloropsis shares few intron positions with any of the three diatoms, while between-diatom species comparisons show higher levels of conservation (Figure 2).

Overall, our analysis has demonstrated that intron conservation decreases with an increase in divergence time. By including Psammoneis as a second pennate diatom, we can conclude that the Psammoneis–Phaeodactylum lineage has likely experienced loss of novel introns. As Psammoneis and Phaeodactylum are in different sub-orders of pennate diatoms (angulates and atazolic, respectively), these results suggest that shallower levels of the phylogeny may be explored to more fully address questions of intron conservation and other patterns of evolution. The centric diatoms can also be divided into radial and polar sub-orders, and investigations within this division should also explore more diversity in diatom traits. Further research including more representation in each order could tell us whether or not our results are generalizable for these major groups.

**Acknowledgements**

We’d like to thank NSF–REU grant DBI-1461007 and NSF Award DBE-1353152 for support. Additionally, I would like to thank my mentors, Matt Parks, Matt Johnson, and Norm Wickett, my research partner Nina Denne, and the Chicago Botanic Garden for their time and assistance with this research project.

**References**