

České Budějovice, 15th of May, 2018

Reviewer report of Bachelor thesis by Tereza Faitová

Thesis title: **In silico characterization of the plastid proteomes of *Chromera velia* and *Vitrella brassicaformis***

It was my pleasure to read and review the Bachelor thesis written by Tereza Faitová. The thesis is written clearly and with very good English and is, for the most part, of appropriate depth. There are a few typographical errors that I have found, but only in the amount expected in such a document. It consists of 64 pages of text and contains 120 references.

The thesis concerns two photosynthetic algae *Vitrella brassicaformis* and *Chromera velia*. These are closely related to Apicomplexa parasites, which makes them extremely important for our understanding of the origin and evolution of this obligatory parasitic group. The author uses sophisticated bioinformatics analyses to predict organellar proteomes of these two algae. One of the results is an extended database of proteins putatively targeted to plastids in *Chromera* and *Vitrella*, which is an extremely useful resource for subsequent research and other scientists.

The amount of work and results presented in Tereza's thesis, in my opinion, more than surpasses the requirements for a good bachelor's thesis. I am, therefore, very pleased to recommend this **high quality Bachelor thesis for acceptance at the University of South Bohemia.**



Mgr. Martin Kolísko, Ph.D.

Questions and Comments to the author:

1. The author uses Blast analyses as a proxy for phylogenetic analyses to “avoid time-consuming tree calculation and inspection for thousands of sequences”. Given the required results and thresholds this is mostly appropriate for the thesis; however, I would like to know how the author would proceed to construct trees for thousands of proteins – i.e. how she would design the “algorithm”.
2. The author states on page 46: “To do so, we set an e-value threshold for the hits so that they were discarded when their e-value was 100 orders of magnitude larger than the e-value of the query to itself.” Query to itself will commonly have e-value 0.0. What was the threshold set to in such cases – e^{-100} or even smaller? If so, that might explain the high number of sequences with no close hits. Would it be possible to use a script or algorithm to analyse e-values of several top hits and “compute” a threshold specific to each analysed sequence?
3. On page 6 it is stated: “However, in most apicomplexan parasites the symptoms are not observable, nor fatal, unless affecting immunocompromised patients, which is the case of Cryptosporidium infection.” This implies that a healthy individual will not be affected when they are infected with Cryptosporidium – Is that true?
4. Are there any lineages that are expected/known to have lost their plastids or mitochondria completely (and if so, which lineage(s))?