

Match and mismatch between dietary switches and microbial partners in plant sap-feeding insects

Louis Bell-Roberts, Angela E. Douglas and Gijsbert D. A. Werner

Article citation details

Proc. R. Soc. B **286**: 20190065.
<http://dx.doi.org/10.1098/rspb.2019.0065>

Review timeline

Original submission: 9 January 2019
Revised submission: 21 March 2019
Final acceptance: 23 April 2019

Note: Reports are unedited and appear as submitted by the referee. The review history appears in chronological order.

Review History

RSPB-2019-0065.R0 (Original submission)

Review form: Reviewer 1

Recommendation

Major revision is needed (please make suggestions in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?

Good

General interest: Is the paper of sufficient general interest?

Good

Quality of the paper: Is the overall quality of the paper suitable?

Excellent

Is the length of the paper justified?

Yes

Should the paper be seen by a specialist statistical reviewer?

No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.

No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?

Yes

Is it clear?

Yes

Is it adequate?

Yes

Do you have any ethical concerns with this paper?

No

Comments to the Author

In their manuscript, Bell-Robert et al conducted a meta-analysis to test the association of dietary traits with the presence of endosymbionts in plant sap-feeding insects. In addition to providing a compilation of information on the phylogeny, symbionts, and nutrition status of the sub-order Auchenorrhyncha, the authors found that the primary endosymbiont *Sulcia* is associated with xylem-feeding, whereas complementary symbionts are not directly associated with dietary switches.

In my opinion, the major contribution made by this study was to comprehensively quantify and trace back the evolutionary trajectory of symbiont loss in a large sub-order of plant sap-feeding insects. However, there are several points that require further investigation:

Major points:

- To further complement their results on the loss of the primary endosymbiont *Sulcia* (L100-105, L122-128); the authors could perform a phylogenetic analysis of available symbiont sequences to confirm a host-symbiont co-diversification (Moran et al, AEM, 2005) with the set of insects the authors focused on.
- I think that the discussion of non-dietary influences in L177-184 falls too short to explore alternative explanatory factors that the authors do not consider with their analysis. Temperature resistance has been for example suggested to favor replacement of primary endosymbionts (Moran and Yun, PNAS, 2015). For cicadas, it has been also suggested that complementary symbionts were potentially outcompeted by pathogenic fungi (Matsuura et al, PNAS, 2018). There might be also differential costs for the host in maintaining symbionts.
- The Method section should be expanded (e.g., with parts of the supplemental text) to explain in more detail how the results were generated.

Minor points:

- The authors largely apply correlative and statistical models, yet they claim in several parts of the manuscript to investigate the causality of symbiont switches (L17, L44).

- The part of the discussion (L141-144) is a strong statement and should be tuned down in my opinion, because the authors focus on the correlation with dietary switches and not other host traits.
- To facilitate reading of Figure 1, the color legends could be arranged under each other in the same order as the bands (from inside to outside). A color legend for the branch lines of the phylogeny would be also helpful.
- The legends of Figure 1 and 2 could elaborate on the meaning and interpretation of the pie charts in the figures.
- The legend in Figure S4 should reference Figure S3 in L148.

Review form: Reviewer 2

Recommendation

Major revision is needed (please make suggestions in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?

Good

General interest: Is the paper of sufficient general interest?

Acceptable

Quality of the paper: Is the overall quality of the paper suitable?

Acceptable

Is the length of the paper justified?

Yes

Should the paper be seen by a specialist statistical reviewer?

No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.

No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?

Yes

Is it clear?

No

Is it adequate?

No

Do you have any ethical concerns with this paper?

No

Comments to the Author

In this study, Bell-Roberts et al. generate a phylogeny and perform ancestral state reconstructions to investigate the relationship between symbiont composition and host diet in Auchenorrhynca plant sap-feeders. I think the general question is interesting to think about given that obligate symbiont function is closely tied with host diet and would be of interest to those studying co-obligate symbiotic systems. I understand that manuscripts to this journal are supposed to be concise, but as written I think some information is lacking. Specifically (1) a broader context for why the work is being done and how this relates to other symbiotic systems and (2) expanded details on some methods since many of the conclusions rely heavily on a few parts of the methods which are unclear and (3) perhaps some adjustments to figures to help readers interpret the results.

It is a little unclear between what taxa were used for trait analysis, sequences and phylogeny reconstruction, and how these data overlap. For further transparency, it would be useful to include the accession numbers for sequence data in addition to the traits and references included for each species. In the event that different gene sequences and traits were pulled from multiple species to represent a single genus, this would be important to know for interpreting the results. And clearly indicate which data were included in the end for the 145 genera ancestral state reconstruction analysis (full vs. pruned? in the methods it mentions having a final database of 165 genera; line 77). In addition to adding this information as a supplemental table that could be included in the electron appendix (or added to the authors' github), please elaborate in the supplementary text (see comments below).

Figures: I generally think the figures are good, though a bit difficult to compare the information across figures and some of the information present is redundant between figures. It might be better to present all of the information in a single figure, which I know may be a lot of information. A suggestion: stick with the standard tree format as in fig 2 (the circle phylogeny is a little difficult to read/interpret with the trait data), but highlight the branches with colors to represent likely companion symbiont status with the pie charts from fig 2 (or a combination of fill color and borders for branches to represent primary symbiont status and predicted diet, with pie charts indicating companion status) and columns along the tips of the phylogeny each colored to indicate the relevant traits (primary endosymbiont, companion symbiont, host diet, xylem and non-xylem (or shapes combined with fill color to indicate the last two)). This way, the phylogeny can be compressed a little to allow space for complete taxa name (if used at the species level) and the entire relevant trait data is easily visualized on the same phylogeny. A more minor point, it might be less cluttered to just use squares/rectangles in columns rather than other shapes that overlap significantly (or don't add much more info) and to not use the combination of red and green in a single pie chart. It would be useful to have at least one phylogeny that was completely readable (a lot of the labels overlap). It looks like only genera were included as tip names and only in Fig S2 & Fig S4. See first comment under supplementary materials regarding sequence data and taxa included

Minor comments:

- Line 44: The authors suggest a causal relationship for symbiont switches based on changes in host diet, but it is really a correlation at this point since it would be hard to determine causality in this system
- line 72. "COI" is listed twice, should be replaced with 18S.
- line 80-81. Figure 1 also lists moss and predatory, but these trait states were not included in the analysis. Can this decision be elaborated in the methods?
- line 130: lower case "A" in analysis
- check reference formatting

Supplementary materials:

- Line 14: "supermatrix of 832" I assume this means 832bp, does this include gaps? How much missing data if any? Did sequence for all five genes for a given taxa represent the same individual or species or were genes "mixed-and-matched" to compile a sort of consensus gene set per genus? If the data were a composition of sequences, this could impact the phylogenetic estimation and ancestral reconstruction.
- Line 14-17. How was the substitution model selected and were the data partitioned in any way using different substitution model parameters especially considering that the data represent mitochondria, ribosomal, and nuclear genes?
- Line 17-18: why weren't any closer outgroup taxa used such as any member of Sternorrhyncha when estimating the phylogeny?
- Part (a) phylogeny: I am curious why the authors didn't use a Bayesian approach, a program such as BEAST (or in RevBayes) can simultaneously estimate phylogeny and an ultrametric tree with time calibrated divergence times. This approach may be better suited for inferring any ancestral states across the phylogeny
- Line 61: when pruning the phylogeny, did tips represent individual species that matched the same taxa for trait data, or were any members within a genus all considered equal for sequence data in addition to trait data?
- Line 94-96: Does this mean that the models for trait evolution were run on the ML phylogeny to provide support for ancestral states or run across each of the 100 trees that formed the ML phylogeny with bootstrapped values? This is a bit confusing as written.
- Line 105-107. This mentions the influence of species, but the analyses presented were conducted on genera. Does this comment refer to the replacement of individual species to represent a genus for the analysis or the removal of a subset of genera for the analyses?

Decision letter (RSPB-2019-0065.R0)

25-Feb-2019

Dear Dr Werner:

Your manuscript has now been peer reviewed and the reviews have been assessed by an Associate Editor. The reviewers' comments (not including confidential comments to the Editor) and the comments from the Associate Editor are included at the end of this email for your reference. As you will see, the reviewers and the Editors have raised some concerns with your manuscript and we would like to invite you to revise your manuscript to address them.

We do not allow multiple rounds of revision so we urge you to make every effort to fully address all of the comments at this stage. If deemed necessary by the Associate Editor, your manuscript will be sent back to one or more of the original reviewers for assessment. If the original reviewers are not available we may invite new reviewers. Please note that we cannot guarantee eventual acceptance of your manuscript at this stage.

To submit your revision please log into <http://mc.manuscriptcentral.com/prsb> and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions", click on "Create a Revision". Your manuscript number has been appended to denote a revision.

When submitting your revision please upload a file under "Response to Referees" - in the "File Upload" section. This should document, point by point, how you have responded to the

reviewers' and Editors' comments, and the adjustments you have made to the manuscript. We require a copy of the manuscript with revisions made since the previous version marked as 'tracked changes' to be included in the 'response to referees' document.

Your main manuscript should be submitted as a text file (doc, txt, rtf or tex), not a PDF. Your figures should be submitted as separate files and not included within the main manuscript file.

When revising your manuscript you should also ensure that it adheres to our editorial policies (<https://royalsociety.org/journals/ethics-policies/>). You should pay particular attention to the following:

Research ethics:

If your study contains research on humans please ensure that you detail in the methods section whether you obtained ethical approval from your local research ethics committee and gained informed consent to participate from each of the participants.

Use of animals and field studies:

If your study uses animals please include details in the methods section of any approval and licences given to carry out the study and include full details of how animal welfare standards were ensured. Field studies should be conducted in accordance with local legislation; please include details of the appropriate permission and licences that you obtained to carry out the field work.

Data accessibility and data citation:

It is a condition of publication that you make available the data and research materials supporting the results in the article. Datasets should be deposited in an appropriate publicly available repository and details of the associated accession number, link or DOI to the datasets must be included in the Data Accessibility section of the article (<https://royalsociety.org/journals/ethics-policies/data-sharing-mining/>). Reference(s) to datasets should also be included in the reference list of the article with DOIs (where available).

In order to ensure effective and robust dissemination and appropriate credit to authors the dataset(s) used should also be fully cited and listed in the references.

If you wish to submit your data to Dryad (<http://datadryad.org/>) and have not already done so you can submit your data via this link [http://datadryad.org/submit?journalID=RSPB&manu=\(Document not available\)](http://datadryad.org/submit?journalID=RSPB&manu=(Document%20not%20available)), which will take you to your unique entry in the Dryad repository.

If you have already submitted your data to dryad you can make any necessary revisions to your dataset by following the above link.

For more information please see our open data policy <http://royalsocietypublishing.org/data-sharing>.

Electronic supplementary material:

All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI. Please try to submit all supplementary material as a single file.

Online supplementary material will also carry the title and description provided during

submission, so please ensure these are accurate and informative. Note that the Royal Society will not edit or typeset supplementary material and it will be hosted as provided. Please ensure that the supplementary material includes the paper details (authors, title, journal name, article DOI). Your article DOI will be 10.1098/rspb.[paper ID in form xxxx.xxxx e.g. 10.1098/rspb.2016.0049].

Please submit a copy of your revised paper within three weeks. If we do not hear from you within this time your manuscript will be rejected. If you are unable to meet this deadline please let us know as soon as possible, as we may be able to grant a short extension.

Thank you for submitting your manuscript to Proceedings B; we look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Best wishes,

Proceedings B
mailto: proceedingsb@royalsociety.org

Associate Editor
Board Member: 1
Comments to Author:

Both reviewers were positive about the question, approach, and results. Nevertheless both have concerns about the presentation of the manuscript and the interpretation of results. In particular, the methods and results are exceedingly brief right now; the main text should be fleshed out more so the reader is not required to tap into the supplement for even a basic understanding of what was done. Given the broad readership of the journal, the introduction can do more to justify how a better understanding of diet switches is important for symbiosis more generally and also motivate the macro-evolutionary patterns tested here with micro-(co)evolutionary mechanisms. Notably both reviewers point out the need for more caution in interpreting causality given the inherently correlative approach taken.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)

In their manuscript, Bell-Robert et al conducted a meta-analysis to test the association of dietary traits with the presence of endosymbionts in plant sap-feeding insects. In addition to providing a compilation of information on the phylogeny, symbionts, and nutrition status of the sub-order Auchenorrhyncha, the authors found that the primary endosymbiont *Sulcia* is associated with xylem-feeding, whereas complementary symbionts are not directly associated with dietary switches.

In my opinion, the major contribution made by this study was to comprehensively quantify and trace back the evolutionary trajectory of symbiont loss in a large sub-order of plant sap-feeding insects. However, there are several points that require further investigation:

Major points:

- To further complement their results on the loss of the primary endosymbiont *Sulcia* (L100-105, L122-128); the authors could perform a phylogenetic analysis of available symbiont sequences to

confirm a host-symbiont co-diversification (Moran et al, AEM, 2005) with the set of insects the authors focused on.

- I think that the discussion of non-dietary influences in L177-184 falls too short to explore alternative explanatory factors that the authors do not consider with their analysis. Temperature resistance has been for example suggested to favor replacement of primary endosymbionts (Moran and Yun, PNAS, 2015). For cicadas, it has been also suggested that complementary symbionts were potentially outcompeted by pathogenic fungi (Matsuura et al, PNAS, 2018). There might be also differential costs for the host in maintaining symbionts.
- The Method section should be expanded (e.g., with parts of the supplemental text) to explain in more detail how the results were generated.

Minor points:

- The authors largely apply correlative and statistical models, yet they claim in several parts of the manuscript to investigate the causality of symbiont switches (L17, L44).
- The part of the discussion (L141-144) is a strong statement and should be tuned down in my opinion, because the authors focus on the correlation with dietary switches and not other host traits.
- To facilitate reading of Figure 1, the color legends could be arranged under each other in the same order as the bands (from inside to outside). A color legend for the branch lines of the phylogeny would be also helpful.
- The legends of Figure 1 and 2 could elaborate on the meaning and interpretation of the pie charts in the figures.
- The legend in Figure S4 should reference Figure S3 in L148.

Referee: 2

Comments to the Author(s)

In this study, Bell-Roberts et al. generate a phylogeny and perform ancestral state reconstructions to investigate the relationship between symbiont composition and host diet in Auchenorrhynca plant sap- feeders. I think the general question is interesting to think about given that obligate symbiont function is closely tied with host diet and would be of interest to those studying co-obligate symbiotic systems. I understand that manuscripts to this journal are supposed to be concise, but as written I think some information is lacking. Specifically (1) a broader context for why the work is being done and how this relates to other symbiotic systems and (2) expanded details on some methods since many of the conclusions rely heavily on a few parts of the methods which are unclear and (3) perhaps some adjustments to figures to help readers interpret the results.

It is a little unclear between what taxa were used for trait analysis, sequences and phylogeny reconstruction, and how these data overlap. For further transparency, it would be useful to include the accession numbers for sequence data in addition to the traits and references included for each species. In the event that different gene sequences and traits were pulled from multiple species to represent a single genus, this would be important to know for interpreting the results. And clearly indicate which data were included in the end for the 145 genera ancestral state reconstruction analysis (full vs. pruned? in the methods it mentions having a final database of 165 genera; line 77). In addition to adding this information as a supplemental table that could be included in the electron appendix (or added to the authors' github), please elaborate in the supplementary text (see comments below).

Figures: I generally think the figures are good, though a bit difficult to compare the information across figures and some of the information present is redundant between figures. It might be better to present all of the information in a single figure, which I know may be a lot of

information. A suggestion: stick with the standard tree format as in fig 2 (the circle phylogeny is a little difficult to read/interpret with the trait data), but highlight the branches with colors to represent likely companion symbiont status with the pie charts from fig 2 (or a combination of fill color and borders for branches to represent primary symbiont status and predicted diet, with pie charts indicating companion status) and columns along the tips of the phylogeny each colored to indicate the relevant traits (primary endosymbiont, companion symbiont, host diet, xylem and non-xylem (or shapes combined with fill color to indicate the last two)). This way, the phylogeny can be compressed a little to allow space for complete taxa name (if used at the species level) and the entire relevant trait data is easily visualized on the same phylogeny. A more minor point, it might be less cluttered to just use squares/rectangles in columns rather than other shapes that overlap significantly (or don't add much more info) and to not use the combination of red and green in a single pie chart. It would be useful to have at least one phylogeny that was completely readable (a lot of the labels overlap). It looks like only genera were included as tip names and only in Fig S2 & Fig S4. See first comment under supplementary materials regarding sequence data and taxa included

Minor comments:

- Line 44: The authors suggest a causal relationship for symbiont switches based on changes in host diet, but it is really a correlation at this point since it would be hard to determine causality in this system
- line 72. "COI" is listed twice, should be replaced with 18S.
- line 80-81. Figure 1 also lists moss and predatory, but these trait states were not included in the analysis. Can this decision be elaborated in the methods?
- line 130: lower case "A" in analysis
- check reference formatting

Supplementary materials:

- Line 14: "supermatrix of 832" I assume this means 832bp, does this include gaps? How much missing data if any? Did sequence for all five genes for a given taxa represent the same individual or species or were genes "mixed-and-matched" to compile a sort of consensus gene set per genus? If the data were a composition of sequences, this could impact the phylogenetic estimation and ancestral reconstruction.
- Line 14-17. How was the substitution model selected and were the data partitioned in any way using different substitution model parameters especially considering that the data represent mitochondria, ribosomal, and nuclear genes?
- Line 17-18: why weren't any closer outgroup taxa used such as any member of Sternorrhyncha when estimating the phylogeny?
- Part (a) phylogeny: I am curious why the authors didn't use a Bayesian approach, a program such as BEAST (or in RevBayes) can simultaneously estimate phylogeny and an ultrametric tree with time calibrated divergence times. This approach may be better suited for inferring any ancestral states across the phylogeny
- Line 61: when pruning the phylogeny, did tips represent individual species that matched the same taxa for trait data, or were any members within a genus all considered equal for sequence data in addition to trait data?
- Line 94-96: Does this mean that the models for trait evolution were run on the ML phylogeny to provide support for ancestral states or run across each of the 100 trees that formed the ML phylogeny with bootstrapped values? This is a bit confusing as written.
- Line 105-107. This mentions the influence of species, but the analyses presented were conducted on genera. Does this comment refer to the replacement of individual species to represent a genus for the analysis or the removal of a subset of genera for the analyses?

Author's Response to Decision Letter for (RSPB-2019-0065.R0)

See Appendix A.

RSPB-2019-0065.R1 (Revision)

Review form: Reviewer 1

Recommendation

Accept as is

Scientific importance: Is the manuscript an original and important contribution to its field?

Good

General interest: Is the paper of sufficient general interest?

Good

Quality of the paper: Is the overall quality of the paper suitable?

Excellent

Is the length of the paper justified?

Yes

Should the paper be seen by a specialist statistical reviewer?

No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.

No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?

Yes

Is it clear?

Yes

Is it adequate?

Yes

Do you have any ethical concerns with this paper?

No

Comments to the Author

The authors addressed my previous comments and concerns satisfactorily.

Decision letter (RSPB-2019-0065.R1)

23-Apr-2019

Dear Dr Werner

I am pleased to inform you that your manuscript entitled "Match and Mismatch between Dietary Switches and Microbial Partners in Plant Sap-Feeding Insects" has been accepted for publication in Proceedings B.

You can expect to receive a proof of your article from our Production office in due course, please check your spam filter if you do not receive it. PLEASE NOTE: you will be given the exact page length of your paper which may be different from the estimation from Editorial and you may be asked to reduce your paper if it goes over the 10 page limit.

If you are likely to be away from e-mail contact please let us know. Due to rapid publication and an extremely tight schedule, if comments are not received, we may publish the paper as it stands.

If you have any queries regarding the production of your final article or the publication date please contact procb_proofs@royalsociety.org

Open Access

You are invited to opt for Open Access, making your freely available to all as soon as it is ready for publication under a CCBY licence. Our article processing charge for Open Access is £1700.

Corresponding authors from member institutions

(<http://royalsocietypublishing.org/site/librarians/allmembers.xhtml>) receive a 25% discount to these charges. For more information please visit <http://royalsocietypublishing.org/open-access>.

Your article has been estimated as being 8 pages long. Our Production Office will be able to confirm the exact length at proof stage.

Paper charges

An e-mail request for payment of any related charges will be sent out after proof stage (within approximately 2-6 weeks). The preferred payment method is by credit card; however, other payment options are available

Electronic supplementary material:

All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

Thank you for your fine contribution. On behalf of the Editors of the Proceedings B, we look forward to your continued contributions to the Journal.

Sincerely,

Proceedings B

<mailto:proceedingsb@royalsociety.org>

Appendix A

Associate Editor

Board Member: 1

Comments to Author:

Both reviewers were positive about the question, approach, and results. Nevertheless both have concerns about the presentation of the manuscript and the interpretation of results. In particular, the methods and results are exceedingly brief right now; the main text should be fleshed out more so the reader is not required to tap into the supplement for even a basic understanding of what was done. Given the broad readership of the journal, the introduction can do more to justify how a better understanding of diet switches is important for symbiosis more generally and also motivate the macro-evolutionary patterns tested here with micro-(co)evolutionary mechanisms. Notably both reviewers point out the need for more caution in interpreting causality given the inherently correlative approach taken.

>> We are grateful for the two very useful and constructive reviews and are pleased that both reviewers are positive about our questions, approach and results. Following their suggestions and the suggestions of the AE, we have (i) expanded the methods, using text previously contained in the supplement (lines 72-165), (ii) further justified why a better understanding of potential links between diet and symbiont status is important (lines 35-42), (iii) further expanded our discussion of potential micro-(co)evolutionary mechanisms and (lines 259-262, 266-268) and (iv) exercised more caution in inferring causality based on our macro-evolutionary patterns (lines 15, 45-46, 219-229).

Please see below for our point-by-point responses to the reviewer comments. Below these responses, we also include a track-change version of our main text, indicating the changes compared to our initial submission.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)

In their manuscript, Bell-Robert et al conducted a meta-analysis to test the association of dietary traits with the presence of endosymbionts in plant sap-feeding insects. In addition to providing a compilation of information on the phylogeny, symbionts, and nutrition status of the sub-order Auchenorrhyncha, the authors found that the primary endosymbiont *Sulcia* is associated with xylem-feeding, whereas complementary symbionts are not directly associated with dietary switches.

In my opinion, the major contribution made by this study was to comprehensively quantify and trace back the evolutionary trajectory of symbiont loss in a large sub-order of plant sap-feeding insects. However, there are several points that require further investigation:

Major points:

- To further complement their results on the loss of the primary endosymbiont *Sulcia* (L100-105, L122-128); the authors could perform a phylogenetic analysis of available symbiont sequences to confirm a host-symbiont co-diversification (Moran et al, AEM, 2005) with the set of insects the authors focused on.

*>> We thank the reviewer for their suggestion. At an early stage in our research, we had considered performing a co-diversification analysis but judged that, given our research question, this analysis would provide only limited new insight beyond previously published work. Specifically, a key study of the bacterial symbiosis in Auchenorrhyncha, Moran et al. (2005), demonstrates strict codiversification with *Sulcia*, using representative insect taxa of all main clades of the Auchenorrhyncha.*

*In our current work, in contrast, we are interested primarily in the relationship between symbiosis loss/gain/identity and diet, rather than in the congruence of host and symbiont phylogenies. While we agree with Reviewer 1 that our current study provides considerable new insight into the patterns of loss of *Sulcia* within the Auchenorrhyncha and its (lack of) association with diet, given the comprehensive sampling in Moran et al. 2005, it is very unlikely we would have come to a substantially different conclusion regarding strict *Sulcia*-insect codiversification.*

*Furthermore, while new sequence *Sulcia* sequences have become available since, these are often for different regions of the 16S gene, and are not available for all host taxa in our study. The same argument applies to potential phylogenetic congruence between companion symbionts and (subsets of the) auchenorrhynchan insects, where there are even fewer data.*

*For all these reasons, we decided to study the patterns of presence and absence of the symbiont (for *Sulcia*), taxonomic identity (for the companion symbionts) and its relation to diet, rather than analyse the potential codiversification of symbionts and hosts.*

We are, however, encouraged by the increasing numbers of whole genome datasets for the symbionts and expect that within a few years, there will be sufficient data to adopt a phylogenomic approaches to investigate codiversification. This will undoubtedly reveal a much richer and more detailed picture of Sulcia evolution than currently possible. We now suggest this as future research (line 285-287). The insect phylogeny we have made available online (on our Github page, and upon publication also on Dryad) will be a potentially useful resource to facilitate these co-diversification analyses from the host side.

- I think that the discussion of non-dietary influences in L177-184 falls too short to explore alternative explanatory factors that the authors do not consider with their analysis. Temperature resistance has been for example suggested to favor replacement of primary endosymbionts (Moran and Yun, PNAS, 2015). For cicadas, it has been also suggested that complementary symbionts were potentially outcompeted by pathogenic fungi (Matsuura et al, PNAS, 2018). There might be also differential costs for the host in maintaining symbionts.

>> We agree, and have expanded this paragraph to include the specific example of replacement by thermally-tolerant symbionts, as recommended by Reviewer 1 (new line 269-261). We also alert readers to the likely costs of symbiont replacement, especially where the incoming symbiont may have pathogenic traits (line 266-268).

- The Method section should be expanded (e.g., with parts of the supplemental text) to explain in more detail how the results were generated.

>> We have moved almost all the previous supplemental text to the main paper, now explaining our steps in detail in the main text. To create the space to do so, we have condensed the original two main figures into a single figure, following a suggestion of Reviewer 2.

Minor points:

- The authors largely apply correlative and statistical models, yet they claim in several parts of the manuscript to investigate the causality of symbiont switches (L17, L44).

>> We have amended the wording at lines 17 and 44 to avoid the assumption of causality (lines 15, 45-46).

- The part of the discussion (L141-144) is a strong statement and should be tuned down in my opinion, because the authors focus on the correlation with dietary switches and not other host traits.

>> We have softened the language at these lines to emphasize that our findings are correlative and to emphasize our results provide only limited support for influence on host diets (lines 219-229).

- To facilitate reading of Figure 1, the color legends could be arranged under each other in the same order as the bands (from inside to outside). A color legend for the branch lines of the phylogeny would be also helpful.

- The legends of Figure 1 and 2 could elaborate on the meaning and interpretation of the pie charts in the figures.

>> We have implemented these changes, and substantially changed the figures – also following comments from Reviewer 2. We now explain the interpretation of the pie charts and figure colours in more detail, and include full legends (figure legends lines 501-513).

- The legend in Figure S4 should reference Figure S3 in L148.

We have now corrected this to refer to the correct supplementary figure.

Referee: 2

Comments to the Author(s)

In this study, Bell-Roberts et al. generate a phylogeny and perform ancestral state reconstructions to investigate the relationship between symbiont composition and host diet in Auchenorrhynca plant sap-feeders. I think the general question is interesting to think about given that obligate symbiont function is closely tied with host diet and would be of interest to those studying co-obligate symbiotic systems.

>> We are glad you find our study interesting. We have amended our introduction (line 62-63) and discussion (line 221-223) to reflect that we agree that precisely this close functional link is one of the reasons that the Auchenorrhyncha is such an interesting model system for this question.

I understand that manuscripts to this journal are supposed to be concise, but as written I think some information is lacking. Specifically (1) a broader context for why the work is being done and how this relates to other symbiotic systems and (2) expanded details on some methods since many of the conclusions rely heavily on a few parts of the methods which are unclear and (3) perhaps some adjustments to figures to help readers interpret the results.

>> (1) We have expanded the introductory text to illustrate that symbiont shifting has occurred in a diversity of different groups and to emphasize the potential evolutionary significance of these events (lines 35-42). We have also reworded the text in our discussion to provide greater emphasis on the open question of the generalizability of our findings to other nutritional symbioses and to protective symbioses (276-282). We trust that these amendments provide readers with a broader context for our study.

>> (2) Following both reviewers' comments we have entirely moved the supplementary methods to the main text (lines 72-165).

>> (3) We have combined both figures into a single main figure (see also detailed comments below).

It is a little unclear between what taxa were used for trait analysis, sequences and phylogeny reconstruction, and how these data overlap. For further transparency, it would be useful to include the accession numbers for sequence data in addition to the traits and references included for each species. In the event that different gene sequences and traits were pulled from multiple species to represent a single genus, this would be important to know for interpreting the results.

>> Reviewer 2 is correct that in our genus-level phylogeny sequences were pulled from multiple species to represent a genus. We used this approach because we set out to create a phylogeny with tip states representing genera, rather than species (see also our reply to your comment on Line 14 for more details). We now provide a full table online indicating all the accession numbers (*Bell-Roberts_Accession_Table.csv* – available on our Github page, and upon publication will be made available on Dryad). Given that the traits we study tend to not vary within genera, we used only a single species to represent each genus. Our trait database is also fully available online (*Master_db_standardised_unique_with_taxonomy.csv*). We now clarify these decision in our new main text methods (lines 74-81, 97-99).

And clearly indicate which data were included in the end for the 145 genera ancestral state reconstruction analysis (full vs. pruned? in the methods it mentions having a final database of 165 genera; line 77). In addition to adding this information as a supplemental table that could be included in the electron appendix (or added to the authors' github), please elaborate in the supplementary text (see comments below).

>> We thank the reviewer for alerting us that this section was unclear. Our full genus-level phylogeny contains 824 tips and our trait database contains 162 genera. The overlap between these was 145 and this is what we included in our comparative analyses in our initial submission (in other words, we pruned both the dataset and the tree to this set of 145 genera). For 142 of these, we had data on primary symbiont (*Sulcia*) status. In order to facilitate representation of our results in a single figure, as suggested below by Reviewer 2, we have now decided to limit our analyses to only those 142 genera – dropping 3 genera compared to our initial submission. This change has affected none of our results, other than minimally changing the log-likelihood and AICc-values reported in Table S2.

We have now clarified the relationships between the different genera numbers and datasets in the new main methods text (lines 115-119). In addition, as suggested by the reviewer, we have added a supplementary table (Table S4), indicating all the 142 analysed genera. In this table, the genera are in the same order as in the main text and supplementary figures, helping the reader to identify genera where these are difficult to read in the supplementary figures. Lastly, we include the full, unpruned, data files and phylogeny online.

Figures: I generally think the figures are good, though a bit difficult to compare the information across figures and some of the information present is redundant between figures. It might be better to present all of the information in a single figure, which I know may be a lot of information. A suggestion: stick with the standard tree format as in fig 2 (the circle phylogeny is a little difficult to read/interpret with the trait data), but highlight the branches with colors to represent likely companion symbiont status with the pie charts from fig 2 (or a combination of fill color and borders for branches to represent primary symbiont status and predicted diet, with pie charts indicating companion status) and columns along the tips of the phylogeny each colored to indicate the relevant traits (primary endosymbiont, companion symbiont, host diet, xylem and non-xylem (or shapes combined with fill color to indicate the last two). This way, the phylogeny can be compressed a little to allow space for complete taxa name (if used at the species level) and the entire relevant trait data is easily visualized on the same phylogeny. A more minor point, it might be less cluttered to just use squares/rectangles in columns rather than other shapes that overlap significantly (or don't add much more info) and to not use the combination of red and green in a single pie chart. It would be useful to have at least one phylogeny that was completely readable (a lot of the labels overlap). It looks like only genera were included as tip names and only in Fig S2 & Fig S4. See first comment under supplementary materials regarding sequence data and taxa included

>> We thank the reviewer for these very useful suggestions. They have allowed us to both represent our key results more clearly, while saving a lot of space. We have followed the reviewer suggestions and combined the two main text figures into a single figure. As suggested, in this new figure, the correlated evolution model of diet and primary symbiont evolution (previously Figure 2) is represented by pie charts, while the branch colours represent the ASR of companion symbiont status (previously Figure 1 pie charts). We represent the seven diet

transitions (xylem-> phloem or xylem-> parenchyma) with star icons. The trait data analysed are represented by coloured bands, similar to previously Figure 1 (although now in the standard rather than in the circular tree format). We have also ensured that red and green don't appear together anymore, providing better colour-blind readability. The three key ASRs (diet, primary symbiont, secondary symbiont) are now each printed with full pie charts in separate supplementary figures in the same orientation, facilitating comparison with the main figure (figures S2, S4 and S5). Lastly, for these supplementary phylogeny figures, we now provide online high definition, and fully zoomable versions (in addition to the ordered genus list now in Table S4).

Minor comments:

- Line 44: The authors suggest a causal relationship for symbiont switches based on changes in host diet, but it is really a correlation at this point since it would be hard to determine causality in this system.

We agree and have corrected this (line 45-46).

- line 72. "COI" is listed twice, should be replaced with 18S.

We have corrected this (line 76).

- line 80-81. Figure 1 also lists moss and predatory, but these trait states were not included in the analysis. Can this decision be elaborated in the methods?

>> We did not include these two states in our model because the focus of our analysis is the diet states as present in the Auchenorrhyncha, and moss/predatory are only present in the outgroup. For purposes of our diet ASR, we treated diet in the outgroup as missing data. We now represent this in the main paper Figure 1 by showing diet as missing data for the outgroup, and clarify this in the new methods section (line 127-128).

- line 130: lower case "A" in analysis

We have corrected this.

- check reference formatting

Supplementary materials:

- Line 14: "supermatrix of 832" I assume this means 832bp, does this include gaps? How much missing data if any? Did sequence for all five genes for a given taxa represent the same individual or species or where genes "mixed-and-matched" to compile a sort of consensus gene set per genus? If the data were a composition of sequences, this could impact the phylogenetic estimation and ancestral reconstruction.

>> We apologise for the word 'genera' missing here. We also found out that we used an outdated genera number in this sentence, and have now updated this to reflect the correct 824 genera and 7,243 bp (line 81-83). In order to create our genus-level phylogeny we used the referenceDownload method in phylogenerator. Within each genus, this method uses a set of reference sequences for each marker gene (now made available online), to select the highest quality sequence representing that genus (see also Pearse and Purvis 2013, ref 39, for more detail). This means that a genus is potentially (but not necessarily) represented by sequences from different species. For full transparency, we now provide a table online indicating all accession numbers and missing data per genus (Bell-Roberts_Accession_Table.csv – available online). We have also made our analysed alignment (full_alignment_analysed.fasta) and constraint tree (till_subfamily_aucho_NCBI_taxonomy_constraint_phylogeny.phy) available online, allowing readers to rerun our analysis using other methods for phylogenetic inference.

- Line 14-17. How was the substitution model selected and were the data partitioned in any way using different substitution model parameters especially considering that the data represent mitochondria, ribosomal, and nuclear genes?

>> Following the recommendations in the RAxML manual and in Stamatakis (2006), we used the GTRCAT substitution model because it is computationally highly efficient, while allowing for rate heterogeneity among sites (important given the different types of sequences in our alignment). We did not partition our data. In line with recent research suggesting model selection is generally unlikely to substantially affect phylogenetic inference (Abadi et al., 2019, Nature Communications), we did not otherwise perform model selection. We note that the Abadi et al. conclusion is even more likely to apply in our case, given that our tree search was constrained by taxonomy (see for details in Methods Section A). We now make our alignment files and constraint tree available for readers wanting to reanalyse our data using different substitution models.

- Line 17-18: why weren't any closer outgroup taxa used such as any member of Sternorrhyncha when estimating the phylogeny?

>> *The high-level phylogeny of the Hemiptera is not well resolved yet. We selected our outgroups based on Cryan and Urban (2012) Syst. Entomol. (ref 20). Here, the Heteroptera and the Coleorrhyncha are the closest two sister groups to the Auchenorrhyncha. In this reference, as well as in Li et al. (2017) Proc Soc B (ref 21), the Sternorrhyncha are inferred to be more distantly-related to the Auchenorrhyncha. For these reasons we decided to include Ceratocombus (Heteroptera) and Xenophyes (Coleorrhyncha) as outgroups.*

- Part (a) phylogeny: I am curious why the authors didn't use a Bayesian approach, a program such as BEAST (or in RevBayes) can simultaneously estimate phylogeny and an ultrametric tree with time calibrated divergence times. This approach may be better suited for inferring any ancestral states across the phylogeny

>> *Earlier on during this project, we have considered Bayesian approaches to our phylogeny building. We have decided in favour of our current approach, for a number of reasons.*

A Bayesian approach would potentially have allowed us to simultaneously estimate ancestral states and the phylogeny itself. However, given our questions, we specifically wanted to be able to use hidden rate models (Beaulieu et al. 2013, ref 50) and correlated models of evolution (Pagel et al. 1993, ref 49) for our comparative analyses. This was important, because given the very long time period of our study (>250MY) variation in the gain/loss rate of the Sulcia symbiosis was highly likely, and indeed found using the Beaulieu et al. method (Figure S3). Similarly, correlated evolution models were important in testing our hypothesis of diet-symbiont type correlations. Since, as far as we are aware, neither of these methods have been implemented in BEAST (or RevBayes), we needed to separate our tree inference from our comparative analyses.

We then decided to infer our phylogeny in a maximum-likelihood framework (RAxML), because Bayesian frameworks are computationally considerably more intensive than RAxML, which has been optimised for large datasets (like our full 824 tip phylogeny). RAxML in combination with r8s, has been well established as a robust strategy to calibrate divergence times and generate an appropriate ultrametric tree, supported by >8000 and >1300 references. We also note that we primarily time-calibrated our trees for purposes of ease of interpretation and visualisation. Analysing our data on the phylogram rather than on our ultrametric tree, results in qualitatively the same conclusions (not shown in the paper). This is in line with previous results, that branch lengths do not always substantially affect the conclusions of ancestral state reconstructions (Cusimano en Renner, 2014), and – alongside our presented sensitivity analyses – further suggest that the specific method utilised is unlikely to fundamentally affect our conclusions.

We now provide our full alignment files to allow readers to replicate our analyses and to apply other frameworks, including Bayesian frameworks such as implemented in BEAST or RevBayes.

- Line 61: when pruning the phylogeny, did tips represent individual species that matched the same taxa for trait data, or were any members within a genus all considered equal for sequence data in addition to trait data?

Both our trait database and our phylogeny were compiled only at the genus level (see also our replies above). We now indicate this in our main text methods (line 74-81, 97-99).

- Line 94-96: Does this mean that the models for trait evolution were run on the ML phylogeny to provide support for ancestral states or run across each of the 100 trees that formed the ML phylogeny with bootstrapped values? This is a bit confusing as written.

We agree this was confusing and have now rewritten this sentence (line 150-152). We meant to say the latter (models were run on each of the 100 trees that formed the ML phylogeny with bootstrapped values).

- Line 105-107. This mentions the influence of species, but the analyses presented were conducted on genera. Does this comment refer to the replacement of individual species to represent a genus for the analysis or the removal of a subset of genera for the analyses?

>> *This refers to the removal of a subset of genera for analysis. We thank the reviewer for highlighting that this paragraph was confusing. Paterno et al. 2018, Methods in Ecology and Evolution (Ref nr. xxx) refers to the concept of 'species sampling uncertainty' because typically in comparative analyses, tips represent species. We simply forgot to change this to genus/taxon sampling uncertainty to accurately reflect our analyses in this paper. We have now changed the text to reflect the fact that our phylogeny and all of our analyses represent only genera as tips (lines 158-161, 215-216).*

References cited in this review

Abadi S, Azouri D, Pupko T, Mayrose I. 2019. Model selection may not be a mandatory step for phylogeny reconstruction. *Nature communications*: 1–37.

- Beaulieu JM, O'Meara BC, Donoghue MJ. 2013 Identifying Hidden Rate Changes in the Evolution of a Binary Morphological Character: The Evolution of Plant Habit in Campanulid Angiosperms. *Syst. Biol.* 62, 725–737.
- Cryan JR, Urban JM. 2012 Higher-level phylogeny of the insect order Hemiptera: is Auchenorrhyncha really paraphyletic? *Syst. Entomol.* 37, 7–21
- Cusimano N, Renner SS. 2014 Ultrametric trees or phylograms for ancestral state reconstruction: Does it matter? *Taxon* 63, 721–726.
- Dolling WR. 1991 *The Hemiptera*. London, UK: The Natural History Museum Publications.
- Li H, Leavengood JM, Chapman EG, Burkhardt D, Song F, Jiang P, Liu J, Zhou X, Cai W. 2017 Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. *Proc. R. Soc. B Biol. Sci.* 284, 20171223
- Pagel M. 1994 Detecting Correlated Evolution on Phylogenies: A General Method for the Comparative Analysis of Discrete Characters. *Proc. R. Soc. B Biol. Sci.* 255, 37–45
- Paterno GB, Penone C, Werner GDA. 2018 sensiPhy: An r-package for sensitivity analysis in phylogenetic comparative methods. *Methods Ecol. Evol.* 9, 1461–1467
- Pearse WD, Purvis A. 2013 phyloGenerator: an automated phylogeny generation tool for ecologists. *Methods Ecol. Evol.* 4, 692–698.
- Stamatakis, "Phylogenetic models of rate heterogeneity: a high performance computing perspective," *Proceedings 20th IEEE International Parallel & Distributed Processing Symposium*, 2006