Response to Reviewers

We thank the reviewers for their comments – their suggestions have greatly improved the manuscript. Below we provide a detailed point-by-point response to their comments and recommendations, with line numbers referring to the revised manuscript.

Reviewer 1

This manuscript addresses the hypothesis that vertically transmitted symbionts should evolve low virulence as detrimental effects of the symbiont on the host would decrease fitness of the host, and consequently also the endosymbiont. The wMelPop Wolbachia strain that infects Drosophila flies indeed show such an expected low virulence, except at higher temperatures. It has been proposed that this is likely non-adaptive as flies are generally not found at such high temperatures. The authors use this setting to conduct an evolution experiment to test if selection for late reproduction results in the evolution of lower virulence. Interestingly, the virulence of wMelPop can be reduced by supplementing Paraquat to the medium. Although this is an interesting idea and setup, it is not always clear how adequate the experiment is to test this idea. The writing and organization of this manuscript is generally highly unclear and very difficult to follow.

We believe the changes made at the request of the reviewers improved the clarity of the manuscript.

1/ Virulence of Wolbachia at high temperatures is reflected by a strong decrease in host longevity. However, the selection experiment was conducted by selecting for late reproduction, and even this selection procedure is doubtful. Why did the authors not simply select for flies with high survival rate?

We selected for late reproduction as the most practical mean of selecting for longevity. The only alternative we can think of would have required to isolate individual female flies following mating and constitute the next generation only with the offspring of the long-lived females. This would have been very time-consuming and would have forced us to considerably decrease the number of experimental lines.

The survival experiment shows that flies indeed start to show (a very low level) of mortality at the time eggs were selected (8 days after emergence). However, how do the authors know that the eggs selected at that time originate from flies with increased longevity? If this is not the case, it should be no surprise that no effects of treatment are observed or if even any type of selection was performed.

We now clarified in the text (lines 125-126) that “the medium was changed regularly, including at the end of day seven, and the eggs already laid on it were therefore removed” The eggs selected on day 8 therefore originated from flies that were still alive and healthy enough to produce eggs at that time.

2/ L231-237: This is very unclear. The first sentence describes that density and mean octomom copy number did not increase during experimental evolution, but the next sentence describes

that it increased? If this latter refers to an increase between the treatments, this should be clearly stated, as well as for which of the two treatments the increase was observed.

The first sentence actually states that density and mean octomom copy number did not decrease during experimental evolution (which is consistent with the next sentence).

I also would expect that the interaction generation x treatment should be the factor demonstrating an effect of selection, but this is not interpreted as such. On L279 the authors state that 'the increase of time in density and copy number….', which refers to a generation x treatment interaction, but this was not significant.

A significant generation x treatment interaction would indeed indicate an effect of selection, as it would indicate that the paraquat and water (control) lines are not selected in the same way (i.e. one is selected and the other is not, or both are selected but selection is stronger in one condition). However, when the generation x treatment term is not significant (which is the case here), we must turn to the generation factor to determine whether there is an effect of selection.

I'm also surprised to note that the significance levels are so high, as the graphs (Fig. 3) hardly show a clear effect because of the very large variance between the lines.

We checked the statistical analyses and graphs and can confirm the stated results. This is not so surprising, as significance levels reflect the statistical power rather than the effect size.

3/ More minor comments:

a. The results and discussion are completely mixed up. For example, the authors interpret the results of the survival rate comparison directly as virulence comparison. It is much more accurate to describe the results of survival as such, and afterwards discuss how this can be interpreted as virulence.

We modified the paragraph on lines 241-258 so that it now describes survival and extinction rate results as such and added a paragraph on lines 259-261 to discuss the implication for the evolution of virulence.

b. The phrasing of the co-evolution between symbionts and hosts is often scientifically incorrect and teleological. E.g. l33-34: endosymbionts exploit their host to maximize their transmission. This phrasing suggests that they evolved this 'on purpose', which is not the case in evolution. Mean trait values evolve as a consequence of fitness difference, but not for any purpose.

The sentence on lines 34-35 was changed to "Deleterious symbionts, also known as parasites, exploit their hosts in a way that maximizes their transmission".

c. L206-207: Clearly state for which response variable a significant effect was found. E.g. "…did not modify the evolution of virulence", but as different response variables are used as proxy for virulence, readers cannot infer about which response variable the authors are writing.

The sentence previously on lines 206-207 (now line 243) was changed to "did not modify the evolution of survival".

Reviewer 2

Review of Monnin et al 2020 for the Peer Community In Evolutionary Biology

This paper is concerned with the evolution of endosymbiont virulence in the *Wolbachia-Drosophila melanogaster* system. The authors used experimental evolution (and artificial selection) to determine how a high temperature affects endosymbiont virulence and consequently host survival. A high temperature was hypothesized to select for decreased endosymbiont virulence (measured as density, as well as octomom copy number). In contrast to the authors' expectations, they find that density, as well as octomom copy number, increased during the course of the experiment in control and treatment lines. It is argued that this outcome could be due to the artificial manipulation of egg number (which is set a 100 for each generation) and/or directional selection for high density/octomom number variants.

This is an interesting study, because it can shed light on the reasons why high virulent endosymbiotic strains can persist in nature and it can give some more insight into the way populations may be affected when exposed to a high virulent endosymbiotic strain, which is of particular interest from an applied perspective, where virulent strains are used to control the spread of disease through insect vectors. It is unfortunate that this virulent strain of Wolbachia does not actually occur in nature. This reduces the ecological and evolutionary relevance of the study. Overall, the results reported are negative, which is disappointing for the authors considering the amount of work that has been done.

It would have been better if the authors could have included treatments where eggs were not artificially selected, and if "outbred" lines would have been included as a control, because I think this would have helped in understanding why you obtained the results that you got. Alternatively, additional experiments could have been performed (backcrossing to the parental generation) to tease apart the different explanations provided for the results obtained.

Thanks for the comments and all these suggestions. We agree that these controls and experiments may have helped in interpreting the results. Unfortunately, the suggested experiments can no longer be performed.

You mention that 'Results contribute to the ongoing debate on the relationship between wMelPop octomom copy number, density and virulence'. I think that more details need to be provided in the introduction as to what is actually debated in this regard. Who has found what and when, how do previous findings contradict each other, and more importantly, how can new light be shed to resolve this debate?

We agree and added more information in the introduction, on lines 68-74:

"Rohrscheib et al. (2016) indeed found that differences in survival between flies reared at 24°C and 29°C cannot be explained by differences in bacterial density. Similarly, comparing flies of the same age but reared at different temperatures, or of different ages but reared at the same temperature, they found no correlation between octomom copy number and either density or virulence. However, to exclude the possibility that octomom copy number has an effect on density, and density on virulence, the effect of these variables should be assessed independently of temperature and age."

What is the likelihood that reduced virulence would/would not evolve in this particular host strain/Wolbachia strain combination?

We do not know how this likelihood could possibly be calculated or estimated. However, based on knowledge of the variability in *w*MelPop copy number and its rapid evolution (Chrostek & Teixeira 2018), we expected the evolution of virulence to be likely in our experimental setting.

As you rightfully point out in your MS, you see an overall increase in virulence, so there is a response, but you have only used 1 Drosophila strain (mixing different isofemale lines) and 1 Wolbachia strain. Even if your results would have been positive, it would have been very difficult to generalize your findings. Moreover, if more Drosophila strains had been tested, perhaps the results would have been different. I think this should be discussed in the manuscript.

We now acknowledge that "we did not vary the experimental temperature – a factor recently shown to impact on the evolution of *D. melanogaster-Wolbachia* associations by Mazzucco et al. (2020) – and used only one host genetic background (*D. melanogaster*^{w1118}) and one *Wolbachia* strain (*w*MelPop). We cannot exclude the possibility that different conditions would have led to different coevolutionary outcomes" (lines 300-305).

Line 23: Explain what octomom copies are (i.e., copies of a Wolbachia DNA region containing 8 genes).

We clarified that octomom is "an 8-genes region of the *Wolbachia* genome" (line 23).

Line 28: I think this statement is rather misleading, because you are also proposing other explanations for your results, and you do not provide any evidence (even suggestive) that intrahost selection could play a role in your experiment. I suggest you just state that you discuss your results 'with respect to the evolutionary causes of wMelPop virulence'.

We agree and modified the manuscript accordingly (lines 28-29).

Line 40-42: This statement requires several references (assuming the theoretical and experimental work were described in separate papers).

References are given on lines 44, 49 (two theoretical papers) and 52 (four experimental papers).

Line 60-61: Why was the repetition of the octomom proposed as the genomic basis of wMelPop high density and virulence? Because selected lines with different octomom copy numbers determined Wolbachia titers and strength of lethality of the phenotype (Chrostek, Texeira 2015 PLoS Biol). Please explain this in the MS.

We added this information on lines 64-66: "Octomom includes genes encoding proteins potentially involved in DNA replication, repair, recombination, transposition or transcription (Chrostek et al. 2003) and was found to be correlated with *w*MelPop density and virulence (Chrostek & Teixeira 2015)."

Line 62-63: And why were these links called into question? Please explain.

We now explain this on lines 68-74:

"Rohrscheib et al. (2016) indeed found that differences in survival between flies reared at 24°C and 29°C cannot be explained by differences in bacterial density. Similarly, comparing flies of the same age but reared at different temperatures, or of different ages but reared at the same temperature, they found no correlation between octomom copy number and either density or virulence. However, to exclude the possibility that octomom copy number has an effect on density, and density on virulence, the effect of these variables should be assessed independently of temperature and age."

Lines 64-65: If the strain is not known to occur in nature, what is the incentive for doing the experiment from an evolutionary point of view? To better understand the link between virulence and vertical transmission? To see whether extreme environmental conditions can mitigate the negative fitness effects from an applied point of view (as this Wolbachia strain is used to reduce virus transmission in vector insects)? This needs to be explained more clearly in the introduction.

We clarified on lines 94-99 that "despite not being known to occur in nature, the *D. melanogaster*-*w*MelPop association can help to understand the real-life dynamics of host-*Wolbachia* associations, especially in their early stages. Indeed, one can expect that the low virulence typically observed in *Wolbachia* symbioses is the outcome of a relatively long process of coevolution between the symbiotic partners. Newly formed associations, by contrast, might be more likely to exhibit levels of virulence similar to what is observed in *D. melanogaster*-*w*MelPop association."

Lines 67-68: The authors have altered the environmental conditions by increasing the temperature (experimental evolution), but have also enforced late reproduction (artificial selection). In experimental evolution experiments, only the environment is altered, and there is no further selection by the experimenter of specific individuals for producing the next generation. I would suggest that the authors describe their experiments as such: experimental evolution in conjunction with artificial selection.

We now state "we submitted *w*MelPop-infected *D. melanogaster* lines to 17 generations of experimental evolution at a high temperature, while enforcing late reproduction by artificial selection" (lines 19-21) and that "we used experimental evolution in conjunction with artificial selection" (line 84).

Line 109: Why were these generations chosen to perform measurements?

Our goal in choosing these generations was, as much as possible, to assess the variables of interest at regular intervals during the experiment. We also waited a few generations after the start of the experiment, as potential effects of selection would have necessarily been very weak otherwise.

Lines 108-111: Why were eggs used either for assessing survival or Wolbachia-related traits? Why not do both measurements at each generation? As you mention 100 eggs are a subset (line 103), this should be feasible by collecting more eggs. Please explain this in the manuscript.

We now explain on lines 134-136 that "these traits were not measured at every single generation, as we do not expect the changes from one generation to the next to be so great as to justify measuring all traits at each generation".

Figure 1: This figure clearly shows when measurements were taken, but aren't the paraquat lines your 'manipulated' control (where you do not expect selection to occur?). The control lines indicated here are actually the lines that should experience selection. This is confusing. Were there also control lines added that did not experience any treatment, because that would be your actual, unmanipulated, control?

We consider that there was no unmanipulated control (as this would imply "not selected for longevity"). We expected less selection to occur in the paraquat lines (lines 231-232), but not necessarily no selection at all. We are now calling "water" instead of "control" the lines not treated with paraquat.

Lines 116-120: The number of lines that went extinct is quite high (n = 8 and 5 for control and paraquat, respectively) considering the number of lines that were started at the beginning of the experiment (n = 20, 10 for control and 10 for paraquat). Would it have been possible to collect more eggs (e.g., after 8 days of age) to at least ensure survival of the lines? Or to have had more replicates within each line (by collecting/distributing multiple egg batches in different vials)?

The extinction of lines is due to the fact that no egg was laid on day 8. Using the few eggs that may have been laid on day 9 or later would probably have only delayed the extinction. Having replicates within each line would have forced us to use less lines, probably making the experiment less informative.

Line 168: The rationale for doing this experiment needs to be explained in the introduction, because it come out of the blue here.

We now reference this experiment in the introduction (lines 92-94) and explain its purpose in the method section (lines 195-197): "To evaluate the possibility that intra-host selection is involved in the evolution of *w*MelPop octomom copy number, we assessed the relationship between the mean octomom copy number and the age of the host."

Line 202: The fact that survival is affected would indeed suggest that there would be some selection pressure, but the question is how strong this selection pressure really is. As you mention later on (line 301), while survival was affected, it still remained high at 8 days. I wonder if selecting 100 eggs at the age of 8 days could have counteracted the effects you would have expected under your experimental evolution regime (high temperature), because it reduced the strength of selection.

It is indeed possible that during the experiment (at 29°C), the selection was inefficient, due to the relatively low number of eggs used to initiate the generations. However, there is no *w*MelPop-induced mortality at 18°C (the temperature at which the flies were kept prior to the experiment), so we do not expect any selection for reduced virulence in those conditions. We discuss this on lines 330-340.

Line 208: There is a trend though in the interaction between generation and treatment, that is interesting. Might be worth to highlight this.

As this is not significant and thus not easy to make sense of (given that we did not observed what we expected in the control condition, it would have been even more difficult to interpret differences between the control and the paraquat conditions), we prefer not to comment on it.

Lines 233-237: How can you explain this increase in density and octomom number in control and paraquat lines? You start explaining this in the next section, but it makes more sense to me to discuss this here (or results and discussion should be reported in separate sections, rather than together).

We reorganised the manuscript so that the explanation of the results follows their description more directly (starting on line 289). We kept the section titles as they were ("test of the predictions" first, and then "potential explanations for the falsification of the predictions") for the sake of clarity. As these explanations partly rely on another experiment (relationship between octomom copy number and age of the host), we do not find any easy way to report results and discussion in separate sections.

The observed differences over time could be due to the fact that genetic similarity increases again during the course of the experiment (isofemale lines were first crossed, but then new isofemale lines were set up for the experiment from the mixed parental generation).

The experimental lines are not isofemale lines. Each experimental line originated from 100 eggs produced by the mixed parental generation (as explained on lines 115-118).

We now state that "the observed increase in relative *Wolbachia* density and mean octomom copy number is not likely to be due to a mere increase in genetic similarity: at G2, no measured fly had a relative density superior to 40, contrary to 21% of the G17 flies (the maximum being 143); similarly, no flies at G2 had a mean octomom copy number over 6, contrary to 18% of the flies measured at G17 (the maximum being 12)" (lines 296-300).

This could have been tested if unmanipulated control lines were included (with or without artificial selection to tease apart the effect of temperature on the one hand and late reproduction on the other). Another way to test this would have been to backcross the selected lines to the parental strain. This would also allow you to tease apart whether the observed changes are due to the host nuclear genomic background or the wMelPop genome (as discussed in Rohrscheib et al 2016 PLoS Pathogens).

We agree that having an unmanipulated control and/or backcrossing the lines may have been helpful. Unfortunately, the suggested experiments can no longer be performed.

Line 284: If genetic drift was at play, you would expect a random outcome (increase/stable/decrease virulence) in different replicated experimental evolution lines (both treatment and control), right? You would expect that the interaction between generation and treatment was significant (and there is a trend). If I am right, this would mean that you can exclude random drift as an explanation for your findings. I would rather think that the specific genetic background of the line affects the course of virulence evolution.

It seems to us that our results (which are indeed non-random) could be explained by genetic drift in conjunction with a mutational bias (mentioned on line 332).

We would expect the interaction between generation and treatment to be significant only if selection is differentially involved in the paraquat and control lines, not if the pattern is wholly explained by genetic drift and mutational biases.

Line 298-301: If you had included unselected lines that went through the experimental evolution experiment, without the selection of eggs at age 8, you would have been able to tease this apart.

The issue with using unselected lines is that their generation time would have been shorter than the one of the selected lines, making comparison difficult: the variables of interest would have to be measured either (i) at the same time, but following a different number of generations, or (ii) following the same number of generations, but at different times.

Reviewer 3

Review: Experimental evolution of virulence and associated traits in a Drosophila melanogaster – Wolbachia symbiosis.

General comments

David Monnin and colleagues investigated virulence evolution of the symbiotic Wolbachia strain wMelPop in Drosophila under high temperature (at which the strain is assumed to be virulent) and in the presence or absence of paraquat. Their hypotheses were well supported by theoretical expectations regarding the evolution of reduced virulence of vertically transmitted symbionts. The introduction is well-written with clear predictions regarding the experiment. These prediction were, however, not confirmed by their experiment and their main hypothesis - the evolution of reduced virulence - was not confirmed. The discussion provides the alternative explanation that intra-host selection might be important here, and I agree that it might possibly play a role here. I do have some major concerns about how the evolution of virulence was measured.

First, from the manuscript it is not clear which survival parameter you compare between G3 and G9. Time until half of the flies died? mean survival time? I think survival analysis would be the best option here.

We now clarified that “individual longevity was used as the response variable” (lines 208-209) and that “in the absence of censored data, a survival model is not necessary, and the gamma distribution is suitable for continuous positive data” (lines 208-209).

I also wonder if using only fly survival of wMelPop infected flies is a good proxy for virulence? I think that virulence is best measured as the reduction in survival compared to an uninfected control. This way you also correct in some way for measuring survival on different days (is it possible that e.g. differences in food quality between days can be a confounding factor?). Ideally you would even isolate different Wolbachia lineages and re-infect flies to test

everything at once (but maybe this is technically not possible?). My feeling is that possibly some patterns were not found because of the way virulence was measured.

We cannot indeed exclude the possibilities mentioned here, and the suggestions to measure virulence are good ones. Unfortunately, these experiments would have been too time-consuming to be performed. Nevertheless, given that density and octomom copy number increased over time, our conclusion that the virulence of our experimental lines did not evolve in the expected way seems solid.

I also have a general suggestion on providing more information on the population dynamics of wMelPop. I think some information is crucial here to understand how selection might act on these bacteria. Is there a population bottleneck at reproduction/how many bacterial cells are transmitted to the egg?, What is the number of generations within a host? Especially in the light of intra-host selection this might be very important.

We agree that these questions are very important to assess the relative strength of inter and intra-host selection. We now indicate that "at 29°C, the doubling time of *w*MelPop was estimated by Duarte et al. (preprint) to vary from 0.88 to 1.38 days" (lines 353-354) and that "the bottleneck experienced by vertically transmitted symbionts" was "estimated to range from 850 to 8000 cells in aphid-*Buchnera* symbioses (Mira & Moran 2002)" (lines 361-362).

Specific comments

Line 60: do you have more functional information on the genes contained in 'octomom'?

We added more information, as requested (lines 64-66): "Octomom includes genes encoding proteins potentially involved in DNA replication, repair, recombination, transposition or transcription (Chrostek et al. 2003) and was found to be correlated with *w*MelPop density and virulence (Chrostek & Teixeira 2015)".

Line 64: Here it is indicated that this Wolbachia strain is not known to occur in nature. Could you then give a little more explanation where it comes from?

We clarified, following Reynolds et al. (2003), that *w*MelPop may have originated in the lab (line 75-76).

Line 91: I think it is better to mention the experimental temperature in the next section.

This was corrected.

Line 103: Why do you use unsupplemented medium for the paraquat lines in the egg-laying phase of the experiment?

Some of the eggs laid during this phase developed into the adults we used for measuring survival, *Wolbachia* density, and octomom copy number. Contact with paraquat would have been a confounding factor. We now indicate that during this phase of the experiment "the medium was not supplemented with paraquat or water, as this would have been a confounding factor for the subsequent measurements" (lines 123-125).

Line 104: Is it known if wMelPop affects reproduction of the flies? If so, this might actually be a stronger selective force for reduced virulence than longevity.

Reynolds et al. (2003, *Genetics*) did not find any effect of *w*MelPop on w^{1118} *D. melanogester* fecundity (experiments performed at 25°C).

Line 168: Maybe best to quickly describe here why you do this (it is only clear when reading the last section of the discussion).

We now make clear that this experiment was performed "to evaluate the possibility that intra-host selection is involved in the evolution of *w*MelPop octomom copy number" (lines 195-196).

Line 202: In the absence on mortality from non-infected flies it is very difficult to say that this mortality is induced by the Wolbachia.

We now clarified that "we can be confident that fly mortality is induced by *w*MelPop, as the strong reduction of host longevity induced by *w*MelPop has been established previously. A study performed in our lab (Monnin et al. 2016), using the same *D. melanogaster* line and the same survival protocol, found a large mean difference in longevity (4.4±0.66 days) between *w*MelPop-infected and uninfected individuals." (lines 232-236)

Line 213: "virulence possibly decreased subsequently" is a bit weird here, as this is shown not to be the case in line 217.

This was changed to: "As the survival was last assessed at G9, we cannot directly test whether it evolved between G9 and G17" (lines 247-248).

Line 215: Can you statistically test differences in extinction rates for different conditions?

We now indicate that "we used Fisher's exact tests to assess differences in extinction rates" (lines 153-154) and that "differences in extinction rate between the early and late generations were not significant (Fisher's exact tests, p=0.15 in the water lines, p=0.61 in the paraquat lines, p=0.16 overall)" (lines 256-258).

Reviewer 4 (Shira Houwenhuyse)

Remarks Monnin et al., 2020 Experimental evolution of virulence and associated traits in a Drosophila melanogaster – Wolbachia symbiosis

Abstract

- Line 17: why is it important to mention "high rearing temperatures". Does the infectivity of Wolbachia depend on temperature?

We clarified that "It was indeed found that *w*MelPop reduces fly survival at 25°C, but not at 19°C (Reynolds et al. 2003)" in the introduction, on line 78-79.

- Line 23: what are octomom copies?

We clarified that octomom is an 8-genes region of the *Wolbachia* genome (line 23).

- Line 23: to which conditions are referred? You mean high temperature and enforced late reproduction?

Yes, and we clarified that (line 24).

- Line 25: are correlated with what? With each other?

We clarified that density, octomom copy number and virulence are indeed correlated to each other (line 26).

Introduction

- Line 34: trade-off. You mention that parasites face a trade-off to maximize their transmission. In the next sentences first a positive correlation is mentioned and then this trade-off is described. For me it was not directly clear what the trade-off was. I think it would be more clear if you directly mention the players in this trade-off before you explain it: "… face a trade-off between the transmission rate and its virulence".

We agree and modified the text thus: "Deleterious symbionts […] usually face a trade-off between instantaneous transmission rate and opportunities for transmission" (lines 34-36).

- Line 45: inter-host and intra-host. Can you explain what you mean with this two types of selection?

We clarified inter-host selection is "between symbionts infecting different hosts" (line 46) and that intra-host selection is "between symbionts infecting the same host" (lines 47-48).

- Line 56: I think it would be a good idea to mention here a bit more about what is known about the virulence of Wolbachia under different temperatures.

We added that "it was indeed found that *w*MelPop reduces fly survival at 25°C, but not at 19°C (Reynolds et al. 2003)" on lines 78-79.

Material and Methods

- Line 99: What is paraquat exactly?

We clarified that paraquat is a pro-oxidant compound and herbicide and that it is also known as 1,1′dimethyl-4-4′bipyridynium dichloride (lines 87-88).

- Line 126: How was the extinction rate calculated? What do you mean with the extinction rate was measured both before and after the last survival measurement (G9)?

We clarified that "the line extinction rate – the ratio of lines that went extinct over the total number of lines – was calculated for both the water and the paraquat lines, both during the 8 generations prior to G9 and during the 9 subsequent generations" (lines 150-153).

- Line 139: "300 µL of AB solution was added" (instead of were added)?

This has been corrected.

- Line 172: "supplemented with water (control) or paraquat (as described above)" (instead of water, paraquat)?

This has been corrected.

- Why was generation 9 chosen to compare survival and Wolbachia density?

Generation 9 is the only generation at which both survival and *Wolbachia* density (and octomom copy number) were measured. It was chosen for a combination of practical

convenience and scientific relevance (i.e. it made sense to wait a few generations after the start of the experiment, as potential effects of selection would have necessarily been very weak otherwise). We now explain on lines 132-136 that "these traits were not measured at every single generation, as we do not expect the changes from one generation to the next to be so great as to justify measuring all traits at each generation. Furthermore, assessing the survival is especially time-consuming, and could not have possibly be done at every generation (15 females per line checked daily for up to 19 days, up to 20 lines)".

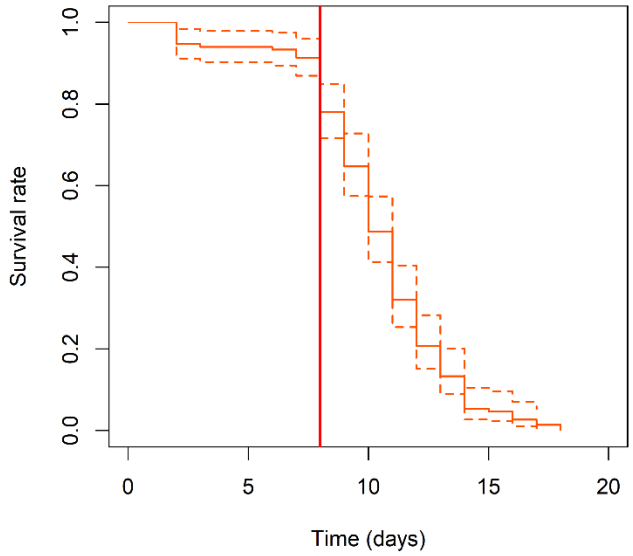
- Is there a reason why specific generations were picked to measure survival (G3 and G9) or density and octomom copy number (G2, G9, G12 and G17)?

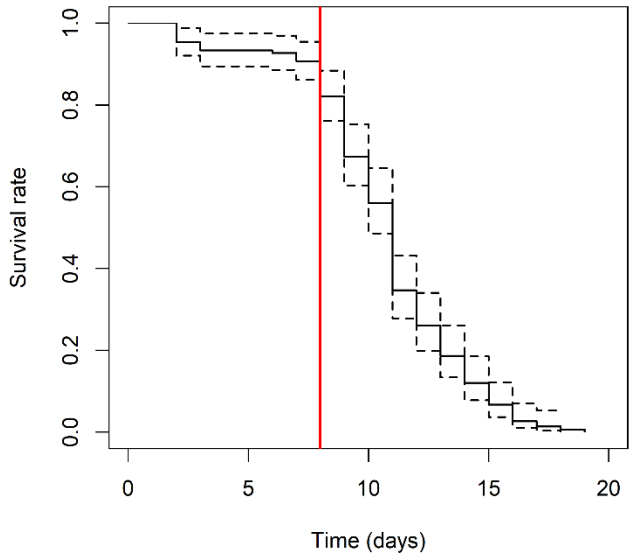
As explained above, our goal in picking these generations was, as much as possible, to assess density and octomom copy number at regular intervals during the experiment. We also waited a few generations after the start of the experiment, as potential effects of selection would have necessarily been very weak otherwise.

Results and discussion

- Line 202 + fig 2a: Is it possible to plot both control and paraquat separately in this figure. I think this figure would be more interesting to show.

As can be seen below, at G3 the "paraquat" survival curve (orange) is very similar to the "water" (black) one. As this similarity is already apparent in Fig 2b, we do not feel that adding the "paraquat" survival curve to Fig 2 would be useful and prefer to provide it as supplementary material.





- Line 294-295: Inter-host and intra-host selection. For me it is not clear what the different selection pressures are in these two types of selection.

We reworked the paragraph in question (lines 343-351).

General remarks

– - More information about inter-host and intra-host selection is needed.

+ - Experiment is well explained. - The surprising results are elaborately explained. Multiple possible explanations are formulated and discussed.