Minor revisions needed

Thank you for your patience. Three reviewers have provided comments on your preprint. All generally agree that the work is sound – after my reading of the manuscript, I fully concur. I greatly appreciated the integrative approach that combined population genomics with functional predictions. Congratulations on this exciting work.

We thank the recommender and the three reviewers for this quick evaluation and the helpful comments and suggestions on our manuscript. We have answered all the comments below and made the corresponding revisions in the main text, associated figures and supplementary material. We specify in our response the lines where those changes have been made in the revised version with tracked changes (**Smadja_et_al_2022_biorxiv_v2_tracked_changes.docx**). We have uploaded on biorxiv the revised version of our manuscript (**Smadja_et_al_2022_biorxiv_v2.docx**).

We have also deposited the supplementary material, scripts and data in publicly available repositories (data deposited in the SRA archive https://www.ncbi.nlm.nih.gov/bioproject/PRJNA603262/ ; supplementary material and scripts deposited in the open Zenodo repositories with a DOI: 10.5281/zenodo.6900875 et 10.5281/zenodo.7225771) (we will publish the Zenodo repositories upon acceptance). This information is provided in the "Data and scripts availability" section and the "Supplementary material" section of the manuscript (**from line 902**). We have added a separate Funding section (**line 932**) and modified the "Conflict of interest disclosure" section (**line 956**) to meet PCI requests.

The reviewers make a number of helpful suggestions for improvements that need to be addressed in a revision. In particular:

1) Reviewer 1 criticizes the methodology applied to identify candidate loci underlying assortative mating as it may underestimate the variability of mating preferences between individuals of the same locality. As it stands, assortative mating is considered a binary trait (Choosy vs Non-Choosy populations), while the genetic architecture suggests a polygenic trait. More information on the genetic variability of this trait in each of the four populations would be valuable.

Response:

We will answer this general comment in two steps: first, we provide additional information on how mate preferences are assessed in natural house mouse populations and the information we have on the variability of this trait within populations; second, we provide several references that justify the analytical strategy we have applied to this study (i.e. analysing a polygenic trait as a binary variable in GWAS to assess its genetic architecture).

The behavioural results and data obtained on the same individuals and populations as the one sequenced in our manuscript were published in a previous study (Smadja et al. 2015). Several previous studies had investigated in detail mate preferences in similar populations (from the border of the hybrid zone in Denmark, and allopatric populations) (Smadja & Ganem 2002, Smadja et al. 2004, Smadja & Ganem 2005). We want to specify one point here because it is specific to our biological system and may not be obvious to the reviewers. In all these studies assessing mate preferences, this trait could only be statistically assessed at the population level in mouse natural populations, since repeated measures on the same individual (to improve the reliability of individual mate preference estimates) could not be obtained due to the habituation of the mouse to the experimental procedure. We therefore estimated an average mate preference ratio for each population based on ~30 recorded individual preference ratios ((time spent with the homosubspecific stimulus-time spent with the heterosubspecific stimulus) divided by the total time spent with both stimuli) and tested whether this average ratio was significantly different from zero, i.e. indicative of assortative mate preferences (ratio >0), disassortative mate preferences (ratio <0), or non-directional (ratio =0). As a result, we can classify populations as Choosy or Non-Choosy statistically, but cannot classify individuals within populations as Choosy or

Non-Choosy. For this reason, the pooling approach at the genetic level is particularly relevant to our biological context, as we cannot link phenotypes and genotypes at the individual level. Having specified those points, we can nevertheless provide information about the variation of mate preference ratios among individuals within populations: the coefficients of variation are comparable across populations and between Choosy and Non-Choosy populations although the distributions of individual mate preference ratios do not overlap between Choosy and Non-Choosy populations, indicating that individuals from the two types of populations express divergent mate preferences (Smadja et al. 2004 Fig.2, Smadja & Ganem 2005 Table 2, Smadja et al. 2015 in the text).

Regardless of the possibility of assessing phenotypic variation at the individual level, the application of pool-seq to mapping quantitative traits was investigated in multiple theoretical and empirical studies (e.g. Bastide et al. 2013, Schlötterer et al 2015, Micheletti and Narum 2018, Inbar et al. 2020) and many published studies have addressed the genetic basis of polygenic traits using a pool sequencing approach (e.g. Gould et al. 2017, Raffini et al. 2017, Carlier et al. 2021, Lirakis et al. 2022). Moreover, many GWAS studies have applied an extreme-phenotype strategy, often coupled with a pool sequencing strategy. This extreme-phenotype strategy relies on the estimation of allele frequencies between pools of individuals representing the two phenotypic extremes (the two phenotypic extremes being considered as a binary variable), enabling the discovery of associations between genetic variants and the quantitative trait of interest (e.g. Yang et al. 2015, Amanat et al. 2020). It does not imply that the genetic basis underlying the trait is monogenetic or simple, as multiple genetic variants can be associated with phenotypic divergence. Given the divergence in mate preference between the two types of populations that we call Choosy and Non-Choosy populations, we can therefore apply this strategy by considering mate preference as a binary trait (two distinct phenotypes) while still investigating the polygenetic nature of its architecture.

Revision: We have more clearly specified in the introduction the fact that sequencing has been performed on individuals that had already previously behaviourally typed (Smadja et al. 2015) (**line 101**). We have also added a sentence in the introduction specifying that we applied a pool-seq strategy and that it is particularly relevant in our case given that mate preferences are statistically assessed in wild mice at the population level and show limited variability among individuals within populations (**lines 105-109**).

References cited :

Gould BA, Chen Y, Lowry DB. 2017. Pooled ecotype sequencing reveals candidate genetic mechanisms for adaptive differentiation and reproductive isolation. *Mol. Ecol.* 26:163–177.

Carlier J, Bonnot F, Roussel V, Abadie C, Wright S, Ravel S, Martinez RT, Perez-Vicente L. 2021. Convergent adaptation to quantitative host resistance in a major plant pathogen.Turgeon BG, editor. *MBio* [Internet] 12:1–18. Available from: https://journals.asm.org/doi/10.1128/mBio.03129-20

Lirakis M, Nolte V, Schlötterer C. 2022. Pool-GWAS on reproductive dormancy in Drosophila simulans suggests a polygenic architecture. *G3 Genes, Genomes, Genet.* 12.

Raffini F, Fruciano C, Franchini P, Meyer A. 2017. Towards understanding the genetic basis of mouth asymmetry in the scale-eating cichlid *Perissodus microlepis*. *Mol. Ecol.* [Internet] 26:77–91. Available from: https://onlinelibrary.wiley.com/doi/10.1111/mec.13699

Schlötterer C, Kofler R, Versace E, Tobler R, Franssen SU. 2015. Combining experimental evolution with next-generation sequencing: A powerful tool to study adaptation from standing genetic variation. *Heredity (Edinb).* 114:431–440.

Bastide H, Betancourt A, Nolte V, Tobler R, Stöbe P, Futschik A, Schlötterer C. 2013. A Genome-Wide, Fine-Scale Map of Natural Pigmentation Variation in Drosophila melanogaster.Wittkopp P, editor. *PLoS Genet.* [Internet] 9:e1003534. Available from: https://dx.plos.org/10.1371/journal.pgen.1003534

Inbar S, Cohen P, Yahav T, Privman E. 2020. Comparative study of population genomic approaches for mapping colony-level traits.Mikheyev A, editor. *PLOS Comput. Biol.* [Internet] 16:e1007653. Available from: https://dx.plos.org/10.1371/journal.pcbi.1007653

Micheletti SJ, Narum SR. 2018. Utility of pooled sequencing for association mapping in nonmodel organisms. *Mol. Ecol. Resour.* [Internet] 18:825–837. Available from: https://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12784

Yang J, Jiang H, Yeh C-T, Yu J, Jeddeloh JA, Nettleton D, Schnable PS. 2015. Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants

by sequencing pools of individuals selected from a diversity panel. *Plant J.* [Internet] 84:587–596. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26386250

Amanat S, Requena T, Lopez-Escamez JA. 2020. A Systematic Review of Extreme Phenotype Strategies to Search for Rare Variants in Genetic Studies of Complex Disorders. *Genes (Basel).* [Internet] 11:987. Available from: https://www.mdpi.com/2073-4425/11/9/987

Besides, I have a suggestion of my own. While the multi-way comparisons across populations are a bright setting to uncover the genetic basis of assortative mating, all differentiated genomic regions between Choosy and Non-Choosy populations are not involved in this trait. For example, neutral clines could produce genetic structure at specific loci between the Choosy populations (introgressed as they are near the hybrid zone) and Non-Choosy populations (not introgressed as they stand in allopatry). The authors could be a bit more critical on this point (it is acknowledged in the Discussion section but not really reflected elsewhere), and shed light on the importance of the functional predictions to narrow down the set of candidate variants.

Response: We completely agree with this comment. We have added some text highlighting this point in the manuscript (**Introduction lines 141-150, discussion lines 466-467**).

2) All three reviewers would like clarifications on the BayPass methodology (see their specific comments). This is key for the reader because it is the first step to identifying putative variants of assortative mating. In particular, Reviewer 3 would like to see more precise expectations for the genomic patterns left by various scenarios (neutrality, selection on a single variant, polygenic selection) on the C2 statistics. There is no need to run simulations; just clarify your expectations, please.

As your genome scan on assortative mating indicates a complex trait, I recommend (if possible) running a method that can detect footprints of polygenic selection around the two candidate clusters on chromosomes 7 and 9. For example, diploS/HIC is a machine learning method that uses unphased genotypes to classify genomic windows in hard and soft sweeps: https://doi.org/10.1534/g3.118.200262.

Response: We reply to the comments on the BayPass methodology in our responses to the referees below. As for your suggestion to run a method that can detect footprints of polygenic selection around the two candidate clusters on chromosomes 7 and 9, we also think that it would be a very interesting perspective. However, we believe that this objective of investigating the regimes of selection acting in these populations and candidate genomic regions should be addressed in a follow-up study, where we would scale up the sampling design (by increasing the number of Choosy populations potentially subject to this regime of polygenic selection) and generate individual genomic data, so that the power to investigate polygenic selection is maximised. Adding this analysis to this manuscript that is already dense might complicate it too much, and the current dataset might be too limiting to properly address this question. In the discussion, we have now highlighted this possible perspective (**line 523-526**).

3) I agree with Reviewer 3 that the Discussion section on the role of receptor genes is a bit long. I would recommend ending the manuscript as they suggest with an outlook on what sort of analyses can be done in the future.

Response: We have significantly shortened this part of the discussion following these comments. We now end the manuscript with an outlook on the possible of follow-up studies that would be interesting to carry out in the future (**lines 605-675**).

As there are no major criticisms (and suggestions are reasonable to tackle), I believe this can lead to a recommendation as soon as it has been revised in response to the points raised. I am looking forward to receiving your revised preprint.

With best regards,

Christelle Fraïsse.

Reviews

Reviewed by Ludovic Claude Maisonneuve, 29 Aug 2022 13:57

The preprint presents a population genomics study of mate preferences suspected to be implied reinforcement. The authors studied the correlation between SNPs/genes and the presence or absence of mate preference leading to reproductive isolation with a closely related subspecies (Choosy vs Non choosy). The authors identify olfactory receptors genes that may implied in reinforcements. Those genes mainly clusters in two areas in the genome.

Reinforcement is an important mechanism because it may be involved in speciation and explain why there is so many species. However we lack of empirical data to determine the prevalence of this mechanism in natural populations. I then think this study has value because it investigates the genetic basis of mate preference suspected to be implied reinforcement. However I am not an expert on genomics, so I cannot judge the pertinence of the method used in this preprint.

We thank the reviewer for the overall good evaluation of our manuscript and the time spent reviewing it. Please find our replies to the specific comments and suggestions below.

My main concern about this study is that the authors did not study the correlation between the genotype and the mating behavior but between the genotype and the locality. If I understood well the study, the mating behavior was inferred from individual locally and not properly tested. This method may underestimate the variability of mating preference between individuals of a same locality.

Response: We did not rely only on locality information to classify the studied populations as Choosy or Non-Choosy ones. We had initially tested mate preferences on the individuals that were subsequently sequenced for the present study, and the behavioural results were published in a previous study (Smadja et al. 2015). Moreover, several other previous studies had investigated in detail mate preferences in similar populations (from the border of the hybrid zone in Denmark, and allopatric populations) (Smadja & Ganem 2002, Smadja et al. 2004, Smadja & Ganem 2005). The combined information from these different studies, therefore, demonstrates a clear pattern of mate preference divergence between populations sampled on the border of the hybrid zone and populations outside of the hybrid zone. We have revised the introduction to more clearly specify the fact that sequencing has been performed on individuals that were already previously behaviourally typed (Smadja et al. 2015) (**line 101**).

Regarding the variability of mate preferences among individuals of the same population, we have to specify first that, in the house mouse, this trait could only be statistically assessed at the population level in mouse natural populations, since repeated measures on the same individual (to improve the reliability of individual mate preference estimates) could not be obtained due to the habituation of the mouse to the experimental procedure. We therefore estimated an average mate preference ratio for each population based on ~30 recorded individual preference ratios ((time spent with the homosubspecific stimulus-time spent with the heterosubspecific stimulus) divided by the total time spent with both stimuli) and tested whether this average ratio was significantly different from zero,

i.e. indicative of assortative mate preferences (ratio >0), disassortative mate preferences (ratio <0), or non-directional (ratio =0). As a result, we can classify populations as Choosy or Non-Choosy statistically, but cannot classify individuals within populations as Choosy or Non-Choosy. For this reason, the pooling approach at the genetic level is particularly relevant to our biological context, as we cannot link phenotypes and genotypes at the individual level. Having said that, we can nevertheless provide information about the variation of mate preference ratios among individuals within populations: the coefficients of variation are comparable across populations and between Choosy and Non-Choosy populations although the distributions of individual mate preference ratios do not overlap between Choosy and Non-Choosy populations, indicating that individuals from the two types of populations express divergent mate preferences (Smadja et al. 2004 Fig.2, Smadja & Ganem 2005 Table 2, Smadja et al. 2015 in the text). We have added a sentence in the introduction specifying that we applied a pool-seq strategy and that it is particularly relevant in our case given that mate preferences are statistically assessed in wild mice at the population level and show limited variability among individuals within populations (**lines 105-109**).

Another concern is about the meaning of the C2 statistic. Much of the study is based on this statistic. However I did not understand how it is computed. At least it would be nice to have an idea of what it represents.

Response: The C_2 contrast statistic has been developed (Olazcuaga et al. 2020) to propose a nonparametric counterpart for the association model implemented in BayPass (Gautier 2015). It is basically aimed at comparing the allele frequency between two groups of populations defined according to a given trait or characteristics. This difference is expected to be maximal at the underlying associated variants (consider for instance and in the most extreme case, the frequency of variants underlying coat colour in domestic animal breeds). However, it is of utmost importance when using difference in allele frequencies in such a way to account for the confounding effect of population genetic differentiation at the whole genome level introduced by their demographic history. Indeed, the higher the genetic divergence between populations (as usually assessed with the F_{ST} statistics), the higher the expected proportion of neutral variants showing moderate to high difference between group of population whatever the way these are defined. The C_2 statistics is precisely constructed to correct (or at least mitigate) the confounding effect of demographic history when assessing the differences in allele frequencies between the two groups of populations by instead considering differences in standardized allele frequencies (i.e., corrected for the "neutral" demographic history of the populations that is captured by an estimated scaled covariance matrix). A more detailed description and evaluation of the performance of this statistic (on simulated data in particular) are available in Olazcuaga et al., 2020).

Also the authors did not explain the relatedness observed in figure 1b. I am surprised by this relatedness as it did not correspond to the spatial structure. It would be nice if the authors discuss about this.

Response: We agree with the reviewer that the relatedness may seem surprising in the first place if we expect only geographic distance or distance to the centre of the hybrid zone to influence population structure in this area. However, the Jutland region is characterised by numerous freshwater streams (rivers and lakes) that can constitute barriers to gene flow among geographically close populations. The spatial structure is therefore quite complex and relatedness among populations can be influenced by other factors than geographic distance. For this study, the observed relatedness is an advantage as it provides more power to detect genetic variants associated with behavioural divergence, which does not follow the background level of genetic differentiation. We have added one sentence in the text to specify this point (**lines 185-187**).

I also have minor comments:

l27: what is a functional prediction ?

Response: The term “functional prediction” is used here to describe the prediction one can make on the nature of the genes underlying the trait of interest, based on *a priori* information on the functions/proximal mechanisms involved in the expression of the trait. In our study, we predict that genes involved in olfactory recognition should be identified as outliers and should be considered as our prime candidates (among all identified outliers), given the information we have that assortative mate preferences in these house mouse populations are displayed based on the olfactory recognition of urinary signals.

l30-31: ‘which contrasts allele frequencies’ it may be because I am unfamiliar with genomics technique but I don’t understand what it means

Response: Because the C_2 contrast statistic is calculated as the mean squared difference of the sum of standardized allele frequencies of the two groups of populations defined according to the binary trait modalities, we say that it “contrasts allele frequencies” between these sets of populations.

l50: ‘to understand how behavioural traits evolve’ it is a bit vague, what do you mean ? To identify the mechanisms underlying behaviors ?

Response: We mean here identifying the evolutionary forces driving the evolution of behavioural traits. In the context of this study, this means understanding whether reinforcing selection could be the source of divergence in mate preferences between the two types of populations. We believe that readers will understand this sentence.

l52: maybe replace ‘play’ but ‘may play’ is more correct, see Servedio 2004 the what and why of research on reinforcement.

Response: We have revised the sentence to “and may diverge rapidly through the action of direct and indirect selective effects” (**line 54**).

l62: why these cases are ‘relatively simple’ ?

Response: Because only a few loci were identified as underlying these behavioural traits – not highly polygenic traits.

l63: similarly what are ‘more complex traits’ ? Do you mean that the genetic basis implied a larger number of genes ?
l66: similarly what is a ‘complex behavioral trait’ ? Does that mean what there are simple behavioral trait ?

Response: We used this term “complex trait”, as in the literature, as a synonym for a quantitative trait. For behavioural traits, some can be highly polygenic (hence the term “complex behavioural traits”) but some can be encoded by a single gene or a very limited number of genes (e.g. circadian rhythm in Drosophila melanogaster encoded by the *period* locus). We acknowledge the fact that this term might not be the most meaningful, so we have revised the manuscript to avoid it (**lines 26, 28, 62, 66, 418**).

l166-168: ‘a situation that makes the identification of outlier divergent loci in the Focal Test comparing Choosy and Non-Choosy populations conservative.’ Can you explain in few words why ?

Response: Overall, we expect fewer genetic differences between Choosy and Non-Choosy populations if the overall relatedness pattern among the four populations does not cluster the two Choosy populations together and the two Non-Choosy populations together. This makes the identification of

genetic variants associated with the behavioural divergence more conservative (fewer false positives) when comparing Choosy versus Non-Choosy populations.

l212: I don't understand what are 'already known variants'

Response: Already known variants in VEP correspond to variants that have already been identified in mouse genomes (that have been previously genotyped or sometimes functionally characterised). We have specified this point in the text (**line 234**).

l512-516: this part is unclear for me. I do not understand why it would correspond to a matching rule.

Response: Under a matching rule hypothesis, choosy individuals will preferentially mate with Choosy individuals if they mate based on a match with their own phenotype (Kopp et al. 2018). Under the scenario where the allele (or set of alleles) conferring the ability to mate assortatively has been acquired via introgression from the nearby subspecies *M. m. domesticus*, the two subspecies share this allele (in populations from the hybrid zone) but express opposite mate preferences (each preferring to mate with individuals of their own subspecies). This implies that this assortative allele should function via a matching rule, i.e. a match between the genetic backgrounds (or mating signal loci) of the individual choosing and the individual being chosen. For example, if a *M. m. musculus* individual carries this assortment allele, it will make a preferential choice towards a mating partner matching its own genetic background (and if a M. m. domesticus individual carriers this allele, it will make a preferential choice towards a *M. m. domesticus* mate). This corresponds to a matching rule. We have revised this sentence in the manuscript to clarify it (**lines 554-558**).

Reviewed by anonymous reviewer, 13 Sep 2022 16:33

This paper investigated the genetic basis of an olfactory-based behavioral divergence in the model organism Mum musculus. To do this, this study compared populations with choosy vs. non-choosy behavior, while also contrasting populations with the same behavior as a control. To further establish the link between genetic divergence and phenotypic divergence, the study also examines divergence across all annotated genes. Moreover, this study also examines possible genomic signature of reinforcement, including selection sweeps among genetic differentiation outliers, and association between outlier genes and candidate hybrid sterility regions. I think this is a well-designed, well-executed, extensive research. The intro and discussion are well written and make extensive connection to relevant literatures in the field. It is an important contribution to the methodology of uncovering genetic basis of behaviors, which is notoriously hard in natural systems. It serves as a rare study that examined genomic signature of reinforcement.

Thank you very much for your positive comments on our manuscript and your time reviewing it!

I only have a few minor suggestions:

1. The exact comparisons in Focal vs. control test in Figure 1 C is not clear. My take is that Focal test means comparing choosy 1 vs. Non choosy 1; choosy 1 vs. Non-choosy 2; choosy 2 vs. Non choosy 1; choosy 2 vs. non-choosy 2 , and that Control test is choosy 1 vs. choosy 2; Non-choosy 1 vs. Non-choosy 2. However, in the figure C, "vesus" symbol is used to show the comparisons, but it means different things in different tests. I suggest to write down each comparison in the figure legend and the result session.

Response: Thank you for this comment, which reflects the fact that the BayPass comparisons were not sufficiently clear in Figure 1. The C_2 statistic of BAYPASS (Olazcuaga et al. 2021) contrasts SNP standardized allele frequencies between the two groups of populations. The advantage of this method is that it avoids running multiple pairwise population comparisons, like the ones described by the reviewer. In the Focal test, we contrast Choosy1+Choosy2 (first group of populations) with (versus) Non-Choosy1+Non-Choosy2 (second group of populations); in the Control test, we contrast two

groups of populations that are grouped independently of their behaviour: Choosy1+Non-Choosy2 with (versus) Choosy2+Non-Choosy1. This is why the "versus" word in Figure 1c is between the two groups of populations. The text in the Material and Methods section presents these different contrasts, but we have revised Figure 1 and its legend to clarify this point (**lines 1348, 1352**; **Figure 1_v2.pdf**).

2. Line 295-306. the title of the section states no signatures of selective sweeps confuses me. In the text, there are some genes showed selective sweeps in both Choosy populations.

Response: We have revised the title to "SFS-based signatures of selective sweeps in genetic differentiation outliers" (**line 320**) and we have more clearly specified in the text that only 3 outlier genes show a consistent signature of a selective sweep in both Choosy populations (**line 330**).

3. Figure 4. I am a bit overwhelmed by different font sizes and different colors of genes in the figure. I understand different color means different functional receptors. However, what do different sizes font mean? The symbol "<", ">" means where the gene starts and ends? No annotations of these symbol meanings are found in the figure legend.

Response: The different colour codes are explained in the figure legend, but we have revised the figure legend to add the meaning of the "<" and ">" symbols (**line 1378**).

4. Because the key findings in this study is built upon program BAYPASS to tease apart the effect of population structure on signals of genetic divergence related to choosy behavior, I hope there is a bit more mechanistic explanation about how this program corrects for the population structure in the introduction. Right now, this information is buried in the method section.

Response: We have added some information about BayPass in the introduction (**lines 114-121**).

Reviewed by Angeles de Cara, 14 Sep 2022 12:28

This is a very nice study and I have enjoyed and learnt while reading it. The goal of the study is to find the genetic basis of olfactory-based mating preferences in mice. For that purpose, the authors sampled and sequenced individuals from the border of a hybrid zone, where strong assortment occurs, and from other locations away from the border of this zone, where no assortment has been identified. By splitting and contrasting individuals into choosy (from the border of the hybrid zone) and non-choosy, the authors obtain candidate regions and genes and then do functional analyses where they identify multiple olfactory and vomeronasal receptor genes as outliers.

Overall, I do think the goals are clearly stated, the paper is well written, and the methods are appropiate and well used.

I really enjoyed the discussion, although the section on the role of receptor genes seemed a bit long. I believe a bit more insight into what sort of analyses can be done in the future, especially for those readers not familiar with the work of Isogai et al (2011), would be useful to close down the manuscript.

Response: Thank you very much for your comments and your time reviewing our manuscript. Concerning the discussion, we have now revised this part to shorten it and we now end the discussion with an outlook of the possible follow-up studies that would be interesting to carry out in the future to further gain insight into the role and modes of divergence of these candidate receptors, including perspectives adapting methods developed by Isogai et al. (**lines 606-677**).

Some issues I have:

It seems there is no reduced diversity in the outlier regions, and no signs of selective sweeps, and all analyses are indicative of polygenic selection. That seems to be the case both for outlier genes in C2_max or in C2_mean, while I would have expected some different in diversity on those two

scenarios. On the whole, I would appreciate whether some findings, like no reduced diversity and no overlap with candidate hybrid sterility genes were expected or unexpected, or what could be the explanation for either of these findings.

Response: C_2_max and C_2_mean gene-based statistics are expected to carry distinct information regarding the genetic architecture of the variants associated with phenotypic variation (either one single main causal variant within a gene or multiple small-effect causal variants within a gene). Under different regimes of selection (hard and soft sweeps), we could expect C_2_max outlier genes to be more under hard sweeps and C_2_mean outlier genes under soft sweeps. However, we are here more probably in a context of polygenic selection acting on all the genes underlying behavioural divergence. The two types of outlier genes may then just reflect different variant architectures within genes rather than different selective regimes. We have clarified this point in the discussion (**lines 518-520**). We have also revised the part of the discussion addressing the limited overlap with candidate hybrid sterility genes, to clarify our expectations and suggest possible explanations for this result (**lines 571-577**).

Are there simulations to know which scenarios of selection / genetic architecture result in outliers of C2_max and which in outliers of C2_mean?

Response: Studies implementing gene-based GWAS methods have run simulations to test the power of gene-based statistics (e.g. Wang et al. 2007, Curtis et al. 2008, Liu et al. 2010, Li et al. 2011, Wang et al. 2017). However, we are not aware of simulation studies addressing the underlying scenarios of selection captured by "max" and "mean" gene-based statistics. Given the genetic architecture captured by each gene-based statistic, the common idea is to combine the two types of statistics on a given dataset to capture the different possible architectures and not miss genes with single causal variants or alternatively genes with multiple small-effect variants.

Minor comments:
- I suppose the order of methods at the end corresponds to the authors' final choice of journal, but I personally find it easier to follow when methods are described prior to results.

Response: We understand this preference, but we decided to keep this format because the Material and Methods section is quite long and detailed and we believe the manuscript will be easier to read in this order.

- l. 215 and l. 221: coding consequences?

Response: We have revised this to "protein-coding variants" and "predictions made on all SNPs…" (**lines 238 and 244**).

- The sentence finishing in l. 270 should be rewritten ("were less found in the very most significantly")

Response: We have revised this sentence to "Comparatively to *Olfr* genes, *Vmn* outlier genes were proportionally less represented among the most differentiated genes …" (**line 293**).

- l. 749: with -> which

Response: Thanks! Done (**line 811**).

In Fig. 3, the three clusters could be magnified to better appreciate the structure.

Response: Yes, but this is why we produced Figure 4 which zooms on two of these most promising clusters and describe in detail their content and structure. We have added a supplementary figure showing these parts of Figure 3 more in detail (**Supp. Figure S7**, in-text reference **line 399**). However, we prefer to keep Figure 4 in the main text for this purpose.

In Supp. Fig.1, text size is too small.

Response: We have increased the size of the two figures (a and b) in Supp. Fig.1 and we hope that now the text and the figure are more readable (**Smadja_et_al_2022_Supplementary_figures_v2.docx**).