

Report on MSc thesis of Jan Brabec entitled: „**Paraphyly of pseudophyllidean tapeworms: testing a phylogenetic hypothesis using sequence data (18S and 28S rDNA)**“

The aim of the thesis was to perform phylogenetic analyses to accurately test for the hypothesis of paraphyly or polyphyly of the order Pseudophyllidea that was previously found in the several studies using molecular data set. When comparing with previously published studies on cestode phylogeny, the main objective of this MSc thesis is to increase the sampling size in order to test the phylogenetic relationships of Pseudophyllidea and closely related groups including monofossate, difossate and tetrafossate cestodes, and try to resolve the phylogenetic relationships within Pseudophyllidea.

This MSc thesis represents an important contribution to the field of molecular phylogeny of Cestoda and the quality of work is excellent.

However, I have several comments and questions regarding to the phylogenetic analyses used in this study.

1. Why was not the analysis of minimum evolution included (approximated by NJ method)?
2. Further justification should be given when including different weights of transition/transversion and different gap opening/gap extension penalties. Why were those different criteria applied for such choice? Why was the eventually ts/tv 1/2 selected for representing the trees?
In my opinion instead of using a wide range of different criteria, the analysis of the saturation curve of the different substitutions could be more appropriated by plotting the p-distances against Tajima-Nei distance as is commonly used.
Moreover, the differences among the phylogenetic reconstructions (topologies) obtained from different analyses and weighting schemes were not tested (Shimodaira-Hasegawa or Templeton tests in MP).
3. It could be a good idea to consider the gaps as the characters and perform the analyses (MP) including the gaps.
4. It is not clear whether GTR model was defined before estimating the parameters for each data set or the evolution and the parameters were estimated together (page 19). It could be precised.
5. The parameters of the model used in the study should be detailed at least for the displayed trees (substitution rate, α parameter of gamma distribution, proportion of the invariable sites; nucleotide frequencies).
6. The justification for a choice of AIC and not LRT criteria for data analyses would be interesting and could be added in the Material and Methods (page 19).
7. The author notes that the position of the individual orders was different in the different phylogenetic trees. Was the similarity in the topologies of phylogenetic trees tested statistically (Shimodaira-Hasegawa or Templeton tests)?
8. Providing BP values <50 does not show much information in the phylogenetic trees.
9. The results of LogDet analyses are not shown.

10. As far as I can see there is no difference in the phylogenetic trees presented in the results following the analyses without and with *Gyrocotyle* and tetrafossates. Maybe some explication would be useful.

11. In the results and discussion chapters, several relationships are highlighted. However, most of them are not supported by BP in most of analyses showed. Caution in the conclusions of those results should therefore be taken.

12. Page 23. First sentence: "Wider spectrum of results" should be more precisely explained.

13. Was the best model fit for the data set recalculated when performing the analyses after the elimination of *Gyrocotyle urna* and tetrafossates?

14. Table 6. The parameters of CI and RC are very low. It could be discussed.

15. Why was the code "5" used for the ts/tv ration 1/2?

16. When comparing the topology of the phylogenetic trees constructed on the base of SSU, LSU and combined data SSU and LSU, the topology of the combined data is similar to that of LSU data (Figure 3 and Figure 5). This result could be related with the fact that the sequences of SSU are twice as long as the sequences of LSU and display more phylogenetic signal. ILD test applied before combining data sets is based on MP criteria, i.e. working on the informative sites. However, when combining data for ML analyses, complementary method would be more appropriated (for instance the method of likelihood sliding window in PLATO).

17. Page 22. last sentence. In BI tree (I suppose that this is a tree showed in Figure 3A) Trypanorhyncha, Diphyllidea, Bothriocephalidea do not form a more derived clade but a sister group to Caryophyllidea+Diphylobothriidea, The Spathebothriidea has the basal position in the phylogenetic reconstruction.

Pages 23 and 34. The provided analyses did not confirm conclusively that Bothriocephalidea are not a more derived group than Diphylobothriidea.

18. Conclusion: The quality of the presented thesis corresponds to the level of MSc thesis and therefore I recommend this thesis for defense and I would like to classify it by the highest degree.

Brno, May 25, 2006

Andrea Šimková, PhD.



Evaluation of MSc of Jan Brabec: "Paraphyly of pseudophyllidean tapeworms: testing a phylogenetic hypothesis using sequence data (18S and 28S rDNA)"

Examiner Dr PD Olson, 28 May 2006

presented to the Faculty of Biological Sciences,
University of South Bohemia in České Budějovice

Jan Brabec presents a molecular phylogenetic analysis of the cestode order Pseudophyllidea based on a combination of complete 18S and partial 28S ribosomal RNA sequences. His work provides the first morphologically independent assessment of the interrelationships of the constituent groups that comprise the 'Pseudophyllidea' as well as assessing their relative positions among the nominal cestode orders. Although paraphyly of the order was suspected from previous studies aimed at elucidating inter-ordinal relationships of the Cestoda, these initial studies were based on single or few exemplar species and a more comprehensive study of this large and ubiquitous group was critically needed to: 1) confirm or reject paraphyly of the 'Pseudophyllidea'; 2) circumscribe monophyletic groups in the case that the 'Pseudophyllidea' is non-natural; 3) resolve the positions of the constituent pseudophyllidean lineages; and 4) to aid in resolving the early evolution of the Cestoda.

Jan's work builds upon a relative small fund of previously characterized ribosomal sequences representing the major lineages of tapeworms. To this he has characterized and added new sequences from a representative set of 'pseudophyllidean' lineages, including both common and unusual forms from each of the six recognized families. He has constructed automated and manual alignments of these data, analyzed them using both character and model-based phylogenetic methods and interpreted his results in the context of historical and contemporary ideas of pseudophyllidean systematics and of cestode evolution more generally. Although not free from grammatical errors, his thesis presented in English shows a high level of competency in a foreign language that will serve him well in his scientific career.

Jan's work addressed the most problematic part of the molecular phylogeny of cestodes: resolving the early branches of the Eucestoda and the interrelationships of the monofossate and difossate groups. Previous studies based on ribosomal data have failed to provide strong support in this part of the tree and one potential reason for this may have been a lack of representative taxa. However, Jan's work shows that although additional taxa are important for revealing paraphyly of groups previously considered natural, the interrelationships among these major lineages remains largely

unresolved by ribosomal data. On the other hand, his work also shows that this combination of data provide very strong signal *within* these major lineages, and his work has thus significantly advanced our understanding of diphyllbothriidean and bothriocephalidean interrelationships with strong support of the hypotheses.

Specific comments and corrections:

Pg 2. You state that microtriches play an important role in the absorption of nutrients, but do not provide any citations to support this. Is there actual evidence of this hypothesis?

Pg 2. You state that Spathebothriidea are non-proglottized; however, because proglottization refers to the serial repetition of the proglottides, they must be considered proglottized—but not segmented (thus distinguishing the somatic separation of proglottides from the formation of the proglottides themselves.

Pg. ‘hermafroditic’ should be ‘hermaphroditic’

Pg. 3. Nineteenth century should be twentieth century.

Pg. 4. ‘homological’ should be ‘homologous’

Pg. 5. ‘apomorphic’ should be ‘apomorphic’

Pg. 6. Fig. 1. I don’t believe this tree is from Hoberg, 1997: in fact, it was not until the results of molecular data began to be accumulated that Hoberg’s subsequent morphological analyses began to reflect the molecular hypotheses. For example, I believe the positions of the Nippotaeniidea, Proteocephalidea and other ‘orders’ were very different in the original cladistic analyses of morphology by Hoberg et al. Thus, subsequent results stemming from molecular data obviously influenced their ideas regarding certain morphological homologies, and thus subsequent morphological analyses came to reflect the molecular data.

Pg. 12. Table 2. No explanation is given for meaning of +, - & *.

Pg. 13. ‘grinded’ should be ‘ground’

Pg. 21. “concatenated data sets constructed under transition/transversion ratio of ½”. What does this mean? What does concatenation have to do with ts/tv ratios?

Pg. 22. ‘basal branches’ would be better worded as ‘internal branches’

Pg. 28. To the extent possible, it would be good to show illustrations of previous hypotheses on the interrelationships of the ‘Pseudophyllidea’

Pg. 31. In the discussion of classification, which Author’s classifications are independent vs. those in which subsequent authors accepted and followed the classifications of previous authors?

Pg. 31. 'within the orders' should be 'among the orders' (or 'between' if comparing only 2 things).

Pg. 32. Although the LogDet transformation can help reduce LBA resulting from nucleotide compositional bias, it is only one reason for LBA to occur and does not affect your data (i.e. the relative frequencies of G, A, T, and C are more or less the same among all of the taxa compared—this can also be tested quite simply in PAUP*. More importantly, however, is that the LogDet model has no applicability to analysis by maximum parsimony. Still, the point made is valid that LBA cannot be invoked as an explanation when the positions of the taxa in question remained the same whether analyzed by unweighted parsimony or by employing a model to 'correct' the data.

It is clear from the presentation of his thesis and from informal conversations I've had with him during the course of his studies that he has a solid understanding of the problems addressed and of the methods employed. I consider his work to fulfil the requirements of the degree in the highest standard.



Dr Peter D. Olson

Department of Zoology, The Natural History Museum, London SW7 5BD UK