AUSTRALIAN MUSEUM

11 May 2020

Dr Joseph Gillespie
Academic Editor
PeerJ

Dear Joe,

Re: Submission 47133

Many thanks to you and the three reviewers for your very rapid, thorough and constructive review of our manuscript “Hiding in plain sight: DNA barcoding suggests cryptic species in all ‘well-known’ Australian flower beetles (Scarabaeidae: Cetoniinae)”.

We have dealt with all of the reviewers’ edits and suggestions and hope the manuscript will now be accepted for publication.

Yours sincerely,

Andrew Mitchell
On behalf of all authors

AUSTRALIAN MUSEUM
1 William Street Sydney NSW 2010 Australia **T** 61 2 9320 6000 australianmuseum.net.au
ABN 85 407 224 698

Note: The reviewers' text is in *italics* and our response follow in blue font. Line numbers refer to the revised manuscript unless stated otherwise.

Reviewer 1

Basic reporting

The background information is mostly clear although could be improved by directly stating the number of diversity of Australian genera in the first paragraph of the introduction, or quantified following statement on 65% of genera on line 56.

This has been clarified in lines 54-55.

The second paragraph is largely a non-sequitur and while the biology is important, this paragraph seems to largely be irrelevant as there are no-discussions or conclusions relating to biology. If the authors decide to include this information, it would be better suited towards the end of the Introduction, and prehaps relating to the utility of generating this barcode library.

We regard this paragraph as important background information, especially relating to the application of barcode data to ecological studies (as pointed out by Reviewer 2) but we take Reviewer 1's point that this paragraph seemed not to link directly to the rest of the paper. We have now simplified and reduced this paragraph, and we incorporated a sentence at the end of the introduction suggesting that a barcode library would aid investigations into the little-known biology of Australian cetoniines, particularly through metabarcoding.

Goals are stated clearly but there is room for a hypothesis that barcodes are likely to identify numerous undescribed species. This could be supported by the limited taxonomic attention in the last 65 years, high levels of endemism, cryptic species complexes already identified and estimates of undescribed proportion of Australian fauna.

We agree. This this is the subject of the 4th paragraph of the introduction, lines 60-69.

English is clear. Citations are sufficient to provide context.

Experimental design

This study provides the first insights in to the hidden diversity of Australian cetoniines and fills an important gap in our knowledge of scarabs as well as the Australian fauna.

The design and investigation is rigorous and methods are described in detail. The authors should be commended that this first investigation provides barcodes for almost 50% of species and all common genera. It provides a clear framework for further investigation and the utility of barcodes for identification.

Validity of the findings

I'd like to see a short discussion of the utility of this barcode library beyond taxonomy, e.g. the potential agricultural importance mentioned in the introduction, or ability to match life-cycles.

There is now a brief mention at the end of the introduction of the possible application of barcoding to the investigation of the little-known biology of Australian Cetoniinae (lines 99-100) and this is expanded on at the end of Discussion.

I was intrigued when I read in the introduction that both larvae and adults were sampled and that one of the goals of this paper were to facilitate life-cycle matching but these larvae didn't get mentioned in the discussion or conclusion. Were they from easily identifiable species or were they matched using these barcodes?

We have inserted a short paragraph in Methods (lines 125-128) explaining larval identification. In all but one case the larvae were reared progeny from identified adult specimens. There was a single case where a wild-collected larva was identified by barcoding and it is now mentioned in Results, lines 277-279.

The taxonomic interpretations were robust. I think its great that when possible, multiple samples were included per species. It would be useful to include locality information either on the trees or in a table so readers don't need to refer to BOLD to decipher locality information from this data set.

A supplementary table (Table S1) is now included which provides specimen sample IDs, BOLD ProcessIDs, GenBank accession numbers, identifications and collection information. For undescribed species, we have released limited locality information (state only) pending formal species descriptions.

Overall, this is an important contribution to the field.

Reviewer 2: Michael Raupach

Basic reporting

In this study Mitchell and co-authors Liu et al. analyzed about 280 specimens of Australian cetoniine flower beetles using the DNA barcode fragment. During the last years, DNA barcoding has become an important and popular tool in modern biodiversity studies as well as taxonomic research. Based on the given data and results, I think that the analysis and results are interesting and likely helpful for other researchers in this research field. In my eyes, a clear, unambiguous and professional English language has been used. The article structure and figures look professional. The literature provided is also context-related, but some important references are missing (see sticky notes in the uploaded pdf).

All sticky notes have been actioned, including the addition of references.

The figures and associated legends are understandable without going back to the main text.

I feel, however, that a discussion of possible pitfalls of using mtDNA in terms of species delineation and specimen identification, e.g., incomplete lineage sorting, the presence of Wolbachia, NUMTS etc. as part of the introduction or discussion will improve the quality of the manuscript. Such effects can cause unexpected mitochondrial diversity as well.

We agree and we have now inserted an appropriate “disclaimer” paragraph in Discussion (lines 394-400), and corresponding references, to address this concern.

Furthermore, it is obvious that DNA barcoding will play an essential role in modern molecular biodiversity research, e.g., the effective use of high-throughput sequencing technologies (meta-barcoding) highly relies on detailed barcode libraries. These aspects should be also discussed.

New paragraphs have been added to Discussion, lines 407-413 and particularly 414-417, to address these concerns.

Experimental design

The research question is well defined, relevant and meaningful. The authors state how their research fills a serious knowledge gap in the taxonomy of Australian flower beetles. Methods are described in sufficient detail to replicate. However, the authors may consider using additional species delimitation approaches (e.g., GMYC, bPP). Do you will get similar results in comparison to the BIN/RESL approach? This would be interesting to know.

This point is related to the reviewer’s previous comment about the limitations of mtDNA data. In the absence of either nuclear gene data or robust morphological studies of more specimens, we stopped short of drawing firm conclusions about species boundaries in this study. This is because of the limitations with COI-based barcoding (see new text in Discussion, lines 394-400). Therefore, while we considered implementing additional species delimitation analyses, we ultimately decided against it, believing that there are diminishing returns to analysing single locus data with multiple methods, particularly for mtDNA gene trees which can differ from species trees for many reasons (as multiple reviewers of this study have correctly pointed out). Given the considerable additional effort required to implement alternative (third or fourth) species delimitation methods, and the fact that these could lead to over-interpretation of the limited data set, we regard it as prudent to simply note the limitations of the current data set and wait until we have nuclear gene data before applying multi-locus species delimitation methods in future studies.

For some species cluster, e.g., Chondropyga dorsalis, you may also think about using statistical parsimony networks as implemented in PopART (http://popart.otago.ac.nz/index.shtml) to visualize observed molecular patterns.

The significant Geographic Distance Correlation tests, on the whole, reflect sampling of very widely separated populations. For *Ischiopsopha wallacei*, this reflects the separation of samples from Sabai and Dauan Islands, within 5 km of Papua New Guinea, and samples from approximately 800km south in Queensland. *Glycyphana stolata* samples were collected from Dauan Island to the Brisbane region >2,200km to the south. In *Metallesthes anneliesae,* the pattern is more subtle as the seven specimens were collected within an 80km radius of each other, some 200 km west of Brisbane, and the most distinct sequence, a separate BIN, is one on the northwestern perimeter of the samples’ distribution. The only highly significant test result was for *Chondropyga dorsalis*, but the 12 specimens of this species were collected within 70km radius of each other in SE Queensland. We have now inserted this explanation in Discussion, lines 317-327.

Nevertheless, we followed Reviewer 2’s suggestion and produced minimum spanning networks in PopART for *Glycyphana stolata* and *Chondropyga dorsalis*, the species above with the largest and smallest sampled distributions, respectively. The samples are grouped (coloured) by

BIN, and the networks have been uploaded for review purposes. However, we do not wish to include them in the paper since they do not show any meaningful patterns not already observed in Figs. 2-7 and S1.

Validity of the findings

I wonder if the results would be different if more range wide samples of different species were included. In some cases the samples sizes were quite small (i.e. one individual). However, the conclusions are well stated, linked to original research question and limited to supporting results. Overall, data and analyses are robust.

Comments for the author

Please check the uploaded pdf for more specific comments.

All edits and suggestions made by Reviewer 2 in his annotated copy of the manuscript have been implemented, except for his comment on line 85 of the original manuscript "*Did you have included these sequences in your data set?*" No. That question is answered and explained in the original manuscript and that sentence has now been modified to clarify why we did not include these sequences, lines 90-92 (in short, none of these sequences added any genetic or species diversity to our dataset, and anyway they could not be morphologically identified to species as we did not have access to the specimens).

Reviewer 3

Basic reporting

The manuscript is generally well written and conforms to the PeerJ standards on structure. There are only a handful of relatively small issues which should be easy to address. The introduction provides a good overview of Australian Cetoniinae and previous taxonomic work on the group. Sequences are made available through both BOLD and GenBank, and detailed specimen metadata and images are provided through BOLD. The figures are of good quality. In particular, the illustrated and color-coded supplementary trees are visually appealing, and I wonder if at least one of them (or parts of one) could be transformed into a proper figure in the manuscript instead of being "hidden" in the supplements.

Our previous Supplementary Fig. S1 has now been moved to the main body of the paper in six parts (Figs. 2-7) while the previous Figs. 2-5 have been deleted as they were excerpts of the complete Bayesian tree.

Some parts of the manuscript should be moved elsewhere to improve the flow of the text and avoid confusing the reader - see detailed comments below under General comments.

Experimental design

The need and motivation for the present study and its goals are clearly stated. The methods are sufficient to achieve the stated goal, i.e. providing an initial DNA barcode library for the focal taxon to serve as a starting point for more thorough taxonomic studies. Some basic details of PCR and sequencing protocols (primers, sequencing platform, etc) should be added in Materials and Methods instead of simply referring to the previous publication.

As the PCR protocol is a complex one, involving eight primers and up to four separate PCRs per sample, we thought it would be more efficient to refer to the paper that describes the primers and the process in detail (it is a 10 page methods-only paper: Mitchell, 2015). We provide an outline of the process in lines 190-194 and we have added basic details of the sequencing method in lines 194-196.

Rooting of the trees is not justified in any way in Material and Methods, but the justification is briefly mentioned in Discussion (lines 297-299). This note should be moved to the last section of M&M where the tree inference methods are described.

We have moved the section on tree rooting from Discussion to Methods, lines 227-230.

Apart from these minor issues, the methods are well reported, and I see no ethical or technical problems.

Validity of the findings

All primary data are made available through both BOLD and GenBank. Data and image quality on BOLD are mostly exemplary, although collection data beyond the country are not reported for many specimens. I assume that this is due to deficient label data on the sampled specimens rather than an omission on the authors' part. Conclusions are not overarching and are supported by the results. A few more specific comments are listed below under General comments.

Comments for the author

Line 71: 4000 is a gross underestimate. A quick search on Web of Science with the term "DNA barcod" as topic resulted in 7262 articles and 1000 or so hits for other document types. The 2003 paper by Hebert et al. (https://doi.org/10.1098/rspb.2002.2218) has been cited more than 10000 times according to Google Scholar.*

We drew our figure of "almost 4,000" papers from Goldstein & DeSalle's (2019) recent review which found 3,720 studies which met their criteria, and we cited their paper. However, it would appear the reviewer is correct as a search of the BOLD Publications database gave a figure of 6,183 barcoding publications. Other publications which refer to barcoding and therefore are found in a Web of Science search, or which cite Hebert et al., 2003, might not have actually generated barcode data or even analysed any. We have updated this figure to ">6,000 studies" and cited the BOLD Publications database (line 70).

Line 76: "…aid taxonomic research in many families" - Consider citing a few studies as examples

Reviewer 2 also requested additional references for this section. We have added three references (lines 75-76).

Line 185-187: A public project on BOLD works fine for data sharing in this case, but in future work, I suggest using the dataset feature on BOLD instead. Specimen records from multiple projects can be combined into one dataset, and the same records can be added into multiple datasets, without any need to move records between projects. A dataset can also be assigned a DOI. I believe the dataset feature will be useful in your future work on Australian Cetoniinae when more sequence data are added to many of the taxa covered here.

We had previously made a dataset (DS-AUCET) and requested a DOI, but as the DOI had not been issued at the time of submission we did not include that information in the manuscript. The DOI has now been issued and is included in Methods (line 200). GenBank accession numbers are also included now, both summarized in Methods (line 201) and provided in detail in Supplementary Table S1.

Line 214: As mentioned above, please move the note on lines 297-299 here to justify your approach to rooting the trees.

Done.

Line 218: How were the larvae identified – by associating them with adults based on barcodes or morphologically? Were some taxa represented by larvae only? Consider adding a note on sampling larvae as well as adults in Material & Methods. If larvae were sampled in order to test identification of larvae based on barcodes, add text appropriately in M&M, Results, and Discussion.

We have inserted a short paragraph in Methods explaining larval identification (lines 125-128). In all but one case the larvae were reared progeny from identified adult specimens. The single case where a wild-collected larva was identified by barcoding was insufficient for a test of barcoding to identify larvae, nevertheless it is now mentioned in Results (lines 277-279) and in Discussion (lines 410-413).

Line 239-246: Mainly a matter of personal preference, but I think the information in this paragraph would be better presented in a table including both monophyletic and non-monophyletic genera.

We agree that this is a matter of personal preference, and we prefer to present this data in a paragraph. We address all cases of non-monophyly in the Discussion, in reference to Figures, which we feel is necessary for clarity.

Line 251-252: Representatives of two genera appearing in the same BIN is highly unusual and is almost always due to operational errors. Is this possibly a case of poorly resolved genus-level taxonomy? Can you exclude the possibility of cross-contamination or specimen mix-up?

We mentioned this discrepancy in Discussion in the original manuscript, but did not provide that information in Results or discuss the issue in any depth in Discussion. We have now provided this information in Results, lines 267-270. In short, RESL cluster analysis incorporated a sequence not included in BIN analysis, and yielded a different result, with each species in an independent OTU. In addition, the placement of *Diaphonia* species within the *Hemichnoodes* cluster, and the complex relationships among these taxa and *Aphanesthes*, warrant further discussion, which is found at lines 378-387 of Discussion.

Line 291-293: This seems to belong to the Results section rather than here, especially since there is no mention of these results in Results in the current version of the manuscript.

Agreed. We have moved the results part to Results but keep a short discussion where it was in Discussion.

Line 306-313: Move this paragraph up before the previous one to make the text more coherent

We feel that this paragraph belongs where it is because it provides context for the discussion immediately following it, i.e., the remainder of Discussion.

Line 314-320: Join to a single paragraph

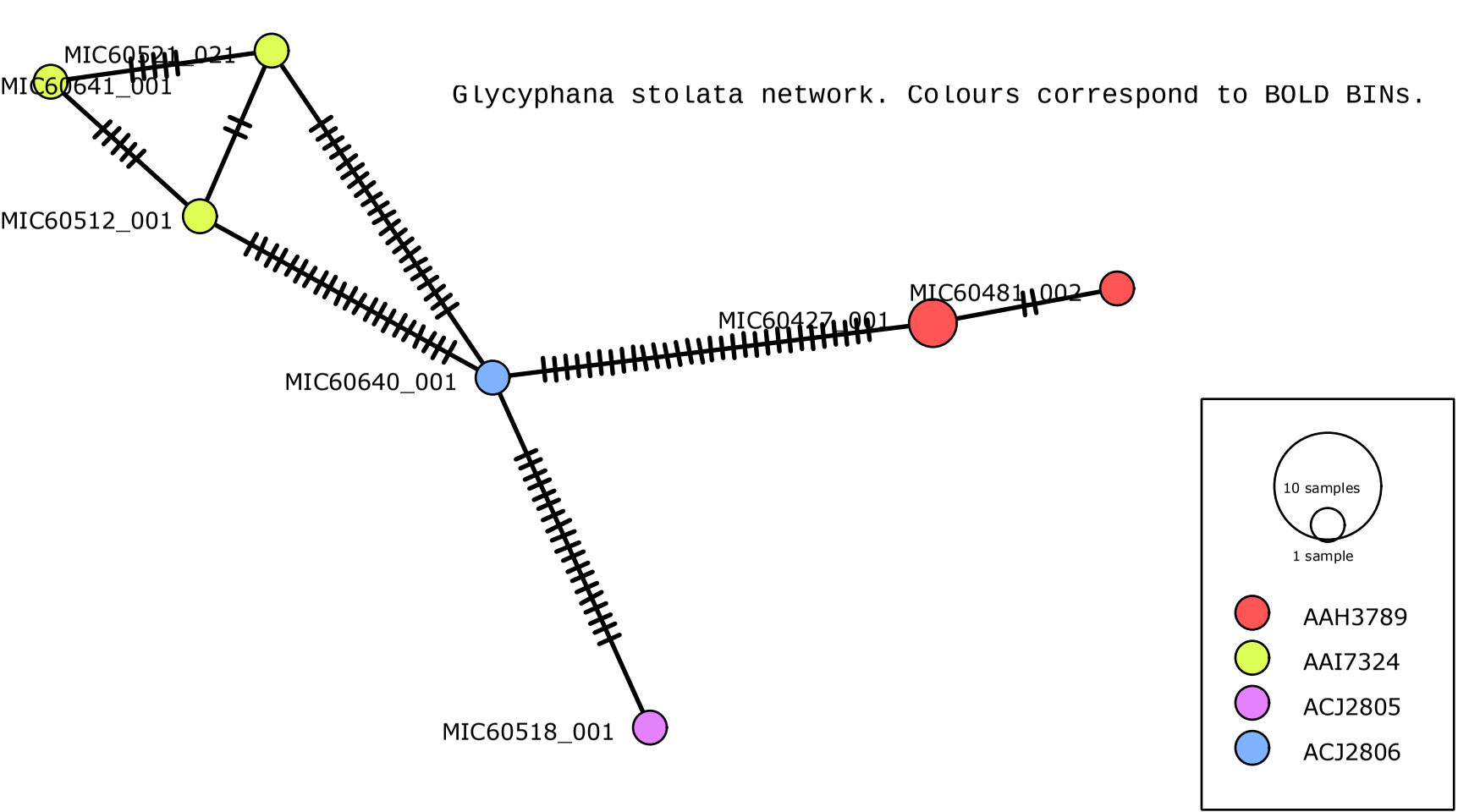
Done.

Line 339: probabilities -> probability

Corrected.

Fig. S1 legend: Baysian -> Bayesian

This legend has been deleted from the image as Fig.S1 has now been incorporated into the paper as Figures 2-7, following Reviewer 1's suggestion.


Glycyphana stolata network. Colours correspond to BOLD BINs.
MIC60521_021
MIC60641_001
MIC60512_001
MIC60481_002
MIC60427_001
MIC60640_001
MIC60518_001
10 samples
1 sample
AAH3789
AAI7324
ACJ2805
ACJ2806

Chondropyga dorsalis network. Colours correspond to BOLD BINs.