

A review of the P.h D. Thesis

Author: Emmanuel Arriaga-Varela

Title: **Understanding the biodiversity boom in terrestrial hydrophilid beetles**

The Thesis contains extensive introduction and summary of Arriaga-Varela's research and four chapters represented by published/accepted studies and three manuscripts formatted for consideration in scientific journals.

First four studies deal with the alpha taxonomy of various genera. The studies are very thorough, with very high standard descriptions and illustrations – a top-quality alpha taxonomy which has already been published in appropriate taxonomic journals, although dealing with a quite low number of species. Three studies have already been published, one has been accepted for publication.

One study reports the morphological investigation into Baltic amber Hydrophilidae. The author used his expertise in the morphology of the group and he studied available fossils with micro-CT scans. Similarly to preceding chapters, the study has a high standard and provides new data for dating analyses which are presented in the following chapters.

The chapters 6 and 7 deal with the molecular phylogeny of selected groups of Hydrophilidae and the author invested rich material, a lot of time, effort, and expertise in finalizing these core parts of his thesis. Although I have some questions about the methods and some comments on the presentation of these manuscripts, I must clearly declare that these studies are well-done, up-to-date methods of analysis are employed and conclusions are fully based on results. The discussion is well-structured. These studies are sure of such standard that they will be published in journals with wide readerships.

Conclusion

E. Arriaga-Varela has demonstrated deep knowledge on the taxonomy and evolutionary biology of hydrophilid beetles. He has been productive over years of his Ph. D. study in Prague and expanded the scope of his work to ask broad questions on the evolution of his group. The conclusions are of general interest and have a potential for general readerships of evolutionary literature. He successfully combined various methodological approaches in his thesis, *i.e.*, the traditional morphological taxonomy, sometimes expanded by the application of the barcode method, the study of fossil records, including high-tech imaging, and molecular studies using multiple molecular markers. His most important work in the thesis presents advanced analytical methods in Sanger-based molecular phylogenetics.

The publication record of Emmanuel Arriaga-Varela is appropriate for the defense of Ph. D. degree at the Charles University, Prague and the acceptance of his Thesis for the defense before the Committee for Doctoral Degree has my strongest support.

I have the following comments and questions to E. Arriaga-Varela:

Of the negligence or haste the work, especially the introductory part contains many errors, misspelling, and poor formatting. There are too many instances to be listed here. As an example, it is difficult to find a correctly formatted reference in pp. 28–32. Comma is used before the year of publication with ~50:50 frequency, the page range is stated with similar frequency with "-" and "-" dash, online publication does not contain a link, "&" and "and" are used in journal titles by chance, some citations are hard to tell if it is a journal or a book – *e.g.*, Jackman (1997), publisher is not fully cited, *e.g.*, "OUP", the place of publication is omitted in cited books, except Oxford in one case, 'Princteon' was cited instead 'Princeton', additionally in italics, as a part of the title of the book, not as a place of publication and it was followed by comma, *etc.*, *etc.* The manuscript with such poor formatting would be returned before review when submitted to a serious journal. Unfortunately, these formatting problems unnecessarily compromise the general positive consideration of the whole Thesis.

The presentations of individual chapters in pp. 18–24 are written with similar negligence.

Chapter 1: the first sentence in the past tense, the second in the present tense. Why do you use the French Article "La" in the English text?

Chapter 2: The features of *C. pulsatum* justifies the description of *Evanesternum*, to accommodate **it** ... (shortened version)

Chapter 3: ..obtained during the survey of pollinators and **visitor** of inflorescences. ...feed on organic material **included** pollen...

Chapter 4: why "Rotten" is capitalized? ... of **the** family... missing article.

Chapter 5: All fossil – a single fossil or all fossils?

... "under conventional tools" ... "do you mean "under light microscope" or "using tools"

... only known fossil Megasternine, ... what is Megasternine – if a family-group name, then Megasternini, if it is an informal designation then do not capitalize.

" ... it really belong in family Phalacridae." should be "it really belongs in **the** family Phalacridae."

Chapter 6: "increase in diversification in the family" much better "diversification rate";

I would prefer 'species diversification' instead 'taxonomic diversification'.

The sentence "We reconstructed the ancestral states for the evolution of habitat preferences between 12 different options." is clumsy and should be definitively reworded. For example: Habitat preferences were coded as twelve discrete character states and their evolution was estimated.

Chapter 7: 'Megasternina, These'; 'are diagnosable based' better 'are diagnosed based'

'their morphological patterns discusses' Can you define 'morphological pattern'?
'intercontinental dispersion in multiple and opposite directions' not very informative expression. I can imagine what you mean but information should be better specified.

The similar expressions are encountered throughout the text which has not yet been published and edited in journals. Although the error frequency is lower in manuscripts (Chapters 5–7), some can also be identified (*e.g.*, p. 217, '...secondary transitions were originated.').

Besides the criticism of the formal presentation, I have some specific questions and a few additional comments:

p. 220

Reference to the list of samples is missing in the text – I found it only later in the Supplements. You should describe the dataset – the number of terminals, taxonomic coverage. Later in Results, you should inform on the number of available data for each fragment and completeness of the dataset. These parameters are important for the reliability of the analysis and can be clearly stated. How biased representation of data in the dataset affects the topology?

Why laboratory methods are described in two lines only (p. 22, lines 9–10). Did you amplify fragments? Was the PCR product cleaned in the lab or only sent for cleaning and sequencing? What methods of laboratory work do you master?

Why do you use ClustalW? It is out-of-date software. Although not so important when protein coding sequences are aligned, it should not be used for length variable rRNA sequences. Have you used other software for alignment?

You do not use maximum likelihood methods for tree inference. Why? I recommend it at least to validate the results of the Bayesian analysis. The method is used in Chapter 7, both chapters are unpublished, *i.e.*, they were not finished before rapid and quite effective methods have been introduced.

p. 225 what is 'specific diversification estimated'?

p. 226 "much higher values of lambda" Could you compare the increase with other publications? It is by 0.006, so biological meaning should be considered with caution.

p. 251 instead 'basal divergences' use 'deep bipartitions' or 'deep splits'. The word 'basal' should be avoided when the topology of the tree is considered.

p. 251 'taxonomic reorganization of the generic concepts is needed'. A simpler sentence would be better: The genera have to be redefined...to be monophyletic.
p. 251 'ancestral area reconstruction estimation show..."

p. 252 'Approximately 570 spp. are described, but this number seems underestimated.' I understand what you mean but you state in fact that you do not count described species well.

p. 258 "divided in the next way:" better "partitioned by gene fragments: "

p. 258 What is the difference between UBF values in IQtree and BS values estimated using such programs like RAxML?

September 10th, 2019



Prof. Ing. Ladislav Bocak, Ph. D.
Dept. Zool. Fac. Sci. UP
17. listopadu 50
771 46 OLOMOUC
Czech Republic