

Evaluation of the PhD Thesis "Taxonomy, phylogeny and phylogeography of selected groups of aquatic beetles (Coleoptera: Hydrophilidae, Hydraenidae) of the Caribbean region", by Albert Deler-Hernández

The PhD Thesis of A. Deler-Hernández consists of a 27 page detailed introduction on the background knowledge of the geographic study area (the Caribbean region, and in particular the Greater Antilles) and its fauna of aquatic Coleoptera, followed by nine papers which constitute the core of the study. In all papers A. Deler-Hernández is the first author, and they have different coauthors according to the subject and the study group. Of these, seven papers are detailed studies on the taxonomy of different genera in two families, Hydrophilidae and Hydraenidae, and have been already published mostly in Entomological or Zoological international journals. The other two are more general papers on the biogeography and evolutionary history of two genera of Hydrophilidae, one is currently in press in the Zool. J. Linn. Soc. (dealing with *Phaenotonum*), and the second (dealing with *Crenitulus*) is a manuscript draft apparently ready for submission.

The PhD of A. Deler-Hernández represents a substantial advance in the knowledge of the Hydraenidae and some genera of Hydrophilidae of the Caribbean region, which was so far most incomplete. Other than the taxonomic revision of the different taxa, the two more general papers have a much broader interest for the understanding of the biogeography of the region, which has largely only been based on the knowledge derived from some well known vertebrate groups. There were very few comprehensive and well substantiated studies on Caribbean invertebrates, and in this sense this PhD is a timely and important contribution. I will focus my comments on these two papers, in part due to their greater complexity but also hoping that they may be useful, especially for the second (the revision of genus *Crenitulus*), as it is the only manuscript which still is not published or in press.

- I have some doubts with respect to the strategy used in partitioning the data for the phylogenetic analyses. In the two papers of the first part the data is partitioned by genes, with the mitochondrial protein coding genes partitioned by codon. Given that in most cases a complex GTR evolutionary model is used for each partition, this results in a high number of parameters to estimate. My concern is that the amount of data may not be sufficient to properly estimate these parameters, as in most cases the number of terminals is relatively low and in some cases there are substantial amounts of missing data, which may lead to overparametrisation. This usually does not affect too much the overall results, but still may be of importance as it may difficulty the convergence of some parameters due to the insufficient data to estimate them adequately. In fact, in some cases there seem to have been some problems of convergence, and the number of generations necessary to reach an adequate convergence seem to have been also very high (up to 500 MY). It may have been convenient to test if the use of this high number of partitions was justified, especially for the codon partitions of the protein coding genes.

- My second main doubt refers to the procedure used for calibrating the phylogenetic trees. In the two papers of the first part of the Thesis the same method is used: fix the date of one of the basal nodes of the phylogeny according to the results of a previous paper presenting

a dated phylogeny of Hydrophilidae (Bloom et al. 2014), and at the same time setting as prior the evolutionary rate of some of the partitions according to other estimates based on a different family (Papadopoulou et al. 2010). A-priori this seems to be a robust, conservative procedure, taking advantage at the same time of the knowledge of the fossil record of the family (following Bloom et al. 2014) and of widely accepted evolutionary rates for some of the mitochondrial genes (Papadopoulou et al. 2010, but the rates are similar to other published estimates for different groups of beetles). However, while the final age of the nodes for which the a-priori date was constrained is very close to this initial constraint, there is no report on the final evolutionary rates. In all cases a relaxed clock was preferred, which means that the initial constraint on the evolutionary rates of the genes was not very strict and the final rate could potentially be very different from that given as a prior. The ages estimated in Bloom et al. (2014) are considerably older than previous estimates - some of them based on rates similar to that of Papadopoulou et al. (2010) -, which suggests the possibility that the final rates of the estimated trees could be substantially lower than the priors used. This does not imply that the estimations are less accurate, but it is something that deserves to be discussed. However, it is interesting to note that in the discussion of the biogeographic interpretation of the evolutionary history of *Phaenotonum* (pages 69 and 70 of the thesis), in different occasions it is noted that the estimated ages are "too old" to fit what apparently is the optimal palaeogeographical interpretation. Additionally, and only as an indirect evidence, it is worth to note that, according to the Introduction, the majority of fossils from Dominican amber represent extinct species, which means that most current species should not be older than 15-20 Ma according to the most accepted estimations. However, most Caribbean species of *Phaenotonum* are estimated to be older than 15-20 Ma, which would imply that the age of the species of the genus is in general older than the average Caribbean species. Again this is no evidence against the estimated ages, but it is an interesting observation to be noted.

Other, relatively minor questions:

- In the first paper the different clock models are compared using Bayes factors: it would have been better to use AICM, as it gives a better estimator. Also, the comparison of a strict vs. relaxed clock was apparently done for both mitochondrial and nuclear genes (or at least there is no comment on that). It may have been more appropriate to compare only mitochondrial genes, as ribosomal nuclear genes are known to rarely vary clock-like, and thus a combined comparison could be bias towards non-clock behaviour due to the only effect of the nuclear genes.

- In the biogeographic analyses it is necessary to assign dispersal probabilities between different geographic settings. The precise values are necessarily arbitrary, but the degree or arbitrariness could be reduced by using a range of values in a sensitivity analyses, and see how results change - or not - depending on the different weight given to the different biogeographical barriers.

- In the first paper, bootstrap values are given as probabilities. They are usually given as percentage (as is done in the second paper).

- Also in the first paper, the key to species of *Phaenotonum* uses the presence or absence of hind wings as a character (although never as the only character). There are many cases of polymorphism in wing development among beetles, so unless there are really a large number of observations to be sure that this is not the case, it may be better not to rely on this character for the identification.

- In the second paper, it is not clear to me the rationale of using separately the genes COI and 16S for species delimitation. They are both inherited as a single haplotype (with no recombination), and thus differences between the two could only be due to stochastic effects or to their different variability - COI is known to be substantially more variable than 16S, and thus in principle more appropriate for a species delimitation analysis, at least in insects.

- Just as a suggestion, it would be interesting to check the identity of the larvae of *Berosus* described in Chapter 7 with molecular data, if suitable material becomes available.

Overall the PhD of A. Deler-Hernández is an impressive achievement, which demonstrates that he has mastered a vast array of methodologies related to systematics in the widest sense. These include field skills to collect sometimes elusive specimens in poorly explored habitats; descriptive morphology of both immature stages and adults, including high-quality taxonomic descriptions; and obtaining and using molecular data to estimate the phylogenetic history of different lineages, with temporal and geographical reconstructions. All these skills resulted in a series of high quality papers in international journals in collaboration with some coauthors, which is itself another merit to be considered as it shows the capability of the candidate to work in an international collaborative context. In my opinion the thesis fulfils the requirements for the PhD program in Zoology, and I recommend it for the final defence.

In Barcelona, 4th September 2017

Ignacio Ribera, PhD