

Plastids are cellular organelles, present in plants and numerous, often distantly related groups of algae. These organelles emerged as a result of endosymbiosis with cyanobacteria and enabled the eukaryotic cells to perform photosynthesis. Plastids carry their own genetic material, which is most often organized into a single circular DNA molecule. Despite being vastly spread across a large portion of the tree of life and, in consequence, among a variety of hosts with different evolutionary history, a majority of plastid genomes exhibits relatively high extent of conservation in terms of the genetic contents and structure. While protein-coding genes are generally present in plastid genomes in single copies, the ribosomal operon, consisting of two genes encoding the small (16S) and large (23S) ribosomal subunit, is usually present in two copies with opposite orientation. The sections of the plastid genomes that contain this operon, as well as two or more tRNA genes and, in some cases, a number of protein-coding genes, are the inverted repeats (IR). These repeats divide the remaining part of the plastid genome into a small (SSC) and large single copy region (LSC), creating a configuration called the quadripartite structure.

Despite the increasing availability of total genome sequencing technology and growth of global interest in the field of genomics, there are still many unsolved problems regarding plastid genomes, their functioning and evolutionary dynamics. Among the major challenges in plastid genome research are explanation of prevalence of conservation of the quadripartite genome structure among a majority of plastids and characterization of contents of the ancestral plastid genome for the plastid-bearing groups. The first of these problems has been addressed in recent literature: a hypothesis states that the inverted repeats play a role in mutation repair, which is supported by an observation that the mutation rate is lower in the genes located within IR than in single-copy genes. It has not been investigated, however, if the same regularity can be observed in non-typical variations of the ribosome-containing repeat regions, such as tandem repeats, where the ribosomal operon is present in consecutive copies with the same orientation. This issue, along with the second of the aforementioned problems, is to be addressed in this project.

The aims of the presented project are to investigate the tandem repeats' potential role in mutation repair, to examine how non-typical plastid genome structures might have formed, and to reconstruct the structure of the ancestral plastid of euglenids. To accomplish these goals, complete plastid genome sequences from a certain set of euglenid strains need to be obtained, which will involve the use of next-generation DNA sequencing techniques, as well as a pipeline of cutting-edge computational tools that will enable successful assembly and annotation of such unusual genomes and, in the following stages of this work, accurate comparison of rates of evolution in the genes of interest and reconstruction of ancestral states.

We have selected the euglenids (Euglenophyta), a lineage of unicellular flagellate algae that acquired its plastid from a chlorophyte alga in an event called secondary endosymbiosis, as our model for plastid genome evolution. The main reason behind this choice is that it is an extensively investigated and well-sampled group of organisms, which provides an abundance of reference data. We believe that the results of this project will bring the functioning of the plastid genomes and their evolution to a better understanding, which could encourage further research in this topic in different model organisms and may be an important contribution to constructing a universal model of evolution of organellar genomes.