Associate Editor (Comments for the Author (Required)):   
  
Your manuscript "Intraspecific variation in responses to extreme and moderate temperature stress in the wild species, Solanum carolinense (Solanaceae)" has now been thoroughly reviewed by two experts. Both reviewers found that the study questions and the results are interesting and may constitute a meaningful contribution. However, they also raise a significant amount of questions related to how the methods (experimental design, statistical analyses, etc.) are presented in the current version. Specifically, a major concern is how the two experiments are related and presented. The need to include the two experiments, how they answer different aspects of the research questions needs to be further justified and clarified. Furthermore, both reviewers agree that many important methodological aspects are currently missing, particularly (but not only) related to how the levels of the treatments were chosen, and how they reflect natural current or future conditions. Finally, clarification on the statistical analyses and proof-reading of the text to avoid mistakes were also pointed out by the reviewers. In sum, this study would require a substantial revision of the manuscript, particularly of the methods section.   
  
  
Referee #1 Evaluations:   
Recommendation: Major changes needed   
  
Referee #1 (Comments for the Author (Required)):   
  
In this manuscript, the authors explore the phenotypic response of the clonal herb Solanum carolinense to different temperature treatments. Genets of this plant were collected from both northern and southern regions of the USA, and thereafter were further divided into ramets to form experimental replicates. The study comprised two experiments: the first involved control and warm treatments in growth chambers, while the second exposed eight traits to control, warm, and extreme cold conditions, both aiming to assess their temperature tolerance. The results indicate that temperature treatments induce phenotypic variation, and that these variations are influenced by the plants' geographic origin. The study is sound and provides insights into plant responses to global warming. However, I have major concerns regarding the clarity and the statistical analysis of the study.   
  
Major comments   
1.- I believe that some parts of the Methods need to be clarified and justified more thoroughly. Currently, some parts are difficult to follow, particularly 'experiment 2,' which in my opinion requires rewriting for clearer comprehension. The rationale behind the temperature conditions used in both experiments needs explanation. Specifically, how were the experimental temperatures chosen, and how do they relate to the natural habitats' temperatures? The designation of experiment 1 as 'long-term' is not clear; I might have missed it, but there does not seem to be clear information about why the timing is a factor here. The rationale for exposing different traits to varying and extreme conditions is also unclear. What is the purpose of the extreme cold treatment in the second experiment, and how do the 'hot' and 'control' conditions differ from those in experiment 1? The experimental cold temperatures reach -18{degree sign}C, which is significantly more extreme than the lowest temperatures in the natural habitats of Minnesota and Houston, ranging from -9 to 6{degree sign}C. Furthermore, in lines 219-220, you mention that pollen plates are subjected to five temperature treatments ranging from 10{degree sign}C to 40{degree sign}C, which seems unrelated to the previously mentioned control, hot, and extreme cold treatments. I suggest reorganizing the experiment descriptions for better coherence and thoroughly justifying the treatment conditions and objectives of each experiment. It would be also helpful to create a schematic figure illustrating the experimental design. This maybe could be effectively presented as a panel alongside the map in Figure 1.   
  
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? 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. In L240-240, you described using a glmer with a Poisson distribution, but it is unclear whether overdispersion was tested. 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. 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.   
  
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.   
  
Minor comments:   
**ABSTRACT**   
L6 (and L9): The term 'long-term' should be omitted unless its use is justified in the methods section. Given that this study spans only one growing season, referring to it as 'long-term' seems inappropriate.

Addressed by omitting the term “long-term.”  
L7: I would include the country

Addressed by adding USA  
L9: Better to directly present that you conducted two experiments.

Addressed by changing from “two-part” to “two experiment”  
L12-13: I think this sentence is a bit repetitive.

Addressed by restructuring sentences  
L13 (and same in L14): Maybe change it to "Plants under moderate heat treatment produced...". Is this effect significant?

Addressed by changing the phrasing to that recommended by the reviewer  
L15-16: To which experimental temperature treatment are you referring to?

Addressed by adding “moderate”  
L19: drought-avoidance strategy? Which trait response allow you to conclude that? See some reference below that can help:

Bowles, A. M., Paps, J., & Bechtold, U. (2021). Evolutionary origins of drought tolerance in spermatophytes. Frontiers in Plant Science, 12, 655924.   
Levitt, J. (1980). Responses of plants to environmental stresses ( 2d ed.). Academic Press.   
Springer, K., Coquery, T., Holland, V., Fitze, P. S., Scheepens, J. F., & March‐Salas, M. (2023). Precipitation predictability drives evolution of drought strategies in the common poppy, Papaver rhoeas. Functional Ecology.

Addressed by changing the term to “heat-avoidance.” Evidence for heat-avoidance was in pollen germination. This could also be a strategy used in drought stress, but here we tested temperature and do not have data to conclude that this is a drought-avoidance strategy.   
L23: But it can specifically be due to plasticity, right? Ovule count, unfertilized ovule count and aborted seed count seems to indicate differences in plasticity.

Addressed by changing “evolutionary processes” to “local adaptation and phenotypic plasticity.”  
  
**INTRODUCTION**   
L27: Check the reference.

Addressed by excluding the second author from the in-text citation.  
L38-39: First, this needs a reference.

Addressed by adding references.

Second, it miss you present the next steps or unresolved questions in this research area, particularly how your study addresses these gaps or introduces novelty.

Addressed by adding a sentence identifying the gap in knowledge and two sentences describing how we address that gap.  
L39-43: For better coherence, consider moving this section to the next paragraph, probably after L44-45. But please think that the brief description of your study should go just before your study questions.

We moved this sentence to a position directly before a description of our study.  
L51: Clarify if you mean "negatively affects"?

Addressed by adding “negatively”  
L51-55: Since these details are elaborated on later, consider removing these sentences to avoid redundancy.

We decided to leave these examples in the manuscript to give the reader sufficient context for our claim.  
L63: For consistency, use 'moderate' instead of 'mild'.

Addressed by changing all reference of “mild” temperatures to “moderate.”  
L64-65: This is repetitive. Clarify what remains unknown in this field and how your work contributes new insights.

Addressed by adding a sentence identifying the gap in knowledge and a sentence describing how our experiment fills that gap.  
L72: Add a supporting reference here.

Addressed by adding three references.  
L87: The genus should be abbreviated here.

Addressed by abbreviating the genus.  
L88: It would be logical to mention the vegetative stage before the reproductive stage.

Addressed by changing the order of the two stages.  
L91: The hypothesis for the second objective is missing and should be stated.

Addressed by adding the hypothesis for the second objective.  
  
**METHODS**   
L99: I assume you wanted to refer to Figure 1.

Addressed by changing the figure reference.  
L106: I would specify the exact number of populations per region.

Addressed by adding the number of population for each region.  
Fig. 2: Consider moving this figure to supplementary material due to its limited relevance. Three points need be addressed: (1) Include the locations in Minnesota. (2) Maintain consistency in temperature units; if {degree sign}C is used in the text, it should be the same here. (3) Clarify if the growing season in Collin County is indeed 12 month

Figure moved to supplementary information. The figure was changed to Celsius and “growing season” was changed to “Temperature above 0C”.

L121: You collected or harvested the plants with rhizomes.

Addressed by changing collected to harvested.  
L122-123: Provide a reference that support that a distance of 1 m is sufficient to obtain independent genets. For instance, I know that other clonal plants require to be sampled at least 10 m far from each other.

We do not have a reference to back this up and many other studies use 2 meters in *S. carolinense.* We added a sentence explaining this caveat.  
L124-125: For better clarity, organize sample sizes by region.

Addressed by including the number of genets per region and dropping the number of genets per population.  
L127-128: The phrase "after one to several weeks" seems too vague to me.

Addressed by adding the collection dates to show that there was a time lag between collection from the two regions and growth in a common garden.  
L141-142: Clarify the time interval between blocks and the duration of each block. The current description of the division among blocks is unclear.

We moved the section from the paragraph below that explains that each block was planted over the course of five weeks, followed by the other blocks.  
L146: Specify what aspect of clarity you are referring to.

We switched the order of the experiments to match the logic that lead to these particular experiments. We first wanted to test for differences in thermotolerance and then addressed the underlying causes of those differences by focusing on reproductive traits.  
L149: The range 3-9 months is very broad.

Added parenthetical about temporal blocking.  
L158: Clarify if you are referring to the 'type' of chambers.

Addressed by adding that we used two different models of environmental chambers.  
L173: Elaborate on what is meant by "represented by northern and southern plants."

Addressed by rephrasing to say “with at least one flower from a southern plant and one flower from a northern plant represented in the pollen mix.”  
L178: Explain why fruits should be "at least one month old."

Addressed by adding that fruit are ripe at one month old.  
L193: As noted in major comments, the experimental design and treatments need to be detailed here.

Addressed by adding a small paragraph summarizing the methods and main points of the study.

L213: In L195, you referred to them as "reproductive variables", so change it here or there for consistency.

Addressed by changing terms to be consistent.  
L217: Confirm if variation among different petri dishes was controlled for in the statistical analysis.

All analysis for pollen germination was based on Tmin, Tmax, and Topt values determined using temperature performance curves. We do not account for variation among petri dishes used for treatments of one plant, but we do include plant (ramet) as a random effect.   
L220: Address the major comments regarding the description of treatments in experiment 2.

Addressed by referencing the protocols we used to examine extreme heat and adding a disclaimer describing that we were interested in the individual variation in temperature limits and thus used values extreme enough to yield differences in tolerance.  
L231: Was the flower date measured in the field? I do not see this variable described in the text.

The first flower date was recorded for plants in the environmental chamber to determine if flowering phenology differed between northern and southern plants. We dropped flower date from our analysis because it wasn’t necessary to include for the sake of this manuscript.   
L238: Explain why overfitting might be a concern for these variables compared to others.

Addressed by adding an explanation detailing that pollen diameter was averaged for one flower per block per genet so genet was omitted as a random effect in this model.  
L239: The lm function is in the R basics, so you don't need to specify the package.   
Addressed by dropping the package information for all functions included in base R.

**RESULTS**   
L269: Confirm if the effects you mention are indeed significant. Apply this to L316, L332, etc.

The sentences in L269 and L316 are overview/summaries of our results. Both are followed by sentences that provide more detail and statistical results. L332: Addressed in the sentence before.  
L271: In which figure/table did you present the results of these two variables?

We decided to cut these results to simplify the manuscript.   
L273-274: These values are not equal to those in Table 1. Why?

Addressed by removing all statistics in the text that are also in a table.  
L281: According to Table 1, there are region differences in Style+Stigma length and in Unfertilized ovules.

Both are included in the text. The Style+Stigma length is in the pre-pollination section and the unfertilized ovule results is toward the end of the post-pollination section.  
Fig. 3: These plots can be misled by the two-way interactions for 5 out of the 7 significant treatment effects, so maybe this figure can be moved to supplements. Also, in both, figure 3 and 4, I would show the plot of pollen germination even this variable is not significant; but you can include the significance above each plot.

We added all insignificant variables to both figures and moved figure 3 to the supplemental information (Fig. S5).  
Fig. 4: Enhance the figure caption for clarity. Describe the colours used, the type of model, and the standard error (SE). Include treatment differences using posthoc contrasts to provide a clearer understanding of the results.

Addressed the figure caption and added posthoc contrasts to the figure caption.  
L327: This seems more appropriate for the discussion section.

Discussion on the block effect for CMS was moved to the discussion section.  
Fig. 6. The legend off panel B is not needed, but Ito be honest I don't understand from which data this figure come from. The model is not described in the Data analysis section.   
Addressed by cutting the legend. Included the term “temperature performance curve” in the figure caption to hopefully clear up confusion.

**DISCUSION**   
L377: Reorganize the content to present the result first, followed by the discussion. This structure will help in maintaining a logical and reader-friendly flow.

Addressed in a couple places throughout the discussion by moving the results up in the paragraph.  
L396: Do you mean "contrasting results"?

There is one study that did not find this result Xu et al. (2017) and one study that did- Din et al. (2015). We changed the wording from “mixed support” to "contrasting support.”  
L396-401: Connect this discussion directly to your results.

Addressed by adding a sentence on what the results in tomato might mean for *S. carolinense*.   
L406-407: Again, specify the results you are referring to. This will help readers understand the connection between your discussion and the specific findings of your study.

Addressed by moving that sentence so it now follows the results of the study.  
L409: Replace the term "dramatic" with a more scientific and precise term.

Addressed by replacing “dramatic” with “pronounced”  
L417: I would write: "Plants from southern regions...".

Addressed by using the reviewer’s wording suggestion.  
L453: As far as I understood, you did not test for the evolution of temperature tolerance (see L464).

Changed to responses instead of evolution.   
L464: Maybe better using "increased thermal tolerance" rather than "evolved thermotolerance".   
Addressed by using the suggested change.  
  
  
Referee #2 Evaluations:   
Recommendation: Major changes needed   
  
Referee #2 (Comments for the Author (Required)):   
  
In this manuscript, the authors touch upon a very timely topic of plant responses to increasing temperatures in a big and complex two-part experiment. Overall, the results are very interesting and worth of sharing. However, the presentation as of now of the two-part experiment is very rough and confusing. I would suggest a re-work of the way the two-part experiment is addressed, potentially in a more "merged" state - right now the terminology around the experiments is very inconsistent (i.e., sometimes referred to as "Experiment 1", "experiment 1", other times as "Experiment one"). Another completely different option that the authors might consider is to split the two-part experiment into two manuscripts. Additionally, greater details need to be included into the method sections, as several aspects are very vague at this point. Some decisions taken during the experimental design also seems questionable. Lastly, it seems like the text overall needs a second pairs of eyes, as several times wrong figure numbers are indicated, and other smaller but consistent mistakes.   
  
**Abstract**:   
L7: You do not use the MN and TX abbreviations anywhere else, so I would just leave them out.

We use MN and TX in a few spots throughout the manuscript, so we decided to leave them in the abstract as well.  
L11: The two regions are MN and TX?

Yes, we changed the wording to more explicitly state that.  
L9: I would make it clear here that you are know talking about Experiment 1, as you make it clear at L16 that you are talking about Experiment 2.

Addressed by adding “Experiment 1”  
L19: Very vague, is there space to clarify "certain traits"?

Addressed by listing out traits.  
L21: Negative consequences?

Addressed by adding “negative”  
  
**Introduction**:   
L31: "relative rapidly" seems a bit subjective.

We just omitted the term completely.   
L30-32: Maybe it would be better to use the more commonly known terms: Drought avoidance, escape and tolerance here.

Addressed by rephrasing to include commonly used terms.  
L41: The standard to test for local adaptation is through investigating G x E interactions in a reciprocal transplant. I would argue that you cannot say anything about local adaptation with your experiments as they are more common-garden approaches with applied treatments.

Addressed by changing the sentence to emphasize that we are testing for “differences in temperature tolerance between plants from different thermal regimes”  
L41-43: Maybe this part works better at the end of the introduction.

Addressed by moving the sentence to the last paragraph of the introduction.  
L74: Remove "range shifts".

Addressed by removing “range shifts”  
L84-85: I would say the temperatures are too specific for your introduction still, save it for the methods.

Addressed by dropping the specific temperatures.  
L89-91: Overall, the introduction reads relatively well, but I still miss some more background information to support the predictions made.

We added some more information on phenotypic plasticity and highlighted how Experiment 2 fits into the context of previous work. Hopefully, this makes our predictions more logical.

**Methods**:   
As of now, when talking about heat treatments in one experiment, it means something different in the other experiment - and sometimes something different within one experiment depending on the trait in discussion. This is all very complex experimental design (one could argue more than necessary).   
Regarding the "Supporting Information: Methods", it seems to mostly be a repetition of the main text methods, but with some vital information on decision making and clarifying details. I would argue that most of this supporting information should be included into the main text instead.

We added more detail to the methods but decided to leave the specifics about each protocol in the supporting information.  
L99: Figure 1?

Addressed by correcting the figure numbers.  
Figure 1: What data is the distribution map based on?

Addressed by adding the citation for the for the data.  
L106: Multiple populations = 5? Please use exact number.

Addressed by including the number of populations in MN and TX.  
L111: Here you state that the average daily temperatures come up to a high of 22 degrees in MN - however, your control temperature is set at 25 degrees? This is a high control temperature for MN.

The control treatment is halfway between the average daily temperatures for TX and MN. We added a sentence explaining this in the methods for Experiment 2.  
Figure 2: Maybe but the temperatures on the y-axis in Celsius degrees so it matches the text. To save space this figure could also go in the Supplementary Information. Please also be clearer about the bars above the plot and which is which.

Figure was changed to Celsius and moved to the supplemental information.  
L123: 1 meter should be enough depending on species but should have a reference to back this up.

We do not have a reference to back this up and many other studies use 2 meters in *S. carolinense.* We added a sentence explaining this caveat.  
L125: N = 6 is very low N if they are then subsequently split up into different treatments?

There were 6 genotypes, but every genotype was subjected to all treatments using ramets. We added a sentence describing that there were four ramets of each of the 52 genotypes (26 north and 26 south) for a total of 208 plants in this study.  
L127-128: Please specify "one to several weeks".

Addressed by adding detail about post-collection processing.  
L128: Please convert one-gallon container into litres (L).

Addressed by changing from one-gallon to 3.8 L container.  
L129: Which standard potting mix was used? Brand?

Addressed by adding the brand.

L129: Please specify "several months".

Addressed by replacing several with the exact number.  
L138: What exact numbers of ramets were planted each week (over the duration of five weeks?)?

Addressed by adding the exact number of plants planted each week.  
L144-146: I would maybe switch around Experiment 1 and Experiment 2 if Experiment 2 did indeed happen "first on the timeline". The presentation of these two experiments is very confusing, and I am not sure how they relate to the current methodology.

Addressed by switching Experiment 1 and 2.  
L162: I don't think you need to repeat here that the study species is andromonecious.

Addressed by removing andromonoecy from the methods section.   
L171: Three additional hermaphroditic flowers or?

Addressed by adding the term “hermaphrodite”  
L172-173: I was very confused about why you would mix the pollen between north and south, and I question how your results on viable seeds, aborted seeds, and unfertilized ovules should now be discussed very carefully with this is in consideration. I later read in the SI that it was mixed to ensure fertilization, but I am not sure if I agree with this decision.

As we were developing this project, our goal was to get estimates of male and female reproduction in response to heat. Female reproductive success is difficult to separate from the male component of fertilization. As you mentioned, the SI can make this system difficult to work with as well because all of our populations are small colonies along the range edge for this species and could have the same S-alleles. Therefore, we decided to mix pollen from the north and south both in the hope that fertilization would occur and that responses to heat would be related to female fitness.   
L173: Why sometimes pollen form 2 flowers? Why sometimes 5?

The plants did not all flower at the same time and often there were only two plants in the control flowering. We preferred to have more than two plants to ensure fertilization with the SI system. We added a parenthetical explaining that it depended on the number of plants flowering at a time.   
L193: Could also go under "ecophysiological traits"?   
L195: Could also go under "phenological traits"?

We were interested in sporophytic (vegetative) and gametophytic (reproductive/pollen) temperature limits, not necessarily the physiology of flowering phenology of the plants in different temperatures.   
L196: I am now confused on where the plants come from here in Experiment 2, these are in fact the first plants that got cut down and regrown for Experiment 1?

We reordered the experiments, so plants used should make more sense now.  
L200-205: All the abbreviations for the different temperature treatments are too difficult to remember.

We dropped the abbreviations in this section.   
L207: When and for how long where they exposed to the treatments between the before and after measurements?

We include the times above – 1hr at 60C for the hot treatment and 1hr at 4C followed by 1hr at -18C.  
L217: They were dispersed over five separate petri dishes, right? I think it was very important to make that distinction, as the pollen as been mixed earlier, and some might assume the same was done again.

We added “one flower per plant” to address this comment.  
L228: Could a difference of 3-9 months storage have any effect on the regrown individuals?

It is unlikely that the different times would affect measures of thermotolerance and regardless, the time stored for individual plants was randomized using a random number generator. So, any effects would simply increase the variance around the means and would not represent confounding.

L249: If you put block effect as random factor, then how can it have a significant effect on your results? It a shame, that space issues have led to so many experimental design flaws, and these should be addressed more in the discussion. However, I believe that you take account for enough of the block effect variation by having it as a random factor.

We were interested in the block effect for our study, since temperatures in the greenhouse were not constant throughout the spring and summer. To look at block effect, we ran the ranova() function, which uses a likelihood ratio test of the full model and reduced models without the random effect terms.

This approach is described in this paper:

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution 24:127-135.

Since blocking did affect our results for cell membrane stability, we decided to run separate paired t-tests for the cell membrane stability to see how it changed with time. Block remained a random effect in all of our final analyses.

L263-264: Please state what the adjusted p-value was, and this is what should be used for reporting the results in Table 1 and Table 2, instead of the "standard" 0.05 p-value threshold.   
The Holm-Bonferroni method was only applied to the correlation analysis that we did between all traits. We added the table with these results to the supplemental material. The p-values reported there are the adjusted p-values.

**Results**:   
It seems the authors jump between referring to North versus South (methods), and now switch back to Texas versus Minnesota. Please just stick to one or the other.

We are sticking with North vs South for the methods and results and then only bring in TX and MN in the discussion and description of the study system.   
L270: Should be Figure 3?

Addressed by changing to figure 3.  
L271: First time you mention the "first flower type" trait?

We decided to drop flower timing and the first flower type (sex) from the manuscript.  
L273-274: I would not report the X2 and p-value in the text when you have a table.

Addressed by removing statistics in text that are also in a table.  
L278: "Style plus stigma length" is considered one combined trait? Like this is seems like two separate traits.

We changed all “style plus stigma” to “style+stigma” to give the impression that it was a combined trait.  
L278-279: I would put the regression values in figure instead of the main text. Furthermore, if previous figure in the text should be figure 3, then now you are mentioning figure 5 before figure 4.

We switched the order of figure 4 and 5 to reflect what is written.   
L281: Specify the traits "Neither number of ovules nor pollen size differed by region"?

Addressed by using ovule count in that sentence.  
L291: What is the definition of an aborted seed? (Should probably be mentioned in the methods)

Addressed in the methods. Aborted seeds were flattened seeds.  
Figure 3 caption: You should not report on the results in the caption, do it in the main text. I.e, delete "Plant development in heat reduced the size of... ".

Addressed by deleting any results from the figure captions.   
Figure 4: Maybe add "a,b,c,d" to the different plot frames to make it easier to refer back to the specific trait of the figure in the main text? Also, clarify that these are based on the significant result from Table 1? Tell what the error bars indicate.

Moved figure to supplemental information but addressed comments in the supplemental figure.  
Figure 5 caption: Incorporate the regression values into the figure itself instead maybe?

Addressed by adding the regression values to the figure.  
L325-326: This also applies to the control conditions. The implications for the results should be discussed.

We moved discussion on reasons for the significant block effect to the discussion.   
L327: Block effect discussion should be in the discussion section, not in the results.

We moved discussion on reasons for the significant block effect to the discussion.  
Table 2 legend: Instead of just omitting a factor resulting in an overfit model, have you tried adding it as a fixed factor instead?

We included genet as a fixed effect and changed the results to match those changes for the three models that were overfit. CCMS no longer has a regional effect.

Furthermore, explain in the legend what dF, F, and P in the table stands for.

Addressed by adding information to the table legend.  
Figure 6: Maybe the bold lines could be a tiny bit thinner?

Addressed by reducing the size of the bolded lines.

Also again, result reporting: "Pollen from the northern region..." should be reported in main text, not in the figure caption.

Addressed by removing results from the figure caption.  
L352: What does "readily" mean in this case?   
Addressed by changing “more readily” to “more likely.”

**Discussion**:   
In general, I think all the experimental design limitations and the probable effect block effects, mixed pollen, etc., might have on results should be discussed and mentioned more in detail here.   
L377-378: If you talk about several studies, you should include several references and not just one.

Addressed by adding more references to back up that claim.  
L396: Maybe it will have more impact if you mention earlier on in the methods that tomatoes are a close relative?

Added a statement to the methods where we introduce *Solanum carolinense*.  
L401: A biological clarification of the negative effects and evolutionary consequences would be good here.

We defined the negative effect here as a reduction in fitness. We also added a few sentences speculating what evolutionary consequences could arise with a decline in reproductive success with increased temperatures, such as temperature tolerance adaptation, range shifts or mating system evolution.  
L405: "among plants" (when it is more than two)

Addressed by changing from “between” to “among”  
L415-420: Too much repetition of the results, it is not necessary.

Addressed by deleting four lines with results.  
L415-416: But how can you distinguish between north and south in the fruit development? Didn't you mix the pollen?

Addressed by adding a sentence pointing out this caveat. However, the number of ovules is a count that takes place before fertilization. So, an increase in the number of unfertilized ovules post-fertilization might just be a factor of an ovary having a larger number of ovules to begin with. We still think this is a legitimate claim.  
L428: Limiting factor for/of what?   
Addressed by adding that “pollen performance is the factor limiting fertilization of viable seeds in the heat treatment.”