Dynamic evolution of inverted repeats in Euglenophyta plastid genomes

Anna Karnkowska[1], Matthew S. Bennett[2], and Richard E. Triemer[2]

1. *Department of Molecular Phylogenetics and Evolution, Biological and Chemical Research Centre, Faculty of Biology, University of Warsaw, Poland*

2. *Department of Plant Biology, Michigan State University, USA*

Photosynthetic euglenids (Euglenophyta) constitute a monophyletic subclade within euglenids. Their plastids enclosed by three membranes arose as the result of the secondary endosymbiosis between a phagotrophic eukaryovorous euglenid and a *Pyramimonas*-related green alga. Euglenophyta is split into two major groups: predominantly marine Eutreptiales and freshwater Euglenales. Euglenales are divided into two families: Phacaceae and Euglenaceae. Despite the fact that many Euglenaceae cpGenomes have been characterized recently, surprisingly little is known about the chloroplast genomes of its sister family Phacaceae.

To achieve better understanding of the evolutionary dynamics of Euglenophyta genomes we sequenced plastid genomes of eight taxa of the family Phacaceae belonging to all three genera classified in this family. A comparative analysis of the Phacaceae indicated considerable diversity in the evolution of their plastid genomes. Genes were arranged in seven clusters in all investigated taxa, and the order of clusters was conserved within the genus. Gene content was well conserved, except for maturases *mat2* and *mat5*. The observed diversity of intron number and presence/absence of maturases corroborated previously suggests a correlation between the number of maturases in the cpGenome and introns' proliferation. Surprisingly, taxa belonging to *Discoplastis* and *Lepocinclis* contained two full inverted repeat (IR) regions containing rDNA operon that are absent from the other Euglenales. By mapping the presence/absence of IR on the obtained phylogenomic tree, we reconstructed the most probable events in the evolution of IRs in the Euglenophyta. Our results suggest that most likely IR regions were most likely lost three times in the evolution of this group.

Our study highlights the highly dynamic nature of the Euglenophyta chloroplast genome, in particular with regards to the IR sequence that underwent losses repeatedly. However, it is necessary to analyze additional taxa from the basal lineages to fully understand the history of the chloroplast genome in the Euglenophyta.