Parasites make hosts more profitable but less available to predators

Loïc Prosnier, Nicolas Loeuille, Florence D. Hulot, David Renault, Christophe Piscart, Baptiste Bicocchi, Muriel Deparis, Matthieu Lam, and Vincent Médoc

<u>Responses to the Recommender, Luis Schiesari</u>

- Both reviewers and myself consider that your manuscript is now improved, in particular with respect to (i) the clarification of the Methods (Table 1, for example, is helpful) and (ii) the reorganization of the material presented in the Results and Appendices, which (iii) increased the emphasis on the joint analysis (MFA) of results of the experimental infection.

  However, there are still several points made by the referees and myself that have to be considered prior to recommendation of your manuscript by PCI Ecology.

We thank you. We expect that, following your comments, the manuscript was improved, particularly about statistical analysis, and the consideration of non-white/healthy/covert individuals.

- From my side, I am still concerned about the reliance on phenotypic analysis for determination of infection status, and I do not think the rebuttal letter did a sufficiently thorough job in addressing my concerns.

  The rebuttal letter is clear in explaining that Daphnia were never actually tested for infection status.

  But I also asked whether the authors could instead provide reflectance data for Daphnia that were exposed to presumably infected Daphnia cadavers (i.e., because had the white phenotype) versus Daphnia that were exposed to presumably uninfected Daphnia cadavers (i.e. because had the non-white phenotype). This would be easier for the readers to accept than presenting reflectance data from wild individuals, as the experimental results were indeed consistent with infection of at least part of the individuals.

We apologize if it was not clear in our previous answer. We do not have the reflectance of experimentally-infected *D. magna* because we did not plan to measure reflectance when we performed the experimental infections. We decided to do it one year later on the naturally-infected individuals. We did not conserve dead *Daphnia* and anyway it is difficult to preserve coloration of on conserved individuals (personal observation). Our aim was not to know whether infection induces the white coloration but to characterize the spectrum of the white phenotype. Finally, although this remains to be formally tested, we expect no difference between the "natural" and experimental white individuals.

- I then asked whether Daphnia from Bercy and La Villette had previously been subject to DIV-1 testing. If a previous study did demonstrate that Daphnia and DIV-1 actually coexist in these two ponds, the readers would feel more comfortable about your studies based on wild caught Daphnia. If no one ever tested Daphnia in these two ponds for DIV-1, the readers would feel a little more comfortable if the authors were able to say ´DIV-1 infection of Daphnia magna is common in ponds surrounding Paris´ (with the appropriate references) or ´DIV-1 is a virus parasite of Daphnia magna that is widespread in ponds in Central Europe, and iridescence in D. magna cannot be attributed to any other parasite or physiological change to date´ (with the appropriate references).

We added: "This white phenotype is highly characteristic to an iridovirus, and only one, the DIV-1, was recently identified by Toenshoff et al. (2018). They used only one Finland population for the determination but found that this highly specific parasite also infects *D. magna* from European ponds (e.g., in France), known to have individuals showing the White Fat Cells Disease. Thus, it is likely that our specimens displaying the White Fat Cell Disease (i.e., the white coloration) were infected with DIV-1." (L133-138)

- On the same line, the authors need to carefully go through the manuscript and adjust the text regarding the infection status of wild caught Daphnia. For example, lines 340-342 read ´Concerning D. magna coloration (Measure 6), we found three peaks in the spectrum that were compared between healthy and infected individuals using Wilcoxon signed-rank tests, because data were not normally distributed´. Likewise, the legend in

Figure reads ´Effects of DIV-1 on reflectance between 280 and 850 nm. Blue (dashed) lines are healthy D. magna and red (solid) lines are infected D. magna.´

In no case you can say that wild Daphnia are infected or uninfected, or healthy or infected. Perhaps you could say ´presumably infected´ and ´presumably uninfected´, or ´white phenotype´ and ´non-white phenotype´ in every case that you refer to wild Daphnia. Of course it does not read as nice, but this is really all you can say.

We agree with you on our inability to formally confirm infection status. Because we assumed that white *Daphnia* host the causative agent of the White Fat Cell Disease (as previous authors did), and that WFCD is caused by DIV-1 infection (according to Toenshoff et al., 2018), we thus believe that white individuals can reasonably be considered as DIV-1 infected. The experimental infection supports this assumption. The concern is also true for the non-white individuals, with the problem of covert infection. For instance, during the experimental infection, control *Daphnia* were exposed to individuals expected to be uninfected based on the absence of white coloration. To discuss these limitations, we added more systematically that the infected and non-infected status were determined by the coloration: e.g., "naturally infected (i.e., white) and uninfected (i.e., non-white)" (L390) or conversely "of white *D. magna* (likely infected)" (L378) (particularly for the measure of coloration). We also added in the discussion "Note that we considered along this study that individuals with the white phenotype (i.e., previously named the White Fat Cell Disease) are infected by DIV-1 (Toenshoff et al., 2018), and that non-white individuals are not infected (but see the discussion about exposed individuals from the experimental infection)." (L410-414) Following a comment of reviewer 2, we also added in the discussion "For instance, here, we found a very low prevalence of DIV-1 (3%) based on individuals showing the white phenotype, suggesting little consequence on ecological dynamics. However, if there is a high prevalence of covert-infected *D. magna* showing (at least) reduced survival and mobility, then consequences on communities should be stronger than expected from the prevalence and phenotype alterations of patent-infected individuals only. Covert infection could explain why our apparently "healthy" individuals are more variable in terms of mobility than the infected ones, with potentially bigger differences between *D. magna* that are actually uninfected and patent-infected individuals." (L572-580)

Responses to the Reviewer 1, Thierry De Meeus

3

- Despite the undisputable quality of this work, I believe that there are still some minor problems that will require being addressed before this preprint can be recommended.

Thank you for your valuable comments and numerous corrections. We expect that the current version is improved, with changes in the statistics – despite our different opinion in some statistical approaches (but leading to interesting thoughts).

- I) About the threshold of 5%. I still feel quite uncomfortable with this. Everybody knows that the conventional threshold is 5%. However, not everybody realize that this is a convention, and that each researcher is responsible for the statistical decision to make with a p-value=0.04999, or 0.0501. This convention, by definition, is not an undisputable law. I still think that the concerned sentence is useless.

We agree with you about the general problem of significant threshold in science. However, we consider that because "this is a convention", a non-official convention, we need to precise it in the M&M. We think that it would be dishonest to use a threshold at 5% without clearly mentioning it (and if we only delete this sentence, then it would be the case). Furthermore, since the use of the term "significant" with a threshold is a convention, we believe that researchers who do not use a fixed threshold should avoid this specific vocabulary.

- II) About parametric/non parametric tests. Authors replied "We did the non-parametric test (Kruskal Wallis/Wilcoxon) and found results that are consistent with the ones we present here…After discussion with statisticians, we kept the parametric tests with log transformation, rather than a non-parametric test, because it is recognized that rank tests lead to a "loss of information", thus are "less efficient or less powerful"; consequently "non-parametric methods are justified when conditions are not satisfied for other methods, after variable transformations" (Dagnelie, 2006, Statistique théorique et appliquée, 2nd ed., de Boeck)".

     This is an odd answer. 1) If the non-parametric tests gave the same results, I do not see where the "loss of information" is.

4

There is a loss of information in the statistical methodology: with a rank-test we loss the exact value because it only orders the values. The parametric and non-parametric tests could give qualitatively the same result (according to our threshold) if this loss of information not affect greatly p-values (but the p-values are not exactly the same) – maybe due to the sample size or to the sample distribution.

- 2) If the log transformed data are not normally distributed and without homoscedasticity, then the result may be not so good. I could not find where the authors tested for the normality and homoscedasticity of their log-transformed data.

You are right that we missed to mention that we tested log-transformed as well as non-transformed data for normality and homoscedasticity. We now detail "normally distributed, sometimes after a log-transformation" (L312) and "a log-transformation, leading to normally distributed and homoscedastic data." (L851) Note that following your next comments we changed some log-transfo by GLM(Gamma).

- 3) In my long carrier, dealing with parasite distributions and other non-normal data, I have met several situations when the statistical analysis undertaken with non-parametric statistics gave a significant result, while the parametric test did not. So, the "loss of information" is not always on the same side.

Our strategy is to find the more suitable test, which does not depend on the p value. In your examples, maybe, the loss of information could lead to false positives. But, of course, researchers need to do with the balance of false positives and false negatives.

- If the non-parametric test provided a non-significant result, while the parametric one outputted a significant one with log transformed data, I would indeed recognize that further argument would have been necessary to explain why. However, the results was apparently the same. So why bother?

We believe that we should use the most appropriate tests (i.e., parametric tests if possible because more powerful from a statistical point of view), according to statistic theories (following statisticians and handbooks), so that we and readers are more confident in the results.

- III) About random effect. I am not convinced by the argument of the authors. What they describe in their rebuttal corresponds more to a nested factor than to a random one. I am not a statistician, but I am not sure that a nested factor would have provided the same result as the mixed model used by the author.

We apologize because we had not fully understood your previous remark about this point. We now use "a mixed model with sampling dates and ponds niched in the infection status and in egg status" (L314)

- IV) About one-sided tests, authors wrote: "It seems not technically possible to do a one-sided test for survival analysis (and it should not affect our conclusions)".

    One little trick of mine when a software does not provide one-sided tests is to check the direction of the response and to halve the p-value if the response is in the expected direction, and compute 1-(p-value/2) otherwise. It should provide an approximate one-sided p-value. This is really a minor remark.

You are right. But because two-sided test has a lower acceptance rate, it will not modify our results.

- 1) Line 106: May be authors could mention here that Daphnia iridescent virus 1 only infects Daphnia and not its predators.

We added "the predators that do not risk infection by this highly specific parasite" (L109)

- 2) Line 313: I would have undertaken a log normal regression (glm with poisson), without log transformation instead of an anova on log transformed data.

Clutch frequency and Mean clutch size are continuous data, thus do not follow a Poisson law. Following your suggestion, we now use GLM with Gamma (residual analysis is quite similar than with our previous log-transfo model, and statistically better than a Kruskal test).

- 3) Line 334: If Holm adjustment method is the sequential Bonferroni, as I think it is, then I would suggest using the less conservative Benjamini and Hochberg (BH in R).

6

Thank you to inform us about the existence of BH, and thus the interesting discussion of "FDR vs FWER". However, we prefer to limit the number of false positive, than to limit the number of false negative. Readers should be more confident with our observed effect if we use a more conservative approach.

- 4) Lines 351-353: "Based on the data obtained (Measures 5 and 7), 100 healthy and 100 infected D. magma were generated using a bootstrapped method (5,000 iterations), allowing for each individual to calculate a profitability."
  Please rephrase, e.g.:
  "Based on the data obtained (Measures 5 and 7), 100 healthy and 100 infected D. magma were generated using a bootstrapped method (5,000 iterations). This procedure allowed computing a profitability for each Daphnia individual."

Modified, thank you.

- 5) Line 357: I think that, if I understood well what it is about, to directly test if the distribution of your p-values significantly deviates from what would be expected if each test had been undertaken under H0, you can undertake a generalized binomial procedure (Teriokhin et al., 2007), with MultiTest (De Meeûs et al., 2009). Nevertheless, combined tests need being independent and testing the exact same H0. I tried to find it in the R supplementary files to get the series of p-values to combine, so that I could undertake it myself, but failed to find those (it probably would not change much things). Where are these data and the associated Kolmogorov-Smirnov test (KS)?

We think we could not use your approach because we randomly generate a null model, thus a new H0, at each iteration.
You are right that we missed to add the R script to the Zenodo data (we added it in the new version). We also missed to add the table C8. We corrected it. Thank you.

- 6) Line 419: Following the previous point, Table C8 is missing. This, with the missing KS test and the series of p-values, represent an important remark regarding the policy of PCIs.

You are right that we missed to add it to the manuscript. Following your previous comment, we corrected it, thank you.

Responses to the Reviewer 2, Eglantine Mathieu-Bégné

- I think most of the comments I raised previously were considered and I appreciate that several points have been nuanced.

Thank you for your useful comments. We expect that the structure, the statistics, and the discussion are still improved following your comments.

- First, while reading the manuscript it is not clear what material belongs to the main text and what material belongs to supplementary. Consequently, the statistical analysis still appears quite redundant especially for fecundity and mortality measures. I also suggest to include a visual for the workflow including the different experiments, the number and the origin of the individual considered and the test used to analyze them in order to help to keep track with the different analysis done.

You are right that we had left M&M of appendix in the main text. We now moved it in the appendix. We expect that the workflow is now clear enough. We also added a visual for the workflow, in complement to the table 1, in Appendix (Fig. C1).

- Second, relying on experimental infection the authors demonstrate that not all the exposed Daphnia develop expected infected phenotype (the so-called white phenotype). Considering that for several phenotype traits those exposed Daphnia exhibit intermediate value compared to healthy and infected Daphnia, could one consider that exposed Daphnia are simply infected Daphnia that do not display all the symptoms of infection with the same intensity and thus don't display the white phenotype? If so, how would that impact conclusions drawn on natural infection on which the infection status was evaluated on the basis of the white phenotype only? Still on this point, it is several times mentioned that the high virulence of the virus could explain the low prevalence in the investigated ponds. If some daphnia could potentially be infected without having the

white phenotype, it is possible that the prevalence is not as low as expected. This point reaches a previous point made by the recommended Editor.

You are right that considering these individuals is interesting and likely very important. That is why we have a paragraph in the discussion to determine what these individuals are, with the assumption of covert infection. Note that based on the MFA, these individuals are not just intermediate (i.e., between heathy and infected individuals). Maybe we have not discussed enough this issue in the "second interesting point" paragraph. Thus, we added: "For instance, here, we found a very low prevalence of DIV-1 (3%) based on individuals showing the white phenotype, suggesting little consequence on ecological dynamics. However, if there is a high prevalence of covert-infected *D. magna* showing (at least) reduced survival and mobility, then consequences on communities should be stronger than expected from the prevalence and phenotype alterations of patent-infected individuals only. Covert infection could explain why our apparently "healthy" individuals are more variable in terms of mobility than the infected ones, with potentially bigger differences between *D. magna* that are actually uninfected and patent-infected individuals." (L572-580)

- L. 178 and 180, it is written that experimental infections were conducted on 23 and 44 juveniles each time obtained from 11 distinct mothers. I am assuming that for practical reasons those juveniles result from clonal reproduction and thus I was wondering if the fact that some of the juveniles could have the exact same genotype caused problems in terms of independency of the data and if so how this was considered in the analysis.

We designed the system to account for the clones in our experiment. But you are right that it must be included in our statistical analysis. Thus, now, when possible (for fecundity, fitness, mobility and size), we used mixed model with the mother (i.e., clone lineage) as a random factor.

- Minor comments
  L. 137 consider adding the duration of the survival experiments

We added: "Experiment lasted until the death of all D. magna, representing 163 days." (L191)

- L. 465: Please explicit here the difference between the two ponds it is referred to

We added: "in terms of speed and carbohydrate content" (L452)

- L. 535-53: Slower daphnia being less parasitized because they encounter the parasite less is appealing, but this does not align with the observation that healthy daphnia are faster than exposed Daphnia.

Yes, although not corroborated by the phenotypes, we briefly mentioned this hypothesis for the (likely) absence of infection because it could come to the mind of readers. We suggest that some daphnia could be uninfected because they are slower (than both control and infected daphnia). We could imagine that, in Fig B3, if control is the initial pool of daphnia, after exposure, the slower are uninfected (exposed are slower) and the faster are infected, but with a little lower speed than healthy individuals due to infection. We also added: "More, due to our setup where microcosms are small and the medium daily resuspended, this escaped explanation seems unlikely" (L528)

- L. 545: a lack of cost to resistance associated with Pasteuria resistant might be linked to the fact that a great part of the resistance to this bacteria is constitutive, which makes the comparison may be not as relevant depending of the basis of DIV-1 resistance.

We added: "the cost of resistance should depend on the immunity system, which differs between fungi, bacterial and virus infection (McTaggart et al., 2009)" (L537)

- L471. It is stated that Daphnia predators were found only in La Villette pond. Could the fact that the interactions the predators were already in this pond have influenced some of the trends on Daphnia phenotypic differences? For instance, predators could have selectively predated bigger infected individuals and hence this phenotype could be less represented in the pond.

You are right that it could affect other characteristic than speed. We added: "The presence of a predator could also affect other phenotypic characteristic as body size: larger individuals in presence of *Chaoborus* but smaller individuals in presence of fish (Riessen, 1999)." (L461)

- Table1: Were individuals from both ponds pooled in the analysis on the size? Ideally the pond effect should be included in the model. If it is the case, then each pond should be named instead of the line called "both" in the table.

We modified it.

- Supplementary Tables: Consider adding caption or reframing the table so there is one table for each dependent variable, the explicative variables in lines and the model statistics in column (df, statistic, p-value).

We are not sure to understand your advice. We choose to have one table by pool of variables analysed and presented simultaneously. We expect that all statistical information is easily available by reader (lines or columns should not affect it), more since we moved, following your previous comments, all corresponding M&M in the appendix.