2.- In my opinion, the 'Data analysis' requires restructuring for better clarity and coherence. For instance, you could start indicating that all analysis were conducted in R. The current description of the models (L232-245) is somewhat disorganized and difficult to follow, lacking a clear explanation of the variables used. It is neither clear what you mean in L234-235 "that different versions of mixed models" were used for the traits measured in the pre- and post-pollination phase. Are you suggesting that the distribution followed by each trait varies?

We reworded “different versions of mixed models” to emphasize that each model for the individual traits had nuances (four traits were count based, one was binomially distributed, pollen size and pollen germination models differed in their structure).

I also have some concerns/doubts about the statistical model you implemented. You mention in L232 and L236 that 'region' is included as a fixed factor, yet multiple populations were sampled in each region. The model should account for this nested design.

We decided to omit population from models for three reasons. First, when using model selection, the AIC jumped substantially for all models when population was included. Second, multiple models provided warnings about multicollinearity or errors about overfitting the model when population was included. We decided to prioritize a simpler model with temporal blocking and ramet over including population. Lastly, we had reviewers from a previous submission question whether we could consider the three southern populations as separate populations. The three populations do have unique morphological characteristics (leaves) but were all within 2 Km. Thus, we decided to pool the populations for each region.

In L240-240, you described using a glmer with a Poisson distribution, but it is unclear whether overdispersion was tested.

We tested for overdispersion using the DHARMa package, which is now included in the data analysis section of the manuscript. The count data were not over dispersed.

Furthermore, the rationale behind the correlation analysis between the mean anther and mean style + stigma lengths is not justified. Why were these specific variables chosen for correlation and not others? This needs clarification. In Fig. 4, it appears that you are examining how mean anther length is affected by style+stigma length across treatments, but this is not explicitly explained in the text. Regarding L262-263, I wonder why correlation analysis do not including treatment as factor that could mediate these correlations, which can be relevant depending on the context of this analysis.

We decided to only examine the correlation between stigma+style and anther length, because changes in relative sizes of those structures could affect reproduction (herkogamy). We test correlations for the control and heat treatments separately.

Finally, the use of temporal blocks (L135) due to insufficient space in the growth chamber is a limitation of the study. It seems that the authors considered this factor in the analysis but this is not clear to me in this 'Data analysis' section. I consider important to include the variation in the response due to the experimental timing, as the storage time of the rhizomes can influence the effects.   
The temporal block was included in the data analysis for Experiment 1. We added a temporal block in Experiment 2. In Experiment 2, we grew all of ramet A and B plants, randomly assigned to treatments in January and then repeated the experiment with all C and D plants six months later. The temporal block (January vs June) was added to all models in Experiment 2 as a fixed effect. We did not include this temporal block as a random effect because there were only two factor levels and only one when plants from a genet did not flower.

3.- To prevent potential misunderstandings, I recommend rearranging a bit the presentation of results to reflect the hierarchical structure of the model. Specifically, it would be more logical to first discuss the effects of the two-way interactions. These interactions, which are currently under-explained, likely represent some of your most significant findings. It is crucial not only to report these interactions but also to clearly describe the direction of the effects. Posthoc contrasts are also missing in figure 4 showing the significant two-way interactions, and they would clearly help to understand the effects. I think these approaches will provide a clearer understanding of the data and reinforce the importance of these results. Please, see below some other minor points regarding this section.

We added a posthoc analysis with p-values for each comparison displayed in a table (Table 3) and a paragraph in the results describing those comparisons. We decided to maintain the format we had (describing region and treatment before interactions) in order start with the simpler fixed effects and end with the more complicated interactions.