DALHOUSIE UNIVERSITY
Inspiring Minds

Biology Department
Life Science Centre
1355 Oxford St.
Halifax, NS. B3H 4J1
902-494-3515 PHONE
902-494-3736 FAX
biology@dal.ca EMAIL
biology.dal.ca WEB

Alastair G.B. Simpson, PhD
Associate Professor, Dalhousie University
Member: CGEB at Dalhousie University
Scholar: Canadian Institute for Advanced Research
(Program in Integrated Microbial Biodiversity)

Phone: (902) 494 1247
Email: alastair.simpson@dal.ca

5 Sept 2008

To: Professor Jaroslav Kulda
Department of Parasitology,
Faculty of Science,
Charles University,
Prague
Czech Republic

Dear Prof. Kulda

Please find attached my review of the PhD thesis of Martin Kostka. Thank you very much for the opportunity to review this work.

I would be grateful if you would extend to Mr Kostka that I very much enjoyed reading his thesis, and that I wish him the best of luck and success in his formal defense.

Yours sincerely,

Alastair Simpson

Report on PhD Thesis by: Martin Kostka

Reporting reviewer: Alastair Simpson

September 2008.

1) Overall assessment

This thesis primarily describes a series of molecular phylogenetic analyses of various protozoan protists. The first two analysis detail the determination of the first legitimate gene sequences from opalinids, one of the more famous 'evolutionary enigmas' in protozoology, plus the first sequences from their hypothesized sister group *Karotomorpha*. Obtaining these sequences was not straightforward since all of the organisms were uncultured, and present in mixed eukaryotic biotae. In the process these works revealed that previously sequenced genes reported as opalinids were actually fungal contaminants. The phylogenetic analyses of the correct SSUrDNA sequences quite emphatically supported a popular hypothesis based on morphology that opalinids and proteromonads are related, and even that the proteromonad *Karotomorpha* is the specific sistergroup of opalinids.

The next set of works is a series phylogenetic analyses of SSUrRNA genes from different groups of protozoa. The first is an interesting analysis that suggests that retortamonads, traditionally considered as monophyletic, may be the paraphyletic ancestors of diplomonads. This is an interesting result that, as the paper indicates, will certainly be interesting to test in the future as sequences from other genes become available for both groups of retortamonads. The other two works deal with determining the phylogenetic positions of two amoebae – as it turns out, both have very close relatives already in the database, and these analyses do not appear to have been particularly challenging, but nonetheless well executed.

The last section details a program written by the author to implement slow-fast analyses – a technique proposed some time ago for use with the phylogenetic analysis of datasets in which long-branch attraction linked with saturation at some sites is an issue. I confess I have not tried to use slowfaster before examining this thesis. From a brief trial it performed as it is supposed to, with very little tweaking. It is extremely user-friendly, and should greatly facilitate researcher employing the slow-fast methodology. This methodology is becoming increasingly used as phylogenomic analyses, having solved the old problem of data quantity (at least on a per-species basis), now focus on issues of data quality. The testing of this program in the applications note includes an analysis of *Blastocystis* phylogeny

The absolute amount of novel data introduced over the course of this thesis is quite modest, although some of the biological material was clearly not straightforward to examine. The phylogenetic analyses, which are the main material, are competently and effectively performed - clearly so, since three of them are published already in a specialist molecular phylogenetics journal. The slopalinid work, in particular represents a solid original contribution to our understanding of protozoan phylogeny from a molecular perspective. The introduction is acceptable, although falls short of excellent in the depth and breadth of its coverage of the background material relevant to the thesis. The introduction is not in perfect English but is absolutely clear and easy to read, as are all the papers. Thank you very much for making this work easy to understand for me, as someone with no second language.

Subject to the candidate performing well at his formal defense, I recommend that this thesis be considered as suitable for an award of a PhD.

2) Specific Critical Comments:

General

I have a couple of minor quibbles about the format of the thesis:

*I could not find a clear statement about how much of the work on each paper was performed by the candidate (except in appendix 6). This is something I had to clarify with the candidate's supervisor by email.

*Page numbers would have been nice for appendices 4, 5 and 6 (partly legible numbers are there for appendix 5).

Introduction:

*The Introduction to the thesis contains quite a number of unreferenced statements for which I would have preferred to see a reference: Here are few examples:

p5: "..there are many isolated, very old lineages of protists and phylogeny-reconstructing methods may have problems with their substitutionally saturated sequences"

p6: "…astomatids – ciliates sharing several features with opalinids: astomatids lack cytostomes and are gut commensals

p7: "The diverse group of stramenopiles…….the actinophryid "heliozoans"

p7-8: "Their phylogenetic position was uncertain until their molecular data were analyzed"

Minor criticisms:

* Page 7, line 16-17: In modern usage the terms 'Heterokonts' and 'stramenopiles' are usually considered to be synonyms. The assemblage of stramenopiles, haptophytes and cryptophytes is sometimes known as "chromists", but not "heterokonts".

*Page 11, line 8: Not that *Entamoeba* is not usually considered a member of pelobionts: instead entamoebids and pelobionts are typically treated as sister taxa.

*Page 21:

"Conclusions" is misspelled as "Conlusions"!

Appendix 2:

This work demonstrates the paraphyly of proteromonads within slopalinids. However, the basal *Proteromonas* rDNA sequence is quite a long branch in the phylogenetic tree.

Q: Are you concerned that the position for* Proteromonas *in the basal position within slopalinids might be an analyses artefact? If so, how could this possibility be tested, in your opinion?

Appendix 4:
*I recognise that the candidate is not the first author on this paper, but I must comment that I find it peculiar that the Introduction includes only a single reference, and for example, no information on *Mayorella,* the actual subject of the manuscript.

*Results, paragraph 3 indicates that the SSUrDNA sequence of *Mayorella gemmifera* is much longer than that of is close relative, *Mayorella* sp., with most of the extra length being concentrated in one region – the 'Helix E10_1' region is 926bp, compared to 277bp.
Q: Did you consider whether this particular expanded region might actually represent an intron? If so, what features/evidence argues that it is not an intron?

Appendix 5
*See comment about the introduction for appendix 4 – it applies equally to this manuscript.

*This paper has a very clear phylogenetic result – species labeled as *Saccamoeba* go in two very different places in the tree. However, I didn't really understand why the paper considers the NTSHR/SC007 clade to be the true *Saccamoeba*, and the LOS7N/I/*'S.limax'* clade to be the 'misidentified' group. What matters most is the nature of the type species, whatever that is.
Q: Were any of the species examined ultrastructurally by Page (1985) the type species of* Saccamoeba*? If not, how do we know that the type species of* Saccamoeba *is NOT a discristate species closely related to LOS7N/I?

Minor comment:
*Page 9: The discussion of mitochondrial cristae is tricky to follow: Firstly the cristae of NTSHR, which are fundamentally tubular, are described as 'unusual', whereas within the context of Amoebozoa, the discoidal cristae of LOS7N/I would be far more 'unusual'. Secondly, it is never said explicitly that LOS7N/I has discoidal cristae - it took me a little time to work this out from the text. Incidentally, Dykova et al., 2002, in which the cristae are of LOS7N/I were reported, seems to be missing from the reference list.

Appendix 6:
Many datasets contain a lot of constant sites, which are nearly/completely uninformative for phylogenetic analyses, and will never be removed in a slow-fast analysis. They will, however, eligible for removal in the jackknifing procedure used as a baseline to assess signal decay during slow-fast analysis. In other words, the jackknifed datasets will always have a lower proportion of constant sites, and hence a higher proportion of variable sites, than the fast-site-deleted datasets of the same size.
Q: Is this appropriate, or might it be more reasonable to jackknife only the variable sites from the alignment?

Minor comment
*Constant sites will often come to dominate the total count of sites in the 'low S number' datasets (e.g. S0) in slow-fast analyses, making total count of site remaining a poor indicator of phylogenetic informativeness. I think it would be helpful, for example, to have another column in Table 1 giving the number of variable sites remaining.

3) General questions:

In general the phylogenetic analyses performed in this thesis uses a relatively small number of included taxa, at least when compared to some related analyses. For example the analysis of relationships amongst phagotrophic stramenopiles using SSUrRNA gene by Cavalier-Smith and Chao, 2006 (J Mol Evol62:388–420) covers similar territory to Kostka et al., 2004 and 2007 (Appendices 1 and 2), but includes several times as many sequences in the reported analyses.
Q: Can you explain the advantages and disadvantages of your approach to taxon sampling, and explain why you ultimately chose to include the number of sequences you did.

For all the analyses reported in the thesis the sequences were initially aligned using an automated alignment program (e.g. Clustal_X). In appendices 1, 2 and 3 the materials and methods sections indicate that these alignments were then edited manually, which I assume means that the actual alignment that resulted from automatic alignment was changed. By contrast Appendices 4, 5 and 6 indicate only that unalignable positions were excluded, implicitly suggesting that the automated alignment was left unchanged.
Q: Were the alignments from Appendices 1,2 and 3 actually treated differently from those in appendices 4,5 and 6? If so, was there a specific reason for treating them differently? In general, does the candidate have an opinion on whether automated alignments of SSUrRNA sequences are generally trustworthy (i.e. is manual correction, when performed, an important step, or a trivial step?)

Some of the analyses included in this thesis recover relationships that might potentially represent long branch attraction artifacts (The position of *Chilomastix* in Appendix 3, and I would contend, perhaps the position of *Proteromonas* in Appendix 2). Appendix 6 describes software to implement a data filtering method - slow-fast analysis - that is specifically intended to reduce the effect of long-branch attraction.
Q: is it feasible and appropriate to use the slow-fast method (and your software) on the datasets in Appendices 2 and 3? Have you already done so, and if so, was the result illuminating?

Slopalinids are closely related to *Blastocystis,* and both groups live within animal intestinal tracts. *Blastocystis* is an anaerobe, with a biochemically reduced mitochondrion (e.g. without complexes III and IV of the electron transport chain and an incomplete TCA cycle (e.g. Stechmann et al. 2008).
Q: What, if anything, is known about the oxygen tolerance of slopalinids? Is there any evidence that they might have biochemically abnormal mitochondria?

Subject to the candidate performing well at his formal defense, I recommend that this thesis be considered as suitable for an award of a PhD.

Signed; _____ Date:_____

Alastair Simpson
Associate Professor, Department of Biology,
Dalhousie University, Canada