

## Decision

by [Joaquín Hortal](#), 2021-01-19 11:29

Manuscript: <https://www.biorxiv.org/content/10.1101/2020.12.07.414326v1>

### Moderate revisions

This is really a superb piece of science; I join both reviewers in their congratulations for one of the best JSDM studies I've seen so far. Really good data, good design, and top-notch interpretation of the results. We all three liked it a lot.

All that said, both reviewers have a number of concerns about the current version of your work. Most of them are minor comments and/or related to clarity, but Carsten Dormann raises a good point about the potential bias in the selection of model parameters caused by the focus on the best models that you may want to consider. As he admits, this is mainly a philosophical question, but if your study is going to set up a higher standard in JSM, a better account of the uncertainty in model selection would help building a stronger discipline.

Also, both reviewers point to the possibility that there may be some spatial and temporal structure in your data that could be due to unaccounted-for processes or factors. I believe that an additional assessment of whether there is some structure remaining in the residuals of the association matrix could be informative about the existence or not of other processes. Indeed, Joaquín Calatayud highlights that these structured effects may be negligible once you account for environment and co-existence processes, but you need to at least discuss that - and having supplementary analyses to support such discussion would round up your work.

In any case, this is really a great work, and I am looking forward for the resubmission of a new version that accounts and/or discuss the concerns raised by both reviewers. If their comments are properly addressed or discussed it is most likely that I can recommend your preprint in *PCI Ecology*.

BF et al: Thank you for evaluating the manuscript and thereby offering us the opportunity to submit an improved version. We now have taken into account and/or discussed the concerns raised by the two reviewers. Following remarks made by other colleagues, we have also slightly modified some parts of the introduction. We hope that you will find our revised version suitable for recommendation.

Looking forward to your response

On behalf of all coauthors,

Regards,

Benoit Facon

## Reviews

*Reviewed by Joaquín Calatayud, 2021-01-13 16:30*

This is a brilliant piece of science: well written, carefully conducted, based on an outstanding dataset, using a thorough and sophisticated methodology, and presenting timely and very exciting results. I really enjoyed reading it! I have only few very minor questions and suggestions.

I could only fully understand the abstract after reading the completed manuscript. I would suggest rewording small details so that the abstract is clearer. Some points where I found difficulties:

“Community structure was mainly determined by...” Here, I found “community structure” to be somehow vague. Moreover, in a previous sentence you mentioned that network inference was used. After reading this, I was expecting you characterized the community structure via network properties, which is not the case. This may nevertheless be a matter of personal bias, but I guess others may have the same problem. I would suggest to change “community structure” by “species abundances”.

BF et al: Done (1.3 p.1, 1.7 p.2, 1.5 p.3).

“The relative importance of these factors was mildly modulated by host plants.” This sentence was also difficult to follow to me without reading the full manuscript. I would say something like: “The relative importance of these factors mildly varied when we used particular host plant groups” or something alike. This may be again a matter of taste.

BF et al: Done (1.9 p.2).

“... specialists and generalists flies almost behaved as separate communities...” I found “behaved as separate communities” difficult to understand here and when mentioned throughout the text. I would try to use a term more clearly connected with the results. Again a matter of taste and totally up to the authors to follow this suggestion.

BF et al: Done (1.10 p.2, 1.4 p.9 and 1.1 p.23).

In the second paragraph of the introduction, it may be worth mentioning that facilitation can also occur between phytophagous arthropods (e.g. Godinho et al. 2016. *Oecologia* 180: 161-167.)

BF et al: Done (1.1-2 p.4).

“Since species interactions mostly occur in/on plant organs, they may be modulated by plant species identity...” Here, it is not totally clear whether “species interactions” refers to intraguild interactions or to fly-plant interactions.

BF et al: Done (1.10 p.5).

Are the 8 species used all the species of Tephritidae present in the island? If so, I would explicitly state it. If not, I think it would be worth mentioning in the discussion the potential

influence of other unevaluated species in the abundance of the used species and model outputs.

BF et al: Over the study period, only one additional fruit fly species was mentioned on the island (Quilici and Jeuffrault, 2001), *Carpomya vesuviana*. There are reasons to think that its presence may have had negligible influence on the abundances of the eight other fruit flies. The species has been described as specialist of *Ziziphus mauritiana* (Indian jujube, Quilici and Jeuffrault, 2001), which is only present in reunion island in dry areas on the western and southern coasts of the island below 200m. The full unfiltered dataset contained 204 samples of *Z. mauritiana* (a total of 1444 fruits, ranging from 1 to 351 fruits per sample), which contained 68 *Bactrocera zonata* and 48 *Ceratitidis catoirii*, but no *C. vesuviana*. The species was absent from the full 12872-sample dataset (while it would have been recorded when found). The species is considered very rare by local entomologists (J. Payet, pers.comm.). This is now mentioned in the beginning of the discussion (1.21-25 p.19).

“Of the 12872 initial samples, only those with GPS coordinates, with at least one individual fly and belonging to one of the 21 host...” As a layman in the modelling used, it is not clear to me why you didn’t use the 0s (i.e. the samples without individuals).

BF et al: The question is indeed not trivial and it has been the subject of extensive discussions among authors during the preliminary analyses of the dataset, because of the important proportion of samples with no fly (~33%). 0s can stem from two distinct processes. (i) Some of these may correspond to technical artifacts. Here for instance, field technicians specifically collected fruits with visible marks of fruit fly stings, hence from environments with active adult flies. Fruits were then brought back to the lab and placed on grids over sand until pupae emerged and jumped in the sand. The sand was then filtered and pupae were reared in cages until adult emergence. During the process a diversity of incidents may cause abortion of all flies (too recently laid eggs, excessive heat during transportation, fungal development...). Such samples would not help in the estimation process and cause zero-inflation currently out of the underlying assumptions of PLN models. (ii) Although all fruits were collected with fly stings, some of them could correspond to real impossibilities of development for all eight species, and these would be valuable data. But first, there is no easy way to distinguish real from artifact 0s. And second, the focus of the paper is on species interactions, and samples with none of the species in it bring very little information on species interactions. Real 0s would help estimating intercepts of the species abundances responses to ecological covariates, but data with very few individuals (e.g., samples with 0 for 7 species and few individuals for one species) already contribute to these estimates.

Overall, considering the risk of including technical 0s and the expectedly small added value of real 0s, we decided not to include them. Following your comment, we nevertheless conducted model choice and parameter estimation on the dataset with 0s and found the results mostly unaffected (not shown, available upon request). We also provided explanation of why the 0s were not kept in the Material and Methods section (1.3 p.7).

Moreover, how many host plants are there in the island? If there are much more than the 21 used, how do you think this could affect subsequent interpretations of assembly mechanisms? This might deserve a line in the discussion.

BF et al: At least 108 plant species have been identified as potential hosts of Tephritid fruit flies on the island. The 21 host plants studied in the lab were considered as most important for

fruit fly dynamics based on previous works (Quilici & Jeuffrault 2001). Yet not all important plants are in the 21, a number mainly limited by laboratory capacities. Missing important plants could in principle amount to missing important covariates, hence resulting in residual variances (if the plant is important for one species) and covariances (if the plant may serve as a resource for several species). As a test of the potential importance of other plants, we conducted the analysis on all samples with GPS coordinates and belonging to a plant with at least 10 samples (with or without laboratory measured traits). This extended dataset is composed of 6434 samples (including those studied in the manuscript) and 36 host plants (15 more than the 21 studied in the laboratory). Results were unaffected (model ranking, species abundances responses' to ecological covariates and plant identity, residual covariance matrix). This is now shown in Appendix S3 (figures S10 to S12 and table S8). This is also explained in the beginning of the discussion (1.1-5 p.20).

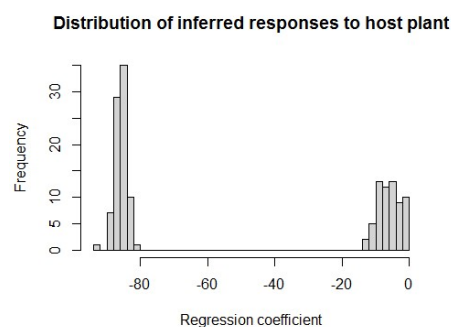
“... all previous models were reevaluated on the datasets excluding *D. ciliatus* (Models 2-0 to 2-6)” To facilitate the reading I would say: “... all previous models were reevaluated on the datasets excluding the species lacking fundamental host use estimates (*D. ciliatus* ; Models 2-0 to 2-6)” or something similar.

BF et al: Done (1.1 p.11).

“Among plants inferred as possible hosts from species abundance patterns (i.e., those with high coefficient values), coefficients correlated positively with fly laboratory-measured fitness for specialists but not for generalists (Figure 3B)” I found this result super interesting! Still, I would better justify why you only used the plants with high coefficients.

BF et al: PLN models rely on the assumption that log-abundances follow a Poisson distribution, which parameter is affected by plant identity and ecological covariates. A species abundance response to a given plant of value  $b$  approximately means that using this plant

increases flies' mean abundance by a factor  $\exp(b)$ . Here all  $b$  values were negative with strong bimodal distribution. 83 values were lower than -80, i.e., so negative that  $\exp(b)$  can be considered negligible ( $<1E-35$ ). The other 64 values lied between -13.8 and 0 ( $>1E-6$ ). We nevertheless looked for a structure in the low values and did not find any. Low values were normally distributed and did not correlate with fitnesses. We therefore considered they were associated to non-host plants. This is now better explained in the result section (1.19-24 p.12).



Moreover, while the modelling approach used seems evident, I would explicitly specify it (perhaps in the figure caption), explaining also the meaning of the shadowed areas in Fig. 3B.

BF et al: Yes indeed. This is now written in the figure legend (cf Figure 3 p.15).

Finally, could this result be mostly driven by only one generalist and/or specialist species?

BF et al: Answering this question brings us close to the statistical limits of the study as we have few points per species (a point is a pair with a fitness estimate and an inferred abundance response to a plant, so there are 64 points for the 7 fly species). If we nevertheless fit a linear model to each species data, all generalists have  $R^2$  values below 0.064. Instead the three specialists have good  $R^2$  values (0.17 for *N. cyanescens*, 0.53 for *D. demmerezi* and 0.83 for *Z. cucurbitae*) and positive slopes. Due to the low number of points, only *Z. cucurbitae* has a significant slope. We would tend to think that all three specialists do show the same trend, but only including additional Cucurbitaceae and Solanaceae plants in the study could help test this hypothesis.

“Accounting for environmental covariates strongly improved model fit and made all residual covariances almost completely vanish, particularly among groups, suggesting that no important environmental factor structuring the community has been missed”. Completely agree. Yet, your data is temporally and spatially structured, which might contain interesting information on assembly mechanisms (e.g. the effects of dispersal processes and interannual and/or seasonal dynamics not linked to climate). While testing this may be out of the scope of your work, I also think that it would deserve a line in the discussion. At least, to avoid criticisms from autocorrelation purists, I would explicitly mention that your results suggest that the influence of temporal and spatial autocorrelation is negligible (besides the influence of other important environmental factors).

BF et al: We agree. Temporal structure has been accounted for by including month and year as ecological covariates. But there could still be spatial structure, which could give interesting insights on the determinants of community dynamics. To evaluate residual spatial structure, we computed variograms with respect to spatial distance (computed from sample coordinates), difference in sampling month and difference in sampling year. No trend was visible on any of the three variograms and the correlations between variance and the evaluated factors were weak. The information is now provided in appendix S3 and mentioned in Results (1.16-18 p.12).

While I really enjoyed “The ghost of competition past” section in the discussion, I’m not totally sure the term perfectly fits here, at least as described by Connell (1980). To my knowledge Connell was referring to the coevolution of competitors and thus to evolutionary changes in the fundamental niches. That is, I agree that by nicely comparing realized and fundamental niches you detected a “ghost of competition”, but I would rather say that it is a current (ecological) ghost rather than a past (evolutionary) one. Perhaps it would be worth to add a few lines explaining this, mentioning also that eco-evolutionary approaches are required to truly address changes in fundamental niches due to competition and their consequences to the assembly of species.

BF et al: You are perfectly right. This is now corrected. (1.6 and 20-22 p.21)

Finally, regarding also the fundamental niches, by looking at figures 3A and 1B, it seems that in some situations flies are able to colonize plant species in which they show a very reduced fitness (close to 0, if not 0). I kept wondering how it is this possible. Perhaps, there are differences in host use among (sub)populations and fundamental niche estimates are based on individuals of a reduced number of (sub)populations. I’m just not sure, but this would perhaps deserve a brief mention in the discussion.

BF et al: This is an interesting hypothesis. There are 10 fly-plant pairs for which laboratory-measured fitness is 0 and both observed field abundances and inferred abundance response to plant show that the plant is within the realized niche. Nine out of the ten cases correspond to zero larval survival and positive female fecundity. Four fly species are concerned (two generalists and three specialists) and six plant species (*Averrhoa carambola*, *Cucumis sativus*, *Sechium edule*, *Solanum lycopersicum*, *Solanum mauritianum*, and *Syzygium jambos*). So there is no easy trend to explain these cases. We may suggest that these cases simply stem from the intrinsic difficulty of measuring fitness in the laboratory. In the survival experiment fruits were transformed into purees to homogenize quantities and conditions as much as possible, and it is possible that the process has gone wrong in some cases. The second hypothesis is related to your suggestion: laboratory measures were made on populations reared in the laboratory for several generations, which guarantees their genetic homogeneity and absence of strong maternal effects, but could cause some divergence with specific field populations. This is now mentioned in Results (1.25 p.12)

Hope this is of any help and congratulations for this excellent study!

BF et al: Thank you for your positive and encouraging review and for your helpful remarks.

Joaquín Calatayud

*Reviewed by Carsten Dormann, 2020-12-11 08:46*

Facon et al.: Joint species distributions reveal the combined effects of host plants, ...

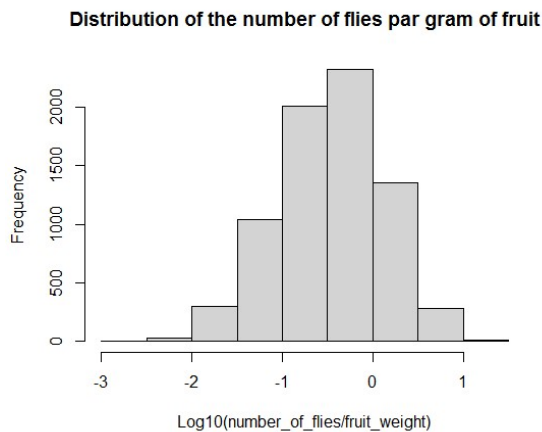
This study investigates fruit fly communities on different host fruits on la Réunion over an 18 year period. Abundances of fruit flies were predicted using both climatic variables and host plant species in a joint species distribution model. This approach yields correlations among fruit fly species as a side-effect (the variance-covariance matrix of the residuals of the  $k \times k$  species, also sometimes called the association matrix), which may represent species interactions.

Accounting for environment, the association matrix became almost diagonal, indicating no apparent associations among fruit fly species. This is in contrast to the raw-data association matrix, which showed a strong difference in host use between generalists and specialists. The distinction between generalists and specialists was based on a fly x host laboratory choice experiment (published elsewhere by members of the same group). In the discussion, the authors carefully interpret the few remaining (all negative) correlations in the residuals as indication of competition among some generalists and some specialists. In particular, they expand on the problem of not seeing many competitive interactions due to evolution having led to niche displacement among similar species (aka the ghost of competition past).

Overall, I find this one of the best jSDM studies I have seen in the literature so far. The relatively small number of species (only fruit fly species) and the huge number of observations (5000 samples with nearly 100,000 individuals) is hard to improve upon. Of course, there are always a few things that remain unclear (at least to me). For example, competition is likely to leave an effect only when resources are scarce, which requires either high population densities of the flies, and/or low availability of fruits. The authors do not provide data on either. Thus, for any given sample, it is unclear whether we would actually expect any trace of competition.



BF et al: Good point. That competition is a central force shaping the community has been a claim in many preceding studies on this community (see for instance Duyck et al. 2006b), based on indirect evidence and experimental competition studies. But you are right that this claim does not rely on clear observation that resources are limited. Unfortunately, there are no good estimates of fruit availability. We may say that this is a subtropical island, with anyway huge quantities of fruits available all year round. But, would we have such estimates, it would not clearly tell us whether there is competition for some specific fruits or at some periods of the year. The present dataset probably constitutes one of the best insights into field densities. Samples differ in terms of fruit number and weight. If we focus on the 7346 samples with a



a single fruit and at least one fly alive (with or without GPS coordinates and including all plant species), we observe that the number of flies per sample varies from 1 to 188 per individual fruit (median 5) and from 0.0011 to 15.8 flies per gram of fruit. Densities are therefore highly variable, with important extreme values. Densities higher than 4 flies per gram of fruit were observed for both specialists and generalists in 11 plant species. Another clue about possible competition is that among these samples 17% contained more than one fly species. Even with these data, it is hard to tell

whether competition may or may not exist. This is why we mainly relied on previous works to assume that competition could play a role.

Also I am not very impressed by the model selection approach used, although admittedly this is almost a philosophical issue and their practice is in line with common analytical strategies (more on this below).

BF et al: We made our best to account for your specific comments below.

I am particularly happy to see the substantial effect of relatively coarse environmental predictors on the association matrix. It could be argued that the small remaining covariance could actually be explained by other predictors, such as acidity of the fruit or something like that. If so, no covariance would indicate no competition, expounding the problem of witnessing the outcome of hundreds of generations undergoing niche separation without being able to see the selection in action. Only extensive laboratory studies with monospecific and paired flies over generations could test for a widening and hence overlapping host niches when released from the invisible competition.

BF et al: Indeed and that would be a terrific project! There have been short-term competition studies, but up to now, no experimental evolution study has addressed the effect of competition on niche displacement in this community.

As far as observation studies go, I don't think there is anything more we could ask for.

### **Model selection bias**

I wouldn't quite call it a secret, but it seems that the well-known effect of model selection causing a bias in model parameter estimates is unknown to ecologists. If, however, we are

interested in model parameters, not only their prediction, we need to be aware of this. In a nutshell, the selection of models leads to a final (few) “best models”, which are then interpreted statistically (“significant effects”) and their parameters are estimated alongside their standard error. Now, since the computer does not “know” that this best model is the result of investigating dozens to hundreds of models *on the same data* before, it naively assumes this is the only model fitted to the data. It “ignores” the model selection uncertainty that results from variability of the data. As a result, the standard errors are too small, as they do not take into account the model selection procedure’s uncertainty. Also, the estimates are biased, as all models that show non-significant effects of this predictor are removed through model selection, making the remaining models more likely to have large (absolute) estimates than small.

It is relatively simple to show this through simulation ([\[http://joshualoftus.com/post/model-selection-bias-invalidates-significance-tests/\]](http://joshualoftus.com/post/model-selection-bias-invalidates-significance-tests/)), but Harrell (2001, Springer: Regression Modeling Strategies) writes about it, and it is found (as obvious introductory statement) in [\[https://pubmed.ncbi.nlm.nih.gov/18189162/\]](https://pubmed.ncbi.nlm.nih.gov/18189162/) or “well known” in [\[https://academic.oup.com/aje/article/167/5/523/212182\]](https://academic.oup.com/aje/article/167/5/523/212182). Parameter shrinkage has been advocated as strategy to counteract the selection bias (see last link for review).

Now, in the present study I do not actually think that model selection bias is a big issue: the data set is large (reducing the problem of model selection leading to different models during bootstrapping); also, the BIC difference between models is very large, indicating little ambiguity in the model ranking. Still, I would have preferred a presentation of few fuller models than the model selection outcome (as indeed Burnham & Anderson themselves argue for).

BF et al: We fully agree with these comments. As you suggest, our results are likely unaffected by the problem (large dataset, large BIC differences) all the more so as the best models are almost always the fullest, i.e., with plant + all ecological covariates. To fully account for your remark, we rebuilt Figure 3 (inferred response to host plant) based on estimates obtained with the fullest model (plant + all ecological covariates + complete residual covariance matrix, model 2-5 instead of 2-6). Figure 4 was already obtained with the fullest model (1-5). We also provide simulated BIC distribution by bootstrapping the dataset 50 times. It shows that the hierarchy of models with different covariates is very stable. It also confirms that models with full residual matrices are hardly differentiable from their diagonal matrix counterpart, confirming the non-detection of important unexplained species abundance covariations (Appendix S3 figure S6). This is also mentioned 1.1-2 p.10 and 1.5-6 p.12.

### **Model diagnostics**

I did not find any statements on how well the model structure meets distributional assumptions and independence of residuals. While I guess that the PLN approach is relatively robust to overdispersion, I think it can be expected of the authors to provide a statement on whether the data were actually well modelled assuming a Poisson distribution. My experience is that fast reproducing and flying beasts, such as fruit flies, tend to clump, requiring a negative binomial to represent the variance in the data. (The DHARMA package may not readily work on this model type, but the principle applies and the authors should be able to simulate the residuals themselves, as demonstrated in that package’s vignette.)



BF et al: Thank you for the suggestion. We now provide DHARMA model diagnostics for the fullest model (Appendix S3 figure S3). As expected, we detected an excess of zeros, which is likely due to the fact fly species do not use all 21 studied plants. This point is one of the reasons why we also conducted the analyses on the three subdatasets. Model diagnostics are now more explicitly mentioned 1.20-22 p.9 and 1.16-18 p.12.

Similarly, the spatial nature of the sampling warrants an assessment of spatial autocorrelation in the model residuals. In fact, it could be that the remaining association signal can be partly explained by spatial effects.

BF et al: This comment was accounted for using variograms (see above), now presented in Appendix S3 figure S7. It was also confirmed using tests provided by DHARMA (not shown).