Annotation-directed draft genome of the non-model species *Euglena hiemalis.*

Magdalena Płecha[1], Halszka Walkiewicz[1], Natalia Gumińska[1], Anna Karnkowska[1], Bożena Zakryś[1], Rafał Milanowski[1]

1. Department of Molecular Phylogenetics and Evolution, Institute of Botany, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, ul. Żwirki i Wigury 101, 02-089 Warsaw, Poland

Euglenids are a diverse group belonging to the Euglenozoa, comprising of primary (*Rhabdomonas*) and secondary (*E. longa*) heterotrophs, and phototrophs (*E. gracilis, E. hiemalis*). Although, euglenids phylogeny has been studied extensively, so far Whole-Genome Sequencing (WGS) has been limited solely to *E. gracilis.* Deepening the knowledge of other euglenids genetic composition (including organellar sequences) shall particularly bring insights into the complexity of gene content and structure. Implications for atypical introns analyses should be parallely approximated. Presented study is a part of the *E. hiemalis* and *E. longa de novo* WGS project, which additionally includes the *E. hiemalis* transcriptome sequence discovery (to compare it to the genome). Apart from seeking out for the novel genetic regions and their regulation mechanisms, we would like to track the general patterns of introns distribution, their origin and types. Total DNA and RNA were isolated from *E. hiemalis* (CCAP 1224/35) culture. A PacBio and Illumina HiSeq DNA pair-end, likewise polyA-selection RNA libraries were constructed. Reads were subjected to quality and contamination check, then trimmed. Hybrid genome assembly was performed with DBG2OLC, mapped against short genomic and transcriptomic reads. Organellar contigs were search by using BLASTN/X. Errors correction, scaffolding and *ab initio* gene prediction (supported by single-copy orthologs analysis and functional annotation with BLASTP, HMMER3) were performed. The hybrid assembly resulted in 34,258 contigs. The total length of the assembly is 0.61 Gbp with N50 equal to 21,268 and average contig length 24,779 bp. Roughly 97% of genomic and 91% of transcriptomic reads were mapped to the assembly. Initially, the chloroplast and mitochondrial genomes were annotated. Further, nuclear genes' models, employing the presence of atypical introns, were developed and implemented in the process of existing genes prediction along with their functional annotation. In the closest future we will be investigating the evolution of atypical introns.