Dr. Ansgar Gruber
Laboratory of Evolutionary Protistology
Biology Centre, Czech Academy of Sciences
Institute of Parasitology Branišovská 1160/31
370 05 České Budějovice
Czech Republic

e-mail: ansgar.gruber@paru.cas.cz
Phone +420 387 775 428

Report on the doctoral thesis of Mgr. Anna M. G. Novák Vanclová

In the thesis entitled "Evolution of euglenid plastid proteome", doctoral candidate Mgr. Anna M. G. Novák Vanclová presents the results of detailed analyses of plastid proteins, plastid protein targeting pathways and plastid proteomes in euglenophytes, represented by the photosynthetic *Euglena gracilis* and the non-photosynthetic *Euglena longa*. This has implications for our interpretation of the evolutionary history of plastids in eukaryotes, for our understanding of the cellular metabolism of euglenophytes, and last but not least for the biotechnological exploration of euglenophytes.

The thesis starts with well-written and suitable introductions to euglenophytes and their relatives, to plastids in general (covering their evolutionary origins, diversity, structure, genetics, and metabolism), and to the methods used in plastid proteomics. Three aims of the theses are then succinctly stated, these are (i) an annotation of a plastid proteome of *Euglena gracilis* in order to aid understanding of its metabolism, (ii) a reconstruction of the plastid protein import pathways in euglenophytes from transcriptomic and proteomic data and (iii) characterization of the targeting signals of nucleus encoded proteins that are targeted to the plastids in euglenophytes.

The actual results are presented in the form of four peer-reviewed publications (one is listed as "under review" in the submitted thesis, but has passed peer review in the meantime as can be seen in the online early content of the journal). The publications are listed along with a clear description of the contributions of Mgr. Anna M. G. Novák Vanclová to each of them.

The list of publications is followed by a summary of the results of the publications, with an emphasis on the contributions by the doctoral candidate. This summary also connects points raised in the introduction with the results of the manuscripts that are part of the thesis, and with other scientific literature. From the summary it becomes clear that the three aims of the study stated earlier are achieved. Finally, the publications are appended.

Vanclová et al. (doi: 10.1016/bs.abr.2017.06.008) provide a comprehensive overview over the euglenophytes and their plastids. The book chapter covers all aspects of plastid biology and presents the respective state of the knowledge. In addition to solid introductions to the open questions that are addressed in the other publications (especially plastid protein targeting, intracellular mapping of metabolic processes and loss of photosynthesis/plastid reduction), the chapter also contains sections on plastid morphologies and plastid genomes of euglenophytes.

Záhonová et al. (doi: 10.1038/s41598-018-35389-1) analysed the transcriptomes of the non-photosynthetic species *Euglena longa* (compared with published transcriptomic data of *Euglena gracilis* and other euglenophytes). One striking finding that was contributed by Mgr. Anna M. G. Novák Vanclová is the absence of TOC and TIC (translocon on the outer/inner chloroplast membrane) complexes from the investigated euglenophytes. TOC and TIC complexes are found in most plastid containing organisms and mediate transport of proteins across the plastid envelope (or the inner two plastid envelope membranes in cells with complex plastids).

Ebenezer et al. (doi: 10.1186/s12915-019-0626-8) present a draft genome, complemented with transcriptome and proteome data of the photosynthetic *Euglena gracilis*. The publication contains a predicted set of plastid proteins (1902 of the estimated total of 36000 protein coding genes) that were screened for the presence or absence of metabolic pathways and for their phylogenetic affiliations.

Novák Vanclová et al (doi: 10.1111/nph.16237) present an experimentally determined plastid proteome (via cell fractionation followed by mass spectrometry), which is analysed with respect to metabolism of the organelle, phylogenetic origin of the genes, and cell biological aspects like vesicle transport and protein transport to the organelle. Furthermore, the properties of the targeting pre-sequences of those plastid proteins for which the N-terminus is complete are thoroughly characterized.

Here are what I think are the strong points of the thesis:

- The introduction in form of a published book chapter, first-authored by the candidate is of very high quality.
- The experimental determination of a plastid proteome is far from being a standard procedure. The work is pioneering, and at the same time carefully conducted and presented in great detail.
- The amount of data analysed is very impressive.
- The experimental results of the plastid proteome determination are a breakthrough for our understanding of plastid metabolism in *Euglena gracilis*, and the characterization of the targeting signals of nucleus encoded plastid proteins provide a precious resource for future studies of euglenophytes.
- The amount and proportion of results published in peer-reviewed publications is very high. Mgr. Anna M. G. Novák Vanclová was first author of one book chapter and one journal article (in New Phytologist, one of the highest ranking journals in plant sciences), and co-author of two more peer-reviewed publications (in the highly reputable journals BMC Biology and Scientific Reports). The publications combine nicely into a rounded body of work.

And here are what I think are the weaker points of the thesis:

- I see the challenge of providing an introduction "on top" of a published chapter that provides an excellent introduction itself (Vanclová et al., doi: 10.1016/bs.abr.2017.06.008), and I appreciate the additional aspects contained in the introduction of the thesis (like the introduction to the evolution of complex plastids in other organisms than euglenophytes). However, I think that the introduction could have been shorter, with a part of the presented information moved to a general discussion, which is not part of the thesis (see other point).

- While the summary of the published chapters is appropriate and easy to follow, it stays close to the published parts of the thesis. I think that a general discussion with a more distant look at the achievements of the thesis and possibly more speculative elements or an outlook to future possibilities, might be interesting as well.

Overall, in my opinion, Mgr. Anna M. G. Novák Vanclová conducted comprehensive studies on an important topic at the intersection of physiology, cell biology and evolutionary biology, and submitted a thesis that very well deserves to be accepted.

Here are some points I would like to discuss at the defence of the dissertation:

- *Cluster analyses of algal groups and traits.* Figure 3 (Introduction, page 39) provides a graphical summary of plastid traits as they are found in the different groups of algae. The groups of algae are labelled on the left side of the figure, along with a cladogram-like graphic that connects them. I was wondering if this graphic is a result of a cluster analysis according to the traits displayed in the figure, or if it is extracted from other published phylogenies? Furthermore, would it be possible to perform a cluster analysis of the traits according to their distribution among the groups of algae? And what might be concluded from such a cluster analysis, especially if compared to the relationships of the different groups of algae?
- *Origin of targeting peptides.* The transit peptide domains identified in bipartite plastid targeting signals of euglenophytes resemble transit peptides found in organisms in which the TOC translocon is present (which is the vast majority of plastid containing eukaryotes), however, no TOC translocon has been identified in the investigated euglenophytes. If indeed plastid targeting in euglenophytes works without a TOC translocon, then the similarities between the transit peptides of euglenophytes and other algae would be due to convergence. One hypothesis on the origin of transit peptides, suggested by Wollmann (doi: 10.1111/tra.12446), is, that they evolved from antimicrobial peptides, which target the endosymbiont/organelle progenitor. Would it be conceivable that a similar process as the one suggested by Wollmann (doi: 10.1111/tra.12446) took place in euglenophytes? Especially with respect to the presence of a second set of genes for iron-sulfur cluster assembly that might have been acquired from Chlamydiae in euglenophytes, as reported by Novák Vanclová et al. (doi: 10.1111/nph.16237)?
- *Horizontal versus endosymbiotic gene transfer.* Multiple new instances of postulated horizontal gene transfers at the origin of plastid-targeted genes of euglenophytes have ben identified in the three journal articles that are part of this thesis. Novák Vanclová et al. (doi: 10.1111/nph.16237) mention that this supports the "shopping bag" (Larkum et al., doi: 10.1016/j.tplants.2007.03.011) or "red carpet" (Ponce-Toledo et al. doi: 10.1111/nph.15965) hypotheses. The red carpet hypothesis is built on the abundance of sequences of red algal origin in organisms with complex plastids of green algal origin. Interestingly, there is also a point to the "converse" case; organisms with complex plastids of red algal origin contain many sequences of green algal origin, which led to the hypothesis of an preceding cryptic endosymbiosis with a green alga in those organisms (Moustafa et al., doi: 10.1126/science.1172983). I think that one explanation of these phenomena is offered by the "limited transfer window hypothesis" (Smith

et al., doi: 10.1093/gbe/evr001), which postulates that endosymbiotic gene transfer is more likely to happen in organisms that contain multiple plastids or multiple endosymbionts. A block of endosymbiotic gene transfer might then increase the proportion of proteins encoded by horizontally acquired genes in the plastid proteomes. How does this relate to the cell and plastid morphologies observed in euglenophytes (figure 1 of Vanclová et al., doi: 10.1016/bs.abr.2017.06.008)? I was also wondering if the high abundance of group I and II introns and the presence of group III introns in plastid genomes of euglenophytes (section 4 of Vanclová et al., doi: 10.1016/bs.abr.2017.06.008) might prevent the expression of genes transferred from the plastid to the nuclear genome in euglenophytes. Could this additionally account for the high abundance of horizontally acquired genes in euglenophytes?

- *What comes next?* The results presented in the thesis are very interesting in their own right, and in addition also provide starting points for further experimental or *in silico* work (like an analysis of *Rapaza viridis* with the same methods, or developing genetic tools for euglenopytes, ...). If there would be no restrictions with respect to resources, what would be the favourite follow-ups of Mgr. Anna M. G. Novák Vanclová?

České Budějovice, November 30 2019

Ansgar Gruber