



HAL
open science

The challenge of delineating species when they are hidden

Fabien L. Condamine

► **To cite this version:**

Fabien L. Condamine. The challenge of delineating species when they are hidden. 2019, pp.100089. 10.24072/pci.evolbiol.100089 . hal-02551069

HAL Id: hal-02551069

<https://hal.science/hal-02551069>

Submitted on 29 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



The challenge of delineating species when they are hidden

Fabien Condamine based on reviews by Christelle Fraïsse, Pavel Matos and Niklas Wahlberg

A recommendation of:

Pilar Alda, Manon Lounnas, Antonio Alejandro Vázquez, Rolando Ayaqui, Manuel Calvopina, Maritza Celi-Erazo, Robert Dillon, Luisa Carolina González Ramírez, Eric S. Loker, Jenny Muzzio-Aroca, Alberto Orlando Nárvaez, Oscar Noya, Andrés Esteban Pereira, Luiggi Martini Robles, Richar Rodríguez-Hidalgo, Nelson Uribe, Patrice David, Philippe Jarne, Jean-Pierre Pointier, Sylvie Hurtrez-Boussès. **Systematics and geographical distribution of *Galba* species, a group of cryptic and worldwide freshwater snails (2019), *bioRxiv*, 647867, ver. 3 peer-reviewed and recommended by Peer Community in Evolutionary Biology. [10.1101/647867](https://doi.org/10.1101/647867)**

Submitted: 25 May 2019, Recommended: 24 November 2019

Cite this recommendation as:

Fabien Condamine (2019) The challenge of delineating species when they are hidden. *Peer Community in Evolutionary Biology*, 100089. [10.24072/pci.evolbiol.100089](https://doi.org/10.24072/pci.evolbiol.100089)

The science of naming species (taxonomy) has been renewed with the developments of molecular sequencing, digitization of museum specimens, and novel analytical tools. However, naming species can be highly subjective, sometimes considered as an art [1], because it is based on human-based criteria that vary among taxonomists. Nonetheless, taxonomists often argue that species names are hypotheses, which are therefore testable and refutable

Open Access

Published: 9 December 2019

Copyright: This work is licensed under the Creative Commons Attribution-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/4.0/>

as new evidence is provided. This challenge comes with a more and more recognized and critical need for rigorously delineated species not only for producing accurate species inventories, but more importantly many questions in evolutionary biology (*e.g.* speciation), ecology (*e.g.* ecosystem structure and functioning), conservation biology (*e.g.* targeting priorities) or biogeography (*e.g.* diversification processes) depend in part on those species inventories and our knowledge of species [2-3]. Inaccurate species boundaries or diversity estimates may lead us to deliver biased answers to those questions, exactly as phylogenetic trees must be reconstructed rigorously and analyzed critically because they are a first step toward discussing broader questions [2-3]. In this context, biological diversity needs to be studied from multiple and complementary perspectives requiring the collaboration of morphologists, molecular biologists, biogeographers, and modelers [4-5]. Integrative taxonomy has been proposed as a solution to tackle the challenge of delimiting species [2], especially in highly diverse and undocumented groups of organisms. In an elegant study that harbors all the characteristics of an integrative approach, Alda *et al.* [6] tackle the delimitation of species within the snail genus *Galba* (Lymnaeidae). Snails of this genus represent a peculiar case study for species delineation with a long and convoluted taxonomic history in which previous works recognized a number of species ranging from 4 to 30. The confusion is likely due to a loose morphology (labile shell features and high plasticity), which makes the identification and naming of species very unstable and likely subjective. An integrative taxonomic approach was needed. After two decades of taxon sampling and visits of type localities, the authors present an impressively dense taxon sampling at a global scale for the genus, which includes all described species. When it comes to delineate species, taxon sampling is often the key if we want to embrace the genetic and morphological diversity. Molecular data was obtained for several types of markers (microsatellites and DNA sequences for four genes), which were combined to morphology of shell and of internal organs, and to geographic distribution. All the data are thoroughly analyzed with cutting-edge methods starting from Bayesian phylogenetic reconstructions using multispecies coalescent models, followed by models of species delimitation based on the molecular specimen-level phylogeny, and then Bayesian divergence time estimates. They

also used probabilistic models of ancestral state estimation to infer the ancestral phenotypic state of the *Galba* ancestors. Their numerous phylogenetic and delimitation analyses allow to redefine the species boundaries that indicate that the genus *Galba* comprises six species. Interestingly, four of these species are morphologically cryptic and likely constitute species with extensive genetic diversity and widespread geographic distribution. The other two species have more geographically restricted distributions and exhibit an alternative morphology that is more phylogenetically derived than the cryptic one. Although further genomic studies would be required to strengthen some species status, this novel delimitation of *Galba* species has important implications for our understanding of convergence and morphological stasis, or the role for stabilizing selection in amphibious habitats; topics that are rarely addressed with invertebrate groups. For instance, in terms of macroevolutionary history, it is striking that an invertebrate clade of that age (22 million years ago) has only given birth to six species today. Including 30 (ancient taxonomy) or 6 (integrative taxonomy) species in a similar amount of evolutionary time does not tell us the same story when studying the diversification processes [7]. Here, Alda *et al.* [6] present a convincing case study that should foster similar studies following their approach, which will provide stimulating perspectives for testing the concepts of species and their effects on evolutionary biology.

References

- [1] Ohl, M. (2018). The art of naming. MIT Press. [2] Dayrat, B. (2005). Towards integrative taxonomy. *Biological Journal of the Linnean Society*, 85(3), 407–415. doi: [10.1111/j.1095-8312.2005.00503.x](https://doi.org/10.1111/j.1095-8312.2005.00503.x) [3] De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, 56(6), 879–886. doi: [10.1080/10635150701701083](https://doi.org/10.1080/10635150701701083) [4] Padial, J. M., Miralles, A., De la Riva, I., and Vences, M. (2010). The integrative future of taxonomy. *Frontiers in Zoology*, 7(1), 16. doi: [10.1186/1742-9994-7-16](https://doi.org/10.1186/1742-9994-7-16) [5] Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E., and Crozier, R. H. (2010). Integrative taxonomy: A multisource approach to exploring biodiversity. *Annual Review of Entomology*, 55(1), 421–438. doi: [10.1146/annurev-ento-112408-085432](https://doi.org/10.1146/annurev-ento-112408-085432) [6] Alda, P. et al. (2019). Systematics and geographical distribution of *Galba* species, a group of

cryptic and worldwide freshwater snails. BioRxiv, 647867, v3 peer-reviewed and recommended by PCI Evolutionary Biology. doi: [10.1101/647867](https://doi.org/10.1101/647867) [7] Ruane, S., Bryson, R. W., Pyron, R. A., and Burbrink, F. T. (2014). Coalescent species delimitation in milksnakes (Genus *Lampropeltis*) and impacts on phylogenetic comparative analyses. *Systematic Biology*, 63(2), 231–250. doi: [10.1093/sysbio/syt099](https://doi.org/10.1093/sysbio/syt099)

Reviewed by [Christelle Fraïsse](#), 2019-11-16 16:39

The authors made a substantial effort to address all my concerns (and those of the other reviewers). I am totally satisfied with this new version, which is clearer, includes many new analyses and much better figures. I think this manuscript is a really nice contribution towards the phylogeography of *Galba* species, and more generally, it addresses the challenges of delineating cryptic species. Therefore, I support its publication in PCI Evolutionary Biology.

Minor suggestions to handle before publication: ●L236 (Material & Methods): “amplification product in step 1” → should this be “step 2” instead? ●L623 (Discussion): typo in “adaption”. ●Figure 4: I found the S-DEC, BBM, S-DVA circles a bit unclear. Could you please directly mention on the figure that these correspond to the probability of a cryptic phenotype in the MRCA?

Revision round #1

2019-07-19

Dear authors,

Thank you for soliciting the Peer Community in Evolutionary Biology to assess your study.

We have received the feedback of three reviewers for your preprint study. You will see that the three referees are positive about the paper but they also bring up very interesting and useful comments as well as suggestions that I am sure will improve the study. Overall, I agree with the reviewers that the study has many

merits and that the findings are interesting. I also think the approach proposed here is original and may be useful for further studies in taxonomy and systematics of difficult groups, especially for invertebrate clades. That being said, the study should be more rooted in the concept of integrative taxonomy (e.g. Dayrat 2005 – *Biol. J. Linn. Soc.*). In addition, the study suffers from some methodological and conceptual issues. I think the main issues concern the phylogeny and dating analyses, but because these results are the cornerstone for the interpretation of species delimitation, the corresponding results may be inconclusive as it stands. The referees also felt that the manuscript suffers from a lack of clarity in several parts of the text, especially in the Methods and the interpretations (Results and Discussion).

To summarize, I have identified six major points raised by the reviewers that you would need to carefully address. This includes the following: (1) Running phylogenetic analyses removing the third position of mitochondrial coding genes from the alignment; (2) Running a second dating analysis using relaxed clocks and compare the results obtained with a strict clock; (3) Performing a multivariate analysis on some quantitative trait(s) and calculating a distance between clusters on this morphometric space; (4) Strengthening the analyses of ancestral state reconstructions, perhaps with the use of other models (e.g. maximum-likelihood models like Dispersal-Extinction-Cladogenesis; Ree & Smith 2008 – *Syst. Biol.*) and by including the uncertainty around the node estimates; (5) Clarifying the rationale used for distinguishing cryptic species and the evolutionary scenarios tested; (6) Taking into account all the comments that aim at improving the text, the hypotheses, and the figures.

Based on the referees' comments and my reading, I believe the manuscript will benefit from a revision and a second round of reviews. If you chose to resubmit a revised paper, please make a point-by-point reply to the comments (like for a traditional journal). For the moment, I do not recommend the study in *PCi Evol. Biol.* but if the revision is thorough (satisfies the reviewers) and the results still support the conclusions, I will be supportive for the paper as being recommended.

Dr. Fabien Condamine, recommender for *PCi Evol. Biol.*

Additional requirements of the managing board: As indicated in the 'How does it work?' section and in the code of conduct, please make sure that:

- Data are available to readers, either in the text or through an open data repository such as Zenodo (free), Dryad or some other institutional repository. Data must be reusable, thus metadata or accompanying text must carefully describe the data.
- Details on quantitative analyses (e.g., data treatment and statistical scripts in R, bioinformatic pipeline scripts, etc.) and details concerning simulations (scripts, codes) are available to readers in the text, as appendices, or through an open data repository, such as Zenodo, Dryad or some other institutional repository. The scripts or codes must be carefully described so that they can be reused.
- Details on experimental procedures are available to readers in the text or as appendices.
- Authors have no financial conflict of interest relating to the article. The article must contain a "Conflict of interest disclosure" paragraph before the reference section containing this sentence: "The authors of this preprint declare that they have no financial conflict of interest with the content of this article." If appropriate, this disclosure may be completed by a sentence indicating that some of the authors are PCI recommenders: "XXX is one of the PCI XXX recommenders."

Preprint DOI: <https://doi.org/10.1101/647867>

Reviewed by [Pavel Matos](#), 2019-06-17 10:37

In this study, Pilar Alda et al aimed to delimit species in the freshwater snail genus *Galba*. In the process, they aimed to clarify the systematics and the evolution of morphological resemblance among species. They developed a 3-step approach where morphology, microsatellites, and DNA sequences of 4 loci informed species identities. By using recent phylogenetic methods including the multispecies coalescent and ancestral state inference, the study suggested the existence of at least 5 *Galba* species. However the inferred gene and species trees seem to support disparate phylogenetic relationships among species. If this gene tree discordance was caused by incomplete lineage sorting, the use of the multispecies coalescent model is largely justified. But other causes, such as hybridization and inter-species gene flow, have not been adequately treated by the methods in the manuscript and should be acknowledged in the text. Finally, the authors suggested that the most recent common ancestor of *Galba* was likely

“cryptic”, and they supported a hypothesis of morphological stasis through time, perhaps driven by stabilizing selection associated to environmental conditions.

This study used a comprehensive taxon sampling, multiple lines of evidence, and state-of-the-art methods. The research questions are relevant to other fields in evolutionary biology and thus the manuscript can be of general interest. I believe that there are still few issues that need to be clarified or strengthened in order to improve the manuscript.

INTRODUCTION Line 98-110: These lines need rewording. The first sentence of the paragraph refers to at least two issues that are problematic with cryptic species. But the following sentences discuss the second issue on biological invasions and a new third issue on disease transmission. Where is the first issue?

MATERIAL AND METHODS Line 193: I suggest moving Fig S2 into the main article because it nicely explains your 3-step approach. The figure can be improved by adding the total number of individuals analyzed in each step.

Line 209: How many microsatellite loci have you targeted in the 1,722 individuals? Line 281: The third codon position of COI was highly saturated. It will be helpful to run the phylogenetic analyses removing such nucleotides from the alignment, and to compare it with the results presented in the manuscript.

Line 287: The mitochondrial gene trees are in fact not independent. These loci are linked and thus share the same phylogenetic history. Line 294: Given the apparent long evolutionary history of the group, ~ 20Myr, the use of strict clock might not be adequate. It will be helpful to run a second comparative analysis using relaxed clocks.

Line 300: It is unclear how the gene trees allowed the identification or validation of species. Are you using any testable, quantitative criterion? It will also be informative to present in Fig. 2 the posterior probabilities on nodes. Line 304: It is unclear how the molecular dataset was used in StarBEAST2. Have you used all individuals and assigned each individual a species identity? Furthermore, StarBEAST2 simultaneously estimates gene and species trees, so it is not clear why you used BEAST2 before to infer gene trees? You could just show the gene trees estimated in StarBEAST2.

RESULTS Line 327: When reading this line, it is unclear the overall rationale for distinguishing cryptic species in this manuscript. Before this line, you acknowledged that morphological similarities among species occur in the genus *Galba*. But when it comes to *G. cousini*, you dismissed this rationale and assumed that the morphological similarity between individuals from Venezuela (known as a separate species, *G. meridensis*) and Ecuador/Colombia are because they are a single species, *G. cousini*. Line 335: It is unclear how you identified specimens using microsatellite loci. How many loci were used? Line 341: What is the criterion to identify clusters using COI sequences? Is it genetic distance? But in Fig. 2 the genetic divergence between clusters II and IV seem to be lower than the intra-cluster divergences of clade V (*cubensis/viator*). In addition, posterior probabilities for the crown nodes of clusters II and IV seem to be low, 0.9 and 0.7, respectively.

DISCUSSION Line 412: You suggest that the specific status of *G. viator* depends on the gene considered, and take a cautious but ambiguous position to consider it part of a species complex or part of a single species with wide diversity. But I encourage you to take advantage of StarBeast2. You could estimate marginal likelihoods of two StarBeast2 runs, one assuming *G. viator* as a separate species from *G. cubensis* and another assuming *G. viator* and *G. cubensis* are the same species. Then, you could compare the marginal likelihoods by computing Bayes factors, and take a clearer position on this matter. Alternatively, you could use your multi-locus dataset and multispecies coalescent methods that estimate species limits, such as BP&P or STACEY. This alternative approach will be a stronger criterion for delimiting species compared to the current approach consisting of identifying clusters. Line 521: You ruled out the “recent-divergence hypothesis” to explain the morphological resemblance among species because a previous study suggested that the genus had a 20-Ma origin. But you have not estimated divergence times among extant species in this study. An old crown age of *Galba* does not rule out very young species divergences. Line 524: You ruled out the “parallelism and convergence hypotheses” based on your ancestral inference of states “cryptic” or “non-cryptic”. But as currently defined, these two qualitative states do not properly inform your hypotheses on parallelism and convergence. You would instead need to estimate ancestral states

of measurable traits to really rule out such hypotheses. Or at the very least, consider and discuss this issue in the light of the Galba fossil record (https://paleobiodb.org/classic/basicTaxonInfo?taxon_no=307489). Were the extinct species also cryptic? Line 529: You seem to accept your “morphological stasis hypothesis” by default (literally written in the text). But again, you could discuss this in the light of the fossil record. Could we see this morphological stasis over millions of years? In this respect, you also consider strong stabilizing selection related to environmental conditions as an agent for morphological stasis. However, given that you recorded habitat associations for sampled individuals (Line 183: “The sampled habitats were characterized ...”), then why not associate this meta-data to back up your hypothesis of selection driven by environmental conditions? Are the cryptic species significantly associated with particular habitats compared to the non-cryptic species? In Line 533 you seem to relate cryptic species to a variety of freshwater habitats, but nothing is mentioned about how you concluded this and if there is another type of associations for the non-cryptic species based on your sampling notes. *Pável Matos-Maraví*

Reviewed by [Niklas Wahlberg](#), 2019-06-26 16:26

The authors report a very detailed study of the genetics of a group of cryptic species of snails in the genus *Galba*, mainly from the New World. The genus does have a worldwide distribution, and the taxonomic situation is not clear anywhere. This study is a first detailed step to clearing the taxonomy of the snails. I found the study to be well done and an excellent contribution to *Galba* taxonomy and systematics. The next step is highlighted by the authors, and that is to expand the specimen sampling to the rest of the world, preferably at the same detailed level. I have no suggestions for improvement, I enjoyed reading the manuscript, despite not being an expert in molluscs.

Reviewed by [Christelle Fraïsse](#), 2019-06-11 16:17

By means of a large-scale sampling and intense literature review (clearly presented in the Supplementary Tables), Alda et al. provide an excellent overview on the systematics and geographic distribution of cryptic freshwater snails of the

genus *Galba*. This primarily descriptive and organism-centred work meets its objectives, and I think it will be valuable for future biogeographic research on the genus.

My main concern regards the assessment of the different scenarios to explain crypticity in *Galba*, which I believe is the most interesting part of the paper from an evolutionary perspective. These are presented from the introduction (L87 – 97) and on Figure S1: (i) recent divergence, (ii) parallelism, (iii) convergence and (iv) morphological stasis.

First, the authors only provide verbal arguments in the discussion to disentangle between the four scenarios (L516 – 537), while their Figure S1 suggests that a more quantitative test could be performed. For example, the level of disparity between clades could be measured by performing a multivariate analysis on some quantitative trait (e.g. shell morphology), and then by calculating a distance between clusters on this morphometric space (e.g. by using the coordinates of the clusters on the first component). This will give a multivariate measure of disparity that then can be compared between different species pair. More generally, I really think that the authors should go beyond a qualitative description of morphology (L507-510) if they want to discuss scenarios on the evolution of crypticity. Given that they have photographs of the shells, it would be possible to perform a morphometric analysis using R packages such as Geomorph (Adams & Otárola-Castillo 2013: <https://doi.org/10.1111/2041-210X.12035>) and Morpho (Schlager 2015: <http://CRAN.R-project.org/package=Morpho>).

Second, the authors do not explain how the species pairs should be chosen in the phylogeny to compare their level of disparity (e.g. they take A1/A2 and A1/A3 for scenario A). Should the pairs being compared (“similar” vs “similar” ; “similar” vs “different”) have similar level of molecular divergence?

Third, the ancestral reconstruction trait analysis, as it stands, is not convincing enough. The root state posterior probabilities shown on Figure 3 (60% vs 40%) indicate rather weak evidence that “crypticity” is the ancestral state. Could you please add 95% confidence intervals around the ancestral values to assess the uncertainty around these estimates? And would it be possible to impose an

informative prior on nodes based on information from the fossil record? Moreover, instead of using a dichotomic trait (“cryptic” vs “non-cryptic”), the authors may gain some power by using the multivariate scores of the morphometric analysis suggested above to reconstruct ancestral states. Finally, by using the scores reconstructed at the different nodes in the past, it may be possible to actually plot the level of disparity as a function of divergence time, as illustrated on your Figure S1.

Minor concerns are listed below.

→ Discrepancies between genes trees and species tree. I totally agree with the authors that conflicting genealogical histories is a major issue for phylogenetic reconstruction, whether it is due to sampling errors, incomplete lineage sorting or introgression between lineages. For this same reason, I think it would be more appropriate to represent the genealogical relationships between alleles of the four genes as genetic networks instead of gene trees (Figure S5 – S8).

→ Evolution of crypticity. - L503: “some habitats possibly favouring the emergence of cryptic species (like caves, Katz et al., 2018).” Briefly explain why please.

- L531: “Selfing in *Galba* might have led to limited genetic variation favouring stasis”. It is not entirely clear to me why this would be the case. Is there any evidence from the literature that this has been observed (e.g. in plants by contrasting outcrossers and selfers)?

→ Figures and typos. - Figure 3: Please, add numbers (I to V) after names to be consistent with Figure 2.

- Figure S1. Maybe remove one species (five instead of six) to be consistent with your biological system. Please, explain in the legend what the colours correspond to (green vs blue).

- Figure S2. The numbers in the legend do not match: “111 individuals of *Galba cubensis* (41), *Galba schirazensis* (41) and *Galba truncatula* (29)”. Please, specify how many individuals were used for *Galba humilis* and *Galba viator* in the third step.

- Figure S4. Please, add a scale for the photographs and the drawings.
- Figure S5 – S8: Please, specify the unit of the scale. Also, I cannot see any highlighting in yellow.
- L100 and L105: I cannot see where the first issue is described in “with regard to at least two issues. The [second] issue is biological invasions” and “[third] issue arises when species”.

Author's reply:

[Download author's reply \(PDF file\)](#)