Biology Centre, Academy of Sciences
Branišovská 31, 370 05 České Budějovice, Czech Republic

Review of the bachelor thesis of Vojtěch David entitled „Assembly and annotation of a mitochondrial genome of kinetoplastid protist *Perkinsela*"

Bachelor thesis of Vojta David is a part of a larger project of collaborating laboratories of J. Lukeš (IoP, BC ASCR) and J. Archibald (Dalhousie University, Canada). Vojta's role was to identify contigs corresponding to the mitochondrial genes of the endosymbiont of *Neoparamoeba pemaquidensis* and characterize the RNA editing.

Vojta did really a great job. He was able to identify certain mt genes of *Perkinsela amoebae*-like organism including 9S and 12S rRNAs. Identification of rRNA genes was really tough nut to crack. He discovered that RNA editing takes place also in these genes, which is the first case of editing of rRNA genes. Scientific level of this bachelor thesis is outstanding and can be compared with very good master thesis.

However, the text part of the thesis is in some parts very weak. The main weakness of the bachelor thesis I can see in the first part of the Introduction:

Paragraph "What is *Perkinsela*?" is very confusing and a reader, by my opinion, can't find the correct answer of this question here at all. Please, can you specify the host of *Perkinsela amoebae*? Logically, it should not be (only) *Neoparamoeba pemaquidensis*, as stated in the thesis, that was described seven years after the description of *Perkinsela amoebae*. The alternative name *Ichthyobodo*-related organism does not originate from Morrison et al. 2005 (what is the correct reference?). Can you explain the term *Perkinsela amoebae*-like organism? - it is not clear from the text. I would also expect at least a half of sentence about the phylogenetic relationship of the endosymbiont in the text, not just reference to the tree in Fig. 5. Phylogenetic relationships among closely related amoebae isolates and their endosymbionts have not been investigated in the cited papers - their co-evolution was investigated but only in Dyková et al. 2008, not in the Morrison et al. 2005. Also the reference to Fig. 6 is erroneous in that context - Fig. 6 shows only relationships of *Neoparamoeba pemaquidensis* strains. Furthermore, in the legend to this figure, there is written that the tree shows relationships of the endosymbionts - it shows actually relationships of amoebae strains. The last issue I have with this paragraph is the naming "*Perkinsela*" for the endosymbiont of *N. pemaquidensis* isolate CCAP1560/4. I understand the reason - simplicity, but my opinion is, that in the bachelor thesis, it is not correct to use the genus name of the symbiont of different amoebae as a nickname for the endosymbiont of particular *N. pemaquidensis* strain.

Paragraph "*Neoparamoeba*: the host" is very similar to the previous one. We do not get much information about the *Neoparamoeba* - there is nothing about morphology, ultrastructure or phylogeny. First column is about the disease that amoebae cause, second about the relationship with *Perkinsela* and the third column starts surprisingly with the sentence "The overall cell structure of *Perkinsela* is interesting for several reasons", however we are not even informed about the cell structure of *Perkinsela* (or *Neoparamoeba*) but about the

Telephone No.: +420 38 7775051
Fax: +420 38 5310338
E-mail: bc@bc.cas.cz
www.bc.cas.cz

Biology Centre, Academy of Sciences
Branišovská 31, 370 05 České Budějovice, Czech Republic

division of *Neoparamoeba* spp. and their endosymbionts. Legend to Fig. 6 does not explain the figure. In the Fig. 7, there is not phase contrast but Nomarski DIC.

The second part of the Introduction focusing on the sequencing, genome and transcriptome reconstruction and description of software is noticeably better written with all the details correctly explained.

Results and Discussion part is well done. It is clear that Vojta was able to understand all the methods he used and was able to synthesize the results. I liked many figures included in the text, however, many of them lacking good quality captions.

Notes:
- species names must be in italics also in References
- when cited a paper, there should not be comma between name and year, when cited the author of species description, there must be comma between name and year

General questions
Is it possible that lack of the genes for NADH subunits is just caused by accidental removing together by filtration of bacterial contamination from the raw genomic data? How was the bacterial genes filtered out? Did you screen raw data? Or is it possible that NADH genes are not detected by BLAST search simply because of strong editing? I missed any discussion about that.

Any idea why editing of mitochondrial 9S and 12S genes occur only in *Perkinsela amoebae*-like organism?

Overall, I find the results of the work of excellent quality and I am sure that Vojta is very important member of the lab team although he is very young researcher. I am recommending the thesis to be approved.

In České Budějovice, 24. 5. 2013

Ivan Fiala, Ph. D.

Telephone No.: +420 38 7775051
Fax: +420 38 5310338
E-mail: bc@bc.cas.cz
www.bc.cas.cz

DEPARTMENT OF BIOCHEMISTRY & MOLECULAR BIOLOGY
FACULTY OF MEDICINE, SIR CHARLES TUPPER MEDICAL BUILDING
HALIFAX, NOVA SCOTIA, CANADA B3H 1X5
PHONE: (902) 494-2480 FAX: (902) 494-1355
WWW.BIOCHEM.DAL.CA

DALHOUSIE UNIVERSITY
Inspiring Minds

21 May 2013

ATTENTION: University of South Bohemia, Faculty of Science, Czech Republic.

RE: Bachelor thesis of Vojtěch David: Assembly and annotation of a mitochondrial genome of kinetoplastid protist *Perkensela*

To whom it may concern,

I am happy to comment on the bachelor thesis Vojtěch David, which describes genomic research on a unicellular organism called *Perkinsela*, a 'kinetoplastid' protozoan living inside a single-celled amoeba. The specific aims of the project are to characterize the mitochondrial genome (kDNA) of the organism and determine the extent to which RNA editing takes place. The thesis is part of a larger project to sequence and analyse the genomes of the host and endosymbiont towards the goal of elucidating their biology and evolution.

Mr. David has successfully identified a minimal set of contigs derived from the Perkinsela kDNA and annotated the handful of genes contained therein. The bulk of the results of the thesis cover the interesting phenomenon of U-insertion/deletion editing in protein genes and even in putative 9S and 12S rRNAs. If true this would be the first example of RNA editing in a ribosomal RNA.

The work is definitely cutting-edge. It uses 'next generation' genomic and transcriptomic data, analyses of which present many challenges. It is abundantly clear from the thesis that Mr. David is functioning at a very high level. The English writing is not perfect but is still very good.

In the sections that follow I discuss what I believe to be the strengths and weaknesses of the document.

Strengths:

-Mr. David is to be congratulated for producing such a detailed and comprehensive thesis. This is most impressive for a student at such an early stage of their scientific career.

-The thesis contains ample background on the methodologies used, from next-gen sequencing to assembly to RNA-Seq read mapping. One the whole these sections of the thesis are well written and, to my knowledge, accurate (see below).

-More generally, the introduction to the organisms involved also makes it clear that the student understands their biology and the significance of the work being performed. Mr. David clearly gets the 'big picture'. This is a clear strength of the thesis.

-There are a large number of figures included, which is very helpful for the reader to be able to take in what is being described in the text. Some of these figures (and their legends) could have been improved (see below), but overall I think this is truly excellent.

Weaknesses:

-The Discussion section is relatively short compared to the sections that precede it. There is a lot of detail in the Materials and Methods and Results sections, and by the time the reader gets to the Discussion a lot of the background information provided in the Introduction has been forgotten. The Discussion could be improved by adding a brief section at the beginning of it to reiterate the importance of the organisms and the specific questions being addressed. This would serve as a way to better contextualize the research that was carried out and how the results that have been obtained lead to your specific conclusions about the structure of the kDNA and RNA editing.

-The thesis could have been more extensively referenced. There were several places in the text where I thought that statements needed to be referenced. For example, on page 22, where does the information about the alternative oxidase acquired by HGT come from? As well, in the last sentence of the first paragraph of the Discussion, the text describing the RNA editing pattern in T. borreli should mention the primary literature from which the data came from. This is something to consider when it comes to writing up work for publication.

-In general the figure legends are relied upon too heavily. For example, on Page 11 the section on MUSCLE is very short and simply refers to Fig. 12 as an "overview". But Figure 12 has a lot of detail and one wonders how much of it is necessary? The figure legend does nothing to help the reader understand it.

-Table 1 could have benefited from a more detailed title and more information. What are the genus / species names in the 'isolate' column? Do these strain designations refer to the host or endosymbionts?

-Are both Figures 14 and 16 necessary? That is, the initial and revised work flows?

Specific comments:

-Page 3: "Pan edited proteins"?

-Page 5: "Pulse-field" should be "Pulsed-field"

-Page 13, Materials and methods: "All the data were processed...". Which data? Need to be specific.

-Page 14 top line: ""sequence data for 2 isolates". Isolates of what?

-Page 19, Figure 18. It's not clear what it is about the data shown in the figure that suggests that these are guide RNAs. In general I think a bit more background about the role of guide RNAs in the whole process would have been beneficial.

-Page 21, near bottom. "this particular region undergoes alternative editing". What does alternative mean in this context exactly? More explanation would have been beneficial.

-Page 23. "... so the actual gene content of Perkinsela remains to be established." Which genome? kDNA or nuclear or both?

-Occasionally it is not clear whether the author is referring to genes or proteins. For example on page 23: "The lack of subunits of NADH dehydrogenase both in mitochondria and nucleus...". Clearly in this case it is the lack of GENES ENCODING subunits. It is always best to be fully clear so that there is no chance of misinterpretation on the part of the reader.

Conclusion:

Overall I am delighted with the quality of the thesis and am happy to recommend that it be approved. The research is of top quality and it is clear that the student is fully engaged with it.

Sincerely,

John M. Archibald
Professor and Graduate Coordinator,
Department of Biochemistry & Molecular Biology, Dalhousie University,
Fellow, Canadian Institute for Advanced Research, Program in Integrated Microbial Biodiversity,
Sir Charles Tupper Medical Building
5850 College Street, Halifax, Nova Scotia
B3H 1X5 Canada