

We thank the reviewers for their thorough comments. Please find attached an annotated version of the main text with all changes made to the original manuscript highlighted in yellow, as well as supplementary material updated according to the reviewers' requests.

Response to reviewer 1.

Main of my comments follow the text, only the last ones on *Eimeria* model and parasitological concepts are given apart. I refer to previous or subsequent comments that complete each other when necessary, to avoid repetitions.

C1. Lines 57-59: **Review the statement:** The experimental works, Moulia *et al.* 1995 and Derothe *et al.* 2004 **do not show** that natural hybrids are less susceptible than parental mice, as they investigate hybrid resistance / susceptibility via experimental crosses. In Moulia *et al.* (1995), only F1 are concerned, which clearly cannot be compared to highly recombined wild hybrids. Indeed, Derothe *et al.*, 2004, showed, via experimental crosses until the F2 generation, that recombination can not be directly involved in hybrid susceptibility (i.e. the genome disruption is not directly involved in the higher parasites loads in the wild). But they do not refute the hybrid susceptibility with these experimental data as they cannot denied either the observed differences in wild populations (Sage et al, and Moulia *et al.* 1991), either those observed with experimental infections of wild-derived strains (Moulia *et al.*, 1993). Derothe *et al.*, 2004 proposed a new hypothesis to be coherent with these previous field results and experimental infections, suggesting that a pleiotropic effect was selected in the hybrid zone (see further).

We agree that the difference between F1 and recombinant hybrids should be made as clear as possible and therefore changed the wording “hybrid” by “inter-subspecies F1” (line 58-59 new). We address general challenges dealing with these previous results in our last comment to this reviewer.

C2. Lines 59-60: “the highest statistical power”. **Quite strong statement without arguments.** The statistical choice of Baird et al. 2012, as well as this of this very study, is not the only available one (see upcoming comments).

We agree that superlatives should be avoided in scientific writing. We mention sample size and sampling design (see following paragraph) as arguments now. We do not develop further, as the argument is thoroughly developed in the study itself (Baird et al. 2012). The sentence now reads “[...] more recent field study, with high statistical power (more mice sampled and more suitable statistical approach) and improved sampling design,[...]” (line 60-62 new).

C3. Line 91 : “prevalence and abundance, the latter defined as parasite load in all hosts”. **Bad definition of abundance.** It is the mean of parasite loads (individual intensities) from the whole sample. It then includes the “zero” parasite loads, given an indication (even if weak) of the distribution: a lot of zero will reduced this potential parasite load of an given lambda individual, few zero will increased the mean value.

This comment highlights a lack of clarity in our original text, in which we refer to “all possible hosts, i.e. the whole sample” as “host”. We therefore replace “hosts” by “animals” (line 97 new).

C4. Lines 71-72: As an important part of the interpretation of the results is linked to the notion of pathogenicity (and virulence /Resistance / tolerance debates - quite unshared points of view in Parasitology/ Microbiology), it will be useful to remember the definition of pathogenicity used in this study.

This is a very useful conceptual point to clarify and we thank the reviewer for bringing it to our attention. We, indeed, had used controversial terminology (pathogenicity) for “impact on host health”. We now replace all occurrences of the term “pathogenicity” by “impact on host health” in the text.

C5. Lines 73-74: **narrow point of view and weakness of the references:** very frequent in all mouse populations. See for instance Behnke, 1975; Behnke, 1976; Kriska, 1993; Ressouche et al., 1998.

We added the mentioned references.

C6. Lines 75-76. Pay attention, this is a **prejudiced idea**, even if commonly shared and supported by the occurrence of an immune response successful to regulate the parasite loads (Jacobson & Reed, 1974, Lewis et al. 1991). This indicates a coevolution limiting parasite virulence. However, some studies prove that high parasite loads occurred in some specific environmental situations, with clear negative effect on health in laboratory strains (Harwell & Boyd, 1968; Eaton, 1972; Jacobson & Reed, 1974 even Taffs). Note also the demonstration that strong parasite pressure could have selected form laboratory strains resistance (Derothe et al. 1997). See upcoming comments on notion of pathogenicity versus impact on host fitness

In the specific sentence, we modified as follow:

“They are often considered to provoke mild symptoms on their hosts, even if in rare conditions (e.g. particularly high burden) they have been shown to affect the health of laboratory mice (Taffs, 1976).” (lines 84-86 new).

As a more general comment, to which we will refer to in later comments:

Our study is designed to jointly study two different parasites. We test if any general hybrid effect on resistance and impact on health could be detected. We chose to focus on two parasites from very different phyla, namely Nematoda and Apicomplexa. We wanted to test the likelihood of hybrid resistance or susceptibility against a null model (hybrid parasite load = average of parental loads), and finding the same results on both these phyla is for sure reinforcing our confidence in the results. As a result of the reviewer's comments, we developed more on the differences that make this comparison valuable, and abandoned the term "pathogenicity" which is too vague to carry all differences. Instead we highlight all differences between the parasites: a) extra vs. intra-cellular, b) targeting Th2 vs Th1 immune pathways, c) possible different impact on host health. We now introduce these aspects in more detail in the **introduction** (line 70-77 new), and in the **discussion** (line 422-427 new).

C7. Lines 79-81: **Review the statement.** That is not really what Al-khlifeh et al., 2019, shows. Especially on weight loss, one of the wild-derived *Eimeria* strains (*E. ferissi* from the mice studied in this very study) show the same level of negative effect than the laboratory maintained strain of the same *Eimeria* species. The other wild-derived strain (*E. falciformis*) does not decrease the weight of the NMRI mice compared to the controls. The Al-khlifeh et al. reference then shows that different species of *Eimeria* could have different effect on the mouse strain NMRI (i.e. one mouse genotype). That is all the authors can say!

We disagree. This statement is correct. The main focus of Al-khlifeh et al. 2019 is on differences in gene expression induced by different *Eimeria* isolates. Indeed *E. ferissi* (as opposed to the *E. falciformis* lab isolate) did not induce a systematic response measurable in the spleen. Nevertheless, mice lost weight in infection with all *Eimeria* isolates from the different species. "The period of patency (oocyst shedding) was characterized by body weight loss in infected mice in all infections (Figure 1b)." (<https://www.biorxiv.org/content/10.1101/611277v2.full>)

C8. Lines 81-84 : parasites in another host species than *M. musculus*! **The extrapolation is far too fast.** Host species, even closed ones (which is not the case here!) have been shown not share the same co-evolutionary history with closed or identical parasites. Therefore, they do not show the same response phenotypes. Look to the very demonstrative models: Humans versus non-human primates, and *Treponema* or HIV /SIV.

We agree, and remove the extrapolation.

C9. Lines 84-86: we **lack data** on the distributions of these 3 species across the HZ, especially since we know that the authors have those data (Jarquín-Díaz et al., 2019). See further specific comments on *Eimeria* model.

That is a very good point. By the time of the analysis, we did not have the full dataset genotyped. We are happy to provide supplementary data on the most prevalent *Eimeria* species, namely *E. ferrisi*. This analysis confirms the results on hybrid resistance at the genus level. We added this information in the text (lines 363-368 new) and as **Supplementary Figure S6**. We moreover added identification at the species level, for those that we could identify, in a column of **Supplementary Table S3** (previously Supplementary Table S1). We hope it is obvious that hybrid differences cannot be tested for parasites with low prevalence due to a lack of statistical power.

C10. Line 87: “a novel transect”. **Specify:** “novel” compared to what? I suppose not from this of Jarquín-Díaz et al., 2019, as it is the same transect and as this one includes the Jarquin-Diaz sampling!

For clarity, we added “in which the hypothesis of hybrid resistance/susceptibility to parasite had never before been tested.” (lines 95-96 new)

C11. Lines 88-91: **Trouble with excluding “zero” loads of the analysis.** The authors argue that they then exclude “ecological and epidemiological factors for differences in loads”. Harboring zero parasites could not be only relied to non-encounter of infesting stages (ecological*). Harboring no parasite indicates that:

- you have not yet encounter the parasites,
- you have still encounter them but they failed to establish (low level of infesting stages, bad gut environment (physiological, microbiotic), previous parasites already occupying the same niche, cross-immune response to other recent infection...)
- you have been infested before those around you and have already eliminate the parasites
- you have been infested in the same time than those around you but you have been successful to eliminate the parasites, while the others no (or not yet)

This comment, as well as several mentioned later on, helped us to improve our article in the following way. We:

- added a detailed paragraph describing our approach (lines 198-213 new) integrating information on the SIR model and how we approach it in our model

- tested that there is no change in the age structure across the HMMZ (proxy for potential differences in mortality of hybrids) (lines 229-244; lines 346-354)
- added a new figure reporting the result of our test on mortality (Supplementary Figure S7)

C12. If the two first items could refer to specific extrinsic ecological parameters (only part of the second indeed), not the last ones! At least, assuming their argument about the absence of ecological differences across the zone, the authors should compare the respective portions of “zero” loads in parental and hybrid mice, has been important in evaluating difference in resistance or susceptibility. See specific comments on “parasitological indexes approach”.

“the authors should compare the respective portions of “zero” loads in parental and hybrid mice” is the exact analysis we are conducting on a continuous scale of hybridization, using a logistic regression. We reworded the section of material and methods to make this clearer. (lines 219-228 new).

C13. *What is the sense of “epidemiological” for the authors? Indeed Epidemiology includes ecological parameters as it describes the dynamic of the parasitism at the individual level (that is, the course of the disease) and at the population level (in time and space). I believe the term is not appropriate as it included intrinsic factors linked to the hosts and the parasites, and not only external ones (ecological and environmental)

We agree and removed the term “epidemiological”.

C14. Lines 101-102 “The trapping was designed to capture both parental and hybrid/recombinant populations.” **Not clear**, which bias between hybrid or parental mouse trapping is (are) expected?

We agree with a possible confusion of the terms, “trapping” could mean “location of trapping” (the way we use it) or “method of trapping” (not expected to differ across hybridization scale). We detail by replacing “The trapping was designed” by “The locations for trapping were selected in a geographical range allowing [...]” (line 108-109 new)

C15. Lines 102-108: Sexing? Pregnancy statement? Age estimation (even categories)? The age could be interfering factor in Body Condition (see upcoming comments)

We have added information in the text (lines 329-331) as well as in the new **Supplementary Table S5**.

C16. Lines 111-112: “a value of the hybrid index (HI) calculated as a proportion of Mmm alleles in a set

of 4-14 diagnostic markers (at least 10 loci in 92% of the mice)". **The difference in reliability of HI determined with 4 or 14 diagnostic loci is too high.** Estimating genome hybridization with only 4 loci is not rigorous. Indeed, most of the studies studying the HZ have work with at least 6 allozymic markers (At this state of the manuscript, we could then interpret that in this very study, these 6 mandatory markers have not been assessed in all tested mice! see next comment and comments on table S1). The authors speak of 92% of mice genotyped with at least 10 loci (thus from 10 to 14), variance much more acceptable. So

- How are the remaining 8% distributed from 4 to 9?

-What are the part of markers shared by each mouse of the sampling to estimate HI ? If some individual HI estimate on totally different markers, do the authors estimate it is rigorous to consider that each marker gives a similar and comparable evaluation of the HI? If so, then why previous and numerous genetic studies use several markers and not one to estimate the HI (see previously).

-Why do not eliminate these 8% or at least all the individuals among these 8% with less than 7 exploitable diagnostic loci.

Moreover, how do the authors explain such a large percentage of missing data (8% below 10 from 14 tested loci (indeed 13 for all the mice, as YNPARG is not present in females))? See further comments on results and table S1.

The reviewer misunderstands the nature of how variation in uncertainty in the hybrid index will affect our results. One could, in a worst-case scenario, imagine a correlation between (a) those mice with high uncertainty in their HI and (b) those mice with most extreme effects. The consequence is not what the reviewer implies. The extreme-effect mice will now be scattered across the HI axis (due to their uncertain HIs), *removing* any pattern of effect at particular HI values. We find clear effects at particular HI values *despite* the (inevitable in any study) uncertainty in HI estimates.

The reviewer misrepresents *using all available data as working with missing data*. The mice involved in the study are typed for a number of purposes. We could make the number of loci considered for HI calculations uniform across all mice by discarding 'excess' data for particular subsets, thus *increasing our uncertainty in many of their HIs for no justifiable reason* (see paragraph above) [but reducing the amount of missing data as defined by the reviewer]. For example, we could reduce all HI estimates to a basis of 4

allozymes in equivalence with the Moulia study repeatedly favourably cited by the reviewer, and work with ‘no missing data’. We chose not to do so in our original submission, and remain firm that more information about the HI is better, not worse.

It is a matter of choice how many loci shall be chosen to estimate inter genomic hybridization. We made a comparison between a HI used in the submitted MS and a dense SNPs array Mouse Diversity Array. In a paper submitted recently we characterized each individual by hybrid index (HI) computed as frequency of *musculus* diagnostic alleles across six autosomal (*Es1*, *H6pd*, *Idh1*, *Mpi*, *Np*, *Sod1*) and six X-linked (*X332*, *X347*, *X65*, *Tsx1*, *Btk1*, *Syap1*) loci. The resulting values are highly correlated ($R = 0.98$) with those based on a high-density SNP array (Yang et al. 2009). We believe the choice of number of loci depends on goals of the study, given the number of mice and estimation of HI based on a reduced number of loci is fully relevant in this context. We also want to highlight that this comment can always be used (even ad absurdum requesting only whole genome sequences) to criticize any study.

For the sake of transparency we provided originally (as supplementary material) our full raw data file, including mice not used for the analysis (N=10) for insufficient data on them, or because they were embryos. We had previously supplied, for each analysis, the number of mice considered. We now clarify this further as we:

- added the histogram of distribution of markers, as well as the distribution along the hybrid index, for each analysis (lines 128-130 new; **Supplementary Figure S1**)
- removed 10 confusing mice (rows) from our supplementary data table (previous Supplementary Table S1, now **Supplementary Table S3**)
- added in this table columns corresponding to each analysis. Entries “yes” and “no”, specify which mouse was used for which analysis (our qPCR technique was not optimised before 2016, hence this was used only for 2016 and 2017 (years and number of mice had mean already mentioned in the text, no analysis had to be modified)).

These modifications did not change any results previously shown, as we never used the full raw data but only the indicated subset. The analysis is openly available in full detail at https://github.com/alicebalard/Article_IntensityEimeriaHMHZ/tree/master/code (indicated in lines 307-308 new). We, however, made the supplementary file self-explanatory without consultation this code. Importantly, the analyses had never been incoherent and the number of mice was indicated in the original version of the manuscript for all analyses.

C17. Line 118: **specify** that H6PD in the text is listed as GPD1 in Table S1.

Thanks for pointing out this inconsistency, we have corrected it. By mistake, we left in Table S1 the former name of the gene, Gpd1, used in our previous papers whereas in the main text we adopted the name H6pd according to the current version of the MGI database. The latter is now used consistently throughout the whole manuscript including Supplementary Table S3 (ex S1).

C18. Lines 122-123 : “...(graphical resolution increased over defaults) based on a subset of the six autosomal markers that were genotyped in all mice” : **contradicts what is written previously** : some mice have been genotyped only on 4 markers so the minimum 6 autosomal ones are not there! See further comments on results and table S1

Thanks for pointing out this error, we corrected it in the text, and mention that we estimate the hybrid zone centre with “all individuals with 6 diploid markers (N=598 mice)” (line 135 new)

C19. Lines 133-134: focus on *A. tetraoptera* and *S. obvelata* only.

1/Need some **justifications** for people who are not familiar with the model and previous parasitological studies (Sage et al., 1986; Moulia et al 1991;). A minimal justification is expected especially as the authors rely quite strongly on a previous study that has chosen another strategy : Baird et al., 2012. Indeed, in this study, the authors considered all the helminths within the digestive tract and the accessory organs (liver) whether they are frequent or not, whether they have a direct cycle or not (which is questionable because of the ecological parameters that may interfere in the analysis).

2/ Perhaps in result part as well as in table S1 **give some elements on helminth species diversity** to evaluate the proportion of the pinworms.

The text has been modified and a supplementary figure added as follows:

“As in this study we need a high statistical power to test our hypothesis, we considered only the most prevalent helminths, the oxyurids *Syphacia obvelata* and *Aspicularis tetraoptera*. Histograms presenting the distribution of counts for other helminths can be found in **Supplementary Figure S2.**” (lines 147-149 new)

C20. Lines 142-144: *Eimeria* primers: specify the source (it seems to be Jarquin-Diaz et al., 2019 and Ahmed et al., 2019) as well as their universality for the three *Eimeria* species present in the HZ.

Lines 144-146: house mouse primers: same details than for *Eimeria* are needed.

We added the appropriate reference and details (lines 161-164 new)

C21. Lines 151-160. **Major lacks of explanation and validation of the protocol.** See specific comments on *Eimeria* model .

This protocol has been validated, we added: “This method was validated in an infection experiment in NMRI mice (Al-khlifeh et al., 2019). We considered $\Delta C_t = -5$ our limit of detection as at this limit it was possible to obtain genotyping data for all samples using independent PCR reactions (Ahmed et al., 2019; Jarquín-Díaz et al., 2019)” (line 172-175 new)

C22. Lines 164-165: “the median of parasite load across all hosts”: that is, one median for all the data set? One median for the studied HZ? Median (for non-parametric distribution), as well as mean (for parametric one), is a weak parameter to represent data when analyzed alone. Why not conserving more parameters of the distribution (percentiles for instance)? Why not sub-dividing the sampling according to the HI or areas across the zone, to define more medians and thus to keep more representativeness of the overall variance?

We thank the reviewer for pointing out the lack of clarity in terminology. We replace “took” by “report”, as we do not use these medians for further analyses (line 187 new). We don’t consider sub-dividing the sampling according to the HI, as we work on a continuous scale of hybrid level which is genetically and statistically more appropriate.

C23. Line 166: “and only median intensity for *Eimeria* spp”. **Lack of explanation.** What is (are) the argument (s) to omit abundance in the case of *Eimeria* and not nematodes?

This is a good point, it is obviously not clear enough based on our previous description in material and methods. We added as explanation: “For qPCR some uninfected samples present technical noise due to unspecific amplification of non-target DNA. We therefore used a qPCR threshold validated by independent genotyping PCRs (see “Fig. 4” of Jarquín-Díaz et al., 2019) to establish the infection status of each sample (and we do not report abundance for *Eimeria*, see Jarquín-Díaz et al., 2019 for details).” (line 189-193 new)

C24. Lines 166 -170 : **repetition of the above lines:** 162-166.

We thank the reviewer for pointing out this mistake, we removed the repetition.

C25. Lines 176 -177. **Review the statement of the epidemiological model basis.** The authors reduce the epidemiological process to two stages of the classically and basically used S(E)IRD model (Kermack, WO. and McKendrick, AG. A contribution to the mathematical theory of epidemics. Proc Roy Soc, London 1927, 115, 700-721- see current use in Ebola epidemiology: T. Berge, J.M.-S. Lubuma, G.M. Moremedi, N. Morris & R. Kondera-Shava (2017). A simple mathematical model for Ebola in Africa, Journal of Biological Dynamics, 11:1, 42-74, DOI: 10.1080/17513758.2016.1229817)

In this classical model, (S) stands for Susceptible, (E) for Infected, (I) for Infectious, (R) for Recovered/Removed, and (D) for Dead/Deceased). Absence of parasites could be the stage of recovery after infection (and it may occur after high parasite load/ intensity). This recovery may be definitive or temporary (if no complete immunization, and/ or if the parasite infra-population disappears because of its aging, leaving the niche to a new generation (as possible in gastro -intestinal parasites). If temporary, then the absence of parasite is a very time-limited stage. Moreover, dead hosts are important to evaluate the level of susceptibility of a host population, but is impossible to estimate directly in wild sampling. How will their models take this into account especially if *Eimeria* sp. induced mortality? See further specific comments on *Eimeria* model.

This has been fully modified and developed, see our answer to **C11**.

C26. Line 180. **Relevance of the reference.** I do not see in which way the work of Poulin 2013 supports the two categories subdivision of the model. Inappropriate.

We removed the reference.

C27. Lines 181 -183. Using a proxy that equate introgression with more than 50% *musculus* genomes and introgression with more than 50% *domesticus* genomes. The authors make **an a priori of equality of the two mouse genomes** regarding their resistance/tolerance abilities. It means equivalence of the two genetic backgrounds when they are the majority in the hybrid genome. This is all the more questionable. 1/ Several studies suggested potential differences in resistance according to the populations of the same species (i.e. the local genotypes) studied along the zone (for instance Derothe et al. 2001 et Derothe et al. 2004); 2 /Jarquin-Diaz et al., 2019 showed a non-uniform distribution across the zone (HI from 0 to 1) of the 3 species of *Eimeria*; 3 / we do not still know exactly how describe and take into account parasite differences/ divergences from either side of the zone (see Baird and Göüy de Bellocq , 2019). Are then the parameters of the association (as resistance /tolerance phenotypes) the same between a given parasite species/taxa/ genotype P0 associated to hosts HI = 0

and the counterpart parasite species/ taxa/ genotype P1 associated to host HI = 1? **At least this simplifying choice must be argued and discussed.**

We added justification for this choice, and tested separately on both sides (finding similar results) as follow: “We considered the predicted probability of infection across the HI as equivalent to the prevalence and modelled a dichotomous response variable (uninfected = 0; infected = 1) by logistic regression. We performed two analyses, one testing for prevalence differences on both halves of the hybrid index separately and a second one with a unified “genetic distance to zone centre” (for individuals with HI between 0 and 0.5 the proxy is HI, for individuals with HI between 0.5 and 1 the proxy is 1 – HI). This means we do not blindly assume equality of prevalence at both ends of the hybrid index, but also maximize power to reject the null hypothesis (esp. in case of a negative result in the separate analysis). Analyses were done in R with the function glm from the stats package (R Development Core Team, 2008) including host sex and interaction terms with the variable representing hybrid genetics.” (lines 219-228 new; in results lines 335-343 new)

C28. Lines 188 à 229: **statistical analyses.** The authors use the global methodology proposed by Baird et al. 2012. I am not expert in this area but the approach seems clearly explained in this previous 2012 work. However, in this study,

1/ for instance, equ 5 – line 199 à 202- is not enough explained (as well as the alpha parameter) whereas this is the main parameter use to compare parasite load!

We did not want to overload the text by repeating the very complete and detailed development given in Baird et al. 2012. We here provide a shortened (maybe more technical explanation detailing all formulas used). We would be happy to take any comment to improve the clarity without falling into repetition but so far we do not see what should be added.

2/ **a more general and basic criticism** : in Baird et al., the only comparison between previous classical statistics (mostly classical non-parametric tests) and their new approach for testing the hybrid resistance/ susceptibility was performed on the data of Sage *et al.* (1986). Indeed, the two statistical approaches went in the same direction, but, **in my opinion, it is insufficiently demonstrated** and regrettable that Baird et al. did not submit their own data to the double analysis. It would have allowed them to "validate" their approach. In conclusion, it seems to me quite interesting to take advantage of this very study to perform this validation.

We strongly disagree with the reasoning behind this statement. One cannot validate a conceptual framework by comparing it with older, less biologically meaningful, less powerful previous one. Baird et al. 2012 developed in details the improvement of this conceptual framework over the older ones (non meaningful and arbitrarily defined categorisation of hybrids vs. parental, developed in length is Baird et al. 2012 “Finally, it should be noted that the definition of a hybrid varies across these studies. A hybrid index (*HI*) can be used to place any mouse on a linear scale from *musculus* to *domesticus* depending on the count of *domesticus* alleles at assayed loci. Sage et al. (1986a) and Moulia et al. (1991) assayed respectively four and 10 enzyme loci for this purpose. However, the interval “hybrids” occupy on this scale differs from study to study. Expressing the *HI* as percentage *domesticus*, “hybrids” have $12.5\% < HI < 87.5\%$ in Sage et al. (1986a), $20\% < HI < 60\%$ in Moulia et al. (1991), and $2\% < HI < 97\%$ in Moulia et al. (1993).”).

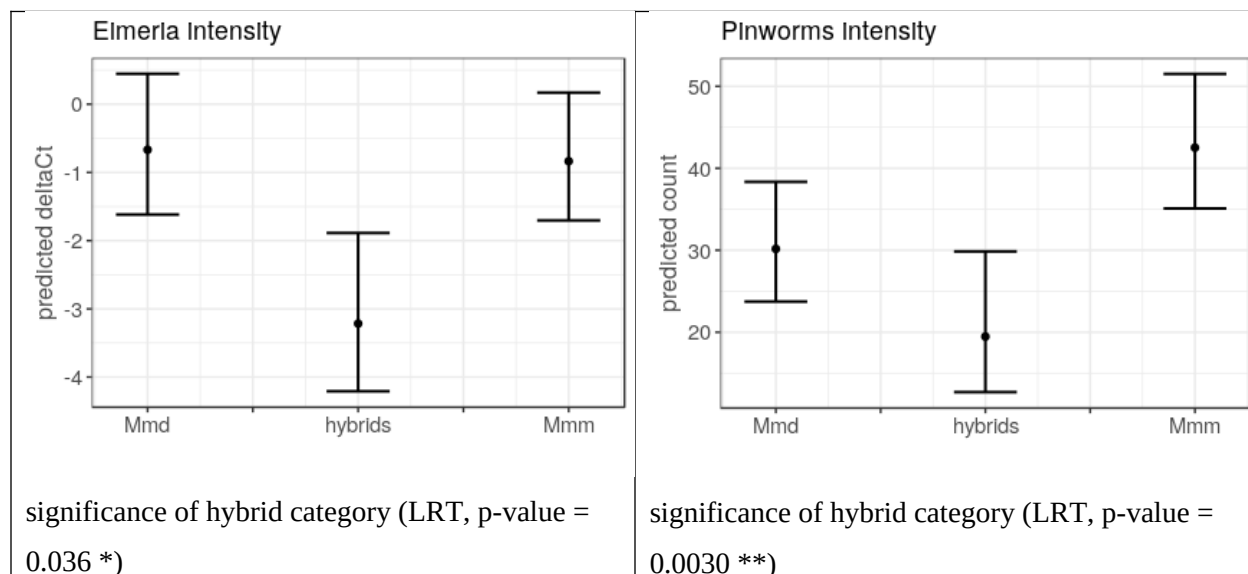
Adding analysis in categories would not validate our approach, but misguide readers seemingly validating an inappropriate approach. Earlier (in their C16 comment) this reviewer expressed at length their worries about uncertainty in the hybrid index estimates. Here, by contrast, they suggest we use an arbitrary categorization of HIs. These two points of view cannot be reconciled. Making uncertain estimates categorical does not eliminate their uncertainty. Rather, it obscures that uncertainty and leaves any downstream inference open to the simple question: What would one infer choosing *different* (arbitrary) category boundaries? This second question is not addressed by the studies recommended by the reviewer. Indeed, inconsistencies in how mice from the same localities are arbitrarily categorized exist across those studies. The reviewer states “the approach seems clearly explained in this previous 2012”. If that is the case, we do not feel it is necessary to re-iterate the inference logic of avoiding arbitrary HI classification.

Nevertheless, for the sake of transparency, we provide here (only in this response to the reviewer) an analysis based on a categorization according to Moulia et al. 1991. For both pinworms and *Eimeria* intensity, we fitted a generalised linear model (negative binomial distribution for pinworms, Weibull distribution for *Eimeria*, as in our analysis) of the parasite intensity depending on the hybrid category. Then, we performed likelihood ratio tests, to test the significance of the hybrid category. Finally, we plot the parasite intensity as predicted by the model. We obtained comparable results to those reported in our present manuscript, namely hybrid increased resistance:

| Genotype | Hybrid index range | Number of infected mice examined | Mean <i>Eimeria</i> intensity (delta Ct) |
|----------|--------------------|----------------------------------|--|
| Mmd | [0 – 0.2] | 28 | -1.32 |

| | | | |
|---------|-------------|----|-------|
| Hybrids | [0.2 – 0.6] | 8 | -3.35 |
| Mmm | [0.6 – 1] | 34 | -1.78 |

| Genotype | Hybrid index range | Number of infected mice examined | Mean pinworms intensity |
|----------|--------------------|----------------------------------|-------------------------|
| Mmd | [0 – 0.2] | 107 | 30.2 |
| Hybrids | [0.2 – 0.6] | 34 | 19.5 |
| Mmm | [0.6 – 1] | 166 | 42.5 |



C29. Line 213: “which was confirmed for helminth **in the HZ** (Baird et al., 2012)”. **Insufficient reference.** Binomial negative is known as relevant distribution for helminths in all kind of host populations (and not only helminths) many previous works and years ago!

We agree, and added “in a previous transect” in the original sentence (line 267 new). More citations were already given at the beginning of the sentence (“The negative binomial distribution should perform well for macroparasite counts (Crofton, 1971; Shaw & Dobson, 1995),[...]”) (orig. line 213).

C30. Lines 214-215: **incomplete.** Argue for Weibull Distribution. It is often regarded as a very flexible law, being able to approximate an great diversity of probability distributions. Is this choice reflect simple and quick way to approximate without further analyses of the real data distribution? Is it relevant? Have the authors verify that binomial negative would not be more relevant (see commenters on line 213)? Moreover, how do the authors “positively shifted” the data? Is the answer in lines 228-229? If so incomplete. Explain a little more the choice “at 7.14”.

We tested our choice of distributions as written lines 209-212 (“Adequate distributions of values for each parasite and detection method considered were selected using log likelihood and AIC criteria and by comparing goodness-of-fits plots (density, CDF, Q-Q, P-P) (R packages MASS (Venables & Ripley, 2002) and fitdistrplus (Delignette-Muller & Dutang, 2015).”). We have now added the corresponding plots for clarity as supplementary material (**Supplementary Figure S4**).

The value of 7.14 is biologically meaningless, it’s just the value that maximises the likelihood while guaranteeing positive values, so we modified the text as follow: “In the case of $\Delta Ct_{\text{Mouse}-\text{Eimeria}}$, the Weibull distribution requires positive values as input. Therefore, we estimated an extra “shift parameter” which was optimized by maximum likelihood.” (lines 279-281 new).

C31. Lines 230 -251. correlating body condition (based on Body Weight) with health status. **No argument is given to support these data as indicator of fitness with or without parasite.** Moreover, literature is divided about the relevance of BC to estimate health state and/or parasite impact on host fitness (see few non-exhaustive examples I have in my personal bibliography: for: in the field : caribous - Hughes et al, 2009 - ; against or with restriction (especially in mice!) see here¹). Note that numerous parasitological studies showed positive correlation between either body size or weight with either parasite richness or parasite load (similar to what the authors referee here as tolerance). This is especially true among fish hosts. **To conclude, I recommend the reading of Wilder et al. 2016.** Functional Ecology 30, 108–115 to question the relevance of BC parameter. Clearly, I am not convinced that this data reflects the health status under infection or not. At least, I would have appreciate the authors develop some arguments! See further comments on pathogeny , parasite cost/ host fitness and induced mortality

¹ Ullman-Culleré and Foltz. 1999. Laboratory Animal Science. Difference in Relevance of BW to assess the health status of males and (virgin) females (variation of organ volumes and increase in tissue mass in response to infection?) - Rao et al. 1989. Fundamental and Applied Toxicology. Volume 13, Issue 1, Pages 156-164. Mice previously infested by virus. If there is no change in body weight between treated and untreated mice, there is lower mortality of the treated ones. Survival one important parameter of the fitness is then not linked to Body Weight. But effect of the infection on survival is clear (mortality) - Effect on age on BC Indices (lot of studies and models for instance humans: Goodpaster and al. The Journals of Gerontology: Series A, Volume 61, Issue 10, October 2006, Pages 1059–1064) **How is differences of age in BC taken into account in this very study?**

This is a notorious complex debate in the community. We chose to not expand on the above, as we obtained negative results (see line 391-394 orig). “We here used body condition as a proxy for the health component of host fitness. We, however, did not find evidence for differences in body condition between hybrids and pure mice upon infection. We conclude that we do not have evidence that lower parasitemia

in hybrids increases their health.”). We are aware that this could reflect an actual similarity of health over the hybrid zone, or the incapacity of our proxy to detect significant difference, and don’t claim otherwise. We, however, have the strong opinion that the absence of an effect on health should be the null hypothesis.

C32. Lines 232-233: **Lake of references!** At least validate the use of the residue with some previous references. Perhaps Jacobs et al. 1996. OIKOS 77: 61-67?

See orig. line 235-236 (“Individuals with a positive residual were considered in better condition than individuals with a negative one, as this index correlates with variation in fat, water, and lean dry mass (Schulte-Hostedde, Zinner, Millar, & Hickling, 2005)”).

C33. Lines 260 -270:

1/ in this part we understand that sampling of mice studied for *Eimeria* and pinworm infections have not been completely identical (*Eimeria* : 384 mice sampled in 2016 and 2017; Pinworms: Between 2014 and 2017, 585 mice ...). **This must be explained before** in M&M

See **C16.**

2/See specific comments on *Eimeria*. The 3 species **are pooled** in this result and the upcoming ones, but their distributions across the zone are not equal (Jarquín-Díaz et al., 2019). Only the most largely distributed species must be taken into account to be rigorous.

See **C9.**

3/ the author have the possibility **to investigate co-infections** but they do not. Evaluation of the susceptibility to co-infections would allow a more subtle investigation of immune abilities (partly under genetic determinisms). So how many mice share *Eimeria sp.* and pinworm infections? When comparing number (prevalence) and intensities of co-infections, are parental and hybrid mice identical?

We agree that co-infection is a major topic of research. We have the option to investigate co-infection with 2 specific parasites. These mice carry a lot of helminths, coccidia, bacteria... So this investigation would be only a small first glimpse at potential of co-infections. We are currently working on a more complete molecular analysis of parasites and the microbiome of mice in the HMMZ, but don’t consider this to be within the scope of the current study.

C34. Lines 271-270. This result agrees with the previous parasitological studies of the HZ. Clearly, we do not expected great variation in encounter and ecological parameters in such a limited area.

This was never tested with this method (logistic regression modelling prevalence) in the HMMHZ
See **C11**.

C35. Lines 279-287 (Results on *Eimeria*), 288-307 (results on Pinworms) and 308-318 (results on Body conditions): these results are coherent only if we agree with the protocol and the previous statements of modeling. See above comments and troubles on this subject.

We are sure that our analysis is coherent. We hope the answers to previous comments and additions to the text allow the reviewer and the reader to come to the same conclusion.

C36. Lines 298-307. Comparison with Baird et al. analysis. The authors do not explain why focusing on the Baird et al. 's alpha is sufficient/ relevant /important to justify a similarity. What does this alpha value concretely represent (too shortly explained in M&M)? Could the authors use the apha value of Baird et al. in this very study? This is too **obscure** to be validated by the reader without more arguments. See more comments on table 1.

All occurrences of “Alpha” in our text are now replaced with “hybridization effect”, we still use the shorthand (alpha) in tables. We added more explanation at the first occurrence: “Alpha represent the hybridization effect, or deviation from additivity between the two parental genomes.” (lines 240-241 new)

C37. Lines 325-327. **Unclear.** “ House mouse hybrids are late generation in the European HMMHZ (Macholán et al., 2007)” What do the author mean and want to argue with this ambigus sentence? Is the reference the good one? Why does it imply the following assertion of the authors: hybrids “should not be considered in categories, but rather on a continuous scale when analyzing parasite infections or any other trait (Baird et al., 2012)? “ The fact is undoubtedly that the HZ is a present and dynamic process of melting and indeed each mouse is a single genetic mixture, as it is more clearly expressed in the sentence 330-331.

We developed for clarity on line 406-411 (new) as follow: “House mouse hybrids in the European HMMHZ are not first-generation crossings, but rather genetically complex “late generation” recombinants. This means that each of their genomes presents a complex admixture of both Mmm and Mmd tracts (Macholán et al., 2007). There is no clear cut-off between hybrids and parental individuals. Therefore, individuals in such systems should not be considered in categories, but rather on a continuous scale of “hybridicity” (a

hybrid index) when analyzing parasite infections or any other trait (Baird et al., 2012).” This also explains our modelling approach.

C38. Lines 331-332: “to de-confound the prevalence and intensity aspects of parasite load”. These two parasitological parameters **are clearly un-confounded and expressed quite different aspects** of the parasite distributions. I could understand what this sentence suggested for the authors when reading the entire document – even if I am reserved (See specific comments on” parasitological indexes approach”). However, this assertion as is, **is a misunderstanding of epidemiology and parasite ecology**. Moreover, in this very context, it is **misused**. Indeed, it does not support in any way the statistical approach on an individual index of hybridization. Each mice shows a specific response both in terms of prevalence and intensity, independently of the way its genome hybridization is quantified: specifically (as in Baird et al. and this study) or by categories (as in Moulia et al. 1991 for instance).

This part of the text has already been modified; see **C11**.

C39. Lines 341-342: “*Eimeria* is very likely **more pathogenic than pinworms** (Al-khlifeh et al., 2019; Fuller & Blaustein, 1996; Hakkarainen et al., 2007)”. **Pay attention to the meaning of the sentence.** These studies **do not compare pathogeny of *Eimeria* and pinworms** but they studied the negative impact on health status (even mortality) of *Eimeria* on various models. The authors could only write that these works argue for a strong impact of these kinds of parasites on health status. Note that Fuller & Blaustein and Hakkarainen et al. (even in a different approach) investigate host survival . This induced mortality, as well as potential induced density fluctuation during the year, have been completely neglected in this very study even in the discussion about sampling or analysis bias. See general comment on *Eimeria*

See **C4. & C6**.

C40/1 Line 343: pinworms frequent in “laboratory facilities “ : not only ! See comments on lines 73-74. And “often considered to provoke subclinical symptoms (Baker, 1998)” : not only! See comment on lines 75-76. Note that Baker 1998 clearly specifies “While infections are usually subclinical, **rectal prolapse, intussusception, fecal impaction, poor weight gain, and rough coat have been reported in heavily infected rodents, although generally without adequate exclusion of other pathogens** (ref475).”Then in case of susceptibility (genetically determined or not) pinworms could then not be regarded as weakly pathogenic.

See **C4. & C6**.

C40/2 Above all, **the symptomatic and medical- centered point of view that is pathogeny makes little sense in parasite ecology**. As C. Combes used to say “you see a bird, I see a community of parasites”. But nobody “sees” that these parasites have a cost for the bird fitness when looking it singing on the tree. Nowadays, numerous studies have showed that this singing and apparently healthy bird could be less bright for females partner, it could be less efficient to build its nest, or less efficient to feed his chicks. But pathogeny is not apparent. **Symptoms as well as pathogenic “signals” do not reflect parasite cost on host fitness**. Occurrence of resistance/ tolerance genotypes, as existing among mice from pinworms (Derothe et al. 1997), prove selective pressure and then potential costs.

We agree, and don't claim that symptoms are correlated with fitness. We wrote “We here used body condition as a proxy for the health component of host fitness” (line 391 orig.), and later “Even if we had found increased health of hybrids, this would not be interpretable as leading to a higher total hybrid fitness, as the parasite mediated health fitness component is only one (likely minor) component of overall fitness.” (line 398 orig.).

C41. Lines 346-359 : I do not understand the relevance of long discussion on this aspect. According to me, demonstration that ecological or epidemiological parameters are not evolved in parasite distributions in the HZ has definitively been performed via experimental controlled infections of wild-derived strains (Moulia et al., Derothe et al.). This field study (i.e. on uncontrolled population sampling) can only bring a new argument if we consider analysis of the prevalence of parasites across the zone. But this is an argument, not a demonstration.

No field study can ever be perfectly controlled, while no lab study can ever claim to represent natural diversity. We therefore consider that a first field study on two very different parasites, showing the same pattern of hybrid resistance, should discuss this interesting finding. Moreover, our study's main characteristic is exactly to consider the “analysis of the prevalence of parasites across the zone”.

C42. Lines 360 – 361 . “a novel aspect of our work compared to previous studies of parasitism in the HMMZ is the separate study of parasite prevalence and intensity”. Pay attention. The previous field studies took into account **abundance**, which is not a mix of intensity and prevalence *sensu stricto*, but a third parasitological parameter giving another information that the two others (“number of infested ones” and “parasite load if infested” respectively). See specific comments on “parasitological indexes approach”.

We agree, but we never claimed that “abundance = intensity + prevalence”. We claim that we test for the first time separately prevalence and intensity in the context of a hybrid zone. See **C11**.

C43. Lines 362-365: See specific comments on” parasitological indexes approach”

See **C11**.

C44. Line 373 : **Incomplete references and discussion.** *Eimeria* prevalence and density. The author refer to Winternitz, et al.2012, with no demonstrated effects, which results going in the same direction as their analysis. But they omit to refer to Hakkarainen et al. (cited by the author for effect of *Eimeria* body conditions). Is it because they have opposite conclusions, as “The presence of *Eimeria* parasites was higher in dense mainland populations than insparseely populated islands”.

We cite Winternitz et al. 2012 to show that lower density does not necessarily correlate with increased prevalence (“This is, however, not a general law [...] Independent of the direction of the effect, correlation between abundance and prevalence could be confounded with intrinsic effects of hybrid hosts.”). For completion of the discussion we added the Hakkarainen et al. citation.

Lines 373-375 “Independent of the direction of the effect (of the density on transmission) , correlation between abundance and prevalence could be confounded with intrinsic effects of hybrid hosts”. See above comments on the definition of these two indexes. Abundance is not more correlated to prevalence that it is correlated to intensity. In other words, **these three indexes are linked as they reflect distributions of parasite**. Zero loads not only reflect similar transmission or density but also potential resistance and clearances (i.e. intrinsic abilities of the host). See above comments on the origin of zero load and upcoming specific comments on” parasitological indexes approach”

See **C11**.

C45. Lines 380 -388: **inappropriate statements for eco-evo-parasitological point of view**. See previous comments about pathogenic versus parasite costs, as well as about selection of resistance and immune mechanisms in mice with pinworms.

See **C4.&C6**.

C46. Lines 389-394. I do not agree with these statements and the conclusion. Parasite load reflects phenotypically the present stage of host- parasite interaction, that is either host resistance or tolerance (which are simply two facets of the same immune or physiological mechanisms), weighted by parasite virulence. Body Condition could be a marker of parasite costs in some models but not in the absolute. Lot of infections would never affect this phenotype, some infections while favor higher body weight (and

consequently body conditions as estimated in this publication). **Do not confound symptoms and parasite effect on fitness.** See previous comments about these points and specific comments on *Eimeria*

See C40/2.

C47. Lines 397-398 : Are there really hybrids with strong heterosis on variable phenotypes in the hybrid zone in which genomes are highly recombined (not those of experimental F1)?

These are general statements about hybrids (not mice in particular). “Intensity of a particular parasite infection is not necessarily correlated with reduced health and a fitness decrease” lines orig. 395-396

C48. Lines 403-405: the authors omit to consider the hypothesis already suggested by Derothe et al., 2004, while it is based on the same first step of reasoning. Indeed, as suggest in this very study, Derothe et al. propose first that genome disruption could affect strongly some functions. The difference between the two hypotheses concerns the hybrid susceptibility (base of the Derothe et al. proposal) and the consequences of the disruption. They propose that a pleiotropic effect acts in the hybrid zone :only disrupted genomes with alleles leading to a somehow compensation are retained (living and reproducing hybrids). But those alleles could impaired some immune functions and then induce susceptibility against some parasites. According to the authors of this very study, there is no hybrid susceptibility. According to me, there is no clear demonstration of their conclusions (no double statistical comparison, no clear modeling design, confusion in parasite parameters and ecology...see all my comments).

We thank the reviewer for all his/her comments that helped us to substantially improve the presentation of our work. We are very confident in our conclusions and hope that, after answering the different comments and correcting our article for clarity, reviewer 1 will agree that we have provided a clear demonstration of our conclusions. To address specific concerns:

“no double statistical comparison, no clear modeling design” answer is addressed comment **C28/2**

“confusion in parasite parameters and ecology” answer is addressed comment **C11**

The Derothe et al., 2004 hypothesis is not based on the same first step of reasoning. For “genome disruption” to “impair some immune functions” requires genes involved in immunity to co-adapt. As pointed out in (Baird 2012) these seem unlikely gene sets in which to see co-adaptation as diversity in immune targeting appears biologically important, favouring recombination which produces new combinations of those genes. Gene sets with high recombination rates are not strong candidates for co-adaptation as particular alleles do not spend sufficient time together to co-adapt.

The reviewer's claim that the current study has 'no clear demonstration of their conclusions [there is no hybrid susceptibility]' might bear inspection if their claims for the *other* studies they repeatedly cite did not rely on exactly the same type of observations: parasite loads in hybrids. To claim that *increased* loads imply important effects, but *reduced* loads (with much better statistical support) do not, appears perverse

C49. Table 1 : Pinworms : the “significantly favored” models in this very study is H3 (potential difference male/ female and between subspecies and H1 for Baird et al. data (no difference male-female but potentially between subspecies). **No comment/ explanation about these differences.** How do you explain this with the same parasite models in the same zone?

We noted these differences (see line 352-355 orig.) without entering into details as our main focus is effect of hybridization. We have now added a brief explanation and modified our text: “We found differences between the Brandenburg and Czech-Bavarian transects in pinworm infection such as distinct loads between males and females and lower prevalence (52.5%) and abundance (18.7) in the former compared to the latter (no significant difference between sexes; prevalence 70.9%, abundance 39.18; Baird et al., 2012). Power to detect distinct male/female load is expected to be reduced by lower prevalence and abundance.” (line 434-443 new)

Figure legends: Lack of legends and explanations

Figure 1

C50. Lines 652-653 “ Spatial organization of the HMMZ was inferred using six autosomal markers (Es1,H6pd, Idh1, Mpi, Np, Sod1)” . See comments on line 122-123 and on Table S1. Precise the complete set of data leading to this map (not the 660!).

See **C16. & C18.**

C51. “ The numbers at the level contours indicate posterior probabilities of population membership for each mouse subspecies”. What do the authors mean by posterior?

This is a specific wording of Bayesian statistics, which uses “prior” and “posterior” probabilities. The software is referenced, and our code modifications are available. Describing this would lead to a lecture on Bayesian statistics, which we think lies outside the focus of our article.

C52. If white dots are mice of the study, very few of them as hybrids with an undeniable part of

recombination (between 0,2-0,8)? This does not seem consistent with the data, especially Table S1.

This software allows to detect separation between two groups and to categorize into this groups, it does not estimate hybrid zone width. We never claim to estimate the width of the zone. For clarity (and to be technically more correct) we modified “the course of the HMMZ” (line 121 old) to “the expected centre” (line 131 new).

Figure 2 .

C53. 1/ Explain the color code in a) and c). Why are some points paler than others?

We added “(points color indicate mice genotype, on a gradient ranging from blue (pure Mmd) to red (pure Mmm); increased number of mice sampled at one point displayed as decrease in opacity).” in the legend.

C54. 2/ “The predicted probability of infection does not differ in more admixed mice (b) for males (green) and females (orange)”. What are the more admixed mice: part of the sampling? Produced by model?

“more admixed mice” are hybrids that are more admixed. See **C37**

C55. “(average observed probability of infection: grey dotted line)” Average of what? Males and females? if so, it could be such an horizontal line. Theoretic average if no hybridization effect?

We added this precision for clarity: “average overall observed probability of infection (prevalence) for males and females considered together: grey dotted line)

C56. 3/ d) what are the white dots? The 70 Individuals with *Eimeria* I suppose.

We added this precision for clarity “(white dots = individual mice)”

C57. When considered HI between 0,25 and 0,75, no mouse show *Eimeria* intensity (as estimated) more than 0. See my previous comments about mortality and *Eimeria* specific comments.

Our sampling locations are selected to capture the broadest possible range of hybrid genotypes, but we cannot know the exact hybrid index of a mouse before capturing and genotyping it. Nevertheless, the maximum likelihood estimation takes into account this difference of density, as shown by the wider confidence interval in the centre of the HI. There is no narrowness of the HZ, as addressed comment **C52**.

Mortality: see **C11**.

Figure 3:

C58. Same questions than for figure 2 1/ and 2/

We corrected on this plot legend as well see **C53**.

C59. 3/remark: it clearly appears from those distributions that mice with HI between 0,25 and 0,75 are less numerous than parental ones (narrowness of the HZ figure 1 and Table S1).

See **C57**.

Comments on table S1 (see xls joint document)

C60. I indeed identify 51 mice from 660 (near 8 %) to which HI was estimated from 4 to 9 loci.

For a very few of them, clearly too much data are lacking and the HI is not valuable (SK 2891, AA_0272)

It must be noted that for AA_0169, HI was determined on 6 loci and not 7 as indicated)

Corrections made; See **C16**.

C61. Note that for a unneglectable number of the 51 mice, as SK_3374 _E2, AA_0304_E1, SK_3153-E1, SK_3153-E2, SK_3153-E3, SK_3153-E4,SK_3153-E5 (underlined in very pale green on .xls table) there is lack of numerous genetic data but no identify as NA : Why? It must be explain too what represent the 5 mice designated as SK 3153? Same mice but considered 5 times?

See **C16**.

C62. For most of the 51, the most troubling is that HI of some groups of mice is **determined on markers entirely different** from those identifying other batches (underlined yellow gold if all, primrose yellow if part of the non-autosomal markers versus orange for autosomal markers). There is no proof that one kind of markers can substitute for the other, but rather already demonstrated that they are complementary and that the use of several markers (mitochondrial, Y or X, autosomal) is needed to a correct genetic estimate of hybridization. It is not seem appropriate to retain these individuals without any directly compared genetic data to establish the distribution of the genomes in the studied zone. The authors write that the 6 autosomal loci were genotyped in all mice allowing to establish the figure 1, **which is false, since for 17**

mice the data are missing for all these markers. Note that lack of data on at least one of these 6 autosomal markers are present for 3 others mice from these very 51 and for 36 of the remaining 609 (if I did not make a mistake counting too quickly)

See **C16**.

C63. Moreover, the authors “use” several of these 51 mice, which showed little supported HI as data for body weight, pinworm load, even *Eimeria* load. The assignment of these parasite data or health status to a weak hybridization level is not without consequence on the analysis of the models and therefore the results of the publication.

I strongly suggest to eliminate those 51 mice from the whole data set, to build the map and the models (load, body weight..) on the remaining 609 mice, and to clarify rigorously the number of complete genetic data leading to the map.

See **C16**.

Comments Concerning *Eimeria* model

C64. A/ Species diversity across the HZ. Species found in the HZ are those of Jarquín-Díaz et al., 2019, as being part of the same sampling (360 versus 660 mice). They reveal that not only *E. ferissi* is the most prevalent in the zone (very few *E. falciformis* and very very few *E. vermiformis*, as mentioned by the authors), but also that *E. ferissi* is the most widely distributed across the zone, unlike the two other species found in only one or two localities of the sampled zone (figure 1a, Jarquín-Díaz et al., 2019). *E. ferissi* is too the species for which no weigh loss has been identified by Al-khlifeh et al., 2019 (strain derived from the same sampling again). **The same result on the same data is in fact use two time and present as a validation**, which is no rigorous.

See **C7. & C9**.

C65. B /Induced mortality. The studies to which the authors referred (Fuller & Blaustein, 1996; Hakkarainen et al., 2007) clearly insist and investigate ones of the main consequence of *Eimeria* in rodent hosts: mortality and population regulation. But the authors do not even mention these results. Let focus on 3 main references about the *Eimeria* species studied in this study: one with *E. ferissi* (Klesius and Heinds, 1979) , two with *E. falciformis* (Stockdale et al, 1985; Marhrt and Shi, 1988). They suggest that these *Eimeria* species can kill their hosts according to the dose of infection, to the host genetic, especially

if immune-“compromised” somehow or other. These studies argue for more drastic effect on fitness than any potential body weight loss. Why do the authors neither mention nor discuss these aspects and the potential biases to estimate them in their one sampling design (same year period²) and in the biological situation (estimations of density differential at temporal scales- The authors discuss density as extrinsic spatial factor, and not from the point of view of under or non- estimation of the mortality). The authors interpret the results of their optimized fit as the absence of high *Eimeria* loads in mice with HI between 0.25 and 0.75. But it can alternatively indicate a mortality eliminating the mice with heavy loads from their initial field sample. Given the narrowness of the area occupied by mice with a highly recombined genome (Figure 1), no difference in population density would be seen!

See **C11**. There is no narrowness of the HZ, as addressed comment **C52**.

C66. C / Quantification of the intensities thanks qPCR

Even after very attentive reading of Jarquín-Díaz et al. 2019 and Ahmed et al 2019, I find not validation of the relevance of the correlation between qPCR and subsequent Δ parameter with individual *Eimeria* loads (i.e. intensity). Not only there is not reference to real validation or argumentations from similar works in the 3 studies (Jarquín-Díaz et al. 2019, Ahmed et al 2019, this very study), but also the authors use sub-sampling of the same data set from the wild, perform QPCR on it, perform insufficiently justified calculation and correlation analysis , and then assess “ it works! we can apply it on all the sample as an estimation of intensity”. This is in my opinion a serious short-circuit of reasoning

The author clearly want to use these non-helminth parasites as a quite definitive element of the demonstration or not of hybrid susceptibility. So, I have trouble understanding why they have not taken the time:

- to build an rigorous experimental protocol on a mouse strain that they handle easily (NMRI for example) , - to realize controlled infections and load data evaluated from histology and coprology (which they know well) - to couple these data with an approach by qPCR they usually performed well too.

This experiment would allow them to validate rigorously and definitively the qPCR estimate of *Eimeria* intensity. In all cases for me, their assertion of the link between their qPCR indicator and the

intensity is actually not acceptable. At least, I least some arguments for mandatory validation in literature and parasites models

²Note that pinworm distributions are not known as season-dependent in mouse populations. It could be verified

A/ Eaton et al. 2016. *Parasites & Vectors* volume 9, 38. *A. lumbricoides* and *N. americanus*. Positive correlations between eggs count and QPCR on human stool samples. More Efficient with *A. lumbricoides*. Note

- that the relationship between number of expelled nematodes after treatment with both DNA quantification and Eggs counts is also evaluated.

- number of references testing the relevance of QPCR versus fecal eggs count, importance of the comparison and standardization : “The fact that the quantitative measures from the KK and qPCR tests are highly correlated indicates that DNA concentrations calculated against a standard curve for qPCR can be used as a quantitative measure of infection intensity.”

- the authors claim for the use of more sensitive qPCR tool for prevalence and even intensity, but with prior rigorous standardization.

See Eaton et al., 2017 . *Parasites & Vectors* volume 10, 256

B/ Sijbrandta et al. 2017. *Parasitology* QPCR *Plasmodium* Bird : Comparison of sensitivity between two PCR methods, complete explanation of the protocol, calibration and assumption to quantify parasite load. “Thus, this study describes the validation and use of a real-time polymerase chain reaction (qPCR) protocol targeting the conserved large subunit ribosomal-RNA (LSUrRNA) gene described by Friedl and Groscurth (2012) to detect and assess infection intensity with *Plasmodium* spp. in NZ birds”.

Based on Friedl, T. W. P. and Groscurth, E. (2012). A real-time PCR protocol for simple and fast quantification of blood parasite infections in evolutionary and ecological studies and some data on intensities of blood parasite infections in a subtropical weaverbird. *Journal of Ornithology* 153, 239– 247. doi:10.1007/s10336-011-0735-9. “Here, we present a relatively simple, fast and reliable protocol based on quantitative real-time PCR to determine the intensity of infections with blood parasites of the genus *Plasmodium* and/or *Haemoproteus* in blood samples of birds, using male Red Bishops (*Euplectes orix*; Ploceidae, Passeriformes) as example. The intensity of infections is assessed by amplification of a specific

85-bp fragment within the plastid-like large subunit ribosomal-RNA (LSU- rRNA) gene, which is conservative across a range of *Plasmodium* and *Haemoproteus* species. By measuring the accumulation of the product during the PCR (in real-time) using a fluorescent labelled oligonucleotide probe, a threshold can be determined at which the fluorescence of the product raises above background level. The starting quantity of blood parasites in the investigated blood samples is then calculated by comparison with thresholds determined for standards of known quantity (clones of a 594-bp fragment within the LSU-rRNA gene from *Plasmodium falciparum* including the target sequence) in the same PCR reaction. With this method, blood parasites were detected in 123 out of 127 samples from male Red Bishops, with a median of 0.059 blood parasites per 100 blood cells (range 0–19.2 blood parasites per 100 blood cells). The method described here produces consistent and reproducible data, can easily be modified and extended to detect and quantify blood parasites at different systematic levels, and thus has broad application to many researchers in the field of evolutionary and behavioral ecology.”

C/ Pett et al. 2016. *Malaria Journal* volume 15, 539. Correlation between estimation of gametocytes in blood / qPCR. Rigorous experimental calibration performed with different dilutions of gametocytes from cell cultures and various *Plasmodium* strains. Clear explanation of their positive threshold and Δ . GLM analysis.

This protocol has been validated, we added for precision: “This method was validated in an infection experiment in NMRI mice (Al-khlifeh et al., 2019). We considered $\Delta C_t = -5$ our limit of detection as at this limit it was possible to obtain genotyping data for all samples using independent PCR reactions (Ahmed et al., 2019; Jarquín-Díaz et al., 2019).” (line 154 orig., lines 172 new). The genotyping PCRs of Jarquín-Díaz et al., 2019 are methodologically independent. We do not compare this to oocyst shedding, as we have evidence from e.g. re-infection (homologous and heterologous challenge of *E. ferrisi* / *E. falciformis*), that intestinal intensity can be high without any oocyst shedding. We argue that the intensity of intestinal parasite stages is a more meaningful measure than parasite reproduction (oocysts in faeces). See also **C11**.

Specific comments on “ parasitological indexes approach”.

If I agree that Abundance could be insufficient to reflect an epidemiological situation, lot of well – design parasitological studies used the three parasitological parameters, in order to describe the parasite distribution. Indeed, the relevance of the Prevalence and Intensity approach (prone by the authors) compared to the Abundance approach is not prove in this study. I agree that prevalence leads the authors to argue (and not prove, see above) against potential difference of transmission. But the comparison of

intensity is performed by fitting to a binomial negative law (or a Weibull law). Indeed, trouble to me is that to fit well with these laws all the data, including the zero loads, are essential. From what I understand from their statistical approach, this omission could lead to misjudge the estimating parameters. If arguments against my analysis, please develop them, and/or explain more precisely.

For an analysis designed for data that does not include zero loads, including the zero loads is not only not essential, it would be non-sensical.

An answer on the technical level is that the fit of our models was carefully optimised, see **C30**. To answer on an almost philosophical level: Natural sciences can only measure our uncertainty, we leave proofs to mathematicians. We thank the reviewer for their rigor and hope to engage in an open and fair discussion.

Supplementary general comment: Could we still compare results from so distant studies in the HZ, as parasite distributions determined 33 years apart? Are the epidemiological situations the same? What about environmental changes (urbanization/ agriculture, climate ...) and their consequence on mouse populations? Have we estimate the disturbance related to all trapping campaigns carried out during these forty consecutive years? Genetic and dynamic on one hand, epidemiological in another hand? I would appreciate discussion on these events to weight all the comparisons and results.

The reviewer's question is most easily answered by reflection: Should studies from distant parts of the HZ *not* be compared? We feel that if we had made no such comparison this reviewer would also have objected.

The epidemiological and ecological situation in the field could obviously have changed in the past 33 years. We, however, find it more likely that earlier studies based on roughly tenfold lower sample sizes, field sampling protocols that obscure their results (mice kept for 2 months allowing infection after capture) and an inappropriate statistical approach, were plainly wrong. This, as our preferred weighting, is clearly reflected in the manuscript, we hope. It is daunting how much more evidence (ours and Baird et al. 2012) seems required to overturn a result only validated by being published earlier (Sage et al. 1986 and Moulia *et al.* 1991). We believe that progress and activity in the field of parasites in hybrid mice (and hybrids in general) is severely hampered by the burden to “disprove” these low quality studies.

Response to reviewer 2.

Comments to the Author

This manuscript tests the hypothesis that hybrid hosts will be more susceptible to infection than parentals. I have a few comments below:

Line 4 – is this the thinking for hybrid mice, or hybrids in general? Would be best to have more broadly conceptual introduction to the experimental work.

Lines 12-14 – Better to show what you did find given the tight word constraints of an abstract.

We thank the reviewer for these comments on the abstract. We updated the abstract and the discussion to start with general points on hybridization.

Line 14 – What is meant by ‘diverse’ parasites?

We now developed this notion including extra vs. intra-cellular, targeting a Th2 vs a Th1 immune pathways, possible different impact on host health. We also introduced in more details this idea in the **introduction** (line 70-75 new), and in the **discussion** (line 422-427 new)

Line 40 – ‘our own species’ should be removed.

We modified it for “humans” (line 23 new)

Line 121 – ‘the course’ of the HMMZ’. What does this mean?

We changed for clarity “the course” to “the expected centre” (line 131 new)

Line 133 – how do the authors define ‘most prevalent’?

We counted all worms, but only include the more prevalent as we need high statistical power to answer our question. The text has been modified and a supplementary figure added as follows:

“As in this study we need a high statistical power to test our hypothesis, we considered only the most prevalent helminths, the oxyurids *Syphacia obvelata* and *Aspiculuris tetraptera*. Histograms presenting the distribution of counts for other helminths can be found in **Supplementary Figure S2.**” (lines 147-149 new)

Line 338 – ‘To our knowledge no studies have previously tested the effect of mouse hybridization on parasites other than helminths in a field setting of the HMMZ.’ Why would there be a different expectation for non-helminths or parasites of higher pathogenicity?

Lines 346-349 – More elaboration and references are required. Why would pathogenicity matter? What is meant by host-parasite and parasite interplay? Also, there should be many references for the statement that immune responses might differ with different parasite species.

Our study is designed to jointly study two different parasites. We test if any general hybrid effect on resistance and impact on health could be detected. We chose to focus on two parasites from very different phyla, namely Nematoda and Apicomplexa. We wanted to test the likelihood of hybrid resistance or susceptibility against a null model (hybrid parasite load = average of parental loads), and finding the same results on both these phyla is for sure reinforcing our confidence in the results. As a result of the reviewer’s comments, we developed more on the differences that make this comparison valuable, and abandoned the term “pathogenicity” which is too vague to carry all differences. Instead we highlight all differences between the parasites: a) extra vs. intra-cellular, b) targeting Th2 vs Th1 immune pathways, c) possible different impact on host health. We now introduce these aspects in more detail in the **introduction** (line 70-77 new), and in the **discussion** (line 422-427 new).

Paragraph starting at 360 – Why are new questions being mentioned at this point in the manuscript?

This is a repetition of our study design, as a reminder of the introduction (“To distinguish between the load interpretations we therefore, in a new transect replicate of the HMMZ, asked if (1) parasite loads are higher or lower in hybrids compared to parentals, and (2) if these loads are consistent, or differ, between prevalent representative helminth and protozoan.”). We think that this repetition is needed at this stage, for clarity. Reviewer 1 stressed this point.

Line 408 – Can authors elaborate on these future experiments and what they might deduce? This is unclear.

We developed as follow:

“We can in future ask whether host (immunity and resistance) parasite (infectivity and virulence) or their interactions are underlying reduced parasite intensity in hybrid house mice. *Eimeria* spp. are suitable pathogens to perform experimental and field studies in this endeavour. An experimental setup investigating resistance (inverse of parasite intensity) and tolerance (impact on host health measured by weight loss) during an infection in mice of pure subspecies and crosses between them could address this question in more detail.” (line 492-497 new)

More information on the state of the literature on hybrids and susceptibility more widely is needed before going into specifics about the study system in the introduction. For example, it seems like Paragraph 2 should come before most of paragraph 1.

We thank the reviewer for this comment and modified the introduction accordingly.

‘Parasites are omnipresent in natural systems and so it is important for biologists interested in hybridization to comprehend the interplay between parasites and hosts under hybridization.’ There is surely better, more conceptually-driven motivation for conducting the study. There should be enough literature available on the connection between hybridization and condition (or susceptibility) to make clear hypotheses and predictions to motivate the study.

The literature in this field is quite controversial. There is disagreement on both the direction of effects of hybridization on parasites in the house mouse system (see Sage et al. 1986 and Moulia et al. 1991 vs. Baird et al. 2012 and our study) and on the interpretation of these findings with regards to host fitness and hybridization (see for example Theodosopoulos et al. 2019; Baird & Goüy de Bellocq, 2019).

<https://www.sciencedirect.com/science/article/pii/S0169534718302325?via%3Dihub>

[https://www.cell.com/trends/ecology-evolution/fulltext/S0169-5347\(19\)30026-6](https://www.cell.com/trends/ecology-evolution/fulltext/S0169-5347(19)30026-6)

We therefore focus on the study system and limit our hypotheses to hybrid resistance vs. hybrid susceptibility without favoring one result *a priori*. We hope that our work will help to resolve the long lasting debate about (even) the direction of hybrid effects on parasites in the house mouse system (see also comments of reviewer 1). We follow Baird & Goüy de Bellocq, 2019 in their view that broader hypotheses on especially fitness effects of parasites on their hosts (or even the maintenance of hybrid zones) should be avoided, especially without evidence for a link between parasites and their impact on host health (if this is what the reviewer means by “condition”).