

PCI Ecology Abbate et al. 2020 (MS#85) Version 1

Response to Reviewers

Dear Recommender & Reviewers: Please find our responses in red, and line numbers refer to the tracked-changes document for your convenience.

We have updated the manuscript with substantial edits based on reviewer comments and suggestions. However, there has been no qualitative change to any of the results, and conclusions remain unchanged, despite a few additional diversity analyses. Scripts and source data for analyses have been provided as supplemental materials and a few errors affecting quantitative values but not qualitative outcomes have been corrected. We believe the process was constructive and reassured us that our work was robust to the methods used.

The official clean version and all supplemental materials will be available on bioRxiv (screening completed) at <https://www.biorxiv.org/content/10.1101/2020.02.09.940494v2.full>

Sincerely,
Jessie Abbate

Reviews

Dear authors,

with two reviews and my own reading of your paper, I would like you to conduct a thorough revision before I can recommend your work. As you will see, both reviewers found a lot of merits to your work, but they also raised some concerns, notably regarding the limitations of the methods and their interpretation, as well as the lack of justification for some methodological choices.

Personally, I also find merits to your manuscript: the questions and data are sound, the general approach is interesting and the results are going to be quite useful for applied ecologists working on zoonotic prevention. However, you I have some issues with some of your methodological choices:

1. A simple one: PCI asks for data and code availability, but I have not been able to find any (or maybe my eyes don't work that well these days). Can you make R scripts and the raw data available for the next version of your manuscript?

Indeed! We provided the data, but failed to realize the R scripts were also to be submitted. We have done so with this revision.

2. For virus and bacterial detection, you use two different procedures (antibody detection and 16S amplification and sequencing). However, these are not only technical differences, but also epistemic ones: with antibodies, you are only going to find

viruses that you were looking for, while with 16S you can find bacteria that you even had not suspected in the first place. For this reason, many of the analyses will not behave in the same way for viruses and bacteria. I suggest that you acknowledge this early on (in the M&M) and discuss it thoroughly.

We fully agree that the 16S metabarcoding approach enables us to detect bacteria with no *a priori* information, while antibodies are specific to the viruses targeted. We would have loved to have been able to use a similar approach of high-throughput sequencing to characterize the rodent virome without *a priori* information. Unfortunately, the developments of these approaches for the study of viruses in wild animals still face many important challenges (from both the molecular and bioinformatics points of view). Viruses are much more diversified than bacteria, so that there is no universal gene marker that can be used for their identification (Mokili et al 2012). Viral loads are often very low in animals, so viral RNA or DNA is strongly diluted in host genetic material. These challenges are detailed specifically for rodents in Drewes et al. (2017).

Moreover, the high per sample cost of NGS/HTS technologies remains problematic for population studies. For these reasons, virome analyses in wild animals still remain scarcely explored compared to the prokaryotic microbiome (e.g., in birds François & Pybus 2020).

By focusing on antibodies, we have no doubt missed many viruses that circulate in these rodent communities. However, we remind the reviewer that we did not test for diversity of all bacteria detected by 16S, only those which we could identify as pathogenic. Furthermore, we only tested one organ (the spleen). This means it is very likely we have missed other bacterial pathogens in this system as well – either those that cannot be detected via the spleen, are rare and missed in 16S sampling, or those with close commensal relatives from which they cannot be distinguished without further sequencing (like *Yersinia pestis*). So, while we have captured a substantial swath of the pathogenic bacterial taxa, we have essentially also chosen specific bacteria to target (analogous to choosing specific viruses).

However, we argue that the violation of assumptions for diversity indices may not be so different between antiviral antibodies and a selection of bacteria from our 16S metabarcoding survey of splenic tissue because neither represents all taxa known to be present in the community. We believe this also means the relationships of both antiviral antibodies and pathogenic bacteria with rarefaction are similar. We have thus stated in the methods that what we were testing here was not total diversity, but relative diversity of the selected pathogens (Lines 238-241) and detailed in the discussion that this means the patterns we uncovered could be influenced, or even driven, by unsampled or excluded taxa (Lines 1015-1018).

3. While I understand your focus on pathogenic bacteria and viruses, it would probably be interesting to also look at other forms of symbiont association (negative or positive), for instance non-pathogenic bacteria with pathogenic ones, non-pathogenic viruses with pathogenic bacteria, etc. You obviously did not search for non-pathogenic viruses like bacteriophages, but at least if you still have blasted data on non-pathogenic bacteria, there might be something to dig in there about microbial associations within rodents (notably, it would be cool to know whether some other bacteria are negatively associated with nasty pathogens).

Again, indeed! We fully agree with the recommender, however, the way that we sampled the mice – dissecting out only the spleen – means there are two biases and two problems. The first bias is that the body will presumably respond differently to non-pathogenic bacteria vs. pathogenic bacteria, meaning that the ability to detect pathogenic species is much greater than that for non-pathogenic bacteria. The second (related) bias is that the location of beneficial bacteria in the body may differ greatly from the mostly blood-borne pathogens that would be found in the spleen. The first problem is that a great deal of the commensal bacteria cannot be distinguished from closely-related pathogenic taxa by 16S. It was not uncommon for some mice to have very specific and highly represented reads of taxa with both pathogenic and commensal species or context-dependent virulence (eg, *Helicobacter spp.*). A great example of this is *Yersinia spp.* We found several mice with high copy number of *Yersinia spp.*, but most (though not all) of those same mice also had an unusually high number of other normally commensal taxa. Which leads to the second problem: there appears to be some level of contamination, perhaps introduced during dissection or through some other immune process or injury. One can see this by the over-represented diversity of typically commensal taxa in some of the spleens. And again, because we could not distinguish pathogenic from non-pathogenic, there was no way to know the nature of the taxa detected. We also cannot just exclude these animals because there is often some level of commensal taxa – and the line between suspect and not suspect is not clear. To avoid making subjective calls, we decided to keep all animals and look only at distinguishable pathogens.

So, I will respectfully say that a more thorough sampling protocol including dissection or swabs of other microbiomes (eg, gut, oral, skin) would be needed to truly address this question in a rigorous way. The associations we have found here may very well be artifacts of such interactions (eg, a commensal protecting against two unrelated pathogens could cause the two pathogens to appear positively correlated), that is to be answered in a follow-up study. We have strengthened the language around this point in the discussion (throughout Lines 958-1024).

4. Quite a large part of the Introduction announces apicomplexan, protozoa, cestods, trematods, etc. as important parasites of rodents, but you chose to focus on bacteria and viruses. Why? (I guess for money reason, but maybe you can elaborate)

Well, it is true that we were limited by resources (in terms of money for pathogen detection and of experienced personnel, e.g. for protozoa, helminth identification), and that is a big reason for the limited scope. However, the main reason is that the original study was focused on investigating rodent-borne zoonotic pathogens. In France, very few helminth parasites of rodents are zoonotic. The main one is *Schistosoma spp* (only found in Corsica) and *Echinococcus multilocularis*, that is mainly present in North-eastern France. As we observed very few rodents infected with this cestode during rodent dissections (results not shown), we did not include a (laborious) systematic search for it in this study. All other helminths described in rodents from the Ardennes were non-zoonotic (see Salvador et al. 2011).

Historically, the clearest evidence for biological processes responsible for pathogen-pathogen associations identified in the wild to date have involved gut parasites (eg, in rodents: Kreisinger et al. 2015, Sweeny et al. 2020), which have huge and well-described impacts on the immune system, offering clear hypotheses that have been

tested. So, it is only natural that the introduction goes into this literature. But we appreciate the recommender's perspective and have tried to place examples from more relevant pathogen groups in the introduction (e.g., Lines 81-89).

5. For the diversity analyses, did you use multiplicative or additive Shannon numbers? (i.e. the exponential version or the log version) As you considered both beta and alpha diversities, which version of the partitioning did you use? (as there are at least two main ones, Jost-Routledge's and Chao's) Did you correct for the usual bias in estimates of Shannon diversity? (using e.g. Chao's estimator) I think the entropart package does this...

We have followed the analyses that are usually applied and recommended for animal microbiome ecology studies (eg: [Haegeman etal 2013](#); [Kim etal 2017](#); [Reese & Dunn 2018](#)).

For alpha diversity we used the log version of Shannon index. The use of Shannon diversity index has been recommended to robustly measure microbial diversity ([Haegeman etal 2013](#)). It is one of the most-used indices of alpha diversity in microbial ecology ([Reese & Dunn 2018](#)). We have not corrected for usual biases as may be done in non-microbial ecology, as such corrections do not seem to be relevant for microbiome communities (eg, [Kim etal 2017](#)). We did, however apply Chao's estimator for host species diversity in a new analysis added following a suggestion by Adrian Diaz (below). This is all now specified in the methods section (Lines 238-253).

We did not apply any partitioning function, as we were focusing on aspects of alpha diversity and beta diversity, but not attempting to get a complete picture of overall gamma diversity.

6. I'm not a big fan of the drop1 method (i.e. likelihood ratio test with the complete model only) because the best model explaining your data can actually have fewer than n-1 variables in it (if n is the total of all tested variables). A very simple trick would be to run your complete model through GLMulti to get the AIC or BIC of all models and ascertain which one is the best. That way, you could have a plausible answer to the question "which factors affect Shannon diversity of pathogens?".

The choice to use the drop1 method was made to be able and understand marginal effects in a systematic way throughout the paper. It's true that this decreases the power of the tests, but it is also true that it could amplify the significance of associations driven by weak but collectively important external factors. We have, however, run the GLMulti where appropriate to show the recommender that it does not qualitatively affect the outcomes. We have not included it in the paper, but do provide the code within the appropriate scripts.

For Shannon diversity: GLMulti Results

Best model = divGRP ~ 1 + AGE + HABITAT + HostSPP						
Single term deletions. Model: divGRP ~ AGE + HABITAT + HostSPP						
	df	Sum of Sq	RSS	AIC	F value	Pr(>F)
			13.027	-130.02		
AGE	1	4.1129	17.140	-109.80	22.7321	9.436e-06 ***
HABITAT	2	2.0627	15.090	-122.11	5.7002	0.005034 **
HostSPP	5	6.5507	19.578	-107.02	7.2411	1.532e-05 ***

7. At some point you mention that you wanted to relate beta diversities with factors, but I did not see that exactly. An option to do this easily is to use dissimilarities obtained from beta diversities (see e.g. Ohlmann et al. 2019) as a distance and then make an adonis analysis on these distances using factors as covariables (adonis is in package vegan). That way, you could explain differences in pathogen composition using some of your factors (habitats...).

We have taken the recommender's suggestion and applied the PERMANOVA (adonis analysis) as recommended. This does not change the main results, that host species was the most important extrinsic factor, but it does reveal habitat to also be significant. We have included these results in the beginning of section 3.2.2, the methods in 2.3.1, and have updated the discussion (Lines 764-766) to reflect this change.

8. Fig. 2 has an issue: I don't think you can test anything using residuals. If you delete one factor in a model and then plot the residuals according to the modalities of that dismissed factor, what you will see is only how bad your model is at compensating this missing factor, not the intrinsic differences between modalities. To obtain the differences between modalities, you need to actually predict the model (with the factor you want to look at) on new data (using predict in R), probably with a randomization (bootstrap) of the initial data.

This was a type-o. It should have read marginal means, rather than residuals. It was always the correct display of marginal means based on TukeyHSD(). Again, apologies for not including the scripts in the first pass, but thank you for catching this misnomer. It now reads "Differences in Shannon diversity index were tested on marginal means of each significant factor in the multiple regression model controlling for all other extrinsic factors. Different letters signify statistically significant differences at $p < 0.05$, with post-hoc Tukey adjustments for multi-level factors." The axis in the figure has also been fixed.

9. In Fig. 3B and around it: when you write "pathogen co-exposure", does it mean "virus + bacteria co-infections"?

This is now more explicitly explained in the methods and throughout the paper. See in particular Lines 238-242. They cannot be called co-infections because the virus may or may not still be present (all have chronic forms), thus we refer to them as co-exposures.

10. Regarding the "MCA -> principal dimensions to be regressed individually by factors" approach, I'm sure you can do better than that by using CCA (Peres-Neto et al. 2006) or distance-based RDA (Blanchet et al. 2014) to obtain the fractions of the total chi² of your host x pathogen table explained by each factor. The individual regression of MCA eigenvector seems a bit pedestrian and it does not give a super-clear answer in the end.

For understanding the overall contribution of extrinsic factors, a dbRDA (using capscale with Bray-Curtis distances) analysis (and its results) is essentially identical to the PERMANOVA from the reviewer's comment #7 (Anderson 2017). The loadings acquired from the dbRDA for unconstrained factors (pathogens), however, do not appear to be advantageous to our MCA methods for identifying potential associations among pathogen species. Thus, we have included this method and re-justified the purpose of the logistic regressions (Lines 254-270).

11. Fig. 4 does not seem that useful -- do you really need it?

True, it is not necessary. It was to illustrate the contribution of extrinsic factors to beta-diversity, but it doesn't really add much that isn't in the text or sup mat tables. We have thus moved it to the sup mats.

12. As stated above about model comparisons through AIC or BIC using GLMulti, I suggest you do something like that for the GLM you do on pathos-pathos associations.

Or maybe better: for this part, you can also shift completely and use your dataset as a network (host nodes connected to pathos nodes). By using any sensible community-detection algorithm or block model, you can find modules/blocks/communities of pathogens that have similar connection patterns, which in a way is what you are looking for, isn't it? This approach is less driven by hypotheses (it's actually super-exploratory), but it might confirm what you know while at the same time finding associations you did not see or did not expect. I suggest Leger et al. (2015) as a good reading on modules/blocks in networks, if needed.

Or you can do both: GLMs to test particular associations and module-search to get a broader view on all associations...

We appreciate the recommender's suggestion here, but again refrain from changing our methods, which were selected specifically for two main reasons: 1) Power. This seems counter-intuitive but we knew going into this that our power was limited, so we felt that removing factors that often play a role in correlations of presence/absence in structured populations was not worth potentially over- or under-estimating the significance of an association; For instance, see the results for Myco~Hantavirus. 2) Structure. With such an exploratory study, we wanted to provide as much consistency in how the analyses were conducted as possible. So, we are keeping our original methods intact.

We again run the GLMulti to show how it affects the outcomes (Appendix 1 at the end of this letter), and provide the code within the appropriate scripts.

We also thank the recommender for suggesting network analyses, and it would indeed be interesting to see how the results would look if these methods were applied. However, we decline to add these methods for two reasons: 1) The clustering analysis should essentially give us another picture similar to what we have done with the MCA, but the MCA gave us a way to evaluate variation in the occurrence data in a multi-variable (rather than pairwise) way. And 2) The paper is already quite complex in terms of methodologies. Newer methods are rapidly evolving (eg, multi-variable associations via Markov random fields, Clark et al 2018) and a future study that compares our findings with those from e.g., network clustering models or hierarchical species distribution models would be very interesting for future work.

13. A very general comment: you mention p-value corrections only for Tukey's tests, but given the amount of tests performed, I suggest you use p.adjust with Benjamini-Hochberg correction overall (so you do that for the vector of all p-values obtained with all analyses on the same dataset).

We have now included a section to cover this in the methods and results (Lines 333-344 and 637-640).

14. The discussion is a bit long (9 pages +), I guess you can shorten the parts where

you repeat the results.

We tried to remove repetitive parts, indeed. However, it's now a bit longer since we were asked to add discussion of various things. We tried to keep that to a minimum, obviously, but it's a very large body of work. Further suggestions are welcome.

Typos/minor things

L. 90 Leptospirosa => Leptospira ? **Corrected**

L. 133 sexually-immature => sexually immature **Corrected**

L. 208 "This produced a ~~down-to-a~~ set of quantitative (...)" The sentence looks warped. **Corrected.**

L. 252 implicated => involved ? **Corrected**

L. 276-277 For M&M instead? **Moved.**

L. 290-297 Problems with punctuation and coherence between propositions: "because they were...", "inability to rule out...", "or because their identity..." **Corrected**

L. 335 pseudo-R2 from deviance ? Please state it.

Done. It now reads "analysis of deviance Pseudo Rsq"

L. 424 "more rarely than expected (p = 0.13..." -- really expected? A p-value above 0.05 more likely means "as expected"

Actually, the way this analysis works is a bit special. It is not a true p-value, but is instead a value based on simulations. When the observed number is at the lower or upper bound, particularly when the envelope is wide and the bound in question is zero, it suggests a trend. We have included the phrase "sitting on the lower bound at zero;" to clarify. Line 550.

I hope you will find all reviews useful for revising your paper.

Yes, extremely helpful. Thank you.

Sincerely, François Massol

Reviews *Reviewed by anonymous reviewer, 2020-03-05 17:44*

The authors present a study entitled "Pathogen community composition and co-infection patterns in a wild community of rodents" in which they analyze rodent-borne pathogen communities in a rural area of northern France, a region known to be endemic for several rodent-borne diseases. Using elegant, stepwise and pedagogically described approaches, the authors propose that the host species is the main determinant of pathogen community composition and that hosts share habitats that may have very different pathogen communities. The authors are aware of the weak points of their study and hypotheses are always made in relation to them. The figures are very clear and very esthetic. All the results obtained lead to a very descriptive study with a good number of

speculations that need to be confirmed by experimental studies. If I go back to the order of the article:

Line 104-107 "We investigate exposuremetbarcoding) : Why distinguish between the two? How can we be sure that there is no old bacterial or recent viral association?

It is true that both of those things may be present. However, it is much more likely that viral antibodies represent past infection, while bacteria are more likely to be current or at least recent. This is why we use the words "current or recent exposure to bacterial pathogens" here, and why we discuss the possibility that the viruses may still be active in the discussion. We have now added a dedicated paragraph to this caveat in the methods, Lines 238-241.

Line 160 : Why choose the spleen as the only tissue to be analyzed?

In short, we had limited resources, and were not convinced that it was a good idea to mix tissues. We focus on the spleen because this organ is known to filter microbial cells in mammals. Several studies have shown that a wide array of pathogen/zoonotic bacteria could be detected using spleen extracted DNA/RNA (e.g, Movilla et al. 2017). We've included this explanation in the methods (Lines 190-192).

Line 203 : "to regroup significantly distinct factor levels" Isn't it the other way around, i.e. grouping levels that are not significantly different? **Done.**

Line 223 : "where evidence suggested" I don't understand what evidence are the authors talking about?

We added the parenthetical phrase "(non-significant variation in the MCA dimension among host species or non-significant host species identity x explanatory pathogen term in logistic regressions)" to clarify. (Lines 283-284)

Line 255-256 : (-2log.....full model) explanation to move line 199 when the function "drop1" is mentioned for the first time?

No, because for logistic regression, the drop1 function uses the Chi-sq distribution (-log likelihood) rather than the F-distribution in linear regression.

Line 256-258: "Despite...sampled" Why this choice when the risk of false positive is high?

We have revised the analysis to include a correction for this risk (see our response to recommender's comment #13 above). That said, p-values on small sample sizes for an exploratory analysis should serve as a guide for other researchers who may have new information suggesting the hypothesis is still worth exploring, rather than definitive yes or no.

Line 270-271 : "No animals...antibodies" This sentence seems incomplete to me

"Tested positive for..." is the verb here, but it is weird, we agree, and replaced it simply with "were positive for"

Line 458 : "antibodies: x2=5.21"to be completed by "antibodies - Bartonella: x2=5.21" **Done.**

Line 471: "non-significant interaction...."In the material and method, it is not indicated that interactions are included.

True, and we thank the reviewer for catching that. We added a statement in the methods: "When the multiple host species were involved, we tested an interaction term (host species identity x explanatory pathogen), and either (if $p < 0.05$) performed separate analyses for each host species or (if $p \geq 0.05$) simply added host species identity as another covariate in the model." (Lines 318-322)

Reviewed by Adrian Diaz, 2020-03-23 14:02

Dear Recommender, The manuscript entitled "Pathogen community composition and co-infection patterns in a wild community of rodents" submitted by Jessica L. Abbate et al to PCI Ecology aimed to characterize the pathogen community harboring by a wild community of rodents in north of France. Through seroprevalence techniques and molecular characterization, authors studied the community composition, pattern of distribution of pathogens in the host species range and habitats. Also, through appropriate statistical analysis, they evaluate how intrinsic and extrinsic factors influence the pathogen activity/distribution. Moreover, they tested a priori hypothesis regarding pathogen interactions already published in scientific literature and a plus of potential interactions detected during the study. Although the pathogen detection/identification methodology is quite assertive they had limitations and those are clearly stated at the end of Discussion section:

General comments:

Overall the manuscript was well written, including a detailed statistical analyses section and a good graphical support of results sometimes, specially in the Discussion section the English is hard to follow. I recommend the authors to revise in order to make it more fluent for readers.

We have attempted to simplify and break up long sentences throughout the Discussion in order to improve both clarity and readability. Since the writing was done by a native english speaker, the revised version was read and edited for clarity by a non-native English speaker.

Abstract, Line 25: Authors expressed they study the incidence of pathogens in the rodents community. Actually they estimate the prevalence or presence/absence of pathogen in the community. Incidence means how many individuals changed their status as a function of time (diseases, infection,)

True! Replaced "incidence" with "occurrence".

If you have seroprevalence data with no assumption about time of exposure. Can you associate that with an acute infection? That could result in a spurious speculation with no biological meaning?

Many pathogen-pathogen interactions are mediated by host immunity, and we have

added that in the introduction explicitly (Lines 81-89). That said, we would argue that in fact, all of the associations we have found statistical support for here are speculative interactions with possibly no biological meaning. That doesn't mean that the root cause of the association is not interesting to investigate.

How did you build your matrix in the pathogen community composition?

The answer to this now lies within the codes provided (it was just an MCA, so presence-absence data matrix of pathogen taxa x host individuals), although we have also added an explicit analysis from the *vegan* package, creating a Bray-Curtis dissimilarity matrix in the *adonis2* function. (Lines 255-263)

Page 19 Lines 427-430. When you ran GLM tests to analyze Myco1, Myco3 and anti-hantaviruses antibodies interactions you stated extrinsic factors as drivers and also as heterogenous host group. It is not clear.

The line now reads: "We therefore ran three reciprocal GLM models on the restricted dataset, one for each pathogen as a function of extrinsic factors *to control for* heterogeneous host groups..." (Lines 553-556)

Did you consider microbiome (OTUs from microbes typically of healthy flora) as an intrinsic factor driving pathogen diversity?

No, we did not. See our response to the recommender's comment #3 above.

Minor editing observations:

Page 11 Line 264. Better state the Number of positive sera and the percentage

Done.

Figure 3 B and 3 C: Is Pathogen equal to virus or mixing viruses and bacteria? In the first case states virus rather than pathogen since you have been using that criteria.

There was a typo, which we believe caused this confusion. It should now make sense. Pathogen refers to both viruses and bacteria.

Figure 5. Better state anti xxxx antibodies to homogenize it use through the manuscript.

Done.

Discussion: Page 23, Line 501: Where were Lyjungan virus and spirochetes detected? In your study? Please clarify. Done.

Page 24, Line 532: Could you please clarify this sentence "has only ever before been"

Done. The details of the experiment in which they were found is not important, so hopefully this is now clear. Lines 719-723.

Poor discussion about why some host species harbor poor number of zoonotic agents.

We have added an explanation in the following lines: "While physiology, genetics, and

behavior can contribute to the number of pathogen species able to infect a given host species, larger geographic range size is highly correlated with higher pathogen species diversity (Dallas, Holian & Foster 2020); this explanation matches the pattern among hosts in the communities sampled here (i.e., *Ar. scherman* and *Apodemus* spp. have small geographic ranges compared to those of *Mi. arvalis*, *My. glareolus*, and *R. norvegicus*)." (Lines 730-734). We have also re-organized the introduction as another paragraph previously lower down had given some explanation (confusingly). We have now integrated it here instead, following the lines added above. (Lines 735-759)

First two pages are describing patterns already described in results. It sounds a little redundant.

We have revised this as much as we felt we could. We've tried to remove parts that repeat details of the results, but much of it is introducing information about the significance of the bacteria found – information that is not in the results section.

Page 26, Line 578: change colom by a period. **Done.**

Do you test any association about diversity of co-infection and diversity of host species at place?

No, we did not test for the relationship between host species diversity and co-infection or pathogen diversity (not clear which the reviewer was asking). This could potentially be an explanation for the differences seen between habitats. We investigated this and it turns out that the proportion of co-infected individuals indeed was correlated – inversely – with the number of host species present in a given habitat, but this relationship entirely disappeared without rats (the sole species on farms). However, when we measured host species diversity against pathogen diversity at each location (year*site*habitat), we found that there was a significant positive correlation. $r=0.62$, $t = 2.5$, $df=10$, $p = 0.032$. There is much discussion around the impacts of host diversity on pathogen communities, and it is an aspect we neglected in our first pass, so we thank the reviewer for this suggestion. It has now been incorporated throughout the manuscript (methods, results, and discussion). e.g., Lines 250-253, 735-759.

Page 28, Line 650: anti-hantaviurs antibody **Done.**

Page 28, line 668: (Previously described by...) **Done.**

Page 29, Line 654: association between *Mycoplasma* and anti-hantaviruses antibodies. Authors stated this positive interaction is because of the cronic diseases that infectious agents developed. In that

Not sure what the question was here. Perhaps because the example given is only for chronic hantavirus? We had already explained extensively that *Mycoplasma* infections are typically chronic.

Page 29, Lines 656-657: why independently of host species identity?

The original sentence read: "This positive association was found in both host species where the majority of hantavirus exposures occurred (*Microtus arvalis* and *Myodes glareolus*), consistent with the generality of association between *Mycoplasma* species

across host taxa detailed above, suggesting the intrinsic ecology of these pathogens contributes to shaping variation in the pathogen community independent of host species identity.” We simply meant that it was consistent across the host species in which the two pathogens were found. It’s true that it’s both redundant and therefore unclear, so we removed this clause. (Line 924)

Page 29, Line 662: anti-CPXV antibodies **Done**.

Page 29, Line 664: delete virus. It is already included in the abbreviation CPXV **Done**.

Page 29, Line 678: reported instead of revealed **Done**.

Page 32, Line 740: undescribed **Done**.

References:

Anderson MJ. Permutational Multivariate Analysis of Variance (PERMANOVA). In Wiley StatsRef: Statistics Reference Online (eds N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri and J.L. Teugels). 2017. doi: 10.1002/9781118445112.stat07841.

Drewes S, Straková P, Drexler JF, Jacob J, Ulrich RG. Assessing the Diversity of Rodent-Borne Viruses: Exploring of High-Throughput Sequencing and Classical Amplification/Sequencing Approaches. *Adv Virus Res.* 2017; 99:61-108. doi: 10.1016/bs.aivir.2017.08.002.

François S, Pybus OG. Towards an understanding of the avian virome. *Journal of General Virology.* 2020; 101(8): 785-790. doi: <https://doi.org/10.1099/jgv.0.001447>.

Haegeman B, Hamelin J, Moriarty J, Neal P, Dushoff J, Weitz JS. Robust estimation of microbial diversity in theory and in practice. *ISME J.* 2013; 7(6):1092-1101. doi:10.1038/ismej.2013.10

Kim BR, Shin J, Guevarra R, Lee JH, Kim DW, Seol KH, Lee JH, Kim HB, Isaacson R. Deciphering Diversity Indices for a Better Understanding of Microbial Communities. *J Microbiol Biotechnol.* 2017; 27(12):2089-2093. doi: 10.4014/jmb.1709.09027.

Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol* 2012; 2:63–77. doi: 10.1016/j.coviro.2011.12.004.

Movilla R, Altet L, Serrano L, et al. Molecular detection of vector-borne pathogens in blood and splenic samples from dogs with splenic disease. *Parasites Vectors* 10, 131 (2017). doi: 10.1186/s13071-017-2074-z

Reese AT, Dunn RR. Drivers of Microbiome Biodiversity: A Review of General Rules, Feces, and Ignorance. *mBio.* 2018; 9(4):e01294-18. doi: 10.1128/mBio.01294-18.

Salvador AR, Guivier E, Xuéreb A. et al. Concomitant influence of helminth infection and landscape on the distribution of Puumala hantavirus in its reservoir, *Myodes glareolus*. *BMC Microbiol.* 2011; 11:30. doi: 10.1186/1471-2180-11-30.

Appendix 1: GLMulti results for pathogen-pathogen association logistic regression models.

For Myco1 v. Myco3 v. Hantavirus

```
#model selection in response to reviewer comments (no qualitative impact on results)
summary(glmulti(mod, family="binomial", level=1))

#best model: "Myco1 ~ 1 + HostSPP + SITE + VirusPUUV"
mod<-glm(Myco1~HostSPP+SITE+VirusPUUV,data=XdatFAC[which(XdatFAC$AGE=="ADULT" &
(XdatFAC$HostSPP=="C_My_glar" |XdatFAC$HostSPP=="C_Mi_arva")),],family="binomial")

drop1(mod,~,.,test="Chisq")
```

```
Model:
Myco1 ~ SPPGrp + SITE + VirusPUUV
      Df Deviance   AIC   LRT Pr(>Chi)
<none>      55.231 63.231
SPPGrp    1  90.252 96.252 35.021 3.262e-09 ***
SITE      1  58.689 64.689  3.458  0.06293 .
VirusPUUV 1  58.668 64.668  3.437  0.06375 .
```

Result: no qualitative change. Myco1 ~ Hantavirus is still only a trend.

```
#model selection in response to reviewer comments (no qualitative impact on results)
summary(glmulti(mod, family="binomial", level=1))

#best model: "Myco3 ~ 1 + YEAR + VirusPUUV"
mod<-glm(Myco3~YEAR + VirusPUUV,data=XdatFAC[which(XdatFAC$AGE=="ADULT" &
(XdatFAC$HostSPP=="C_My_glar" |XdatFAC$HostSPP=="C_Mi_arva")),],family="binomial")

drop1(mod,~,.,test="Chisq")
```

```
Model:
Myco1 ~ ANNEE + VirusPUUV
      Df Deviance   AIC   LRT Pr(>Chi)
<none>      102.20 108.20
ANNEE     1  102.43 106.43  0.2356  0.6274
VirusPUUV 1  104.91 108.91  2.7119  0.0996 .
```

Result: potentially inaccurate loss of significance. Myco1 ~ Hantavirus is only a trend when extrinsic factors are not taken into account.

```
#model selection in response to reviewer comments (no qualitative impact on results)
summary(glmulti(mod, family="binomial", level=1))

#best model: "VirusPUUV ~ 1 + HABITAT + Myco1"
mod<-glm(VirusPUUV~ HABITAT + Myco1,data=XdatFAC[which(XdatFAC$AGE=="ADULT" &
(XdatFAC$HostSPP=="C_My_glar" |XdatFAC$HostSPP=="C_Mi_arva")),],family="binomial")

drop1(mod,~,.,test="Chisq")
```

```
Model:
VirusPUUV ~ HABITAT + Myco1
      Df Deviance   AIC   LRT Pr(>Chi)
<none>      30.464 38.464
HABITAT  2  40.075 44.075  9.6112 0.008184 **
Myco1    1  36.263 42.263  5.7996 0.016030 *
```

Result: no qualitative change. Hantavirus ~ Myco1 is still significant.

MCA 3 (Myco2 vs Myco4)

```
#model selection in response to reviewer comments (no qualitative impact on results)
summary(glmulti(mod, family="binomial", level=1))

#best model: " Myco2 ~ 1 + AGE + Myco4"
mod<-glm(Myco2 ~ 1 + AGE + Myco4,data=XdatFAC,family="binomial")

drop1(mod,~,.,test="Chisq")
```

Model:

```
Myco2 ~ 1 + STAGE2 + Myco4
      Df Deviance   AIC   LRT Pr(>Chi)
<none>    57.150 63.150
STAGE2  1  62.065 66.065 4.9156 0.02662 *
Myco4   1  59.598 63.598 2.4483 0.11765
```

No change.

```
#model selection in response to reviewer comments (no qualitative impact on results)
summary(glmulti(mod, family="binomial", level=1))
```

```
#best model: " Myco4 ~ 1"
```

No change (but now not even a trend).

Bart vs OPXV (e.g., just in *Arvicola scherman*)

```
#model selection in response to reviewer comments (no qualitative impact on results)
summary(glmulti(mod, family="binomial", level=1))
```

```
#best model: " Bartonella ~ 1 + YEAR + VirusOPXV "
mod<-glm(Bartonella ~ 1 + YEAR + VirusOPXV, data=XdatFAC[which(XdatFAC$HostSPP=="C_Ar_sche"),],
family="binomial")
```

Model:

```
Bart1 ~ 1 + ANNEE + VirusOPXV
      Df Deviance   AIC   LRT Pr(>Chi)
<none>    50.247 56.247
ANNEE    1  57.540 61.540 7.2935 0.00692 **
VirusOPXV 1  55.320 59.320 5.0727 0.02431 *
```

No change.

```
#model selection in response to reviewer comments (no qualitative impact on results)
summary(glmulti(mod, family="binomial", level=1))
```

```
#best model: " VirusOPXV ~ 1 + YEAR + Bartonella"
mod<-glm(VirusOPXV ~ 1 + YEAR + Bartonella, data=XdatFAC[which(XdatFAC$HostSPP=="C_Ar_sche"),],
family="binomial")
```

Model:

```
VirusOPXV ~ 1 + ANNEE + Bart1
      Df Deviance   AIC   LRT Pr(>Chi)
<none>    82.027 88.027
ANNEE    1  90.452 94.452 8.4253 0.00370 **
Bart1    1  87.099 91.099 5.0727 0.02431 *
```

No change.

HM vs HC

```
#model selection in response to reviewer comments (no qualitative impact on results)
summary(glmulti(mod, family="binomial", level=1))
```

```
#best model: " MH ~ 1 + YEAR + SITE + HostSPP + MC "
mod<-glm(MH ~ 1 + YEAR + SITE + HostSPP + MC, data=XdatFAC, family="binomial")
```

```
drop1(mod, ~., test="Chisq")
```

Model:

```
MycoHM ~ ANNEE + SITE + SPPGrp + MycoHC
      Df Deviance   AIC   LRT Pr(>Chi)
<none>    177.53 191.53
ANNEE    1  180.59 192.59 3.060 0.080253 .
SITE     1  185.74 197.74 8.204 0.004179 **
SPPGrp   3  252.91 260.91 75.376 3.009e-16 ***
MycoHC   1  187.26 199.26 9.726 0.001817 **
```

No change.

```
#model selection in response to reviewer comments (no qualitative impact on results)
summary(glmulti(mod, family="binomial", level=1))
```

```
#best model: " MC ~ 1 + HostSPP + MH "
mod<-glm(MC ~ 1 + HostSPP + MH,data=XdatFAC,family="binomial")
```

```
drop1(mod,~,test="Chisq")
```

Model:

```
MycoHC ~ SPPGrp + MycoHM
      Df Deviance   AIC   LRT Pr(>Chi)
<none>      146.36 156.36
SPPGrp  3   169.27 173.27 22.908 4.22e-05 ***
MycoHM  1   156.77 164.77 10.411 0.001252 **
```

No change.

For Bartonella spp vs. hemotropic Mycoplasma spp.

```
#model selection in response to reviewer comments (no qualitative impact on results)
summary(glmulti(mod, family="binomial", level=1))
```

```
#best model: "Bartonella ~ 1 + HABITAT + YEAR + HMyco"
mod<-glm(Bartonella ~ HABITAT + YEAR + HMyco,data=XdatFAC,family="binomial")
```

```
drop1(mod,~,test="Chisq")
```

Model:

```
Bart1 ~ HABITAT + ANNEE + MycoHF
      Df Deviance   AIC   LRT Pr(>Chi)
<none>      288.39 300.39
HABITAT  3   347.25 353.25 58.858 1.031e-12 ***
ANNEE    1   298.66 308.66 10.266 0.001355 **
MycoHF   1   290.81 300.81  2.417 0.120008
```

No change.

```
#model selection in response to reviewer comments (no qualitative impact on results)
summary(glmulti(mod, family="binomial", level=1))
```

```
#best model: "HMyco ~ 1 + HABITAT + SITE + HostSPP"
mod<-glm(HMyco ~ 1 + HABITAT + SITE + HostSPP,data=XdatFAC,family="binomial")
```

```
drop1(mod,~,test="Chisq")
```

Model:

```
MycoHF ~ 1 + HABITAT + SITE + SPPGrp
      Df Deviance   AIC   LRT Pr(>Chi)
<none>      266.04 286.04
HABITAT  2   271.64 287.64  5.600 0.0608110 .
SITE     1   279.72 297.72 13.689 0.0002157 ***
SPPGrp   5   369.35 379.35 103.314 < 2.2e-16 ***
```

No change (not significant enough to enter into the model).