Peer Community In Genomics

Identification of distinct YX-like loci for sex determination and self-incompatibility in an androdioecious shrub

Tatiana Giraud *and* ***Ricardo C. Rodríguez de la Vega*** *based on reviews by 2 anonymous reviewers*

A recommendation of:

Open Access

Genetic mapping of sex and self-incompatibility determinants in the androdioecious plant *Phillyrea angustifolia*

Amelie Carre, Sophie Gallina, Sylvain Santoni, Philippe Vernet, Cecile Gode, Vincent Castric, Pierre Saumitou-Laprade *(2021), bioRxiv, 2021.04.15.439943, ver. 7 peer-reviewed and recommended by Peer Community in Genomics*
10.1101/2021.04.15.439943

Published: 20 July 2021

Submitted: 04 May 2021, *Recommended: 07 July 2021*

Cite this recommendation as:
Tatiana Giraud and Ricardo C. Rodríguez de la Vega (2021) Identification of distinct YX-like loci for sex determination and self-incompatibility in an androdioecious shrub. *Peer Community in Genomics, 100011.* ***10.24072/pci.genomics.100011***

Recommendation

A wide variety of systems have evolved to control mating compatibility in sexual organisms. Their genetic determinism and the factors controlling their evolution represent fascinating questions in evolutionary biology and genomics. The plant *Phillyrea angustifolia* (Oleaceae family) represents an exciting model organism, as it displays two distinct and rare mating compatibility systems [1]: 1) males and hermaphrodites co-occur in populations of this shrub (a rare system called androdioecy), while the evolution and maintenance of purely hermaphroditic plants or mixtures of females and hermaphrodites (a system called gynodioecy) are easier to explain [2]; 2) a homomorphic diallelic self-incompatibility system acts in hermaphrodites, while such systems are usually multi-allelic, as rare alleles are advantageous, being compatible with all other alleles. Previous analyses of crosses brought some interesting answers to these puzzles, showing that males benefit from the ability to mate with all hermaphrodites regardless of their allele at the self-incompatibility system, and suggesting that both sex and self incompatibility are determined by XY-like genetic systems, i.e. with each a dominant allele; homozygotes for a single allele and heterozygotes therefore co-occur in natural populations at both sex and self-incompatibility loci [3].

Here, Carré et al. used genotyping-by-sequencing to build a genome linkage map of *P. angustifolia* [4]. The elegant and original use of a probabilistic model of

segregating alleles (implemented in the SEX-DETector method) allowed to identify both the sex and self-incompatibility loci [4], while this tool was initially developed for detecting sex-linked genes in species with strictly separated sexes (dioecy) [5]. Carré et al. [4] confirmed that the sex and self-incompatibility loci are located in two distinct linkage groups and correspond to XY-like systems. A comparison with the genome of the closely related Olive tree indicated that their self-incompatibility systems were homologous. Such a XY-like system represents a rare genetic determination mechanism for self-incompatibility and has also been recently found to control mating types in oomycetes [6].

This study [4] paves the way for identifying the genes controlling the sex and self-incompatibility phenotypes and for understanding why and how self-incompatibility is only expressed in hermaphrodites and not in males. It will also be fascinating to study more finely the degree and extent of genomic differentiation at these two loci and to assess whether recombination suppression has extended stepwise away from the sex and self-incompatibility loci, as can be expected under some hypotheses, such as the sheltering of deleterious alleles near permanently heterozygous alleles [7]. Furthermore, the co-occurrence in *P. angustifolia* of sex and mating types can contribute to our understanding of the factor controlling their evolution [8].

References

[1] Saumitou-Laprade P, Vernet P, Vassiliadis C, Hoareau Y, Magny G de, Dommée B, Lepart J (2010) A Self-Incompatibility System Explains High Male Frequencies in an Androdioecious Plant. Science, 327, 1648–1650. https://doi.org/10.1126/science.1186687

[2] Pannell JR, Voillemot M (2015) Plant Mating Systems: Female Sterility in the Driver's Seat. Current Biology, 25, R511–R514. https://doi.org/10.1016/j.cub.2015.04.044

[3] Billiard S, Husse L, Lepercq P, Godé C, Bourceaux A, Lepart J, Vernet P, Saumitou-Laprade P (2015) Selfish male-determining element favors the transition from hermaphroditism to androdioecy. Evolution, 69, 683–693. https://doi.org/10.1111/evo.12613

[4] Carre A, Gallina S, Santoni S, Vernet P, Gode C, Castric V, Saumitou-Laprade P (2021) Genetic mapping of sex and self-incompatibility determinants in the androdioecious plant *Phillyrea angustifolia*. bioRxiv, 2021.04.15.439943, ver. 7 peer-reviewed and recommended by Peer Community in Genomics. https://doi.org/10.1101/2021.04.15.439943

[5] Muyle A, Käfer J, Zemp N, Mousset S, Picard F, Marais GA (2016) SEX-DETector: A Probabilistic Approach to Study Sex Chromosomes in Non-Model Organisms. Genome Biology and Evolution, 8, 2530–2543. https://doi.org/10.1093/gbe/evw172

[6] Dussert Y, Legrand L, Mazet ID, Couture C, Piron M-C, Serre R-F, Bouchez O, Mestre P, Toffolatti SL, Giraud T, Delmotte F (2020) Identification of the First Oomycete Mating-type Locus Sequence in the Grapevine Downy Mildew Pathogen, *Plasmopara viticola*. Current Biology, 30, 3897-3907.e4. https://doi.org/10.1016/j.cub.2020.07.057

[7] Jay P, Tezenas E, Giraud T (2021) A deleterious mutation-sheltering theory for the evolution of sex chromosomes and supergenes. bioRxiv, 2021.05.17.444504. https://doi.org/10.1101/2021.05.17.444504

[8] Billiard S, López-Villavicencio M, Devier B, Hood ME, Fairhead C, Giraud T (2011) Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. Biological Reviews, 86, 421–442. https://doi.org/10.1111/j.1469-185X.2010.00153.x

Reviews

Toggle reviews

Revision round #2

2021-06-29

Author's Reply

Download author's reply (PDF file)Download tracked changes file

Decision round #2

The authors have throroughly revised the manuscript according to the referees' comments and we have only a few minor suggestions before we recommend the paper in PCI Genomics.

-L39 : Modes of sexual reproduction are strikingly diverse

-L97 : add a reference

-L99 : italics for « P. angustifolia »

-L108 : explain briefly here what is this stigma test and in details in M&M

-L110 : coma after the bracket

-L115 : unclear what « its » refers to

-L112, L356, 358, 370, 372, 373, 397, 422 and elsewhere : tense should be homogeneous within sentences ; I would keep the past tense all along

-L145 : M allele

-L152 : it has not been clarified that this represents experimental data, and it is unclear as the previous sentence mentions a model

-L177 : delete specific

-L212, L297 : no capital within a sentence even for explaining acronyms

-L234, L244, L723 : no plural when a name is before another name, so either just SNPs or SNP markers

-L277 : it is still not clear what is the principle of this method (ant not only its goal), i.e., how it differs from just classical association genetics

-L321, L323 : give the P values and N

-L325 : no number at the beginning of a sentence or written in full letters

-L331 : revise sentence, the two numbers and two comas are unclear

-L377 : fewer instead of less

-L304 : I would recommend using sequence similarity instead of homology, which has a different meaning in evolution, ie with a notion of shared ancestry

-P11 : is it possible to plot the differentiation between alleles in the scaffold of interest to assess whether the differentiation is much higher at the SI and sex deterining loci than alsewhere, or even if there is a pattern of evolutionary strata ? Even if the contigs cannot be fully ordered, the differentiation levels may add further strong support to the findings and interpretation.

-P12 : it could be interesting to discuss the mating-type system in oomycetes that also resembles XY sex-determining systems (DOI:https://doi.org/10.1016/j.cub.2020.07.057)

-L471, L484, L493, 497, 530 : I would use the term homologous instead of orthologous ; Orthologous is used for genes, to distinguish paralogous and orthologous genes among homologous genes, but you have not studied or found paralogs here, and it cannot be used with « functionnally » to my understanding

-L478-479 : I do not understand this sentence, it seems to have a syntax issue

-L493 : sexes or sexual morphs instead of sex

-L496 : « such as » is redundant with « eg »

-L14 : I find frustrating not to have some speculations about what can cause this distortion… meiotic drive ? more complex genetic system than a single locus ? selection for balanced male and female functions in the population, which can be different from the sexual morph census numbers ?

Tatiana Giraud and Ricardo Rodriguez de la Vega

Preprint DOI: ***10.1101/2021.04.15.439943***

Revision round #1

2021-06-09

Author's Reply

Download author's reply (PDF file)Download tracked changes file

Decision round #1

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.

In addition, please also address the comments below:

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

-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?

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

-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

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

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

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

-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).

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

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

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

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

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

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

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

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

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

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

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

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

-L221: explain briefly the principle of the method.

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

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

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

-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?

-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".

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

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

-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)?

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

-L416: missing closing parenthesis.

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

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

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

-Figure 1: Remove the lines framing the figure.

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

-Figure 3: Adding as scale of SNP density to the points (e.g. color by factor, color= in ggplot2’s aes). Could you comment on the interwoven forward and reverse synteny between olive tree Chr18 and LG18?

-Figure 4: lines are blue, not green.

-Figure 5:

- 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?

*Preprint DOI: **10.1101/2021.04.15.439943***

Reviewed by anonymous reviewer, 2021-05-31 17:19

This ms addresses the question of the genetic architecture of SI and sex in the plant Phillyrea angustifolia. This ms is a follow up of all the work done by the same group on studying the unusual mating system of this plant. P. angustifolia is androdioecious with hermaphrodite and male individuals. Hermaphrodite selfing is prevented by a homomorphic SI system with two alleles only. Androdioecy and bi-allelic SI thus co-exist in this plant. The ms reports the construction of a high-resolution GBS-based genetic map from a cross and the identification of the sex-linked and SI-linked markers using the tool SEX-DETector. Two loci (one for sex one for SI) were found on two different linkage groups. Comparison with the Olive tree reference genome (a relatively close relative) suggested the SI locus is shared between the two species.

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. 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.

- 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.

References

Badouin H, Velt A, Gindraud F, Flutre T, Dumas V, Vautrin S, Marande W, Corbi J, Sallet E, Ganofsky J, Santoni S, Guyot D, Ricciardelli E, Jepsen K, Käfer J, Berges H, Duchêne E, Picard F, Hugueney P, Tavares R, Bacilieri R, Rustenholz C, Marais GAB. The wild grape genome sequence provides insights into the transition from dioecy to hermaphroditism during grape domestication. Genome Biol. 2020 Sep 7;21(1):223. doi: 10.1186/s13059-020-02131-y. PMID: 32892750; PMCID: PMC7487632.

Käfer J, Lartillot N, Marais GAB, Picard F. Detecting sex-linked genes using genotyped individuals sampled in natural populations. Genetics. 2021 Mar 25:iyab053. doi: 10.1093/genetics/iyab053. Epub ahead of print. PMID: 33764439.

Reviewed by anonymous reviewer, 2021-06-07 15:42

The manuscript by Carre and colleagues describes a new genetic map, constructed using RAD markers, in the androdioecious plant Philyrea angustifolia, maps the self incompatibility and sex loci and compares them with the olive tree genome, with which synteny is generally conserved. The manuscript is clearly written and the results are well-supported and add necessary data points in the evolution of separate sexes and the self incompatibility system in this interesting phylogeny.

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.

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. 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.

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.

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.

Minor points

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

Line 61: also can -> can also

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.

Line 172: random -> randomly

Line 173: were -> was

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?

Line 187: to assembly -> using assembly

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

Line 194: reads cover -> read coverage

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

Line 202: 23 linkage -> 23 linkage groups

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

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

Line 241: for the synteny -> for synteny

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.

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..."

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.

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

Line 310: delete "overall"

Line 316: single small -> single

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

Line 351: indeed expected -> expected

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.

Line 367: A first -> One

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

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

Line 422: perfectly -> perfect

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.