

January 26th, 2020

RE: Report on the PhD thesis by Lukáš V. F. Novák

Dear Members of the Examination Committee,

I was a pleasure to read the thesis submitted by Lukáš V. F. Novák. Please find below my comments on this thesis. I am very much looking forward to the examination and to discuss the thesis with the candidate and the other members of the examination committee.

(1) Short overall assessment of the thesis (e.g., scientific merit, originality, experimental design, choice and mastering of the methods, interpretation of results, quality of documentation)

This is overall a well-structured PhD thesis that present data and their interpretation addressing fundamental questions on the diversity and evolution of microbial eukaryote at the genomic level by investigating species members of the Preaxostyla lineage and the phylogeny of specific genes. The thesis presents the biological interpretation of genome sequence and transcriptome sequence data from several species of Preaxostyla and selected gene phylogenies. It led to four publications, one as first author, one as second author and two additional publications as middle author. Overall the quality of the presented document indicate that the thesis is suitable for awarding the candidate the Ph.D. title. This will be further investigated during the examination of the candidate.

The thesis report is structured into three major sections. The first section is made of 88 pages that makes up 7 chapters that include a general introduction presenting the background and rational to the PhD project and its aims, list of publications directly associated with the thesis, a summary of the project and the list of cited publications. The second section is made of the four research articles (main articles only as it does not include supplementary material referred to in these articles) that integrate the data generate by the candidate. These papers are in high profile scientific journals and have already attracted, as of today, a total of 290 citations, highlighting their high quality. The third section consists of two supplements that present (i) an ongoing project (work in progress on additional relevant species broadening the comparative study already published) and (ii) two outreach articles for the general public.

(2) Specific critical comments

tel: +44 (0) 191 208 6963 (direct line)
fax +44 (0) 191 208 7424 (central office)

www.ncl.ac.uk/camb

The introduction provides an overall excellent and broad historical background, including typically overlooked papers such as Cheissin (1965), on the biology of microbial eukaryotes that at one time or another where members of the Archezoa and how the Archezoa hypothesis was originally put forward and, with time, falsified. It also provides the broad background on the genomics and transcriptomics sequence data of a range of microbial eukaryotes and the important insights these molecular data provide into the organisms' evolution of their cellular systems and individual genes with a particular focus on the diversity of mitochondria and metabolic pathways. There is also an important section providing key background information on the group of organisms targeted by the candidate, the *Preaxostyla* (section 1.3).

There was however no mention of proteomics survey of mitochondria and their variants, an important set of data that provide global perspective on the protein complement of mitochondria and their derived organelles, including the mitosomes from *Giardia intestinalis* and the hydrogenosomes from *Trichomonas vaginalis*.

An introduction to genome annotation and phylogenetics was also expected as this represent an important aspect of the thesis of the candidate and justify his co-authorship on the paper by Kamkowska et al. (2016) and the other papers co-authors by the candidate, including his first authored paper Novak et al. (2016). The papers included in the thesis do not provide much details on this - e.g. genome annotation is described in the "Supplementary Experimental Procedures" file, which was not part of the submitted thesis.

Some specific comments relating to the published papers

Kamkowska et al. (2016): This is a paper published in the format of an Article in Current Biology. The annotations of amino acid metabolic pathways and mitochondrial genes in the draft genome of *Monocercomonoides* sp. represents an important and central aspect of this paper, clearly justifying both the inclusion of this important paper in the submitted thesis by the candidate and his co-authorship on that paper. These data strongly support the notion that eukaryotes can lose the mitochondria organelles, in contrast to all previous studies that investigated candidate Archezoans, where one form or another of the mitochondria were always identified. The phylogenetic position of the *Monocercomonoides* sp. clearly indicate that this was a secondary loss and that the absence of the mitochondrion organelle in that lineage cannot represent an ancestral state in the evolution of eukaryotes. I am looking forward to discuss with the candidate some of the important points made in that paper, as well as some technical details pertaining to the sequence analyses and their interpretation.

Novak et al. (2016): This represents a central aspect of the submitted thesis that was specific to the project of the candidate, with both him and his supervisor being listed as the authors who conceived the research. The candidate did in particular contribute to phylogenetic analyses and writing up of the manuscript. This is a very interesting paper that provides very detailed and global analyses of the genes encoding the enzymes of the ADI pathway, clearly expanding the analyses of related published papers.

In my opinion the paper would have benefited from a synthetic presentation of the different inferred origins (such as in the format of a figure combined with a table) for the different analysed genes and sampled species as I found it difficult to get a synthetic picture of these different analyses from reading the text and the available figures. Some details seem also to be missing concerning some aspects of the performed phylogenetic analyses (e.g. AU tests). The paper could have also presented more detailed analyses of the protein sequences of the different enzymes (e.g. alignment of the different enzymes to identify (i) key functional residues and (ii) characteristics of the N-termini). These points represent some of the issues that I am looking forward to discuss with the candidate.

Vacek, Novak et al. (2018): This short paper in the form of a letter, was published in MBE. The candidate is the second author and he contributed to key aspects of the paper, which presents a detailed analysis of the genes encoding the enzymes involved in Fe-S cluster biosynthesis among 16 studied *Preaxostyla* species. It identified the SUF genes and also the absence of any of the gene characteristic of the mitochondrial ISC machinery expanding the analyses for *Monocercomonoides* sp. described by Kamkowska et al. (2016), representing an interesting and important extension of the data and their interpretations published for a single *Preaxostyla* genome. I am looking forward to discuss some of the detailed analyses presented in this manuscript including the mentioned potential contaminant genes and candidate LGTs.

Kamkowska et al. (2019): The fourth paper listed in this thesis was published as an Article in MBE. The authors present an extended annotation of protein coding genes of *Monocercomonoides* sp., expanding the range of biological inferences made in relation to the loss of the mitochondrion organelle across the conserved cellular systems. The candidate contributed to key aspects of this paper, including in particular to the annotation relating to amino acid metabolism, selenium utilisation machinery and calcium homeostasis regulation). I will be interested to discuss with the candidate some of the specific interpretations of the absence/presence of specific genes and how this might inform us on the nature, the origin and the evolution of the first eukaryotes.

(3) More general questions to the defendant dealing with the subject, methods or interpretations that would allow the candidate to show in the reply how he/she is mastering the discipline.

The following questions and topics will be discussed with the candidate

3.1) See specific points identified earlier (under point 2) for the different published papers.

3.2) I cannot recall reading any specific justifications for the evolutionary models used for phylogenetic analyses presented in the papers associated with this thesis. It will be interesting to discuss the importance of this point with the candidate and how this might impact on the interpretations of molecular phylogenies

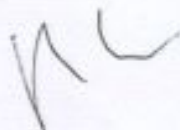
3.3) A number of references are made for specific genes to represent LGT or potential contaminants in some of the datasets analysed by the candidate. I am looking forward to the hear the views of the candidate on how to tackle these different scenarios and differentiate them from each other analytically and experimentally

3.4) Different sequencing technologies have been used for the different projects relevant for this thesis. I am interested to know the opinion of the candidate about the potential implications of these different technologies on the quality of the sequence data and their potential impact on downstream bioinformatic analyses, including the annotations of protein coding genes (included 454, HiSeq2000, PacBio, MinION, HiSeq and MiSeq technologies/chemistries).

More detailed points about specific statements and figures

Some questions will relate to some details of the text or Figures – e.g Figure 3. Could you explain the meaning of the different symbols and what does EMP refers too?

Best Regards,



Robert Hirt - PhD
Professor of Evolutionary Parasitology
E-mail: Robert.Hirt@ncl.ac.uk
Home and Lab page:
<https://www.ncl.ac.uk/medical-sciences/people/profile/roberthirt.html>