Responses to Reviewers

Fitness costs and benefits in response to artificial artesunate selection in *Plasmodium*

Villa M[1], Berthomieu A[1], Rivero A[1*]

Dear Prof Huijben,

Our heartfelt apologies for the delay in getting back to you with the revised version of the manuscript. We would like to thank you and the two reviewers for the very thorough review of our manuscript and for your pertinent and useful suggestions. We have carried out a substantial revision of the manuscript, including new statistical analyses, figures and tables. We reply below to each of the different referee comments (in blue) detailing, when appropriate, the changes that have been made. We'd be happy to make further changes should these be deemed necessary.

Decision for round #1 : ***Revision needed***

Dear authors. I have now received two reviews to your manuscript. Both the reviewers and myself agree this is an important piece of work that addresses a critical component of the evolution of drug-resistant malaria parasites. The reviewers' comments are mostly centered around the clarity of the methods used, the data presentation and some of the data interpretation. In addition to the reviewers' comments, please also consider the following:

Experiment 2: What is the sample size for the data presented in figure 4? Since unfed mosquitoes were removed, how many were remaining that this data is based on?

R1. Yes, sorry, sample size was missing. This data was based on 60 mosquitoes per group (30 per bird): uninfected, control, AS1, AS2. This is now stated in the figure legend.

Also, only 29% of females were infected, but this data is, I presume, based on all mosquitoes regardless of infection status. It would be important to address this caveat, as it might be masking a true effect if it were there.

R2. Not sure we understand this comment. The 29% infection prevalence is the overall prevalence based on blood fed mosquitoes exposed to the parasite (i.e. excluding mosquitoes fed on the uninfected birds). As stated in **Line 293**, this figure is independent of the parasite line.

Additionally, it seems that for some strains, it takes longer to establish an infection in the mosquito (AS2, figure 5) and this strain would not have any (detectable?) oocysts at the time of egg laying. It is unclear to me what the hypothesis was to expect to observe differences on day 3-6 when the parasites are unlikely to have an impact on the mosquito at that point in time.

R3. We had no parting hypothesis on what we would expect to observe on days 4-6 (first dissection day was day 4, **Line 209**). The aim of this experiment was to try to observe differences in the oocyst/sporozoite dynamics between the different lines, so we started our dissections on day 4 to make sure that we didn't miss the first (detectable) oocysts to make it to the gut. As it turned out, no oocysts were observed in any mosquito in any of the lines on day 4 so this time point was taken off the analyses. This is now mentioned in the text (**Line 311**)

An argument is made that the phenotype of the selected lines is a delayed clearance time, similar as seen in the field. As the data is currently presented, I do not believe we could make this conclusion. In the untreated infections, there is also a 'delayed clearance' where AS1 and AS2 persist for longer. An alternative hypothesis could be that these are just be differences in growth dynamics (e.g. the reference strain peaking and crashing sooner than the resistant strains), and perhaps these differences are enlarged by the treatment? As also mentioned in the review by Sarah Reece, it would be important to show the full parasite dynamics (also before the onset of treatment) to be making any comparisons based on treatment.

R4. The delayed clearance refers to the differences found between AS-selected and control line on day 13, which disappears from day 14 onwards (**Figure 2A**). In the untreated lines no such difference is found between day 13 and subsequent days. We agree with Sarah Reece that we should have given more information about the pre-treatment dynamics, which we now do. Please see full reply below (**R8**).

Slightly increased fitness was observed in the in selected lines: was the untreated reference strain selected at the same time? Could these differences feasibly have emerged as part of the selection process itself, rather than the pressure that the drug provides?

R5. The control (reference) line was passaged in parallel and using the same protocol (see **Lines 139-140**)

No fitness cost were found in the vertebrate host in these experiments. It would be useful to compare these to results found in in vitro experiments. While these are referenced as having been conducted, the results from these experiments are not mentioned. In most of these in vitro competition experiments they do demonstrate a fitness cost for ART resistant parasite (in the vertebrate host). It would be useful to discuss this discrepancy as it may throw the extrapolation from this P. relictum model to P. falciparum into question and therefore the conclusion on costs observed in the mosquito host.

R6. Agree. These costs have been observed when resistant lines are in competition with susceptible lines. Which we could not do because we had no markers to distinguish them (**Lines 451-452)**. We now discuss the discrepancy between our and previous (in vitro and in vivo) experiments (**Lines 373-382**)

It would be nice to see the raw data of the weights and RBC dynamics as well.

R7. We have carried out an extensive re-analysis and done new figures for the weights and RBC dynamics.

The analyses of weights reported in the first version of the manuscript were carried out using *raw weights*. Here we analyze *weight change* (i.e. increase or decrease in weight with respect to the weight on the first day of the experiment). We believe that this is a more accurate way of estimating weight,

because the heterogeneity in bird weights on the first day of the experiment (which varied between 17.9 – 24.8 grams, see also **R22**) adds noise to the analyses of raw weights, obscuring the changes that take place once the birds are infected. When we analyze the data in this way, differences between the lines do appear both in treated and control birds (birds infected with the AS-selected lines tend to have lower weights, these differences disappear during the AS treatment). We have added these results to the manuscript (**Lines 275-286**) as well as a new supplementary figure (**SF2**).

The analysis of RBC were also repeated using *rbc change*. No significant differences in rbcs between lines were observed in these new analyses. The results section has been reworded (**Lines 284-286**), and a new supplementary figure added (**SF3**).

Minor comments:

Line 108: "will suffer higher fitness costs than their unselected counterparts", change to "suffer fitness costs in comparison to their unselected counterparts"

Corrected, thank you

Line 119: "occasional passage through the mosquito", is it possible to give more details on frequency to get a sense on how frequently 'occasional' is?

Done (see **Line 120**)

Lines 132-137: Were lines AS2 and AS3 obtained as true replicate lines or isolated from the same birds (somewhere in the passage history)? Were other parasite lines obtained as well or were these the only ones that survived the passages? Some additional information on the process (perhaps in supplementary information) would likely be of interest to some.

Yes, all lines were true replicate lines. Please see also **R20**.

Line 271: add 'was'

Corrected, thank you

Figure 1: Update the caption, it refers to different names of lines (C, R1, R2 and R3) and dpbi instead of pbi, etc)

Corrected, thank you

Figure 2: brief explanation of px/pb would be helpful in the figure legend. It would also be helpful if the terminology around the reference line would be consistent, it's mostly referred to as 'reference line', but in figures and sometimes in the text as 'control'.

Thank you. Now called "reference" throughout.

Bibliography: Fix author 'Huijben' initials from 'U' to 'S'

Sorry! Corrected

Review by Sarah Reece

Overall summary

This manuscript considers a very important but overlooked topic – whether fitness costs of drug resistance exist in the between-host transmission component of the malaria parasite lifecycle. I found the presentation the concepts and the experiments clear and easy to follow on the whole (Fig 1 is very helpful). The study is thorough and finds interesting results, despite small sample sizes. My comments mostly concern clarifications.

General comments

Analysis/results

Line 244 – I think more needs to be done with the pre-treatment data. It is important that the dynamics of each line are similar in both the untreated and treated groups so they can be compared without confounding effects of e.g. different parasitaemias at the point of treatment. Thus, combining data across the untreated and treated groups for each line is interesting but not verifying that the key assumption of the experimental design is met.

R8. We disagree that comparing *treatments within lines* is of major relevance for our conclusions, which are not so much based on treated vs untreated comparisons (beyond showing that the artesunate treatment works) but on comparisons of unselected vs AS-selected lines in treated and untreated birds. Please note that different parasitaemias at the point of treatment are accounted for in the analyses (see **Line 165-169**).

We agree, however, that the pre-treatment data is crucial for understanding what happens afterwards, and that more should have been done with this data. In our view, the crucial comparison here is *between lines within a treatment group*. This is important because the differences we observe on treated birds on day 13 (**Figure 1A**) could potentially simply be the by-product of the parasite dynamics on days 2-12. We have changed the Supplementary Figure **SF1** to show the parasite dynamics of each line (reference, AS1, AS2 and AS3) in treated (A) and untreated (B) groups before the treatment (days 2-12) and done a new statistical analyses.

We fitted a full model that included day, line and treatment (and bird as a random to control for temporal pseudo-replication). There was no significant 3-way interaction and no significant differences in parasitaemia between birds infected with the different lines, or subsequently allocated to the treated or control treatments, but as you can see variance between birds is high (a pervasive problem in this system, see **R22**). Visual inspection of the data (to be taken for what they are worth) suggest that in birds subsequently allocated to the artesunate treatment, parasitaemia of AS1, AS2 peaked (on average) on day 10 so that by the day of the treatment (day 12) the parasitaemia had started its downward trajectory (**SF1B**, **SF1C**). In contrast, the parasitaemia of the reference line was still increasing by day 12 (**SF1A**). The significantly lower parasitaemias observed in the reference line compared to the AS1 and AS2 lines 24h after the beginning of the treatment (day 13) are in our opinion therefore unlikely to be explained by the pre-treatment dynamics. In contrast, the dynamics of AS3 were different (maybe due to the lower infective dose) and were following a clear upward trajectory at the time of treatment (**SF1D**). It is therefore not impossible that this upward trajectory may partly explain the higher parasitaemias in AS3 observed on day 13. We have modified the results (**L250-256**) and discussion (**L352-354**) sections accordingly.

Line 220 – was sampling day also accounted for in the expt 2 oocyst samples? It doesn't appear so from table ST2? I feel this is important given that expt 2 stimulated more detailed consideration of temporal dynamics in expt 3.

R9. Dissections took place over two days because it was impossible to dissect so many mosquitoes in a single day (and frozen mosquitoes cannot be dissected). Adding dissection day to the model does not change the results of the statistical analyses. Supplementary Table **ST2** has been modified.

Figures

Fig 2 – I was expecting this to be presented with lines (connecting mean points and SE bars/shading) to demonstrate the dynamics, not a bar chart. Such data are not normally plotted in this way and I found it hard to connect the day-to-day patterns. Also, I wonder if the during and after segments for A and B should be split with different y axis scales to aid interpretation?

R10. Figure 2 has been redone using connected means and standard errors**.** We have, however, kept the same scales for the 'during' and 'after' because in our view having different scales would be confusing (difficult for example to see the relapse that takes place in AS1 and AS2 after the end of the treatment).

And why not include the pre-treatment dynamics for the reason in line 244?

R12. Sorry, there must be a line number error here, as Line 244 simply states that we started by analyzing differences in parasitaemia before the treatment.

Expt 1 - Given that there are no figures of the virulence data included in main text or SI, then some effect size info should be included in the results text to give a sense of the differences in anaemia (pre and during/after) at least. It's probably worth pointing out there is no evidence that AS3 followed different dynamics due to its 50% lower infective dose?

R13. New figures and analyses done. Please refer to **R7** above. Reference to the different infective doses has been added (**Line 357-358**).

Expt 2 – include some readouts for longevity, e.g what was median / quartiles for lifespan across groups?

R14. We have added a table with the median longevity (the time at which 50% of the mosquitoes in each treatment is still alive) and the proportion of mosquitoes that survived till day 14 (peak sporozoite production) per treatment and bird – Supplementary Table **ST4**.

Interpretation

Line 360 – as well as selection on life history traits that underpin within host replication, AS selected parasites might selected to adopt reproductive restraint (via genetic evolution and /or plasticity), see Schneider & Reece Mol Biochem Para 2021, Schneider et al PLoS Paths 2018.

R15. Good point. Added, thank you (**Lines 388-391**)

Line 363-70 – I found this rather confusing and sort of contradictory. Maybe first state that the data do not support selection for virulence evolution, and then explain why this is the case?

R16. These lines have been re-written, we now mention the changes in weight we observe in AS-infected birds (see **R8**), how this is, surprisingly, not correlated with a higher anaemia, and we call for more studies on this issue (**Lines 394-400**).

The AS selected lines are not cloned so each passage could contain a mix of sensitive and resistant

parasites [unless the number of parasites in each 100ul passage was quantified and found to be very low?]. This means the AS lines used for the experimental infections might contain a mix of AS-altered and ancestral parasites? If so, I imagine this is likely to act to erode the differences between the AS lines and the unselected line, which makes results conservative. This is probably worth pointing out?

R17. Good point, we had not thought about this. This is now mentioned in **Lines 454-455**.

Furthermore, the number of infections/group for experiment is in line with sample sizes of similar studies, but the transmission experiments rely on only n=2 infections /group. Thus, the differences between the unselected line and the AS lines might be vulnerable to stochastic differences between the 2 unselected line infections in birds and the (4) AS infections rather than the effects of selection. I feel this should be considered in the discussion. Do the unselected line infections used for transmissions follow those of the n=6 in experiment, and previous equivalent infections?

R18. We wholly agree that more birds/line would have been better. As you know, many/most malaria papers out there infect mosquitoes using a handful of hosts (this is particularly striking in human malaria studies, where the number of donors is rarely mentioned). But this is no excuse. We didn't do it because logistically, we simply could not manage more. We have added a few sentences to the discussion to acknowledge this caveat (**Lines 428-431**).

The observation of slower developing oocysts in the AS selected lines is really interesting. For a future study, measuring their size would also help understand their developmental rate and fecundity. In the discussion, I was hoping to read some exploration of this result and potential reasons for how DR could link to, for example, reduced ability to acquire/process resources needed for within-vector replication.

R19. Thank you. We've added a sentence to the discussion (**Lines 422-423**).

Detailed comments

Line 24 -27 – could be more concise and just a single punchy sentence? "research on the costs of DR focusses on interactions with vertebrate host, yet whether they are also expressed in the vector has been overlooked'

Corrected, thank you.

Line 70 – could add some more recent observations of costs of DR, such as resource limitation exacerbating poor competitive ability [https://www.pnas.org/content/114/52/13774]?

Thank you, added.

Line 107 – it's a bit confusing to refer to both birds and mosquitoes as hosts when in other places, mosquitoes are referred to as vectors. Change to '…in untreated hosts (birds) and vectors (mosquitoes)…' ?

Corrected, thank you.

Line 116 – add 'France' to the location of collection?

Added, thank you.

Line 146 – its not accurate to say that parasite density was followed because the data are % parasiteamia. This is an ok metric to use (though subject to being skewed by variation in anaemia), but should be clarified.

True, this has been corrected, thank you.

Expt 1 - I couldn't see any info on the sample size of birds used, nor their age or sex, in the main text.

The information was in the Legend of Figure 1 (6 birds per line). Bird age is now mentioned (1-2 years old). Sexing canaries based on morphology is only possible during the breeding season and is a complex skill reserved to a chosen few (our success rate was ~50%, so we may as well have been allocating sexes at random). However, when we first started working with this experimental system, we used molecular biology methods to sex canaries (and to verify our, as it turns out nonexistent, morphology-based sexing skills) and we found no effect whatsoever of sex on any trait (parasitaemia, virulence, etc), so we don't bother sexing them anymore.

Line 201 and 220 – previously, uninfected birds were referred to as the control, so call the control group in experiment the unselected reference line to clarify its purpose?

Done, thank you.

Line 222 - Clarify that bird ID (i.e. cage ID) was also fitted as a random effect for the mosquito data

We changed "Birds" for "Bird ID" in the line (is this what you meant?).

Line 379 – I thought dissections occurred on day 8 and day 9?

Yes, the wording was a bit weird and may have led to confusion. This has now been changed to "half of the mosquitoes on day 8 and half on day 9"

Line 552 – 'were' is a typo?

Yes, sorry, the whole sentence was incomprehensible. This whole legend has been corrected, as there were issues elsewhere (Expt 2 and 3 were inversed). We've also clarified the number of replicates. Thank you.

Review by Marianna Szucs

It is a very well-written paper that was easy to follow despite my lack of expertise in disease biology. I am an entomologist with experience carrying out artificial selection experiments. Overall, I found the study well-designed, the methods and analyses sound. The paper appears quite straight forward but some of the data do not align with the main line of interpretation. It is normal to have outlier replicates that do not conform with the main conclusions, but I feel that the discussion largely omits why that might have happened.

Firstly, I was wondering why different methods were used to obtain the three replicated artesunate-selected lines. The AS1 replication was obtained using different concentrations of artesunate than AS2 and AS3.

R20. We started many different selection lines using different drug concentrations, to account for the fact that nobody had tried giving AS to birds before, so we did not know how it would affect their wellbeing. The three lines were the ones that we managed to keep long enough to do the experiments.

For experiment 2 the AS3 line was not used. It is mentioned that 'following the results of the previous experiment' it was decided not to use that line. However, I could not find any explanation in the results

why it was dropped. A bit of clarification on these points would help to better understand the reasoning behind these experimental choices.

R21. We decided to drop AS3 because it did not have the same phenotype of resistance in the bird (delayed clearance followed by recrudescence) as the other two lines. This is now mentioned in **Lines 176-178**.

Given the variation in results among replicates for a few metrics I have kept wondering how the measured parasite density in birds or oocyst density in mosquitoes compare to levels in nature. I think a brief paragraph in the discussion that talks about what kind of variation there is in these metrics in natural populations of birds and mosquitoes would help to place this study in better context. It could also be used to explain the variation seen among replicates that do not necessarily show the expected response.

R22. We believe that variation between birds is due to canaries not being experimental animals that can be bought from a scientific supplier as, say, mice. We get our canaries from different breeders whose main purpose is to select for plumage colour and song (we buy the brown ones that don't sing). Although we strive to get canaries that are as homogeneous as possible (bred inside, under similar conditions, of same age etc) we believe that there's still a great degree of heterogeneity in these birds due to eg bird condition on arrival, history of previous infections etc (on arrival, all our birds get tested for preexisting *P. relictum* infections and returned to the breeder if infected, but they may be harboring other infections that we are not able to detect). So when we infect several birds with the same number of parasites from the same isolate, the resulting parasitaemias can vary by an order of magnitude, which is why we often have to work with relative measures such as the ones used in Expt 1. It is one of the disadvantages of this experimental system.

As for the heterogeneities in parasitaemia in mosquitoes, this follows the binomial (overdispersed) distribution characteristic of most parasitic infections, whereby most mosquitoes carry no or very few parasites, while a few mosquitoes carry dozens, sometimes hundreds of parasites, even though they are highly inbred (as ours are) and have all fed on the same bird. The mechanistic reasons for this universal pattern of parasite distribution are poorly known, but we have demonstrated in a recent paper (Isaïa et al, *Proc R Soc* 2020) that part of the explanation lays with the order in which the mosquitoes bite (the first mosquitoes to bite stimulate the parasite so that later mosquitoes become more infected).

We do not have any measurement of parasite density in the wild (though we know that prevalence varies between 5-20% depending on the time of the year, Zélé et al, *Par. and Vect.* 2014). In birds, the infection goes through an initial *acute* stage (high parasitaemia) lasting a few days, followed by a very long *chronic* stage (low to v low parasitaemias) that can last months or years. Most birds in the wild are therefore in the chronic stage, and most mosquitoes in the wild get infected from feeding in these birds. Like most other *in vivo* experimental work of malaria, we fed our mosquitoes during the acute stage of the infection (this ensures that a large proportion of the mosquitoes become infected), and in this respect this may not be realistic. We see however no real need to explain all of this as our experiments are not expected to mimic what happens with birds in the field: bird malaria infections in the wild are not treated (except in the case of penguins in zoos which are treated with a different drug: primaquine, but thus far no resistance evolution has been reported).

A few minor comments:

Line 144: were

Changed, thank you.

Line 157: the other half

Corrected, thank you.

Line 271: was lower

Corrected, thank you

Lines 552, 574, 597: grammar

All corrected, thank you

Legend to Figure 1

Corrected, thank you.