# THE UNIVERSITY OF BRITISH COLUMBIA

**Professor Beverley R. Green,**
**Department of Botany**
#3529 – 6270 University Boulevard
Vancouver, B.C. Canada V6T 1Z4
Tel: (604) 822-2349
Fax: (604) 822-6089
E-mail: brgreen@interchange.ubc.ca

October 14, 2009

Report on Ph.D. Thesis of Ms. Kateřina Jiroutová

This thesis concerns the evolution of enzymes of the tryptophan biosynthetic pathway, including contributions to the annotation of the *Phaeodactylum tricornutum* genome, and some experimental work on the fate of a gene that is in the process of moving from the chloroplast to the nucleus.

Overall, the thesis is well-written in a straight-forward fashion. It is organized in manuscript format, with two papers already published and another ready for submission to a scholarly journal.The two published papers have appeared in high profile journals. The amount and quality of the research would certainly be considered suitable for the award of the Ph.D. degree at the University of British Columbia.

Following the format of an External Examiners report for my university, here are some specific comments and some questions that could be asked of the candidate during the oral examination, if they haven't already been covered by the examining committee.

1. The Introduction is quite thorough in giving the background of the research: the chromalveolate theory and biology of diatoms, and what is known about the evolutionary origins of several of the major metabolic pathways studied in diatoms by the supervisor and others. There are a lot of mistakes in English, but that is not surprising considering that the candidate's native tongue is Czech. The candidate is to be commended for her efforts.

Question: If the Introduction were to be the starting point for a general review article on the evolution of metabolic pathways, what other pathways have been studied and should be included? Is any sort of general picture emerging?

2.Tryptophan biosynthesis paper: This is a very solid piece of work, with many good controls. It illustrates both the strong and weak points of various methods of phylogenetic analysis. It is an important paper for showing that some important pathways found in the chloroplast were not acquired from the cyanobacterial or red algal ancestors, but from the host.

Note: in both the published paper and the thesis, a phosphate group has gone missing from N-5-phosphoribosyl anthranilate.

Question: InGPS seems to be the only enzyme with apicomplexan homologs. If tryptophan is synthesized in the cytosol in apicomplexans, what else could this enzyme be doing in the plastid?

Question: A number of papers have now shown that the transit peptides are often predicted as mitochondrial rather than chloroplast targeting signals by prediction programs such as TargetP and SignalP. What are some possible reasons for this? Do you have some suggestions for development of better algorithms?

3. Phaeodactylum genome paper: The candidate's contribution to this work was in annotating all the genes for the tryptophan biosynthetic pathway (and possibly some other pathways?). These annotations are now part of the Genbank record. This shows that she is already a member of a scientific community. The success of large expensive genome projects is largely determined by such contributions from the scientific community.

4. A gene in the process of endosymbiotic gene transfer in *T. pseudonana* (manuscript). This paper provides experimental evidence that both the plastid and nuclear copies of psb28 are transcribed, and that the nuclear-encoded protein is synthesized and correctly targeted to the plastid. Transformation of the nuclear gene-YFP construct utilized another diatom, *P. tricornutum* for technical reasons, but it is also interesting because *P. tricornutum* only has a plastid copy of the gene.

Question: *T. pseudonana* is a centric diatom, and *P. tricornutum* a pennate. These two groups are believed to have diverged at least 80 Mya. But the recently released genome of *Fragilariopsis cylindrus*, another pennate, has both nuclear and plastid copies. What does this tell us?

In summary, I find this to be a good solid thesis and worthy of the award of the Ph.D. degree.

Yours truly,

Beverley R. Green

Beverley R. Green, Ph.D.
Professor Emerita of Botany

Institute of Microbiology
Academy of Sciences of the Czech Republic
Department of Autotrophic Microorganism
379 81 Trebon - Opatovicky mlym
Czech Republic

# Review of the doctoral thesis of Katerina Jiroutová 'The origin and localization of selected metabolic pathway in marine diatoms'

Diatoms are beautiful, which used to be a good reason to explore these organisms by microscopical pioneers long time ago when just aesthetic view was enough for scientists to pay attention on a object. Today we know diatoms to be extremely important for life on this planet producing perhaps twenty percent of oxygen we breath and the beautiful silicon frustule of diatoms is taken by modern researchers as one of key factors of their ecological success. However, even the importance of diatoms is apparent for quite long time, the biology of diatom cell is surprisingly poorly understood. It should be noted that this is the true for all 'exotic' algae, where exotic means all algae except the model green alga *Chlamydomonas reinhardi*. However, situation is quickly changing now, particularly after genomes of the first diatom species *Thalassiosira pseudonana* became available in 2004. Last year the genome of *Phaeodactylum tricornutum* was published as the second genome of diatom organism. Katerina Jiroutová participated in this project and she is co-author of the Nature paper that is included as a part of her PhD thesis. Thanks to this sequencing effort it is starting now to emerge a new picture of diatom – the picture of the cell undergoing very fast evolution and possesses unique mosaic of genes and metabolic pathways unusual for other photosynthetic organisms. Diatom complex metabolism thus appears to be very important factor ensuring their frequent dominance over other autotrophic microorganisms.

The thesis is focused on the origin of enzymes involved in biosynthesis of aromatic amino-acids in diatoms and on the gene transfer from plastid to nucleus. The work is based mostly on *in silico* approaches, however the author utilized in her work also recently developed genetic tools for diatom *Phaeodactylum*. In the first part of the thesis author summarized in detail important aspects of the diatoms evolution discussed mostly in the frame of evolution of the whole group of Chromalveolata. The second part 'Results' consists of a collection of two publications and one manuscript. All chapters are well written following the current state-of-the-art approaches and knowledge. In my

opinion, some chapters, e.g. origin and protein import into secondary plastid and evolution of selected metabolic pathways, could be published after some supplementation with additional data in the form of mini-review in good quality journals. In the chapter 6.2. 'Biosynthesis of photosynthetic pigments', there is an interesting note that sequenced diatoms lack the light-dependent protochlorophyllide oxidoreductase enzyme (POR). In my opinion the absence of the light-dependent POR could be related to the fact that chlorophyll *c* synthesized by diatoms together with chlorophyll *a* is a competitive inhibitor of this enzyme. Is there any evidence that diatom genomes coding for so called dark POR, multi-subunit enzyme related to nitrogenase and presented in cyanobacteria and green algae? My other question also concerns tetrapyrrole metabolism: is there any suggestion based on *in silico* data that enzymes of this pathway are targeted into mitochondria? Regarding number of membranes in diatoms the transport of heme from plastid to mitochondria seems to be difficult.

Katerina Jiroutová is the first author of one article and one manuscript and the co-author in the important paper describing analysis of the genome of diatom *Phaeodactylum tricornutum.* These papers clearly demonstrate author's experience in the data collection, processing and writing the articles. As two papers were already published in high quality peer-reviewed journals, I focus my review only on the enclosed manuscript entitled 'A gene in the process of endosymbiotic gene transfer'. Here, the authors deal with noteworthy finding that in diatom *Thalassiosira pseudonana* putative Photosystem II assembly factor Psb28 in coded both in the plastid and in the nucleus. Authors follow attractive hypothesis that the nucleus gene resulted from the duplication of the *psb28* gene coded in the plastid and it is actually in process of endosymbiotic gene transfer. This is supported by a detail phylogenetic analysis of the *psb28* gene and by experimental data demonstrating the targeting of the nucleus coded Psb28 protein into plastid. A bit unexpected is the accumulation of fused nuPsb28-EYFP protein inside the pyrenoid; probably in a soluble form. This location does not seem to reflect the function of Psb28 protein, which was demonstrated to be associated with membranes. Is it known where is localized the non-fused EYFP protein targeted into the *Phaeodactylum* plastid? Generally, the topic of the study is very interesting and the data are thorough done and discussed. However, the title of the paper appears to be a bit preliminary. For more significant paper it would be important to found evidences about the coexistence of both putative Psb28 proteins in the plastid e.g. by specific antibodies or same mass-spectrometry approach. Of course, even the presence of both proteins in plastid does not

mean that they share the same function and number of different amino-acids residues in both proteins might indicate rather distinct roles. Nonetheless, to query the question, whether both proteins (if both exist in the cell) operate as the Psb28, is very hard task taking into account that the function of Psb28 in chloroplast is quite unclear. So, let's take this last note as a dream of reviewer and rather speculative than factual comment.

Finally, I would like to emphasise that all my comments do not reduce the very good scientific quality of the presented thesis written in good English and containing a minimum of formal errors. The thesis fully complies with all general demands for the doctoral thesis and I fully recommend it for the defence.

Here my questions are summarized:

1. Is there any evidence that diatom genomes coding for so called dark POR, multi-subunit enzyme related to nitrogenase and presented in cyanobacteria and green algae?
2. Is there any suggestion data that some enzymes of tetrapyrrole pathway are targeted into diatom mitochondria?
3. Is it known where is localized the non-fused EYFP protein when targeted into the *Phaeodactylum* plastid?

Trebon, 10.10. 2009

Roman Sobotka, PhD

☎ +420 384 340 434 ☎Fax +420 384 340 415

In Prague 16th of October 2009

Review of the Ph.D. thesis of Kateřina Jiroutová:

**THE ORIGIN AND LOCALIZATION OF SELECTED METABOLIC PATHWAYS IN MARINE DIATOMS**

The Ph.D. thesis of Kateřina Jiroutová consists of a review on 39 pages, 3 original paper (one in the form of submitted manuscript) and brief conclusions. The thesis is written in good-level English. The review consists of 6 parts: Introduction, Frustule of diatoms, Origin of the diatom complex plastid, Protein import into the complex plastid, Diatom genomics, and finally Evolution of metabolic pathways. The review generally reads well and displays the knowledge of the topic and abilities to sum relatively diverse primary literature into a compact text. On page 8 is probably a part of the text missing, as there is a free space and the text on the next page does not seem to connect to the previous page. The last chapter about the metabolic pathways is relatively difficult to digest. This part if very complicated and technical with the descriptions of all the enzymes involved and their confusing relationships and requires quite experienced reader with blazing mind. In general, I think, the last part could be published as a review paper in the future, but before I would suggest making the text as simple and transparent as possible.

My particular questions and comments to the review are following?

1) On Page 8 you state: *„Supergroup Chromalveolata was postulated on the basis of molecular phylogenetic analyses that unite particular members of these morphologically disparate lineages among alveolates and chromists (figure 3 and 4)."* Could you name a phylogenetic analysis that did so?
2) Page 10, *"Sixteen genes of plastid origin were identified in ciliate genomes, which suggest that even ciliates could have photosynthetic ancestors..."* Is there an alternative explanation for this observation?
3) Page 10, *"...Apicomplexa represent in addition to ciliates, dinoflagellates and chromerids the fourth group belonging to alveolates"*. Please do not forget on *Perkinsus*, *Colpodella* and *Colponema.*
4) Page 11, *"In other chromalveolates [besides cryptophytes] ... the nucleomorph is not present."* Here I would like to draw your attention to picobiliphytes (a member of the haptophyta+cryptophyta clade) that likely contain a nucleomorph.
5) Page 11, *„The outermost membrane of the plastid is connected to the endoplasmic reticulum of the host in haptophytes and stramenopiles..."* Is the original membrane really connected to ER or has it been lost and replaced by ER? By the way, the same applies also to cryptophytes.
6) Page 14,"*...secondary plastids are surrounded by more than two membranes, usually three or four."* Is there a secondary plastid with more than four membranes?
7) Page 14, "*The targeting signal thus consists of at least two parts, the ER signal peptide .....and conventional chloroplast transit peptide...."* Are there targeting signals with more than two parts?
8) Page 16, Could you please explain, what you think is the argument that "*the hypothesis in which Cavalier-Smith (1999) proposed relocated Toc complex as a translocon within the second outermost membrane is correct."*?
9) General question to the protein transport. I heard a murmur that the transport between the 2nd and 3rd outermost membrane of the complex plastid may be performed by vesicles budding from the 2nd a fusing with the 3rd membrane. What is your opinion on such hypothesis?

10) Page 17, "*...phytoplankton in pelagic and benthic habitats.*" Benthic organisms are in my opinion not plankton.

The first attached paper was published in Journal of Molecular Evolution. The aim of the paper is to disentangle the evolutionary origin of tryptophan biosynthesis pathway in Stramenopiles. As the title says, most enzymes originated from the nucleus of the secondary host. The paper contains a lot of precise work; however, I would like to cast doubt on one-sided interpretations of some results.

1) Let me consider the case of APRT enzyme. In APRT tree, all eukaryotes formed a clade. Stramenopile sequences grouped with green and red algae in the way that red alga – *Cyanidioschyzon* – broke into Stramenopiles with bootstrap 52. Authors take this as evidence that the stramenopile gene originated from red algal endosymbiont. Let me propose alternative scenario. Bootstrap 52 is not high, *Cyanidioschyzon* is famously strange sequence-wise and I bet that the topology in which it groups with the green algae would not be rejected by tests (not performed). If *Cyanidioschyzon* moved to the green algae the topology would be in absolute agreement with the eukaryotic phylogeny as we understand it nowadays. Therefore, no endosymbiotic gene transfer need to be postulated and stramenopile could inherit APRT by simple vertical descent from their eukaryotic ancestors – in other words, APRT would originate from the secondary host nucleus.
2) In case of PRAI tree, the phylogeny looks quite similar to APRT with one difference Stramenopila are sister to Fungi. You comment this phylogeny in the following way: "*This suggests that plants and algae have retained their original eukaryotic gene, while stramenopiles appear to use PRAI derived from the secondary host*". I do not understand the sentence because I do not understand the difference between *retaining original eukaryotic gene* and *deriving the gene from secondary host*; secondary host gene is the original eukaryotic gene, isn't it? Could you explain it?
3) In many enzyme phylogenies I see a close association of Fungi and Stramenopiles. In most cases (AS, PRAI, TSα, TSβ) Fungi are or could be (because the nodes are poorly supported) nested within Stramenopiles as a sister group to oomycetes. Richards et al. (2006) noticed that many fungal genes were laterally transferred to oomycetes. Could you results suggest that the tryptophan biosynthesis pathway was laterally transferred in the opposite direction from oomycetes to fungi?

The second paper was published in the most prestigious journal – Nature – and presents the second genome of diatom - *Phaeodactylum tricornutum*. The comparison of the two diatom genomes with other eukaryotes brought up many interesting findings. One of the most surprising for me is the high proportion of genes acquired by lateral gene transfer from various (supposedly food) bacteria in both diatom genomes in comparison to other eukaryotes. Could you speculate about the reason for this? What in the diatom biology could make them so prone to LGTs?

The third attached work is a submitted manuscript. It investigates the cases of gene for protein Psb28. In *Thalassiosira* one copy of the genes is coded in the plastid genome and the other quite altered copy is coded in the nucleus, i.e. this gene is apparently in the process of endosymbiotic gene transfer. Nuclear copy contains targeting presequences and, as the author experimentally showed, it is transported into the plastid. I would like to congratulate to this work that I find very interesting. The history of this gene is worth of further investigation. In context of this study, I would like to ask three questions:

1) Does the nuclear copy contain introns?
2) Did you investigate the flanking regions of the nuclear gene to see if they are derived from plastid.
3) I suppose that the targeting presequences originate by duplication of the existing ones. Would it be possible to trace the origin of the targeting sequences by comparison of their sequence or their 5' flanking region sequence with other targeting sequences in *Thalassiosira* nucleus?

In conclusion, I would like to state that the thesis has in my opinion high scientific quality is concise and well written and clearly fulfils criteria for award of a Ph.D.

Vladimír Hampl