Sexual selection improves population fitness: a meta-analysis

Response to Decision Letter

Dear Dr. Jones and colleagues,

Many thanks for your time, and for the very helpful comments on our manuscript. We have considerably reworked the manuscript in light of the many insightful comments and suggestions provided by the reviewers. Please see below for a detailed list of responses to each of the reviewers' points. We also attach a revised manuscript and a revised copy of the supplementary material, as well as a copy of the revised manuscript in which the significant changes and additions have been highlighted. Please let us know if we can do anything further to assist you.

Yours Sincerely,

Justin Cally and co-authors

Comments from Reviewer 1 (Prof. Jacek Radwan)

> The manuscript presents results of a meta-analysis assessing the effect of sexual selection on population fitness. This topic has been increasingly studied over the past decade or so, and the time seems ripe for a synthesis. The authors have done good job searching the literature – it includes all relevant papers I could remember, and the results should be of interest to a broad range of evolutionary biologists. I cannot expertly judge on state-of-the-art methodology, but I do have several general concerns about how the analyses were performed.

Many thanks for your time and this valuable feedback.

> Firstly, it seems that several male traits like to be under sexual selection, but not related to population fitness were included. I don't think this is correct given the question being asked (i.e. whether sexual selection increases population fitness). I think traits which are directly under sexual selection (eg. attractiveness, reproductive success) should not be pooled with traits which may respond to manipulation of sexual selection indirectly (eg. male development time, survival) and can affect population fitness. Distinguishing between both types of trait could actually be revealing – e.g. could expose trade-offs between sexually selected traits and fitness components unrelated to reproductive competition (see eg. Radwan et al. 2015 Evol Biol), a thus potentially explain lower effect of sexual selection on (pooled) male traits than on female traits.

The reviewer is correct that several traits are likely under trade-offs with other traits as sexual selection not only affects mean fitness but alters life history traits (*i.e.* sexual selection may increase male body size or attractiveness, but decrease development rate). In essence this is why our meta-analysis sought to investigate the *net* effect of sexual selection on population fitness across many traits, rather than one that may be traded off in a given direction. Importantly, to account for the variable and uncertain relationships certain traits have with fitness (as raised by the reviewer). We provide an expacded table on how each trait is measured and why they are classed as ambiguous, indirect and direct in their relationship with fitness. Given that this information was in the Supplementary Material we have revised the manuscript to clarify the distinctions more thoroughly.

Additionally, the reviewers comment that male attractiveness may not be related to population fitness is correct. Because male attractiveness may be positively or negatively correlated with average fitness (although in many condition-dependent traits a positive association is assumed) we have re-classified *Male Attractiveness* as ambiguously related to fitness, rather than indirect. All subsequent models, statistics and figures now take

into account this re-classification. However, given the small sample size for male attractiveness effect sizes (n = 6), **we find only marginal changes in the results and no key changes in the key findings of the study*.**
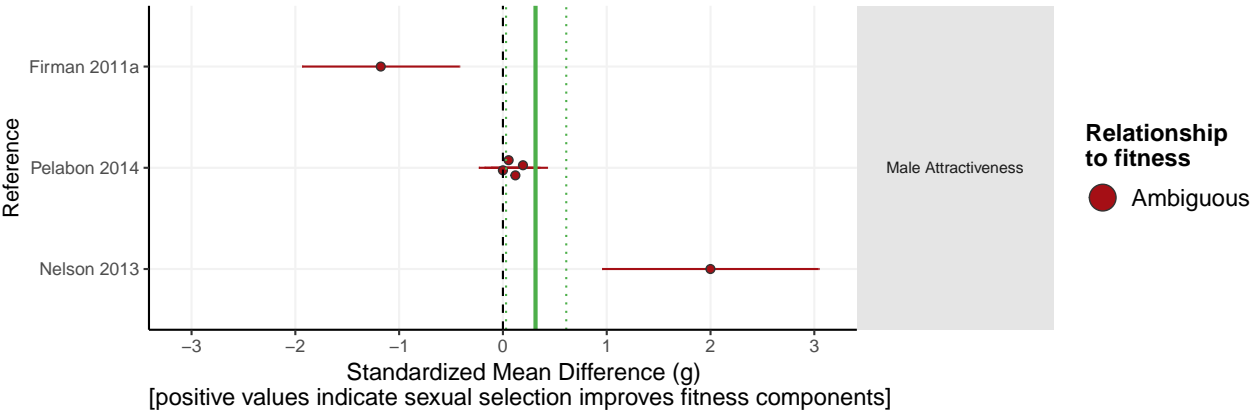
> *Secondly, looking at Figure 1, one notices apparent anomalies, for example significantly positive slope for male attractiveness, based on studies with average effect size close to zero. I think this (and few others) surprising estimates may result from fitting random terms across all trait types (fitted as a fixed factor, second model); I guess fitting interaction (random slopes) would not be feasible for some categories including few data points, but some of them could easily be pooled in wider categories.*

The reviewer raises questions of the analysis that produce seemingly suprising results. We agree that some of *significant* predictions in Figure 1 (whereby 95 % CIs do not overlap zero) appear surprising based on the distribution on effect sizes. In regards to the trait "male attractiveness", the 95 % CIs do not overlap zero. This is the case even though the higher weighted studies are very close to zero. However, this seeming anomaly can be explained by various reasons.

Firstly, althought the effect size prediction for the trait has 95 % CIs not overlapping zero, the overall magnitude of the effect size is small (see Table exerpt), and the Bayes Factor (an estimate of the likelihood that the effect size is actually positive) is also small. The full table for all predicted effect sizes used in Figure 1 can be seen in **SXXX**. Below we provide the exerpt of this table and forest plot for "Male attractiveness".

Secondly, the effect size predictions were made using a single model that incorperated all 459 effect sizes, thus model predictions for each trait have narrower confidence intervals as the fitted values do not incorperate measurement error. In regards to our analysis, this means that the predictions that appear in Figure 1 are sourced from using `fitted.brmsfit` as opposed to `predict.brmsfit`. Given the reviewers comments we have clarified how we obtained model predictions used in our results. Additionally, some meta-analyses investigating the effect of a variable on more than one outcome model each outcome (in this case the *fitness component*) seperately. Arguably, this method provides more variable model estimates as models will always have lower power due to reduced sample sizes and the incorperation of measurement error in the predictions. We have added a table to the Supplementary Material where models are independently conducted on all fitness components where n > 3.

Fitness Component	Bayes Prediction	Bayes SE	Bayes LCI	Bayes UCI	n	BF	REML Prediction	REML SE	REML LCI	REML UCI
Male Attractiveness	0.314	0.15	0.0302	0.61	6	5.5e+01	0.306	0.115	0.081	0.531

Standardized Mean Difference (g)
[positive values indicate sexual selection improves fitness components]

> *Thirdly, type of trait measured explained 35% variance, but the authors do not explore this any further. However, examination of Fig. 1 suggest that some indirect/ambiguous fitness measures account for much of this heterogeneity and they generally have higher average effect sizes that direct measures (except for immunity). I'd like to see if the authors recover their main result if*

2

they only direct measures.

Add section in discussion about the variability amongst traits and why some might be subject to trade offs (e.g. immunity).

Summarise results for direct, indirect and ambiguous

The reviewer provides an accurate shortcoming of discussion on the observed heterogeneity between fitness components. We have extended certain parts of the discussion related to this aspect...

In regards to conducting a meta-analysis on just direct fitnesss components, the reviewer provids a valid question of our use of direct **and** indirect fitness components for the meta-analysis investigating the effects of sex and stress. To some degree the use of direct **and** indirect is arbitrary. But based on the question we asked we were interested in the effect of sexual selection on many components of fitness.

Nevertheless, we conducted our analysis using just direct fitness measures. We find the same key associations; that sexual selection generally elevates fitness, especially for females, with the effect magnified in stressful environments. Below we present the findings, which we have incorperated into our supplementary material and referenced within the manuscript and Figure 2 caption.

Additionally, in order to further distinguish the effects of sexual selection on ambiguous, indirect and direct fitness components, we have structured the second paragraph of the results to more consistently reflect our classification system.

Table 2: Model using only direct measures of fitness

Sex	Environment	Prediction	SE	CI.lb	CI.ub	n
Male	Benign	0.067	0.099	-0.127	0.26	43
Both	Benign	0.125	0.112	-0.095	0.34	15
Female	Benign	0.173	0.096	-0.015	0.36	84
Male	Stressful	-0.024	0.109	-0.237	0.19	13
Both	Stressful	0.176	0.112	-0.045	0.4	12
Female	Stressful	0.378	0.098	0.186	0.57	33
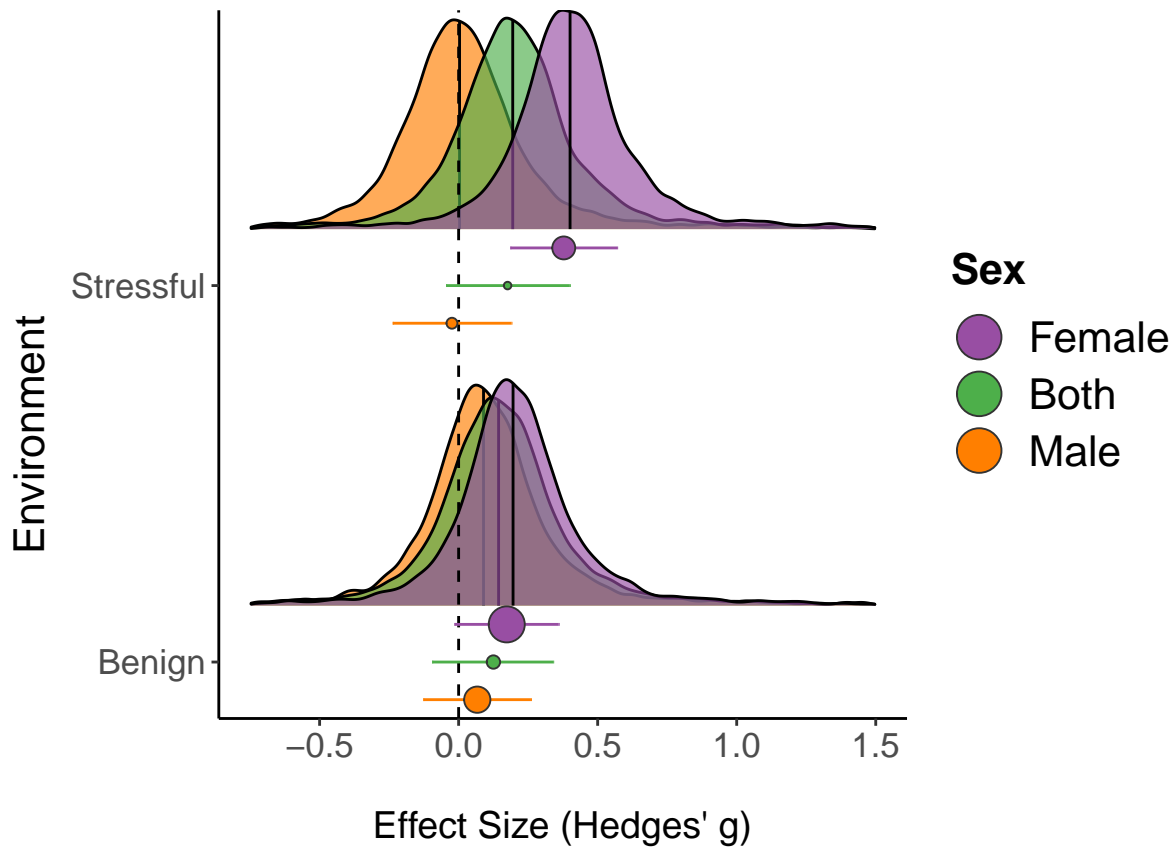
	I2__Est.	2.5% CI	97.5% CI
Study.ID	84	74.4	90.4
Outcome	0.987	0.0204	3.67
Taxon	4.79	0.468	12.8
total	89.8	85.3	92.9

Another major problem I have with the manuscript concerns interpretation of the very intriguing finding that the response to manipulation of sexual selection was stronger for female traits compared to male traits. I'm confused by the authors' explanation: do they assume sexual selection acted directly on females, and not only indirectly, via males? Only then things like mother to daughter heritability, or hard selection on females, should matter.

So this is a tricky one. We have the data to seperate the groups for where SS can be acting 2 ways (ie > 1 female) because we noted the ratios down. We can compare this against the treatments where SS could only be acting on males (1 female)... That might reveal findings that we can include in text as explanation.

Good sugggestion!

Perhaps the effect on females is indeed direct, and results from stress imposed by polygamous treatment, which magnifies direct, hard selection on females? This would be an important finding, and perhaps the authors could test it with their dataset by contrasting middle-class-neighborhood-like studies from those which allowed for female evolution. But if correct, this explanation is not exactly the effect of sexual selection, but rather enhanced selection of females due to enhanced (male induced) stress, so the interpretation of results should change.

Answer

Other comments: l. 30 – reviews on sexual conflict are OK to cite here, but there are empirical

papers actually demonstrating correlation between male sexual selected traits (Harano et al. 2011; Plesnar et al. 2014) which should also be cited.

Thank you, we now cite XXX as suggested.

l. 151 – the authors discuss beneficial effects of sexual selection on direct fitness measures such as reproductive success or offspring viability, but estimates for both of these measures actually overlapped zero! Perhaps joint analysis of direct fitness measures, as I suggested above, could support this conclusion, but currently this is an overstatement.

Answer

In the discussion the authors say they included the number of experimental evolution generations, but I could not find this information in methods.

Answer

Fig. S1 is not referred to in the main texts, is it different from Fig. 2, except that the latter contains predicted average values for fitness components?

Answer

Comments from Reviewer 2

This meta-analysis investigated the consequences of sexual selection experiments on trait mean and variance (comparing sexually selected groups vs control groups). Overall, the authors observe sexual selection usually increase mean and reduce variance (especially in females). The authors conclude sexual selection's benefits outweigh its detrimental effects. This meta-analysis is extremely well conducted (the use of both likelihood-based and Bayesian models for robustness), and although I am not an expert on this topic, I really enjoyed reading it, and it was very clear. Especially, I am impressed with the detailed supplement which showed the code and analysis. However, I have several comments regarding their analysis, which will increase the robustness of results and thus conclusions.

Many thanks for your time and attention, and for the valuable feedback.

1 — The use of Hedges' g. I understand that Hedges' g was probably used because it can take interval measurements as well as ratio measurements (lnCVR can only take ratio measurements). However, g cannot really deal well with heterogeneity between two groups (i.e. experimental and control groups having different variances). This is why Hedges' g (Cohen's d) was criticized earlier.

Answer

Osenberg, C. W., O. Sarnelle, and S. D. Cooper. 1997. Effect size in ecological experiments: the application of biological models in meta-analysis. *American Naturalist* 150:798-812.

Answer

As a response, they come up with log response ratio (lnRR) - see

Hedges, L. V., J. Gurevitch, and P. S. Curtis. 1999. The meta-analysis of response ratios in experimental ecology. *Ecology* 80:1150-1156.

I recommend that the authors use lnRR for their effect size for mean comparison as well as Hedge's g to see the robustness of their conclusions for the mean.

Answer

2 — The authors may consider also doing another set of meta-analyses using lnVR (log variability ratio - proposed in Nakagawa et al. 2015).

Nakagawa, S., R. Poulin, K. Mengersen, K. Reinhold, L. Engqvist, M. Lagisz, and A. M. Senior. 2015. Meta-analysis of variation: ecological and evolutionary applications and beyond. *Methods in Ecology and Evolution* 6:143-152.

Answer

As one can see in Figure 2, the mean results (Hedges' g) are a mirror image of the variance results (this makes sense CV controls for means). I would like to see what the absolute change in variances. Probably the authors can put the analysis using lnVR in the supplement. The results of this analysis can be discussed. Also, the mean-variance relationship between mean and variance (sd) should be verified (e.g. plot log mean and log sd or log variance).

Answer

3 — I^2 needs to be explained. I^2 is proposed originally here:

Higgins, J., and S. Thompson. 2002. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* 21:1539 - 1558.

But later expanded in here for mixed models (hierarchical models)

Nakagawa, S., and E. S. A. Santos. 2012. Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology* 26:1253-1274.

This is what the authors use. Also, it will be good to put the degree in I^2 in the context. To do this see this paper:

Senior, A. M., C. E. Grueber, T. Kamiya, M. Lagisz, K. O'Dwyer, E. S. A. Santos, and S. Nakagawa. 2016. Heterogeneity in ecological and evolutionary meta-analyses: its magnitude and implications. *Ecology* 97:3293-3299.

Answer

4 — Publication tests have been conducted on, I think, "meta-analytic" residuals results as suggested in Nakagawa and Santos 2012. One could use such residuals to conduct the trim and fill method and see how much the mean could move (see Nakagawa and Santos 2012). One should remember the funnel asymmetry could be caused by the presece of heterogeniety. See:

Egger, M., G. Smith, M. Schneider, and C. Minder. 1997. Bias in metaanalysis detected by a simple, graphical test. *Br Med J* 315:629 - 634.

Answer

5 — Figure 3 - are these grey envelopes 95% CI?

Answer

6 — the title - do the authors include "a systematic review" - so "a systematic reveiw and meta-analysis" - these two things are diferent - see:

Nakagawa, S., and R. Poulin. 2012. Meta-analytic insights into evolutionary ecology: an introduction and synthesis. *Evolutionary Ecology* 26:1085-1099.

Answer

Hope my comments are useful.

They were extremely useful, many thanks!

Comments from Reviewer 3

This manuscript addresses the question of whether, on average, sexual selection has a net beneficial or detrimental effect on population fitness, using a meta-analysis of experimental evolution studies comparing the fitness of populations under different intensities of sexual selection. The analyses consider the effects of sexual selection on both the mean and variance of population fitness, and test whether these effects differ for stressful vs. benign environments, or depend on the measure of population fitness.

These are issues of longstanding interest, and as there is now quite a wealth of experimental evolution studies addressing these questions a meta-analysis synthesising their findings is timely.

Many thanks for your time and attention, and for the valuable feedback.

The results of this meta-analysis broadly concur with current theoretical predictions, and are likely to interest a wide readership. Across all studies, environment types and fitness measures there was a small positive effect of sexual selection on mean population fitness, although effect sizes varied with the fitness measure used. The positive effect of sexual selection on fitness was strongest for females in stressful environments, who also showed reduced variance in fitness under sexual selection.

Answer

The authors discuss these findings thoughtfully, although they focus more on the (weaker) effects found for 'direct' measures of fitness while somewhat neglecting the effects seen for 'indirect' fitness measures (e.g. of attractiveness, mating latency, lifespan, ejaculate traits). These potentially deserve more consideration, especially as many appear to be male-limited traits or measured mainly in males, that would appear to have a direct bearing on male mating success (if an indirect link to overall fitness), yet despite this the overall effect of sexual selection on male fitness is not significant.

Answer

I have only a few more specific comments:

Does your approach consider variation in the intensity of sexual selection exclusively on males? This is worth noting (e.g. statements such as that on line **98-99** might be amended to "Sexual selection on males significantly improved female fitness.")

Answer

Experiments manipulating the sex ratio to alter sexual selection might simultaneously decrease sexual selection on males while increasing sexual selection on females. Would this be classified as reduced sexual selection in your analyses? I don't think you included this aspect of study design as a moderator variable in any of your analyses – given that close to half of the effects you include come from "alternative manipulations" (line **70**) perhaps you could test whether this affects effect size?

Answer

It seems possible that sexual selection on females, and not only on males, could affect population fitness – potentially even more directly than sexual selection on males. And might differing extents of sexual selection in each sex interact with the differences you saw in fitness measured in males vs. females?

Answer

L111-118 It seems a little odd that there is no residual heterogeneity in your I^2 estimates. From the supplemental information it is not entirely clear to me how you have adapted the

function that is under development in "metaAidR", but can you double check that you have appropriately incorporated the residual variance into these estimates?

Answer

L307-311 Please clarify here that lnCVR was calculated as the ratio $ln(CVfitness_{SS}/CVfitness_{noSS})$

Answer

Fig. 3 Please explain the dashed red and black lines in the figure caption.

Answer

Tables 1, 2 Please check – the parameter estimates for 'Female sex' and test for 'Female > Male' are bolded inconsistently between the REML and Bayesian models where these estimates are identical.

Answer