**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.

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?

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

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

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 intra-host 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'.

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

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

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.

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.

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

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.

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?

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)?

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

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.

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

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). 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). 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).

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.

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.