**Subject:** Decision on your PeerJ submission: "Biogeographic barriers drive co-diversification within associated eukaryotes of the Sarracenia alata pitcher plant system" (#2015:09:6562:0:0:REVIEW)

|  |
| --- |
| **Peer**J |
| Thank you for your submission to PeerJ. I am writing to inform you that in my opinion as the Academic Editor for your article, your manuscript "Biogeographic barriers drive co-diversification within associated eukaryotes of the *Sarracenia alata* pitcher plant system" (#2015:09:6562:0:0:REVIEW) requires some minor revisions before we could accept it for publication.  ###  Thank you for your consideration of our manuscript for publication in PeerJ. We sincerely appreciate the efforts and comments from you and the reviewers, and have revised the manuscript following the suggestions. To facilitate the revision process, we have recorded our responses just below the reviewer’s comments (separated by pound signs) and have indicated the location in the manuscript where they have been addressed.  ###  The comments supplied by the reviewers on this revision are pasted below. My comments are as follows:  **Editor's comments**  The major issues raised by Reviewer 1 will likely require the most work to address. What can be inferred regarding co-diversification without a quantitative comparison of differentiation relative to the host plant and (more challengingly) relative to the external biota? I might suggest at least adding the results for the host plant to Figure 4, assuming they are already at hand.   ###  Results from the chloroplast marker from *Sarracenia alata* have been added to Figure 4, with all values being significant. This is a challenging question to address, especially in the absence of a quantitative framework for community diversification and without data on the external biota. *Sarracenia alata* displays strong population structure, and recent studies have highlighted the landscape as a major driver of diversification in the plant (Zellmer et al. 2012), to the point where data suggests that populations on either side of the Mississippi River are evolutionary lineages (Carstens and Satler 2013). From the 454 amplicon resequencing of pitcher plant fluid, many OTUs show congruent population genetic structure with the pitcher plant. The challenge is teasing apart whether this is solely due to response to landscape processes or if ecological association between the taxon and plant have helped shape the shared population genetic structure. At a minimum, a chi-squared goodness of fit test shows that more taxa have significant values than would be expected by chance along. We have included this analysis in the manuscript (Lines 267–271, 344–346; Table 2). We infer that ecological association has played a role in shaping population structure in the taxa found within the pitcher fluid, and have revised the manuscript to be more clear about the conclusions that can be made from the results.  ###  Some of the minor comments are also worth pointing out for special attention. I do agree with Reviewer 3 that the use of the term 'metagenomics' for environmental sequencing of a 275bp locus is misleading. I am also not clear why the analysis ignores taxa with less than 3 representatives on one side of the river; doesn't this just downwardly bias estimates of structure at the regional scale, and unnecessarily discard evidence for differentiation?  ###  We have reworded the term to say 454-amplicon based metagenomics, as suggested by Reviewer 3 (Line 36). In regards to discarding taxa with less than 3 sequences on either side of the river, we acknowledge that this is an arbitrary cutoff, but wanted to analyze OTUs that had at least multiple representative sequences and enough (cutoff of 10 sequences total) for various population genetic analyses (to assess both within and among population genetic variation). This may have biased our results, but we also wanted to make sure we analyzed taxa with enough sequence representation for downstream analysis. We now discuss this further in the manuscript (Lines 225–231).  ###  In my own reading, I had a number of additional comments and questions that I would ask the authors to address.  1. Why not exclude the April-May samples from Lake Ramsay, or at least segregate in the analyses, in order to have comparable diversity measurements across sites? See also the comment on sampling times from Reviewer 1.  ###  We thank the AE for the comment, and agree that sampling should have been consistent across sites to have comparable measures. We have now clarified the sampling scheme (Lines 150–158) and highlight that for within locality sampling, we follow the AE’s suggestion and only analyze the samples from the same sampling times (June and August samples; Lines 181–182; 285–289). We reanalyzed the reduced data set from Lake Ramsey; the results can now be seen in Figures 2 and 3. For the global analysis, when we required an OTU to contain sequences on either side of the Mississippi River, we retained all sequences from Lake Ramsey, as any OTUs sampled at other times of the year at LR restricted to those times would not be recovered at other sites, and would be subsequently discarded (Lines 220–225). Any OTUs that contained sequences during these additional time periods at Lake Ramsey were then stable over a longer period of time in these habitats, so their inclusion in our global analyses was warranted as that suggests they are stable members of this community.  ###  2. There is no surprise at seeing a negative Tajima's D when pooling data across populations, because allele frequency differentiation results in an excess of segregating sites. So, significance tests for D=0 do not say much; these values are best interpreted relative to D in the host plant. I believe this can be addressed together with the response to Reviewer 1's first major issue.  ###  We thank the AE for this comment and do agree that if these taxa comprise multiple populations, this could be one of the reasons why we have a large number of OTUs with negative Tajima’s *D* values. This is now stated in the text (Lines 323–324). We also present the Tajima *D* values for *S. alata* in Table 1, so comparisons can be made with the sampled OTUs.  ###  3. I am unclear as to why the allelic sorting analysis is applied to populations on either side of the Mississippi separately. Please clarify the rationale and/or show values for E & W combined.  ###  Since the Mississippi River is a known biogeographic barrier, and corresponds with the deepest split in the population tree of *S. alata* (Zellmer et al. 2012), we wanted to see if allelic sorting on either side of the river was higher than by chance. Our reasoning is that it takes a long time for alleles to sort to monophyly, and if this was the case, or if a reasonable amount of sorting was found (as indicated by the permutations), then that would provide evidence that the OTU has been isolated for a long time on either side of the Mississippi River, a pattern that we see with the plant (Lines 249–256).  ###  4. Lack of significance in phi\_CT seems counter to idea that a substantial amount of biogeographic differentiation is explained by the Mississippi River Instead, it seems that most of the differentiation is between populations instead. Is that not contrary to the conclusion on line 358?  ###  We thank the AE for pointing this out and have now clarified the text in the manuscript (Lines 361–364, 372–374). In a little fewer than half the taxa, we see population structure, both between populations (based on AMOVA) and on either side of the Mississippi River (based on GSI). Surprisingly, we don’t recover any significant values for the AMOVA when comparing region within total distribution, something that we might expect given the influence the Mississippi River has had on numerous species within this region. We now clarify this in the text so the inferences are better supported by the data.  ###  5. Please be sure to also include supplemental Table S2 in the next submission, so that it is possible to inspect the quantitative results that are shown schematically (at only one significance threshold) in Fig 4.  ###  This supplemental table is now included. We apologize for the error in our original submission.  ###  If you are willing to undertake these changes, please [submit your revised manuscript](https://peerj.com/manuscripts/6562/edit/" \t "_blank) (with any rebuttal information\*) to the journal within 45 days.   |  | | --- | | **\* Resubmission checklist:**  When resubmitting, in addition to any revised files (e.g. a clean manuscript version, figures, tables, which you will add to the "Primary Files" upload section), please also provide the following two items:   1. A rebuttal Letter: A single document where you address all the Editor and reviewers' suggestions or requirements, point-by-point. 2. A 'Tracked Changes' version of your manuscript: A document that shows the tracking of the revisions made to the manuscript. You can also choose to simply highlight or mark in bold the changes if you prefer.   Accepted formats for the rebuttal letter and tracked changes document are: DOCX (preferred), DOC, or PDF.  PeerJ does not offer copyediting, so please ensure that your revision is free from errors and that the English language meets our standards: uses clear and unambiguous text, is grammatically correct, and conforms to professional standards of courtesy and expression. |   Todd Vision  Academic Editor for PeerJ  **Reviewer Comments**  **Reviewer 1 (Anonymous)**  **Basic reporting**  This manuscript addresses whether pitcher plants and their associated inquiline community have co-diversified across their range. Although not explicitly addressed in the introduction, this research is closely related to the field of community genetics, in which each genotype of host plant may have an extended phenotype consisting of an associated community. The topic is an important one, as we are increasingly finding that genetic diversity in one species plays a large role in determining the dynamics of other associated species. The authors here are using novel molecular approaches to address this question, as they use genomics to examine all eukaryotes within pitcher plants.  **Experimental design**  I am satisfied with the experimental design.  **Validity of the findings**  Unfortunately, I do not believe that the data necessarily support the authors’ conclusions. This study would benefit greatly from the use of a null model to determine whether the observed patterns exceed what we would expect by chance alone. The evidence presented here suggests that about half of the eukaryotic species show evidence of diversification with the plant. Is 50% more than we would expect? Certainly we don’t expect zero association. If both the plants and the eukaryotes were distributed randomly, we would expect some areas of overlap between them. So what level then is considered significantly greater than random? A null model based on the distribution of both plants and the eukaryote community would help to determine this.  ###  We thank the reviewer for this comment and agree that this is a complicated pattern to explain. The American southeastern landscape is dominated by major rivers, and the Mississippi River is a well-characterized biogeographic barrier, driving isolation of lineages across many disparate taxonomic groups. In our study, we find a little less than half of the OTUs to share similar population genetic structure with the plant. Teasing apart the role that landscape processes and ecological association play in driving this isolation is challenging, and it may be that several of these OTUs are responding to changes in the landscape, and not necessarily because of interaction with the plant (Lines 378–390). We agree that developing a null model to help explain these results is ideal, but what that null model should be or how to test a null model with simulations is unclear.  At a minimum, we can ask the question of what number of OTUs would we expect to show significant results similar to the host plant if it were based on chance alone. Under a null model of no correlation between the host plant and the Eukaryotic community members we would expect to get a significant results based on chance 5% of the time (assuming α = 0.05). We thus used our observed number of significant OTUs to test whether the observed number was greater than the expected null value of 5%. To do this, we used a chi-squared goodness of fit test, and have added this to the manuscript (Lines 267–271, 344–346; Table 2). In particular, notice that for ΦST and GSI values on either side of the Mississippi River, there are more OTUs that show significant values than would be expected based on chance alone. This suggests that there is some significant association between various taxa and the plant (as discussed in the manuscript), and we infer that ecological interaction is helping to drive these patterns.  ###  It would also be helpful to know what the eukaryotic community is like in the neighborhood outside of the plants. Are these eukaryotes specialists in pitcher plants (probably true of some of them) or can they also be found in other habitats? If the latter, then eukaryotes and plants could have diversified in response to similar habitat effects, but not necessarily in response to one another.  ###  These are great points brought up by the reviewer, but unfortunately, we don’t have a great understanding of how these eukaryotes are distributed outside of the pitcher plant habitat. An issue confounding this study is that our taxonomic designations are made from BLAST searches, which precludes us from going into more detail on the specific organisms and where they are and are not found. For example, there are at least two different mite species that are restricted to the pitchers, but they do not contain any sequences on GenBank. Although a portion of our OTUs match sequences from mites, we cannot be certain that they are from either of those two species, although we suspect they are. This makes interpretation more challenging, as we know a community of organisms are restricted to within the pitchers, but without representative sequences on GenBank, we can only place the OTUs within their likely taxonomic group.  ###  Additional comments:  Line 33: what is the evidence that the inquiline community is mutualistic? There is some evidence that bacteria are beneficial to pitcher plants, but if so, then consumers of bacteria are likely to be parasitic.   ###  We have removed this “mutualistic” from the sentence, as many of the relationships displayed by the inquiline community and the plant would be characterized under a variety of terms (Lines 30–32).  ###  Line 75: I would argue that pollinators and plants are mutualists, but not symbiotic, as pollinators spend most of their time unassociated with the plant.  ###  We thank the reviewer for this comment and adopted this change in the manuscript (Lines 69–71).  ###  Line 94: Again, I would argue that this statement applies to the bacteria, but not necessarily the eukaryotes.  ###  We have removed the term “mutualists” from the sentence (see Lines 90–93), and updated it to show that many of the inquilines form complex relationships, rather than attributing these relationships in a specific group (as in mutualists).  ###  Line 164: The 28S subunit is only present in eukaryotes, so I am surprised to see bacteria listed as a taxonomic group in Fig. 2. As such, the figure is misleading, not only because of the presence of bacteria, but also because if one DID sequence the bacteria, their species diversity would likely greatly outnumber the eukaryotes in the community.   ###  We agree that this is a surprising result. Upon further inspection, 13 OTUs contained a closest BLAST match to bacteria (see Supplemental Material for details). The average number of sequences per OTU is 1.46, with a median value of 1, indicating most of these OTUs are singletons. The average percent identity is 84.99%, suggesting that sequence wise, these are not very close matches. Based on our BLAST criteria, these are matches and are presented as such in the paper, but we would believe this to either be (i) error in using BLAST and a taxonomic database for assigning taxonomy to a genetic cluster, or (ii) sequencing error. We would lean towards this being an issue with trying to assign taxonomy through a BLAST search, and maybe the OTU does not have a close relative in the NCBI database (as evident by the low percent sequence identity). Overall, this is a bit of a challenge to interpret, but since these were the results we received from our BLAST searches, felt it was important to report in the paper. We do note, however, that none of these OTUs were used in downstream population genetic analyses, and thus do not make a substantial contribution to the paper.  ###  Line 318-320: Please explain what these values mean biologically.  ###  We clarify the meaning of these values in the text (Lines 318–324). There is a 50 fold range in pi values among the OTUs, and an excessive number of segregating sites in the data sets (as seen from the Tajima *D* values). As discussed in the manuscript, possible explanations for these values include a rapid demographic expansion, or purifying selection on the marker (or linked gene). As pointed out by the AE, the Tajima *D* values could also be due to combining multiple populations prior to analysis within an OTU, and we now discuss this in the manuscript (Lines 323–324).  ###  **Reviewer 2 (Marianne Koller-Peroutka)**  **Basic reporting**  This article is written very clearly; the relevant literature is cited and introduction & background explains the particular importance of this study; Figures are appropriate;  **Experimental design**  The experimental design is planed and carried out very carefully and accurate.  **Validity of the findings**  This study is very innovative and represents a framework in co-evolution of eukaryotic inquilines in phytotelms of carnivorous pitcher plants. I highly recommend the acceptance of this article.  ###  Thank you very much!  ###  **Comments for the author**  I highly recommend the acceptance of this article. I only have one formal suggestion and one question: Line 397: to avoid confusions with the genus Sarracenia, the genus Solenopsis should be fully announced if mentioned for the first time in the text  ###  The full species name is now spelled out in the text (Lines 402–409).  ###  Line 400: S. invicta and S. invictus (female/male form) are these scientific names used synonymous?  ###  The species name has been changed to *Solenopsis invicta*. Thanks for catching this!  ###  **Reviewer 3 (Devin Coleman-Derr)**  **Basic reporting**  The article is exceptionally well written with a sufficient introduction and background section that appropriately cites the relevant literature.  ###  Thank you very much!  ###  **Experimental design**  The experimental design is original and well thought through. The research question is clearly identified and the methods of analysis of the eukaryotic 28S data are appropriate; I cannot comment on the methods of analysis used for assessing genetic variation in the host as these are largely outside of my area of expertise.   A few comments:  Line 162: It would be helpful if the authors listed sampling times for the five locations, as they have indicated that sampling occured in June and again August. As the summer months in the South East are typically accompanied by regular precipitation, one might expect that if sampling on either side of the Mississippi occurred in different months, there may be an environmental affect that has influenced community data structure. I doubt this is a serious issue, but it would be useful to have it addressed.  ###  We thank the reviewer for raising this point, and have clarified the sampling design in the text (Lines 151–158). We sampled from all five localities in June and August; five samples from each site (except Lake Ramsey, where ten samples were taken). In addition, we collected ten samples from Lake Ramsey in April, May, and July. For our within locality analyses (taxonomic diversity and rarefaction curves), we restricted the samples to only those collected in June and August (so reduced the number of sequences analyzed from Lake Ramsey; see Lines 181–182), which were taken around the same time from each site. Environmental affects should have affected sampling in the same way, given that samples were taken at roughly the same time during the same months.  ###  **Validity of the findings**  The findings are presented clearly; the only criticism of the manuscript in this section is the depth of sequencing achieved, which is considerably lower than is standard for research in the field and reduces the level of statistical significance that can be assigned to comparisons between the community data.  ###  We agree that greater sequencing would have improved the data set. These data were collected back in 2010, and the samples only took up a portion of a 454 run.  ###  **Comments for the author**  I thoroughly enjoyed this research article, and have only minor comments.   ###  Thank you!  ###  Line 37: The term metagenomics is more commonly reserved for cases in which shotgun metagenomics have been employed. I would alter this statement to state that you employed 454 amplicon-based metagenomics.  ###  This is now changed in the text (Line 36). “We used 454 amplicon-based metagenomics…”  ###  Line 163: "DNA was extracted", rather than "Fluid was extracted"  ###  This has been changed in the text (Line 158).  ###  Line 177: It is unclear what is meant by the removal of redundant sequences; are the authors referring to dereplication at 100% identity? If so, I would rephrase as these sequences are not removed but rather compressed.  ###  We thank the reviewer for this comment, and have removed the sentence from the manuscript as it is unclear and redundant. This is now discussed at the beginning of the Results section (Lines 276–278). We initially recovered 26,399 sequences, and after demultiplexing and initial quality control steps, we retained the unique sequences from each sample (of the 90 separate samples) from the pitcher plants, and used those in our downstream analyses (resulting in 9,045 sequences, reduced to 8,991 after trimming to a universal length).  ###  Line 180: I have not heard this term used to describe the methods used in UPARSE; I would suggest that authors check to make certain this is in fact the terminology they intend to use. It is also sufficient to state that "sequences were clustered into operational taxonomic units using UPARSE".  ###  Yes, we did intend to use this. We’re highlighting here that we used a pure clustering algorithm without any reference sequence data. Edgar (2013) mentions this in the Nature Methods paper, “I have developed a pipeline…for constructing OTUs *de novo* from next-generation reads…” This is to be explicit that we did not inform our clustering with databases or any other reference sequence data, but just purely from the sampled sequences themselves.  ###  Line 185: I am a little unclear on the clustering approach; my interpretation is that sequences were clustered in two ways: 1) separately for each location first, and 2) combined for all locations. It could be helpful for the authors to put a statement earlier in the methods indicating their use of two different approaches to analyzing the data.   ###  Yes, this is correct. We first clustered sequences based on locality (e.g., Lake Ramsey, Cooter’s Bog, etc.) for diversity statistics and to get a sense of taxonomic diversity within each sample site, and then we combined all sequences and reran the data through UPARSE in a global analysis, regardless of where the sequence came from, to generate a comparative data set where OTUs spanned the Mississippi River. We have now clarified this in the text (Lines 177–179; 218–222).  ###  Line 228: It is unclear why the authors chose to ignore OTUs that may be specific to one side of the Mississippi, as they are trying to test the hypothesis that the position with respect to the river is a major determinant of community structure. Perhaps there were not many OTUs that were unique to one side of the river but were found in more than one sample or location, which could be spurious sequence artifacts.  ###  We wanted to test for OTUs that contain population structure across the Mississippi River. We see strong structure in *S. alata*, suggesting that the river has driven population isolation in the host plant. By investigating only those OTUs that contain sequences on either side of the river, we could see if these micro-eukaryotes also show a similar pattern of population genetic structure. This reduced the total data set (from 323 OTUs to 31 OTUs), but it also allowed us to fully investigate a major goal to the paper, evaluating how a host plant may contribute to population genetic structure of associated species.  ###  Line 361: See also Taylor, 2006 in Biological sciences, which speaks to issues of dispersal limitation in fungi.  ###  We thank the reviewer for this reference and now include it in the manuscript (Lines 364–366).  ###  Line 422: grammar  ###  This has been fixed.  ### |