Dear editor, dear reviewers

First, on behalf of all the co-authors, I want to sincerely apologize for the long delay between receiving your feedback and resubmitting the manuscript. Initially, I, as the first author, was unable to implement the necessary revisions shortly after the reviewers' feedback. Then, very sadly, Laurence Hibrand Saint-Oyant, one of the two leaders of this study, passed away suddenly in mid-November. This loss was deeply emotional for all of us, especially given Laurence's pivotal role in this work. As a result, we decided to further postpone the resubmission. In light of this profoundly sad context, we have included a dedication section at the beginning of the supplementary information to honor Laurence's significant contributions as a leading scientist in this study and as a cherished everyday colleague.

Second, my collaborators and I would like to sincerely thank you and the reviewers for your careful evaluation and valuable feedback on our manuscript. We have particularly appreciated the quality and constructiveness of the reviews, as well as the fact of having considered all the work already performed in the former round of reviewing.

The three new reviewers have provided valuable and diverse feedback, all of which has been extremely helpful. While we must first apologize for not being able to fully implement every suggestion, we would like to emphasize that we have made significant efforts to improve the manuscript by addressing as many points as possible and incorporating substantial revisions. We therefore hope that the changes will meet your expectations, as well as those of the three reviewers.

Once again, thank you for your time, patience, and insightful suggestions. We truly appreciate your support and look forward to your comments on the revised version.

Best regards,
Thibault Leroy, on behalf of all the authors

Revision round #1

Decision for round #1 : Revision needed

Decision on Leroy et al. manuscript

The manuscript by Leroy and collaborators details the evolution of rose breeding in Europe using a newly generated (and impressive) resource consisting of genotyping and whole genome resequencing data from accessions (>200 and 32 for genotyping and WGS data, respectively) representative of varieties cultivated in the 19th century, the period of varietal expansion. The authors illustrate a genetic transformation in modern roses and their ancestors following interbreeding between ancient European and Asian gene pools. They identify specific genomic regions under artificial selection, in particular a significant region harboring a recognized gene associated with repeated flowering. In addition, through extensive SNP array analysis and multi-year phenotyping, the authors compile a comprehensive GWAS dataset encompassing diverse traits such as disease resistance and floral scent components. Overall, I enjoyed reading this manuscript and appreciated the efforts of the authors to carefully address all the issues raised in a first round of review for another journal.

The manuscript was reviewed by three other colleagues, all of whom were positive. Nevertheless, a number of relevant suggestions were made, which in my opinion should help to further improve the manuscript. As an aside, I might also suggest the authors to perform a random allele PCA (see e.g. section 7.1 of the poolfstat R pcakage vignette: https://cran.r-project.org/web/packages/poolfstat/vignettes/vignette.pdf or pca_mds program of the ANGSD package: https://www.popgen.dk/angsd/index.php/PCA_MDS), possibly on all SNPs (i.e. not sure if LD pruning is really needed), in particular to ensure that the assumption of tetraploidy for all individuals (including the coding diploids as tetraploids) does not affect the PCA results presented in Figure 2.

I would be, of course, happy to consider a revised version of the manuscript for possible recommendation.

Dear PCI recommender, dear Mathieu Gautier, thank you very much for your interest regarding our manuscript.

We have followed your suggestion to directly perform a PCA on allele counts. Concretely, we generated a matrix of counts from the AD field of the vcf with an in-house script and then generated a treemix-formatted input file from a set of 65 million biallelic SNPs and then subsequently imported the data in poolfstat thanks to the genotreemix2countdata. Contrasting our former PCA and the new PCA based on read counts (Figure 1 (letter) below), we observed results that are highly consistent, regarding both the relative position of

the individuals on the different axes, as well as regarding the explained variance. This is now included in the results of a new SI Note (Supplementary Note 2).

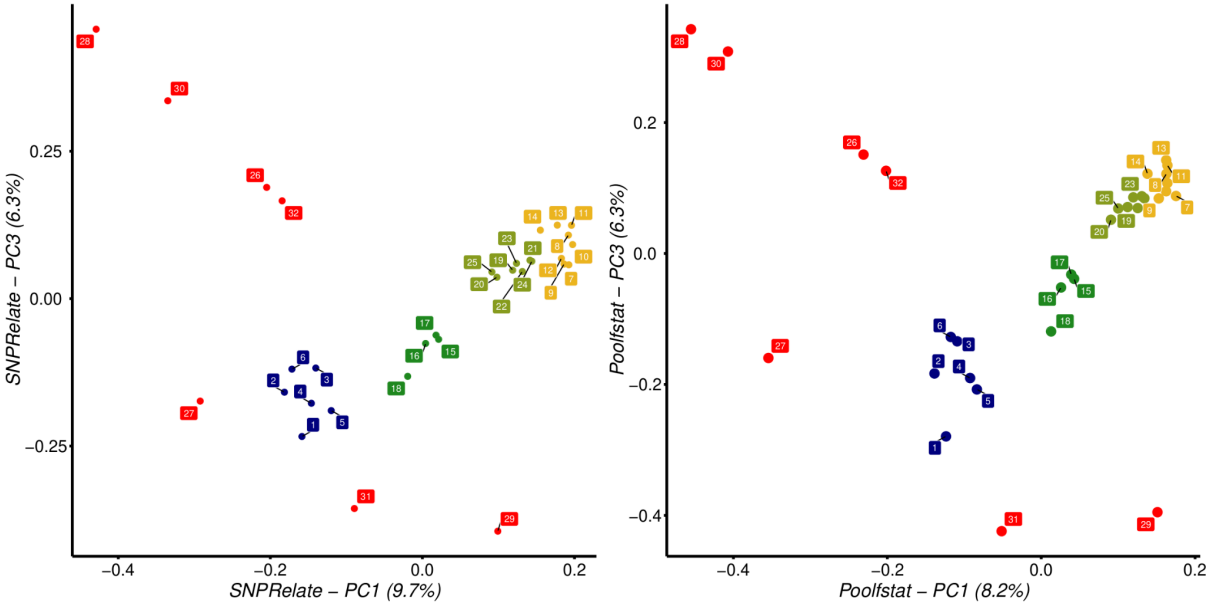


Figure 1 (letter): Comparison of the results of PCA assuming the diploid calls (Fig. 2B) or directly based on the allele counts (random allele PCA from poolfstat) for the first and third components.

Regarding the use of LD pruning, there are pros and cons. Our initial dataset used for population structure was based on a random sampling of SNPs, but without LD-pruning. Based on a previous round of reviewing (outside PCI), LD-pruning was asked. Regarding the benefits, it remains true that SNP pruning reduces the impacts of linkage disequilibrium avoiding the fact that linked SNPs overly influence too much the genome-wide population structuration and in practice, it makes analysis faster and less memory-intensive (which remains a challenge, since the poolfstat on 65 million SNPs required ~ 200Go of RAM to run). While it remains manageable to use tens of million SNPs for PCA, but for the use of other methods (e.g. Bayesian clustering methods), it is still a relevant strategy. We acknowledge that SNP pruning may sometimes mask subtle genetic structure due to uneven marker sampling across the genome. To maintain consistency with our manuscript's version and history of developments, we have decided to keep the LD-pruned dataset for population structure analysis (PCA and fastStructure). However, we also provide complementary PCA results based on read counts from the list of 10 million SNPs, clarifying the fact that both 1) the diploid calls for the first round of genotyping and 2) the limited number of SNPs used after SNPs pruning, do not substantially bias our results.

Review by Pierre Nouhaud, 21 Jun 2024 14:33

In their preprint, Leroy et al. document the history of rose breeding in Europe, showing the impact of recent domestication on both phenotypes of interest and patterns of genetic variation in rose accessions. Authors show in modern roses and their ancestors a shift in the genetic composition after admixture between the two ancient European and Asian gene pools. They also pinpoint candidate genomic regions under artifical selection, including one large region harbouring a known candidate gene associated with recurrent blooming. Finally, combining SNP array data and phenotyping over multiple years, authors provide a large GWAS catalogue for many traits of interest, including disease susceptibility and floral scent components.
Overall, I enjoyed the paper and found it carefully written, concise, and richly illustrated, and I congratulate authors for the impressive amount of work carried. I also appreciated how they rightly addressed some limitations of their GWAS approach in the discussion.

Dear Reviewer 1, dear Dr Pierre Nouhaud, thank you very much for your feedback, as well as your nice summary of our work and your positive assessment regarding our manuscript.

My main criticism is regarding clarity in some places, and I believe it stems partly from the shift between results first in the original version available on biorxiv, and methods first in this second version I reviewed. For instance, some parts of the results read like discussion and might be moved either there, or in the introduction (authors sometimes point to prior assumptions, eg L585 or L682-685, without mentioning them earlier in their manuscript). In the introduction, precisely, the section on rose domestication L103-127 is quite short and could be expanded to provide readers with a state of the art regarding rose domestication. Some of my minor comments below aim at addressing this general (light) issue at specific locations, and I hope they will help authors improving the overall clarity of their manuscript.

We agree that we quite intensively reshaped the ms before to send it for evaluation by *PCI Evol Biol*, which has probably contributed to this feeling. In this new version we have made some efforts in order to not provide too many interpretations in our results section, keeping only interpretations that are absolutely needed to understand the rationale of the analyses that are immediately downstream. In addition, we have included more information about rose domestication in the introduction.

Minor comments:
L175-184: Consider giving more details on the sampling in relation with the overall genetic structure depicted in the introduction L103-127. From Fig. 2B it appears sequenced samples span the major lineages, and mentioning it already here would ease understanding later on.

Our sampling section of the Materials and Methods now starts by a paragraph indicating that we integrate ploidy as well as the different groups in our sampling (see l. 178-182).

L330: From Fig. S2 (where colors are missing, by the way), it seems several samples peak between 0.1 and 0.2, but such values are not discussed in the main text. Results from Table S2 indicate hexaploidy was not considered for these samples while if I understood correctly the analysis, it would fall within this range (hexaploidy is also indicated for several accessions in the zenodo repository linked L330). Could authors elaborate a bit more on these peaks?

Thank you very much for this important point.
First, we have fixed the issue with the color code, thank you for the suggestion. We have now made a Fig. S2 (SI p. 15) that also follows the same color code as the rest of our figures.
Second, we have selected rose accessions for which we expected to have a ploidy level lower or equal to 4, which was partly based on this work on microsatellite data, in order to not fall in this issue. To be even more precise, we tried to have almost exclusively individuals with a ploidy level of 2 or 4 during our sampling campaign. The only triploid in our SNP data is La France, a genotype for which data was already available through literature and for which we decided to not consider among the four groups (2nd round of calling). Importantly, it should also be noted that the results shown in table S2 from Liorzou *et al.* 2016 are based on microsatellite data, for which ploidy estimation remains challenging (e.g. stutter bands), this is also why we decided to not only consider this specific layer of information.

Third, it is crucial to understand that the coverage is highly variable among the different individuals and that this has a huge impact on the accuracy of our inferences. To provide a more precise view on the variation of the depth of coverage, we summarized the values at 5% of all SNP positions for the 32 individuals. The coverage greatly varies, from 5.2 (20_Victor_verdier) to 44.5 (04_Rosa_gallica_officinalis). Of course, we were fully aware of this limitation and therefore we decided to estimate allelic balance for the minor allele only for SNP that has a minimum coverage of 20, but this provides two limitations: i) the number of SNPs fitfulling a minimum coverage of 20 and use to draw the distributions of allelic balance are not at all of the same degree of magnitude and ii) the accuracy of the precision of the allelic balance is not at all the same depending on the coverage (see: https://rosegwasbrowser.github.io/PDF/Extended_FigS2_ploidy_chr_per_chr.pdf). Typically “04_Rosa_gallica_officinalis” (mean_cov=44.5) and “19_Yellow_island” (mean_cov=33.9) exhibit near perfect tetraploid profiles (see Fig. 2 (letter) and https://rosegwasbrowser.github.io/PDF/Extended_FigS2_ploidy_chr_per_chr.pdf for details), while other tetraploid individuals exhibit much lower genome-wide coverage (all tetraploids: 5.2 - 27.3, mean=10.6, non-Botanical tetraploid accessions: 5.2 - 15.7, mean= 8.3), penalizing the quality of the local model fitting (loess). Nevertheless, for all except two individuals, the most frequently observed allelic balance value among all values corresponds to 0.25, 0.33 or 0.5 (see github: ./popgenomics/ploidy/script_infer_ploidy_from_data.R and https://rosegwasbrowser.github.io/PDF/Extended_FigS2_ploidy_chr_per_chr.pdf for details), consistent with expectations for tetraploids, triploids and diploids, respectively. The only two exceptions are Victor Verdier and Lady Waterlow (0.2), but it doesn’t mean neither that these two individuals are pentaploids, since these two individuals are also the two with the lowest number of SNPs fitfulling the coverage >20 for which allelic balance was used (see the associated R script), contributing to an even more degraded context for ploidy inference for

these two individuals (see Fig. 2 (letter)). We therefore consider that these two individuals are also more likely to be tetraploids. This would also be consistent with the results of Mathilde Liorzou and collaborators, since they reported these two accessions as tetraploids based on their microsatellite data.

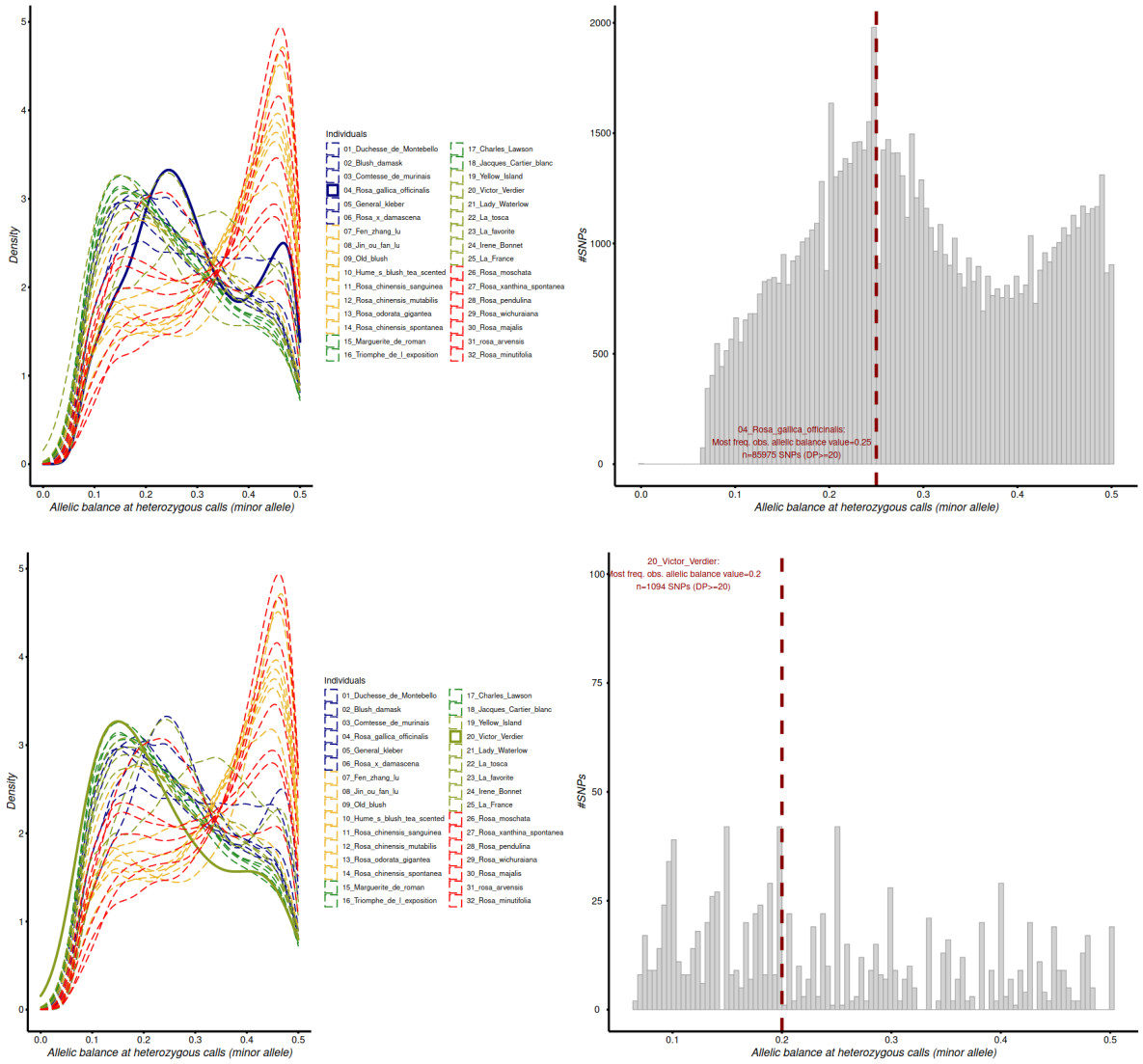


Fig. 2 (letter): DIfference of accuracy in identifying the ploidy level depending on the depth of coverage. This figure highlights the most contrasted individuals in the dataset, with Rosa gallica officinalis (top, mean_cov=44.5) and Victor Verdier (bottom, mean_cov=5.2). Most frequently observed allelic balance corresponds to the exact allelic balance among the SNPs with DP>20. Deviations from 0.25 for individuals with low coverage should not be overinterpreted given the stochasticity in the distributions. For all individuals, see https://rosegwasbrowser.github.io/PDF/Extended_FigS2_ploidy_chr_per_chr.pdf.

Of course, ploidy could be further evaluated in the future with deeper sequencing or with dedicated methodology (e.g. flow cytometry). This is especially the case for some botanicals for which the inference remains unclear in our analysis (as typically observed for *Rosa arvalis*, with a relatively tetraploid pattern, but with a mode at 0.5, see https://rosegwasbrowser.github.io/PDF/Extended_FigS2_ploidy_chr_per_chr.pdf). Of course, this is not an issue in our study since botanicals are then ignored in the rest of our analysis, but it remains important information to indicate for future work. It should be noted that polyploidy in roses is extremely complex (e.g. possibility for segmental allopolyploidy, see Koning-Boucoiran et al. 2012 Theor Appl Genet; Bourke et al. 2017 The Plant Journal; Cheng et al. 2024 Horticultural Plant Journal), which further complicates the story. We have tried to make this information more explicit in the text, as well as providing this information in Table S2 regarding the coverage of all individuals (see also below). All readers should be aware of the limits of this inference of the ploidy level and of the bias associated with low coverage data, in order to not (over)interpret the shifted distributions as evidence for higher ploidy level than 4. As a consequence, a part of this answer has been used to provide more information to the reader in a dedicated section (Supplementary Note 2).

L394: I don't fully grasp the rationale of the RoD analysis strategy, which may be linked to my misunderstanding of the domestication history. My problem is two-fold:
1/ early European x Asian roses are mentioned for the first time L392, and their relationship to the hybrid tea rose group is unclear at this stage of the manuscript (eg ancient diverged lineage? Proxies for the ancestors to hybrid tea roses?).
Early European x Asian are expected to correspond to the first generations of intercrossing between European and Asian. Such new varieties were then subsequently backcrossed with Asian genotypes to generate the first hybrid tea roses. Note that there are many lacks of knowledge in the description of this history (see Sup Note S3) and our present study contributes to providing fundamental knowledge about this history, in particular regarding the fact that the first "European x Asian" roses were indeed found to be the ancestors of hybrid tea roses. Given we focused on hybrid tea roses that predate 1910, the hybrid tea roses used in this study are derived from a very limited number of generations. Our result is consistent with such a limited number of generations of breeding.
2/ I don't see how the RoD index is informative as it is currently done, since from L103-127 hybrid tea roses result from admixture events between ancient Asian and ancient European, and I would assume such admixture events might have an impact on the range of possible RoD values. For instance, Fig. 2D shows that pi is smaller in ancient Asian compared to ancient European. As these values were also used to detect candidate sweep regions (L427, but see my next comment), this requires further justification (possibly some hints given L631-634 could be moved to the introduction?). Why not using weighted ancient Asian and ancient European pi averages, as it was done later in the section "Genetic diversity erosion"?
Empirically estimating reduction of diversity over the last decades is crucial to estimate the anthropogenic damages on biodiversity. So the answer is probably two-fold.

First, we would like to know how the level of genetic diversity in varieties has evolved in Europe or in Asia. In Europe, ancient roses had higher diversity compared to hybrid tea roses, which means that a complete turnover of rose varieties in Europe would have led to a

local reduction of diversity of 27.5%. Of course, this worrying estimate should be considered in the context of roses, for which it is still possible to maintain ancient roses in rose gardens through grafting. Our objective is to raise awareness on the importance of maintaining such collections and therefore to highlight the situation of ancient rose gardens, which are threatened by many factors, including climate change among other anthropogenic deleterious impacts. Note that we have also considered the potential situation in Asia, with potentially the hypothesis of a genetic gain in Asia through the exportation of hybrid tea roses from European roses breeding programs. Given that the genetic diversity of Ancient Asian was indeed low, Hybrid tea with around 25% of European background would have potentially contributed to a burst of diversity in Asia. Our empirical estimate is however not supporting this hypothesis, with a very marginal gain (~ 1%).

Second, we indeed estimate the average reduction of diversity as compared to the expectation assuming 1) the specific genetic make-up recovered from the diagnostic alleles, 2) the genetic diversity in ancient European roses, as well as in Asia. In hybrid tea roses, the estimate is 8.1%. This value is informative here about the genomic impact of breeding, meaning that we expect on average a reduction of 8% of the diversity throughout the genome. However, given that this breeding is unlikely to have occurred in a single generation (see the comment above and Sup Note S2), it is likely that some regions have been more massively impacted than the rest of the genome. This is why we then

To summarize, our ROD analysis allowed us to draw two different conclusions regarding the evolution of diversity in 1) different geographical contexts and 2) along align the genome.

I think both points can be answered by clarifying the rationale behind the RoD analysis carried with these admixed samples, especially as authors actually computed RoD between several lineages (see L690 and Fig. S11). To do so, it could be helpful to provide a tree displaying evolutionary relationships between the different accessions (eg a phylogenetic network based on genetic distances). I think moving L586-593 in the introduction would also help. Finally, on this topic of loss of diversity, an interesting addition might be to compute runs of homozygosity (eg with Plink) to contrast inbreeding regimes between ancient and modern accessions.

We have indeed tried to provide more information regarding the fact that the three summary statistics do not fully overlap (l. 716-717, see also our answer to the next comment). We also thank the reviewer for all the suggestions for the new analyses. It is important to note that phylogenetically-oriented investigations, as well as the use of ROH are already planned in the lab to be part of a new study that will be based on a larger WGS dataset. Given that our manuscript is already relatively ambitious with regards to the diversity of the methods we used, our objective is not to extend too much the content, in order to leave room for future investigations.

L427: In hybrid tea roses, putative sweep regions are identified as being simultaneously outliers in diversity, RoD index and Tajima's D, but this might be circular since diversity is

also used to build RoD (and low diversity regions should by definition display high RoD). Does removing diversity (or RoD) from this pipeline significantly change results?

Nucleotide diversity and RoD do not exactly capture the same information. For instance, regions can have low diversity just due to the impacts of low N_e regions (e.g. centromeric regions). The advantage of RoD is that it accounts for the change of nucleotide diversity between two groups (*i.e.* “two timepoints”), irrespective of the initial levels of nucleotide diversity (and therefore potentially for the long-term effects of linked selection). Even without considering this confounding factor of linked selection, it is also important to notice that breeding occurred in the ancient European and Asian gene pools prior to the 19th century. By using the RoD in complement, we tried to detect footprints of selection that have likely occurred during the 19th century. That having said, to more explicitly answer the comment of the reviewer, we compared the number of windows detected based on each summary statistics and visualized the results with Venn diagrams in order to identify the overlaps between the three methods (Figure 3 (letter) & Fig. S14 of the SI). We see that each method has its own interest in the detection.

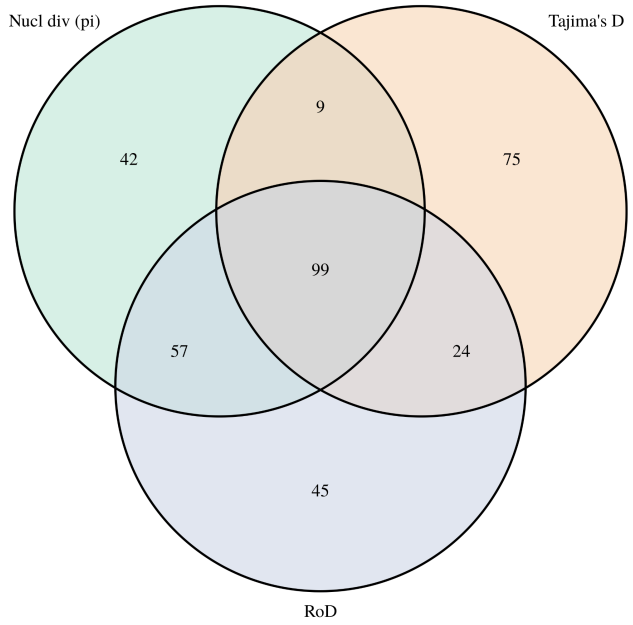


Figure 3 (letter): Venn diagram of the number of windows detected with the different parameters considering the less stringent criteria (top 5%).

L575-476: As is, I find this result puzzling and while I trust authors' expertise, I would like to know more. Is the explanation that "a large proportion of the diversity is indeed only present in the botanical accessions" (L576) the most parcimonious? Is there any evidence regarding accessions #27 & #29 (outliers in Fig. S8) that would corroborate this hypothesis (eg their origin, or some phenotypic trait values)? Alternatively, was there any issue (eg contamination) with genomic data from these samples?

Botanical roses regroup species that have extremely varied phenotypes and most of them are generally named as different species from the *Rosa* genus. Phenotypically, they are very far from the popular perception of what a rose is expected to look like! Genetically, the

percentage of variance on the axes is unlikely to be overestimated, but very likely to be underestimated. This especially the case for the SNP array data since the array was designed on modern roses, underestimating the unique diversity of botanical roses. To some extent, the WGS data provide more direct access to their diversity, even if a part of this diversity is probably hidden by the difficult balance between the exhaustiveness and the accuracy in the SNP detection during the SNP filtering (private polymorphisms). Future dedicated work on the genomics of botanical roses are clearly needed.

Of note, percentages of variance, sample coordinates in PC space and sample IDs displayed along with mapping rates in Fig. S3 don't match with those of Fig. 2B, making any sanity check impossible. Additionally, consider adding sequencing depth information for each accession in Table S2.
Thank you very much for having spotted this mistake between our different bioRxiv versions. We have now fixed this issue (see Fig. S3, SI p. 16). Table S2 (SI, p. 33-37) now also includes information about the coverage based on the average coverage at 5% of the SNP positions.

roseGWASbrowser.github.io: This looks like an great resource. To ensure data reuse, providing SNP association metrics for each analysis would be helpful. GitHub storage space may be a limiting factor, however these tables could be stored on another repository (eg zenodo or data.inrae.fr) with links provided on GitHub. Minor point, circular PDF plots are not displayed on Safari (sorry), but it works fine with Chrome and Firefox.

Thank you for this suggestion. We initially chose not to multiply repositories, instead favoring a single GitHub repository for all results. Following your suggestion, we have decided to extend our GitHub repository which now includes all the results, including the raw results of the GWAS and the qqplot (see also comments form the reviewer #3). Importantly, we a github release was then created and archived on Zenodo. Even if we did our best to follow the architecture of the website, it is important to notice that it represents a lot of information and a very large number of files. All readers interested in our results are invited to contact us, especially Thibault Leroy (thibault.leroy@inrae.fr), in case of difficulties to identify the correct file. We have also made explicit the use of Chrome/Chromium and Firefox (l. 327-329).

Some small remarks / typos below:

L206: As a naive reader, does petal number vary for a given accession?
The number of petals is near-perfectly consistent in single flower roses (5 petals), but then the variance increases with the number of petals and can be substantial for roses with a very large number of petals (>100). That’s why, the number of petals was at least estimated on

two flowers. Environmental conditions, such as temperature, can indeed modify the number of petals for a variety.

L242: Consider defining briefly DQC.
We now explicitly indicate that the DQC corresponds to the Dish quality control, the recommended QC metric for the Axiom SNP array and that the DQC captures the extent to which the distribution of signal values is separated from background values, with 0 indicating no separation and 1 indicating perfect separation (l. 260-263). The default threshold is 0.82, filtering out samples lower than or equal to this value. In our study, use a more stringent threshold value (0.85, l. 265).
L339: Missing "s" (17,669 SNPs)
Edited
L405-408: While this is results material, it would be informative to already give the number of genotypes per group here.
Edited
L409: Consider rephrasing: "a reference allele frequency of 0 in one group, and 1 in the other group".
Changed to “We then subsampled SNPs exhibiting diagnostic alleles, defined as those for which the reference allele frequency is 0 in one group and 1 in the other.” (l.447-449)
L419: "170,637 diagnostic SNPs with our sampling" may be more correct?
Probably yes, but to further improve the clarity, we have preferred an even more explicit sentence by excluding the use of the word “sampling”. We proposed the following alternative: “In total, 170,637 diagnostic SNPs were identified among the 54,481,222 SNPs (0.31%).” (l. 621-622)
L477-497: Some parts may be moved in the discussion, but I don't have any strong feeling on the matter.
L615: Should it read "containing up to"?
Sorry for the typo. The four groups have exactly the same number of chromosome sets (16), corresponding to 4 to 7 individuals depending on the ploidy of the individuals included in each group. Here the sentence was changed as follows: “each containing the same number of complete chromosome sets (16) after considering the ploidy…”(l.641-644)
L747: Split "impactfuldiseases".
Edited
L749: "varies".
Edited
L900: Split "isespecially".
Edited
Fig. 1A: Consider justifying in the legend the choice of time intervals on the y-axis. Are they purely illustrative, based on some historical facts, or to split accessions and traits evenly in time?
To ensure a sufficiently large number of varieties per group, we grouped the varieties in 20-year intervals starting from 1910 (1890-1909, then 1870-1889, and so on). For the final group, beginning in 1829, we could have made the choice to end this group in 1810. However, we also chose to include the few varieties developed between 1800 and 1810 since our objective was to cover the 19th century as a whole. That having said, it is also true that some of the periods were also considered because they were expected to capture some

expected changes among breeders based on literature review efforts (e.g. 1850, 1870). For instance, 'La France', a rose variety which is considered as the first Hybrid tea rose, was registered in 1867 (in France), which could be consistent with the opening of a new period from 1870. This is also a part of the answer. However, this should not be considered too strictly since breeding in roses is all except directional (as compared to wheat or highly domesticated species). Rose breeders were interested in different aesthetical characters, explaining why the variance of trait values tends to increase for the last periods (e.g. petal counts, recurrent flowering etc).

Data availability section: mention that the raw array data was made available on GitHub, and consider archiving the GitHub repo (eg https://docs.github.com/en/repositories/archiving-a-github-repository/referencing-and-citing-content#issuing-a-persistent-identifier-for-your-repository-with-zenodo).

Thank you. All our scripts and GWAS are now archived in a Zenodo repository: https://doi.org/10.5281/zenodo.14450241

Review by anonymous reviewer 1, 08 Jun 2024 12:41

In their work on the breeding history of roses, the authors have built up a very nice resource that makes it possible to study rose breeding over time. Indeed, a large number of samples have been sequenced and phenotyped for a wide variety of traits. For a non-rose specialist like myself, the study therefore seems original, with a number of different phenotypes (recurrent blowing, scent components). The species is also interesting because it is a domesticated species, but selected for non-agronomic reasons, with a different mode of reproduction and with a more recent history. All this gives originality to the study and makes it pleasant to read.

Despite this nice resource, the description of the GWAS results is quite short and not very informative. The authors claim that this resource can make breeding more effective, but it is not clear how? Similarly, there are few results on selection signatures. The high levels of relatedness and 'low' numbers of individuals might reduce the resolution and power of these analyses?

We sincerely thank the reviewer for the interest and positive feedback. Regarding the relative emphasis on GWAS and the detection of selection footprints, we aimed to maintain a balanced manuscript in terms of length across sections. However, it is true that certain sections, such as the one on GWAS, required more intensive work. Our approach was to avoid overloading the main text with details regarding the associations for different traits, in order to make it accessible to a broader audience. Consequently, we have made the choice to provide a case study example in the main text, while directing readers to additional results for other traits of interest. We have tried to provide a bit more information in our new version. But even more importantly, to address Reviewer #2's suggestion along with those from other reviewers, we have decided to make all GWAS results accessible in our GitHub repository (which is archived in a Zenodo repository available at https://doi.org/10.5281/zenodo.14450241). This significantly expands upon the content of the

previous version, enhancing transparency and completeness for further exploration, which will more efficiently contribute to make breeding more effective.

On the topic of relatedness and the relatively small sample size, we completely agree. This can be seen as a limitation stemming from the restricted number of breeding selections, given that selection occurred over only a few generations within a narrow genetic pool (in both European and Asian roses). This has led to (i) a limited number of varieties and (ii) relatively high linkage disequilibrium due to the small number of recombination generations. These factors hinder the detection of associated regions, which is why we employed a combination of strategies from both population and quantitative genetics. Given the degree of relatedness due to recent selection, increasing the sample size would likely yield limited benefits, as the inherent challenge lies in the species' recent domestication and the remaining predominance of the signal associated with genetic structure.

An original aspect of the study is that there is variable ploidy among the samples. At least some of the analyses take this into account, but I still wonder whether ploidy is correctly taken into account in all analyses (I'm not an expert in variable ploidy data analysis myself). When ploidy is taken into account, the authors simply refer to some tools without clearly explaining how ploidy is taken into account. In some parts, the tools seem clearly adapted for different ploidy levels, but in other analyses it remains less obvious (e.g. population structure, relatedness, signatures of selection).

Thank you for this comment. It is important to note that we did not aim to focus our work on polyploidy, but rather that the variation in ploidy is to that point inherent to the rose model, with ancient European varieties being almost exclusively tetraploid and Asian varieties almost exclusively diploid. As a result, any genetic analysis of the varieties requires accounting for this variation as much as possible. In this new version, we have made considerable additional effort to make more explicit the way we performed our investigation (see https://rosegwasbrowser.github.io/PDF/Extended_FigS2_ploidy_chr_per_chr.pdf and comments to reviewer #1). It is important to note that in many respects, our article addresses ploidy variations more explicitly than many other studies on this model (e.g. ploidy-explicit sequence reconstruction for nucleotide diversity); however, this task remains challenging and demands a certain humility, as working with species exhibiting variable ploidy levels is highly complex. We have written a new Supplementary Note dedicated to this topic (see Supplementary Note S2). We have attempted to bring more clarity to potential limitations and hope that the remaining points of concern not covered in this article will be addressed at the community level in future investigations, with more detailed analyses. Our sequencing data is also publicly available to reach this long-term objective.

The title sounds nice, but I did not see a direct link to the results of the paper.

We are unsure to correctly understand the reviewer #2's concern here, since our work focuses on the reconstruction of breeding during the 19th century thanks to large-scale genomic data, as indicated in our title. The start of the title "dark side of the honeymoon" also suggests some (neglected) potential drawbacks (e.g. evolution of susceptibility to blackspot, reduction of diversity, etc).

Why is only a subset of SNPs selected in the population structure analysis when you have full sequence data? The number of final SNPs (17,669) seems extremely small compared to the total number of SNPs available.

The rationale lies in the ability of a handful of genetic markers to provide a representative view of the levels of population structure. We could have basically recovered a similar story with far fewer SNPs, or even three dozens of microsatellites (Liorzou et al. 2016). Our objective was to confirm the previously detected population structure, in order to show that our data is consistent with the literature, allowing us to then perform the rest of the investigations. It should be also noted that the use of a pruned set of SNPs was a request from a former reviewer (this former response to reviewers is available at https://rosegwasbrowser.github.io/PDF/Leroy_PCIEvolBIol_letter.pdf). For this new version, we have followed the suggestion of our PCI recommender to use allele counts instead of calls and to use more SNPs to complement our first PCA analysis. This PCA was performed based on 65 million SNPs and is highly consistent with the one provided on the main text (see Supplementary Note 2).

Is considering individuals with n=2 as n=4 correct for PCA and faststructure?

Thank you for this question. It is important to note that the main advantage of PCA, and its derived approaches, e.g. DAPC, is associated with its independence on a specific population genetics model, making it free from assumptions regarding Hardy-Weinberg equilibrium or linkage disequilibrium. Consequently, PCA is known to be a valuable tool, regardless of their ploidy and rate of genetic recombination (e.g. see Jombart et al. 2010 BMC Genomics). This is a potential difference with methods based on a Bayesian framework such as faststructure. This result explains why we have decided to highlight the results of PCA in the ms (Figs. 2A for the SNP array but also 2B for the WGS data) and shows the results of the model-based methods as supplementary figures (Figs. S7 and S9 for SNP array and WGS, respectively). Follorwing, the suggestion of the editor to use allele counts rather than called genotypes, we have indeed confirmed the robustness of our analyses to this variation of ploidy (see Supplementary Note 2, SI, p. 9).

KING is designed for diploid individuals and not for different ploidy levels. Some assumptions of the method may be violated here. In addition, the rules for declaring individuals as 1st, 2nd, 3rd degree are relatively simple and may differ according to the structure of the population (they were first designed for humans). Therefore, I'm not sure if the KING analysis is completely correct (definition of relatedness for samples with different ploidy levels) and if the classification (in 1st, 2nd degree) is still 100% accurate?

We acknowledge the limitation, and agree that the classification should be approached with caution. We have made this point more explicit in a dedicated supplementary Note (Supplementary Note 2, SI p. 9-10). As introduced in this new SI Note, it is important to note that we empirically observed high kinship in both ancient Asian (almost exclusively diploid) and ancient European (tetraploid) roses. Notably, all but one of the identified ‘potential first-generation’ relationships corresponded to tetraploids. This suggests that any bias present would likely lead to an overestimation of relatedness in tetraploids. Given that our aim was to exclude closely related individuals as much as possible for the rest of the analysis, this would indicate an even more conservative strategy applied to the

tetraploid samples compared to the diploid ones and therefore even more robust subsequent analyses.

For clones, you use a threshold of 0.354. If the individuals are identical the kinship should be 0.5 (only new mutations would create differences).
Thank you for this comment. 0.5 is indeed a correct theoretical expectation for clones, 0.25 for first generation. The cutoff could therefore typically be expected to be near 0.375. But the methods implemented in KING accounts for real-world variation, including genotyping errors. The values we used are the ones recommended by the authors of KING to account for these real-world deviations from theoretical expectations (See KING's manual for instance https://www.kingrelatedness.com/manual.shtml). Part of the deviation has been first documented in human data as indicated by the reviewer in his/her previous comment, however it is important to note that deviations are commonly observed in other animals and plants, explaining why the same threshold values are commonly used.

Base Quality Score Recalibration. If there is no set of highly reliable SNPs, it is recommended to use this tool iteratively, improving your high quality set at each iteration. Also, what have you done for indels?
We have indeed used the tool iteratively, by using the filtered set of SNPs of the first round of calling (32 accessions), which was the recommended strategy by GATK, at least at the time of the study (see l. 349-352 and then l. 405-413). Both SNPs and indels were called in GATK, however indels were subsequently ignored in our study to allow the rest of our investigations (keeping perfectly aligned sequences = same genomic coordinates). Ignoring called indels has further contributed to improve the quality of our final list of variants, since correct indel calling is known to be more challenging with GATK, as well as with many variant callers.

For estimating nucleotide diversity, the number of chromosomes per group is standardised. However, I wonder if using one tetraploid versus two diploids is really equivalent for estimating diversity? It might depend on the timing of the 'duplication' event (when were the extra copies acquired?).
When a species inherited sets of chromosomes from different species and is at an advanced evolutionary stage of allopolyploidy (typically wheats), the genome is "stabilized" and subgenomes should be indeed analyzed independently to investigate nucleotide diversity patterns (e.g. Pont, Leroy et al. 2019 Nature Genetics). Here, given that rose breeding can be performed between varieties of different ploidy levels, such a situation seems extremely unlikely. Such an evolutionary history would have led to specific patterns in our ploidy investigation. However, we do agree that polyploidy is complex in roses, with reported segmental allopolyploidy (Koning-Boucoiran et al. 2012 Theor Appl Genet; Bourke et al. 2017 The Plant Journal; Cheng et al. 2024 Horticultural Plant Journal) and that therefore fundamental studies are still needed to know how rose chromosomes segregate during meiosis. Here, we have made our best effort to account for the variable ploidy as much as possible in our analysis, from the sampling (focusing on diploid and tetraploid roses) to the final analyses (e.g. ploidy-aware demographic inferences). We however made

more explicit the fact that this fundamental work on polyploidy in roses is needed in our new SI Note (Supplementary Note S2, SI, p. 10).

On line 401 - 'Diagnostic alleles': perhaps you could start by explaining what these diagnostic alleles are and why you need them. This only became clear later.
Thank you for this suggestion. The rationale of the use of diagnostic alleles is now clearly indicated in both the Materials and Methods (l. 438-440) and Results sections (l. 619-621).

GWAS: why did you not fit a traditional genomic relationship matrix to account for population structure?
We are not sure to have fully understood the comment from the reviewer here. Our GWAS were performed considering both a Q+K model, as proposed by GWASpoly, which is a traditional way of accounting for both population structure and kinship in the data. We have made this information explicit (l. 311-313).

Rose GWAS browser: some examples I tested were missing the legend in the middle.
Thank you for pointing this out. In fact, the situation is somewhat the reverse. For several key traits related to our research, we have manually added legends directly within the figures, complementing the general descriptions provided in the survival guide accessible on the homepage of the GWAS browser (https://rosegwasbrowser.github.io/help.html). Re-running all the GWAS simply to add legends would require significant effort and is outside the primary scope of our article. Therefore, we recommend that users of the GWAS browser begin by reviewing the survival guide, which also offers essential additional information.

Review by Vincent Segura, 02 Jul 2024 09:33

This manuscript aims at characterizing the genetic and genomic changes during rose breeding, focusing on the 19th century. The choice of this particular period is justified by the fact that it corresponds to a strong increase in the number of varieties.

The paper represents a huge work with the release of genotypic data corresponding to more than 50k SNPs on more than 200 varieties and full genome sequences (short reads) on 15 varieties. These full genome sequencing data were combined with publicly available data of 17 varieties for detailed population genomic analyses, including the search for artificial selection footprints. The genotypic data on the entire set of varieties was further combined with phenotypic data on key traits of interest to perform genome-wide association studies, providing a large catalog of associated SNPs.

The manuscript is well-written and easy to read.

The authors should further be complimented for their effort to release the data, scripts, and results. This last point is particularly illustrated by the dedicated website built to release the GWAS results, which is not usual in the plant community.

This manuscript has previously been reviewed by 3 reviewers for a journal. I have read the revised manuscript as well as the reviews and responses made by the authors. Overall I think that the authors have done a good job in responding to the issues raised by the 3 reviewers, which were quite focused on the degree of ploidy of the species.

Dear Reviewer 3, dear Dr. Vincent Segura, we thank you for the positive assessment and the important feedback. We also acknowledge you for mentioning the fact that the release of the GWAS results is not that usual in the plant community. This gives an important context about the fact that we made a first attempt here, even if some limitations remain because it would then require additional. Finally we especially appreciated the time you took to consider our former 20-page reply prepared for another journal, along with the manuscript version for PCI Evol Biol.

I am not a specialist in population genomics nor polyploid species, so here I will focus my review on the GWAS part on which I have several comments/suggestions, even if I understood that it was intentionally not presented in a very detailed manner in the manuscript because the scope of the work is broader with the genomic diversity part. I think that focusing GWAS results for a particular trait (resistance to blackspot disease) is fine with all the other results being available on the dedicated website. The choice of the trait is well justified.

This was exactly our attempt. People interested in roses can be interested in very different traits (e.g. fragrance, number of petals, prickles, color etc). We have decided here to focus on another trait that is probably a bit less attractive at first glance, but is expected to be increasingly important in roses. In the case of cut flowers, pesticide exposure has received less media attention, but this has changed recently, particularly in France, with the increasing awareness of the massive exposure of florists to pesticides, with a first florist being supported by the compensation fund for pesticide victims. It is highly likely that this will drive new societal demand regarding cut and garden roses in the coming years. Our role is therefore to highlight these emerging breeding targets as much as possible, which represent one of the key pathways for advancing practices towards more sustainable systems.

Even if this part (GWAS) is quite small in the results section, I think that if it is presented, more details should be provided in the method section regarding the statistical models used to test associations. This is particularly important because the work is on a polyploid species which is not so usual, consequently leading to several ways of modeling the effect of the SNPs. So I would suggest to provide more details on the model used.

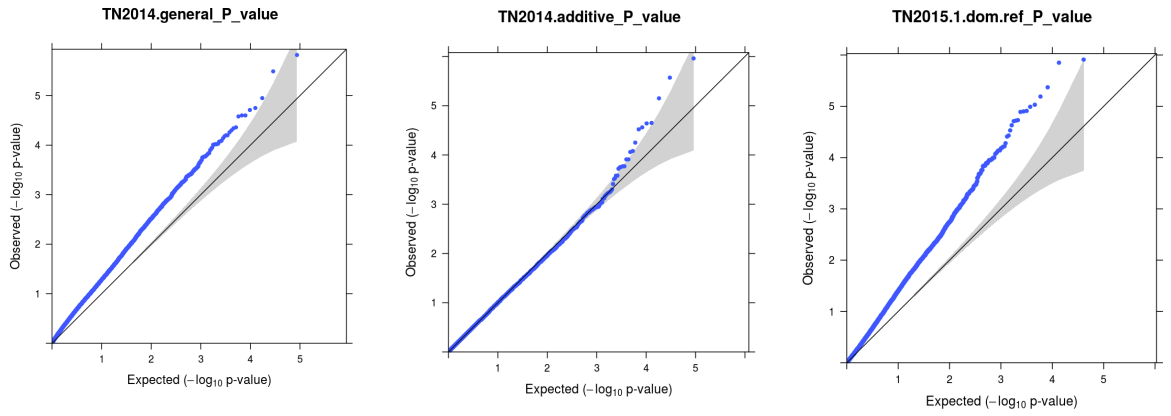
Thank you for this suggestion, we have tried to more explicitly introduce why there are different models that are implemented in GWASpoly regarding the various polyploid gene actions (l. 313-320). We hope that the information now provided in the Materials and Methods section, as well as in the results (l. 762-769), contributes to clarify this section. It is also important to note that additional information is available online in our survival guide (https://rosegwasbrowser.github.io/help.html).

Also, I think that some results are lacking, such as qqplots which have been generated according to the method section (or maybe I missed them?). Although there are a lot (number of traits times the number of models), they could inform the fit of the GWAS models, especially regarding the confounding usually attributed to population structure which seems, by the way, a concern for some phenotypes under study. I would recommend providing them as supplementary information.

The reviewer is correct. We generated QQplots, but until now, these plots were not available to readers. Based on this comment, we have decided to make all QQplots accessible through our Zenodo / GitHub repositories (./GWAS/ALL_GWAS_RESULTS/).

Notably, the results reported in the manuscript (Fig. 4) are based on the most general genetic model, allowing for specific effects across each genotype class (i.e., {AAAA, AAAB, AABB, ABBB, BBBB}). This model is the least restrictive regarding signal interpretation, while alternative models, such as the additive model, assume the SNP effect is proportional to the minor allele dosage, or in the simplex dominant model, that all heterozygotes are equivalent to one of the homozygotes. Consequently, QQ plots from the general model typically show the greatest inflation in p-values (see Fig. 4 below, which is now integrated to the ms as part of the Fig. S16).

More broadly, we recognize that our QQplots indicate some deviation from the expected distribution, supporting a degree of p-value inflation, which varies depending from one model to another (Fig. 4 and Fig. S16). This inflation likely results from the high level of linkage disequilibrium in our data due to the limited number of breeding generations, where neighboring SNPs tend to exhibit similar associations. While our QQplots may not fully meet ideal standards, we believe that our GWAS results nonetheless offer meaningful insights into associations with some traits. We have clarified this information throughout the main text.



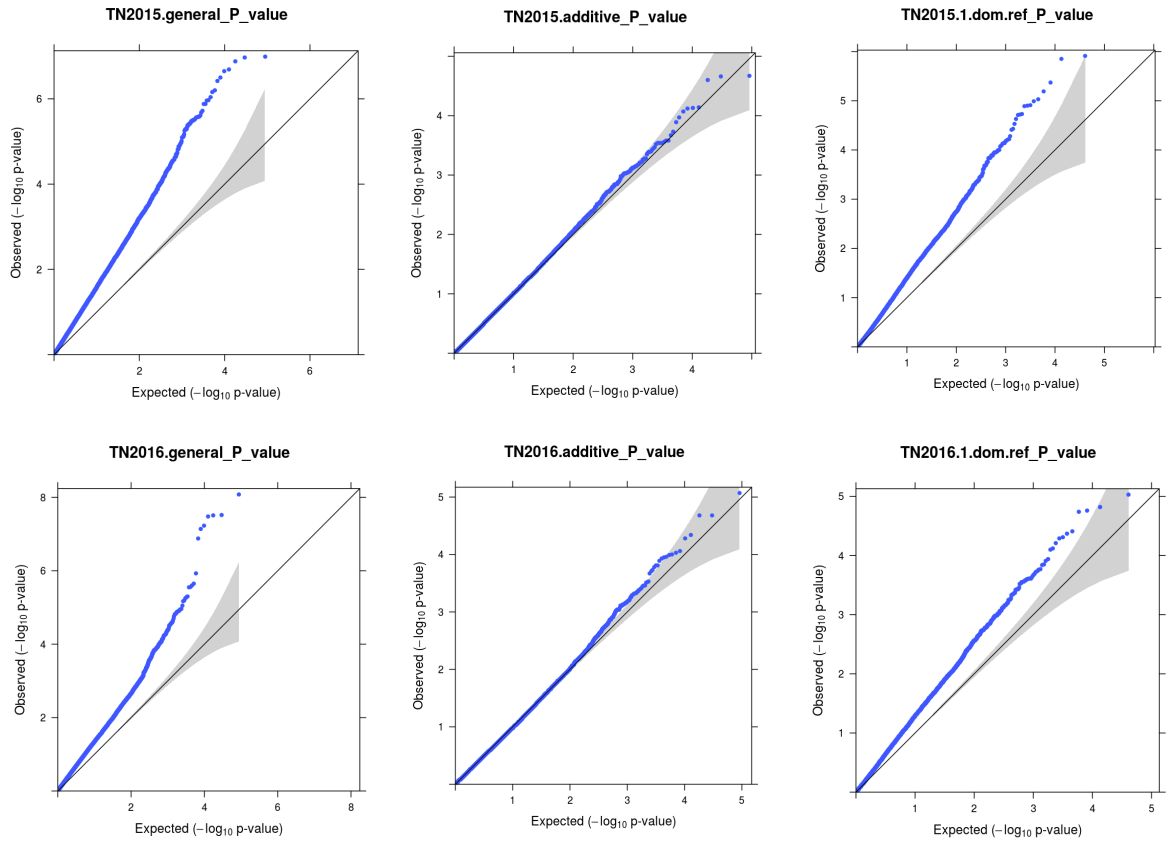


Fig. 4 (Letter): QQplots for the GWAS for the scoring of the black spot disease (3 years: 2014 (top), 2015 (middle), 2016 (bottom)), for 3 GWASpoly 3 models (columns), with general (left, as shown in Fig. 4), additive (center) and single dominant for the reference allele (right). Depending on the models, no (e.g. additive models) to substantial inflation of p-values can be observed (e.g. general). All the QQplots are available on our github repository.

I must also admit that I dislike the circular Manhattan plots which do not help appreciating if this usual problem has properly been handled by the model. But I also understood from the author's response to previous reviews that it is not possible to generate and make available the classic Manhattan plots for all the models times traits investigated. If they can instead provide the qqplots as a supplementary, I think it would be great. Because I am not a specialist in polyploid species, I am curious about the models' fit and I would like to check if there could be a particular pattern in the results according to the model (additive vs. dominant...).

In addition to the QQplots, we have also decided to release 1) all the GWAS results (p- and q-values; see also our reply to reviewer #1 Pierre Nouhaud), 2) all our non-circular Manhattan plots. However, given it represents months of work, we cannot make changes associated with these files or the website. We therefore continue to encourage all the readers to use circular Manhattan plots through the website, since the non-circular

Manhattan plots were initially generated to 1) only be working files (e.g. there are some limitations, e.g. no color code, positions on the x-axis are based on an index and not the exact genomic coordinates etc), 2) the information is less integrated than in our circular Manhattan plots (e.g. there is no information regarding the local densities in SNPs and in associated SNPs) and 3) given the large number of files available on GitHub, it seems easy to make confusions.

Outside our effort to provide far more information than before, it is also important to notice that there is another main limitation with roses, which is associated with the very limited number of generations of breeding, inducing large regions of high linkage disequilibrium. From our perspective, the fact that some QQplots are sometimes ugly should be also interpreted in that respect.

Other points:

- Regarding the traits under study, it seems that GWAS have been made for several architectural traits, but I did not find the description of these traits in the material and method section.

The traits associated with plant architecture were relatively simple, considering height and circumferences in meters, or binary traits evaluated in the field regarding the form of the plant (e.g. bushy or not). We acknowledge certain limitations with these traits, in particular regarding human management in the historical rose garden. Although we do not exclude that rose size may influence some of these architectural traits, we hypothesize that the overall architecture remains little affected.

- I did not like so much the first paragraph of the discussion (l767-780) which is a bit out of the scope.
Thank you for this suggestion. The first paragraph has been reduced and reshaped in order to keep it more focused on the present study. However, we still conserved the general idea of the crucial importance of investigating the evolution of diversity in the wild and domesticated. The paragraph now ends by the lack of knowledge in ornamentals, making more explicit the link between this paragraph and the rest of the discussion (see l. 798-808).

Typos

- l672: there are 2 « also » in the same sentence, maybe remove the second one

Edited

- l696: « target breeding » -> « breeding target » ?

Edited

- l747: a space is missing between « impactful » and « diseases »

Edited

- l900: a space is missing between «is» and « especially »

Edited