Dear Pr Giraud,

The manuscript has been evaluated by two referees, who agree that this manuscript on the genetic mapping of sex and self-incompatibility determinants in an androdioecious plant is sound, elegant and interesting for the evolutionary biology community. Both referees nevertheless provide a list of excellent suggestions and questions, such as testing the new version of sex-DETector, providing the list of candidate genes and their predicted functions, and comparing genetic maps between sexes. I would encourage resubmission if you are able to revise the manuscript along all these lines.

We thank you and the referees for this very positive evaluation and the detailed reviews. We have revised our manuscript based on these constructive comments. We now in particular provide a complete list of candidate genes in a supplementary table, and we compare the genetic maps between the two parents (they are largely similar). We have carefully considered the possibility of using the new version of Sex-DETector. However, as we explain in details below in the response to reviewer 1, we have opted to stick to our current version, which we already adapted to ensure appropriate SNP calling for genomic rather than RNAseq data. Please find below a point-by-point response to the decision letter.

-L33-34: I would delete the two occurrences of "more" here

done

-All along the text, I would avoid abbreviations (such as DSI); abbreviations are rarely useful for readers and only render the text harder to read. In addition, some syntax using DSI is incorrect, eg L89, it should be DSI system or determinant?

We agree that the DSI abbreviation is not common and rendered reading more difficult. We have replaced it with the more transparent "diallelic SI". We have also carefully checked the use of the terms "system" and "determinant".

-L76-77: clarify, are S- and L-morph flowers the same as "pin" and "thrum"?

We have specified that "pin" corresponds to the L-morph and "thrum" to the S-morph.

-L82: there is no such thing as a "basal" clade in a phylogenetic tree, see for example: https://doi.org/10.1111/j.0307-6970.2004.00262.x

Thanks for catching this imprecise wording. We have removed the term "basal".

-L83: clarify, is the diallelic self-incompatibility the expected system for heterostyled phenotypes?

We have modified the sentence to explain that diallelic SI is indeed the expected system for heterostylous species.

-L87-88: clarify, not clear if this applies only to dioecious sex systems or to all three. What's "x" in "2n=2x=46"?

We have simplified the notation (e.g. 2n=26) to make it more clear.

-L97-98, L167: explain what is the "stigma test" also clarify if the "more distant species" are in different tribes.

We have now defined what the "stigma test" is in the preceding sentence. We have also replaced "more distant species" with "other species of the same tribe".

-L103: not all orthologs are "identical by descend", I reckon you want to test whether they are orthologs (here a locus already committed to this function in the last common ancestor) and if so whether they have remained identically positioned (see comment on use of "syntenic" below).

We have simplified the sentence to limit our assertion to the question of whether the determinants are orthologs.

-L111: name some examples of lineages in which separate sexes have unfolded independently of diallelic self-incompatibility

We have modified this sentence to make it more general and highlight the fact that *P. angustifolia* is special for its decoupling of sex and mating types.

-L117: here and elsewhere use common names at first mention (here European olive)

We have carefully checked the use of common/scientific names throughout the manuscript. Because we refer alternatively to either the wild oleastre/olive tree (*Olea europea var. sylvestris*) or the domesticated olive tree/European olive (*Olea europea*), we still specified scientific names at several places whenever needed.

-L132: is "segregates" rightly used here? I would think that alleles segregate in a progeny but a progeny does not "segregate"?

We have modified the sentence to clarify that it is the sexual phenotypes that segregate, not the progeny.

-L135: is this really an "observation" as the sentence suggests, or a result or the model as the previous sentences let think?

It is indeed a direct observation of genetic segregation of sexual phenotypes (Billiard *et al.* 2015).

-L138: a reference is missing for the observed "departure".

We have added the reference: Billiard *et al.* 2015.

-L164, L171, L255: "sex" usually means genetic mixing, here shouldn't it be "gender"?

We rather used "sexual phenotypes" to avoid introducing an additional term.

-L166: no number at a beginning of a sentence, or spell it out.

Done

-L185-188: I found too cursory the description of how the "de novo catalog" was created. I'm not sure what "catalog" means here.

We have rewritten this paragraph to better explain how the catalog of reference sequences was constructed.

-L193: Make sure the "custom script" is available as supplemental or in the public domain. Same at L203.

All scripts used can be accessed at https://github.com/Amelie-Carre/Genetic-map-of-Phillyrea-angustifolia

-L193-194: correct typo and split the sentence "(..), after removal of SNPs markers with read cover <5.The script combines(...)".

done

-L195: homogenize the typography of "Lep-Map3" (written as "Lep-MAP3" elsewhere).

done

-L218: correct the typo, should be "hemizygous".

done

-L221: explain briefly the principle of the method.

We now explain that SEX-DETector is "a maximum-likelihood inference probabilistic model initially designed to distinguish autosomal from sex-linked genes based on segregation patterns in a cross "

-L241: justify why 110bp was considered enough to determine whether the (reciprocal?) hit is syntenic.

The length of reference sequences in our catalog spans from 140 to 280bp. The 110bp threshold was conservatively chosen to allow a good match over a substantial part of the smallest sequences.

-L268: remind the reader that this logarithm of odds score was chosen so 23 linkage groups are obtained.

done

-L277 no plural when a name qualifies another by being before it, so either "locus identification" or "identification of loci"

done

-L285 and L297-298: I reckon the loci showing "autosomal inheritance" are worth discussing, how these could be in a region where other loci follow a "XY segregation" pattern?

We now explain that the autosomal loci "possibly [correspond] to polymorphisms accumulated within allelic lineages associated with either of the alternate SI or sex specificities"

-L313-314: loci do not "find" homology between genomes, a more precise wording would be "365 loci have good/non ambiguous matches in the European olive tree assembly".

done

-L350: comment, is the olive tree's S-locus less confidentially assembled? e.g. is it scaffolded with long/many N tracks?

We now mention that Mariotti et al. 2021 observed structural rearrangement in the region.

-L373-374: I think "Identification of sequences that have remained linked over these" reads better.

done. Thank you for the suggestion.

-In the discussion, I am not sure "syntenic" is rightly used and clear (L344, 347, L364): I do not see how synteny (i.e. similar gene order) provides support for the hypothesis that the two systems are orthologous? Do you mean instead that the two studies mapped the locus genomic regions with orthologous genes? Unclear L344 what is "identified" (add "as the region controlling SI)?

We have modified the sentences to avoid improper use of the word "syntenic".

-L408, it is unclear what "fully aligned" means here?

We use the word "confounded" to convey the idea that a given sexual phenotype is typically always associated with a given mating type.

-L416: missing closing parenthesis.

done

-L418: it is unclear what "sexual specialization" means here? Separate sexes? Sexual dimorphism? Anisogamy?

We now clarified that we refer to sexual phenotypes.

-L419-420: unclear what you mean within brackets, make a separate sentence and explain the logical relationship with the preceding sentence

We have made two separate sentences.

-The following reference may be cited in the last paragraph on P18: 10.1111/j.1469-185X.2010.00153.x.

done

-Figure 1: Remove the lines framing the figure.

We can see no lines framing the figure in our online or pdf versions.

-Figure 2: Lines are colored according to the linkage map, right?

Correct. This is now clarified in the legend.

-Figure 3: Adding as scale of SNP density to the points (e.g. color by factor, color= in ggplot2's aes).

Each point in this figure corresponds to the best blast hit of one GBS locus. The size of these GBS loci is relatively uniform (140 to 280 bp) so confidence in the blast hit should be relatively constant regardless of how many SNPs they had. For clarity, we stuck to our original representation.

-Could you comment on the interwoven forward and reverse synteny between olive tree Chr18 and LG18?

We mention in the text that "the order of hits along that chromosome suggested a large number of rearrangements".

-Figure 4: lines are blue, not green.

done

- Could you comment if the non-recombining region around the sex and self-incompatibility loci have been extended in P. angustifolia? Do the small contigs of olive tree match other regions in P. angustifolia?

There is currently no assembly of the *P. angustifolia* genome (we are working on it!) so at this stage it remains impossible to determine whether the small contigs in the olive tree genome match other regions. We hope that a future assembly will shed light on the dynamics of these two genomics regions.

Reviewer 1

This is a very nice work and an important step towards understanding the genetic architecture of androdioecy and SI in P. angustifolia. The GBS-based map is of high quality (many markers, many syntenies with Olive tree chromosomes). The SEX-DETector analysis is very elegant. This tool is meant for detecting sex-linked genes in dioecious species but it was used here for detecting sex-linked markers in an androdioecious system (comparing males and hermaphrodites) and SI-linked markers (comparing hermaphrodites S1S1 and S1S2) in a very original manner. The map and the localization of both sex and SI loci will be very helpful for the next step: finding the sex-determining and SI genes. I am thus very positive about this work and I think that it should be recommended by PCI Genomics.

We thank the reviewer for their positive evaluation.

Please find below my (minor) comments:

- The authors have used a version for SEX-DETector meant for RNA-seq data if I have understood correctly the M&M. In particular, they have used reads2snp to genotype the individuals prior to the SEX-DETector analysis per se. A new version of SEX-DETector (SD++) has been recently released. This version takes vcf files as input and any genotyper can thus be used (more here: https://gitlab.in2p3.fr/sex-det-family/sex-detector-plusplus). The results of the SEX-DETector analysis are already very neat. The most significant sex-linked and SI-linked markers map to two different (and relatively small) loci on two different LGs. I do not know whether there is a lot of room for improvement. But the authors might want to try a different genotyper plus SD++ to see if the results get better.

We thank the reviewer for their suggestion. The version of SEX-DETector that we use was indeed developped for RNAseq data. To take into account the fact that we used genomic data, we ran reads2snp without the -aeb (account for allelic expression bias) option, making it fully appropriate for the genotyping of genomic data. It is unclear whether and by how much GATK outperforms read2snp in this context, but we already have a large number of SNPs available (44,565 for the male versus hermaphrodites comparison), resulting in a powerful analysis of segregation patterns across the genome. We confirmed with one of the developers of SEX-DETector (Jos Käffer) that the rest of the algorithm of the new version (SD++) was unchanged beside optimization of the computational efficiency. In this context, we therefore feel that the gain of replicating our analysis with SD++ would be marginal.

- It is known that SEX-DETector tends to overestimate the size of the non-recombining region (Muyle et al. 2016). Here the number of individuals was much bigger than in a typical SEX-DETector analysis using RNA-seq data from 10-20 individuals. This issue might not be a serious one here. The authors could study the sex and SI markers they found in unrelated individuals from natural populations. This is a good way to get the correct boundaries of the sex locus (e.g. Badouin et al. 2020). The use of SDpop, another version of SEX-DETector that deals with genotyping data from individuals sampled in the wild, could help (Kafer et al. 2021). I am aware that this is a substantial amount of work and I understand that the authors cannot do it for this ms, but they might want to keep it in mind for future work.

We cannot agree more that the study of unrelated individuals from natural populations would be a way to evaluate more finely the association between genetic markers and the sexual phenotypes. This is indeed the topic of a different chapter of the PhD work of A. Carré, that we leave for a future manuscript.

Reviewer 2

An obvious improvement to the paper would be to provide information on the genes that are identified in the locations of the sex and self-incompatibility loci. These data are discussed but are not found in any supplementary file.

> **We now report in the new Table S1 and S2 the list of genes in these genomic intervals, along with their designation and Gene Ontology categories.**

In addition, the sex-specific maps should be compared and the region of no recombination in males should be briefly discussed. How big a part of the genome is it likely to represent? I would favour an additional figure showing how the sex-specific maps compare, which could be a supplementary file.

> **We now provide a supplementary Figure (S1) comparing the maternal and paternal maps.**

In addition the data used to make the map, along with the sequence of the RAD markers should be made available so that other researchers can compare the map to other genomes that become available in the future.

> **We now provide a supplementary file containing the catalog of GBS markers, as well as a complete genotype table that can be used to compare with other genomes.**

The figure of the genetic map shows markers that are clearly outliers in the end of some linkage groups, and have orders of magnitude more recombination to any marker than the typical average. These markers are usually removed from the final genetic map. Their assignment to a linkage group may still be correct, so they could be retained with a note on the supplementary file presenting all markers, but they should not be plotted. Similarly, the statistics on the genetic map such as distance between markers and total length should be reported after excluding such outliers.

> **We have updated Figure 1 and Table 1 to remove these outlier loci. The full map (with the outliers) is moved to supplementary material. We have added a column in Table 1 to allow direct comparison of the statistics of the genetic map with and without the outliers.**

Finally the discussion refers to segregation distortion but it is unclear whether it was observed in the described cross. Line 401: it is unclear whether SNPs showing segregation distortion were observed in the cross of this paper. Is it not possible to use your data to study segregation distortion? Such analysis would strengthen this part of the discussion.

> **We do indeed observe segregation distortion at the phenotypic level in the described cross (619 hermaphrodites but only 402 males). We have not specifically looked for segregation distortion at the SNP level, as the power would be limited with only 196 individuals genotyped. In addition, the perfect association we find between some SNPs and the phenotypes would make this analysis redundant.**

Line 34: more collinear -> more collinear between the genomes

> **done**

Line 61: also can -> can also

> **done**

Line 87: F. excelsior and F. chinensis - the full name of the genus is unclear, it reads as it could be one of Fontanesia, Forsythia or Fraxinus.

done

Line 172: random -> randomly

done

Line 173: were -> was

done

Line 178: rare-cutting - replace with the number of based or the recognition sequence. Even better, is it possible to estimate how many times it cuts given the genome size or the genome sequence of a close relative?

We have added the recognition motif for each of these enzymes.

Line 187: to assembly -> using assembly

This sentence has been modified according to comments above, so these words are no longer used.

Line 193: remove SPNs markers -> removal of SNP markers

done

Line 194: reads cover -> read coverage

done

Line 195: to the good Lep-Map3 format -> used by Lep-Map3

done

Line 202: 23 linkage -> 23 linkage groups

done

Line 219: to the SI -> to the two SI

done

Line 240: only loci with unique - how many were they?

This piece of information is reported in the results section (*about half (49%) of the 10,388 P. angustifolia loci used for the genetic map had a significant BLAST hit on the olive tree genome).*

Line 241: for the synteny -> for synteny

done

Line 271: These statistics should exclude the outlier markers in the end of some LGs. It is useful to report the average distance (in cM) between markers and, perhaps, how it translates to bp given the genome size.

Done. See response above

Line 278: I was confused when reading this because it studies the SI system, but refers to an XY system (which sex-detector uses). It is eventually clear. One way to avoid the confusion is to more this paragraph below the paragraph discussing the XY system. Alternatively, you could start with the last sentence of the paragraph. eg "we found evidence that the region on LG18 is associated with the SI phenotypes..."

We thank the reviewer for this useful suggestion. We have clarified the paragraph by starting with the sentence "We found evidence that a region on LG18 is associated with the SI phenotypes, with H_b hermaphrodites having heterozygous genotype, akin to a XY system".

Line 295: 2.216cM - does this refer to the sex-averaged map? I presume they are 0cM appart on the male map. If so it would be helpful to also mention the male and female recombination map distance for these markers.

We have specified that these ditances correspond to the sex-average map, which we feel are more reflective of their genomic extent.

Line 306: did not find hits in particularly clustered regions on other chtomosomes -> did not cluster on other chromosomes

done

Line 310: delete "overall"

done

Line 316: single small -> single

done

Line 332: 545.128 bp - in some parts of the text the "." should be replaced with "," for consistency.

done

Line 351: indeed expected -> expected

done

Line 359: in more distant - the discussion on phylogenetic distance would benefit for more specific ages, so that the reader knows what more and less distant means.

We have replaced “more distant” by the more generic “other Brassicaceae species” as we feel the exact phylogenetic distances are not relevant here.

Line 367: A first -> One

done

Line 383: that happen to have been activated along the P. angustifolia specifically -> that have been activated specifically in P. angustifolia

done

Line 394: is an open question -> is still open

done

Line 422: perfectly -> perfect

done

Figure 3: What is the point of doing a linear regression when the data are clearly not linear? These figures would be more clear without plotting the regression lines.

We agree that the linear regression made the figure unclear. We removed them.