

Recommender

Dear Dr. Van Steenberge and colleagues, your manuscript has been reviewed by 3 reviewers. Most of them think your manuscript merits publication but still have some issues. Therefore I suggest a revision of your manuscript integrating the issues the reviewers have brought up. Kind regards, Ellen Decaestecker

A 0.1 We thank the recommender for handling our submission.

Reviewer 1

In the paper “The initial response of females towards congeneric males matches the propensity to hybridize in *Ophthalmotilapia*”, Van Steenberge and colleagues test differences in initial behavioral responses of female *Ophthalmotilapia ventralis* and *Ophthalmotilapia nasuta* towards conspecific males (relative to heterospecific males, conspecific females, or no fish). As predicted from previously reported introgression asymmetries found in natural populations, female *O. nasuta* are “pickier” than female *O. ventralis*. The manuscript reads very well, is well organized and the discussion is balanced (although a bit lengthy). The methods and statistical analyses of the data are detailed, well-founded and produce clear results. Nonetheless, I have one important suggestion regarding the analyses and one correction of the figure that illustrates the experimental setup. I have a few minor additional comments and suggestions that I detail next:

- Asymmetries in pre-zygotic isolation are not only observed in several species (as the authors indicate and for which several references are provided) but they are also expected. This is an important nuance but I think it is worth making and referencing. Coyne and Orr (2004) discuss this but more references should also be available.

A 1.1 This is indeed correct. We changed the paragraph in which case studies were listed by starting the paragraph by “Asymmetric propensities towards hybridization are expected in the intermediate stages of reproductive isolation (Arnold et al. 1996). This has been observed in a variety of animal taxa...”

Arnold SJ Verrell PA Tilley SG 1996 The evolution of asymmetry in sexual isolation: A model and a test case. *Evolution* 50: 1024-1033

- I feel that Figures 2 and 3, which explore focal female behavior before and after visual contact is established with stimuli, would be more informative if the analyses (PCA and CVA) would be ran on the combined (before and after) dataset (ON experiments and OV experiments analyzed separately). The authors would still plot ‘before’ and ‘after’ separately, but both would be on the same coordinate system and, importantly, any change in female behavior in response to stimuli would be easier to perceive.

A 1.2 We agree that performing these analyses on the combined data has the advantage that behaviour before and after exposure can be better compared within groups. However, the research question of the study was whether behavioural differences are present between specimens from different treatments, within a certain timeframe. This is the data that is currently visualised and this is also the data on which permanova was performed.

Although there is an advantage of combining tracking data from different time frames into the analyses, this is less the case for the point events. Most of these events were not recorded in the 'before' frames. These events were, however, important in measuring the difference in behaviour after presentations to the different non-focal species.

Hence, if we would have used the combined dataset to calculate PC and CV axes, these events would have been underrepresented in the factor loadings. This would have created a difference in what was visualised and what was statistically compared. For this reason, we prefer to calculate PCA and CVA on different datasets.

- I detected one mistake in Figures 2 and 3 (and supplementary files) where the different experimental comparisons are depicted. In particular, only conspecific females seem to have been presented to focal females. Thus, in the figures for the OV experiments, the non-focal part of the figure should depict *O. ventralis* females as well.

A 1.3 This is correct, we changed this and made new versions of Figures 2 and 3 and of supplement 7.

- Line 101. I suggest "feeding schools"
- Line 177. I suggest "monospecific aquaria"
- Line 273. I suggest "terracotta flowerpots"
- Line 403. I suggest "We carried out CVAs on the same datasets"

A 1.4 These suggestions were all implemented.

- Line 450-451. Do non-focal females and conspecific males differ in weight from each other or from the focal females? Please specify.

A 1.5 The difference referred to here was between non-focal females and conspecific males used in the OV experiment. We changed the sentence hoping that this has become clear. We did not test the difference in weight between the non-focal and focal specimens, as we only compared the difference in behaviour amongst focal specimens, and amongst non-focal specimens, respectively. The relevant data is also presented in Supplement 4.

- Line 489-490. I suggest "'flee' behaviour when presented to an *O. ventralis* male"
- Line 518. I suggest "did not"

- Line 527. I suggest “more” or “more strongly”
- Line 649. I suggest “stereotypical”
- Line 665. I suggest “does not”
- Line 674. I suggest “ecological range”

A 1.5 These suggestions were all implemented.

Reviewer 2

This experiment aims to examine whether asymmetric hybridization discovered in wild populations has a behavioural basis that can be uncovered in the lab. The aim is to test a case of prezygotic, behavioural isolation. There is a nice experimental design, and some clear predictions that one could make about female behaviours in response to conspecific and heterospecific males (and females). One notable limitation of the study is the lack of comparison between conspecific and heterospecific males and females in both species, acknowledged in the Discussion. Another important limitation acknowledged by the authors is that their experiment seemed to “observe(d) the routine behaviour of a (isolated) female that encounters a conspecific individual, rather than sexually motivated behaviour”. I have some concerns that this rather undermines the main premise of the work, since sexually motivated behaviour would be crucial to understand in order to successfully address the main question the authors identify.

The paper says “we may expect that the early response to conspecific and heterospecific mates will predict the outcome of the mating process to a substantial degree. We test this hypothesis using a cichlidN model.” but this is contradicted by the acknowledgement that they are not assessing sexually motivated behaviour.

A 2.1 We thank the reviewer for his appreciation of the study. We acknowledge that the lack of heterospecific females in our experiments implies that we cannot say that the observed difference in behaviour towards con- and heterospecific males is sexually motivated. However, we cannot exclude it either. We therefore rephrased the sentence: “This suggests that we observed the routine behaviour of a (isolated) female that encounters a conspecific individual, rather than sexually motivated behaviour” with “Hence, we cannot say whether we observed sexually motivated behaviour, or just the routine behaviour of a (isolated) female that encounters a conspecific individual.” In doing so, we remove the contradiction pointed out by the reviewer.

From the point of view of the statistical analysis, I found the approach lacking a clear direction and hypothesis test. There is a very clear prediction here - that specific behaviours should reduce. Why is the exploratory analysis needed? Why “without defining a priori in what variable specimens would differ.” ? I would consider replacing it with a GLMM which tests specific defined behaviours as a dependent variable, and includes gonad weight, before/after1/after2, treatment and species in the model (as well as non-focal male behaviour, which could be

confounding the experiment). If you think that multivariate analysis is important, PCs could be used instead of single behaviours of interest.

A 2.2 The reviewer suggests performing a GLMM to test our main prediction: that certain specific behaviours reduce in hetero- compared to in conspecific encounters. However, for these species, we had no prior information what the relevant specific behaviours could have been. It is for that reason that we choose the definition of species recognition as “a measurable difference in behavioural response toward conspecifics as compared to heterospecifics.” Hence, we were not able to define a dependant variable for this model.

Although the suggestion raised by reviewer 2, to use PC axes as variables could have circumvented this problem there are two additional problems with this approach. Foremost, there is not a single PC axis that explains the differences in all of the contrasts tested. For example in the ON experiment PC 1 explains the difference between encounters with heterospecific and conspecific males whereas PC2 explains the difference between control and non-control behaviour. Additionally, a GLMM would reduce the statistical power since it will include testing a lot of comparisons that are not biologically meaningful and it will reduce the number of replicates in the tested contrasts. For example, for the first contrast tested, we could compare all non-control trails with the control trail. In a GLMM, this would be obtained by comparing all three non-control trails with the control trail.

For this reason, we opted for Permanovas, which were only performed for those contrasts that were biologically meaningful and which could be performed on all of the data combined.

In general, although the PCs might be helpful for exploring the behaviours, I don't think they are very helpful in presenting the results. There were some specific aspects I found confusing and which might benefit from more clarity:

“However, heterospecific males were (somewhat) separated from all other specimens by their higher values for PC1 (ON experiment) or PC2 (OV experiment). This difference was due to a more active swimming behaviour (Sp, SpX, SpY) higher up in the water column (height) for *O. ventralis* males (ON experiment) and a higher number of point events (ram, sand, bite) performed at the floor of the aquarium (height) for the *O. nasuta* males (OV experiment), prior to their presentation to a heterospecific female”

Are these separate plots but from the same PC analysis? Or different PC analyses for each species?

Perhaps these data would be more clearly visualised as boxplots showing the (lack of) difference between species at specific behaviours?

A 2.3 The descriptions of the PCA plots stem from interpretation of the factor loadings. We made this clear in the text by adding “The loadings of the main PCs (Supplement 5.1) suggest that this...” All data are accessible in supplement 4.

There are a couple of other points or small issues that might be helpful to address:

- “which raises the question how can they coexist.” needs a question mark.

A 2.4 This is an indirect question, and hence does not require a question mark.

- hybridise/hybridize needs a consistency throughout

A 2.5 we used UK spelling ‘hybridise’ throughout.

- “ the question remains what mechanisms keep incipient species separated.” needs a question mark as well I think?

A 2.6 This is an indirect question, and hence does not require a question mark.

- Are these really incipient species? Evidence needed.

A 2.7 This statement was part of the introduction to the paper and therefore does not directly relate to our study system. However, Great Lake cichlids are well-known examples of rapid, explosive speciation and these radiations contain several incipient species. However, as indicated on line 92, we specifically choose a case study of mature species to examine the role of species recognition at the later stages of the speciation process.

- “In scenarios of sympatric, closely related species, the ability to correctly distinguish between conspecific and heterospecific mates is probably crucial (Sullivan 2009).” I would add something like “if the speciation process is to be complete” to clarify what it is crucial for.

A 2.8 We added “to maintain the integrity of the species.”

- “By snapping at the egg dummies, which are situated close to the genital opening of the male, the intake of sperm is facilitated, increasing the fertilisation rate of the eggs within the female’s mouth (Salzburger et al. 2007).” - this reference doesn’t evidence this claim. It might be better to add “the intake of sperm is thought to be facilitated” since to the best of my knowledge there isn’t a study that has directly tested this (if there is, please cite it here instead!).

A 2.9 We implemented this suggestion

- “we predict to see an interspecific difference in female response to conspecific and heterospecific males.” - but not females? Surely this is an important control, if this really is about mate choosiness?

A 2.10 In this paragraph, we explain why females should be considered the choosy sex in *Ophthalmotilapia*. Additionally, we base our hypothesis on the premise that the asymmetric pattern of hybridisation is caused by asymmetric

acceptance of males by females. Hence, we think it would break the flow of the text if we would also talk about female-female recognition.

We admit this is an important control, which is we included it in our experimental design. However, since no heterospecific females were included in the design, we can't really formulate a hypothesis on species recognition independent of sex.

- “that a sex change did indeed took place in several specimens” = take place

A 2.11 We changed this accordingly

- “The later was conducted to maximize” = latter

A 2.12 We changed this accordingly

- Before, after1 and after2 - clarify how after1 and after2 were defined and what the biological meaning was.

A 2.13 The time periods corresponding to before, after 1 and after 2 are mentioned in the paragraph starting on line 317.

We choose these time periods, as we aimed to study the immediate response following presentation to another fish. Teleosts respond to changing social situations at two levels: by an immediate change in behaviour, and by a modification of neural circuits. The latter is driven by the regulation of immediate early genes (IEG). Following Oliveira (2012), we assumed that the transcription of IEG only occurs roughly 15 min (after1), and their translation 45 min (after2) after the stimulus. This justification is added to the same paragraph.

- “For this, in view on the size of the dataset” = of

A 2.14 We changed this accordingly

- “*O. nasuta* specimens, on average, spent more time closer to the bottom whereas *O. ventralis* specimens were more often found higher up in the water column” presumably because nasuta is a bower builder?

A 2.15 Because we also observed a higher number of ‘sand’ and ‘bite’ events in the *O. nasuta* males, this might indeed be caused by their bower-building behaviour. However, the reason why *O. nasuta* spend more time closer to the bottom might also be because, in the Lake, *O. nasuta* can be found in deeper waters than *O. ventralis*. We added this interpretation in the results section.

- Figure 1B is a good opportunity to visually outline the experimental design, but I think it could perhaps be improved so that actual fish pictures are used and the reader doesn't have to rely so heavily on the text?

A 2.16 We opted for a schematic overview because this allows us to use 'placeholder fishes'. If we would have used photographs of the actual fishes, or of the actual specimens, we would only be able to show females of one of the two species. In that case, we would have needed to add in the text that a similar design was followed for the other species. Additionally, for contrast A, the placeholder represents conspecific females, conspecific males and heterospecific males. Explaining this would also require additional text. Therefore, we think that an adapted visual outline would not result in a shorter text. However, if the reviewer is of the opinion that using photographs of actual fishes would have rendered the design easier to understand, we can change the figure.

Overall, if there is good evidence that these are incipient species, and that sexual behaviour can be adequately quantified, I would think that this is a nice study (if not optimal, given some limitations of the design).

A 2.17 Thank you for your appreciation.

Reviewer 3

This preprint by van Steenbergen et al. presents a laboratory behavioural study of females of a pair of closely related species of Lake Tanganyika cichlid fishes. Female *Ophthalmotilapia nasuta* showed very different responses to conspecific males (and females) in comparison to males of the closely-related *O. ventralis*. By contrast, *Ophthalmotilapia ventralis* female showed similar responses to conspecifics and heterospecifics. This was consistent with a previous study (Nevado et al. 2011) of wild-caught specimens that showed that *O. nasuta* populations included occasional individuals bearing mitochondrial DNA (mtDNA) sequences more characteristic of related species, including *O. ventralis*. As mtDNA is normally maternally inherited, this suggests that male *O. nasuta* occasionally mate with female *O. ventralis*. However, the scarcity of characteristic *O. nasuta* mtDNA in populations of related species suggests that female *O. nasuta* show strong preferences for conspecific males, which was consistent with van Steenbergen et al.'s laboratory trials. The emphasis on female choice in this study system is likely to be justified, as the cryptic-coloured female *Ophthalmotilapia* brood their eggs and young in their mouths for several days, while the flamboyant males provide no parental care, but compete to attract females. With females taking several weeks or months to complete a reproductive cycle and be ready to produce another clutch, it is likely that there is a strongly male-biased operational sex ratio. The divergence in patterns of female mating preferences is likely to help explain patterns of hybridisation seen among groups of closely related cichlid fish species, which in turn may influence the detailed course of events during adaptive radiations. This is likely to be of wider significance, given the increasing evidence of introgression even among well-differentiated non-sister lineages during adaptive radiations of cichlid fishes and many other taxa. Thus, the study has considerable significance and has results that are likely to advance our understanding of the role of behaviour in adaptive radiation and speciation.

Overall, the experiments seem to be well-designed and analysed, the results presented clearly and succinctly and the interpretation seems generally robust. Occasionally one or two things might be clearer and a couple of statements seemed to be a little too definite in relation to the evidence presented.

A 3.1 We thank the reviewer for his appreciation of our study, and for the clear summary of the rationale of our experiments.

Lines 79-80: 'Given the importance of the initial contact, we may expect that the early response to conspecific and heterospecific mates will predict the outcome of the mating process to a substantial degree.' Perhaps this might be expressed a little more clearly in terms of the justification of the experimental design employed. The suggestion seems to be the initial behavioural response by females to conspecific and heterospecific potential mates is proposed to be a good predictor of actual mate preferences. It could be made clear whether this is being assumed or tested. Maybe we could be given a bit more background to explain why this might be a worthwhile thing to do, for example by pointing out the difficulties with other methods reported in the literature. This section might be most appropriate at the end of the introduction (lines 166-167), as it seems out of place where it is currently.

A 3.2 We removed the last two sentences of this paragraph and changed behavioural response into 'initial behavioural response' in the first sentence of the closing paragraph. Additionally, we added an entire paragraph to the introduction to discuss the differences between our method and the commonly used method in such studies: dichotomous mate choice trials. We hope this answers the reviewers concerns.

Lines 229-238 (also lines 553-557): the idea that focal 'females' had changed sex seems to be concluded rather too strongly. The evidence for sex change in cichlids is fairly weak (see lines 553- 557) and taxonomically narrow, and there is nothing specifically on *Ophthalmotilapia*. The alternative explanation is that sex determination by examination of the genital papilla is not straightforward. Perhaps this ought to be mentioned as an alternative. Solid evidence for sex change would really be when known (isolated or tagged) individuals were seen to spawn (as females) and then post-sex change were found to be able to fertilise eggs of females successfully and when dissected, had male gonads with sperm visible on microscopic examination. The present study falls well short of this level of evidence: immature males and females basically look the same in most cichlids. The illustrations on Supplement 2 show that 'apparent females' that were euthanised after developing male phenotypes (colourful) had male gonads and those that retained female phenotypes had female gonads, but this is also consistent with some of the apparent females being late-developing males.

A 3.3 This solid evidence was indeed not provided. Hence, we agree with the reviewer that we should discuss both possible scenarios. In the introduction, we changed the relevant sentences into: “A comparison between papillae of the same individuals after two weeks revealed that some could be identified as males. This either confirmed that a sex change did indeed take place in several specimens, or showed that the examination of genital papillae was insufficient for sex determination in subadult *Ophthalmotilapia*.”

In the discussion, we included both possibilities and added: “As no female reproductions were recorded when these specimens still had a female morphology, we cannot rule out that they were males that delayed the development of the conspicuous male colouration and the elongated pelvic fins. Hence, we did not provide solid evidence of sex change in *Ophthalmotilapia*.” We also changed the legend of supplement 2.

Lines 281-285: “Each parameter was calculated three times: once using coordinates obtained for 721/871 seconds before (before) the removal of the visual barrier, once using coordinates obtained during 721/871 seconds (after1) after the visual barrier was removed and finally using coordinates obtained during 2186/2685 seconds (after2)”. Does 721/871 mean that a period ranging from 721 to 871 seconds was analysed, depending on how much suitable footage had been filmed? Or that sometimes it was 721 seconds and other times it was 871 seconds (as these were the minimum ‘before’ frame counts in the two experiments). It might be clearer to just give a little bit of an explanation here.

A 3.4 The two time frames were collected in the OV and the ON experiments, respectively. We placed the explanation for this “(OV/ON experiment respectively)” in the beginning of the sentence instead of at the end. We hope that this removes the confusion. Footage of 871 seconds were minimally available for all OV trails, whereas for ON trails, only a minimum of 721 seconds was available. This information is given in the first paragraph of the section ‘Collection of tracking and qualitative behavioural data’.

Line 315: Permanova is method that may not be familiar to all readers, although it does seem appropriate and useful. It might help to briefly drop a few hints about it. For example, it might be useful to point out that Permanova is a non-parametric technique, because otherwise the use of non-parametric MW tests on the same dataset seems a bit odd. Overall, it might have been interesting to see a bit more of the data distribution shapes to get some idea of whether it was really non-parametric and why no attempt could be made to transform the data for parametric analyses, most of which are generally considered quite robust to non-normality and more statistically powerful.

A 3.5 The reason for using permanova was explained later on in the same paragraph. However, we now also added a rationale as to why this test was chosen in the beginning of the paragraph. Although parametric tests are indeed quite robust for deviations of normality, this robustness decreases when datasets are small. In our analyses, some contrasts were tested between sets that only

contained 3 trails favouring non-parametric test. It is true that some comparisons could have been tested with parametric tests. However, we refrained from using parametric tests in some, but not in other comparisons as this would hinder comparing different contrasts. Given that 16 variables were recorded, for four time frames, showing the distribution for all of them could be too much. However, all data are available in the supplements and can be explored at will. The spread of multidimensional data can be visualised through PCA, which is performed in this study.

Lines 459-460: “Although we could not exclude that these differences stem from dispersion effects”. It might be clearer to say what is meant by ‘dispersion’, since it is used in a discussion about significant differences in average values. This also crops up in Table 2, and seems to reflect an analysis analogous to a test of homogeneity of variances as would normally be used in parametric tests. This would usually invalidate the analysis in parametric statistics, but it is not treated this way here. Maybe this could be briefly explained in the methods section.

A 3.6 This is indeed a non-parametric analogue of (lack of) homogeneity of variances. It is theoretically possible that permanova results misreport a location event (i.e. difference in average value), due to a dispersion effect. However, often a dispersion effect is present together with a location effect. This can be seen when visualising the data, eg. through PCA.

Dispersion was tested using the *betadisper* function of *vegan*. It was mostly present in contrast A of the ON experiment, and to a lesser extent in contrast C of the same experiment. Fig 1C visualises how the dispersion effect in the former contrast can be explained by the (much) smaller variation in the control group compared to all other focal females. However, the same plot also shows the distinction of the average values, as values for control females and those of other females don't overlap. In the OV experiments, no dispersion effect was retrieved for contrast A and figure D indeed shows how the large variation of values of control females.

For comparison C of the ON experiments, we observed only a (weak) dispersion effect in some of the time periods. Visualisation of the data shows that this could be due to the larger variance within the females that were presented to a con-than to a heterospecific male.

We added this interpretation to the manuscript as follows:

In the material and methods section, we added: “A difference in dispersion across groups does, however, not contradict a difference in their average values. Hence, the data was visually explored, using PCA, to assess whether dispersion could be present together with a difference in average values.”

In the results section, we added: “However, these can be explained by the (much) lower variation in values for control females compared to those of other focal females. As visualisation of the data showed a clear separation of both groups, we can assume that a difference in dispersion is present together with a difference in average values (Fig. 2C) ” in the results section.

Lines 647-649: “Males therefore evolved morphological characteristics, build conspicuous bowers and/or perform stereotyped displays to distinguish them from

sympatric congeners.” This seems an overly strong statement to the effect that species differences in courtship signals are the result of selection to reduce interbreeding (or at least time wasting courtship) between closely-related species. In theory, it could also be driven by purely intraspecific processes. The relative importance of these two factors can really only be estimated empirically and not determined a priori. Maybe make a slightly less definite statement here.

A 3.7 We changed this into: “*Ophthalmotilapia* females could use the morphological characteristics, conspicuous bowers and/or stereotypical displays of males to distinguish them from sympatric congeners.”