**Responses (will go into interactive review site):**

Editor #1:

Dear Dr. Brillante,

Thank you for the rapid review process and the opportunity to reply to the reviewers. We understand the reviewers' concerns regarding our lab-based design and our lack of replication at the variety-level. Unfortunately, it is not possible for us to collect additional data. However, we have worked to address the reviewers concerns by: (1) Re-drafting much of the text in the abstract, introduction and discussion to provide context for our results, including how they highlight the challenges of variety-rich studies as well as how to approach future variety-rich studies, (2) relatedly, providing an additional literature review to the paper (summarized in a new Table 1), which we believe shows how our results are in-line with previous findings and thus suggest methodological issues are unlikely to be the main cause of our findings.

We provide point-by-point responses to the reviewers concerns separately.

Thank you.

Lizzie & Nicole, on behalf of our co-authors

Reviewer #1:

Q1 (general review of paper): No reply need

Q2: The limitations in this study are significant. The idea to correlate phenology in the greenhouse with phenology in the vineyard is a good one, but the study was conducted in the off-season and it isn't clear that the greenhouse conditions were appropriate to mimic the vineyard. Only a fraction of the vines flowered. The authors state that the subsequent chamber study had "little to no replication" and indeed there was only one vine of Cabernet Sauvignon, Gewurtz, and Vinhao.

We completely agree with the reviewer that our study was limited in replicates per variety, which is why we did not evaluate variety-specific responses. We now clarify this in the methods, where we state, “Given limited replicates per variety all analyses of the growth chamber study were done across varieties,” and in the results where we state, “Given the limited number of replicated per variety, we do not report variety-specific estimates and all statistics are done across varieties.” Further, we have re-worked the abstract, and discussion to emphasize this limitation and focus on what it may mean for studies such as ours, that attempt to study many varieties. In our abstract, for example, we now write: “Indeed, we found high variability in flowering success across varieties in the lab (28 out of 50 varieties had no flowering), which made it impossible to study variety-specific response to temperature. Across varieties ….”

Q3: The results are interpreted appropriately. I appreciate that the authors were upfront with the limitations to their study, but unfortunately these limitations result in little useful information coming from this manuscript. As the authors note in the introduction, there have been previous studies that demonstrate flower abortion in higher temperatures.

We agree with the reviewer that flower abortion in higher temperatures has been previously shown, but we believe our results---growing 50 varieties in the lab and finding their vegetative phenology correlated with field results, as well as the challenges in getting enough varieties to flower---are relevant and important to the field. We have re-worked our abstract, introduction and discussion to highlight these findings and added a brief literature review to show that few studies have examined many varieties. Given this, we think publishing results, which outline our methods and challenges, could advance the field by providing a template for future studies to consider, as we now clearly discuss.

Q4 Checklist

e. Are the methods sufficiently documented to allow replication studies?  
- No

We have reviewed our methods and attempted to fill in any gaps. If the reviewer has specific areas we should further document we would be happy to do so if directed.

Q5: There were not enough vines/not enough replication in this study due to the limited number of vines available. Due to that limitation, I do not believe the results are publishable. My suggestion to the authors is that they re-run the study with more vines.

We appreciate the reviewer’s concerns. Unfortunately, it is not possible for us to re-run the study at this time, and while the number of vines was low in the growth chamber portion of the experiment, our replication level in the greenhouse was quite high (351 plants). We believe our Brief Report highlights the challenges in such variety-rich studies and provides valuable information on how to approach these types of studies in the future. We have worked in this draft to better capture this message of the work and hope the reviewer may find it suitable for publication in this light.

Reviewer # 2

Q1 (general review of paper): No reply need

Q2: The main limitation of the experimental plan of the research is the very tricky and unphysiological way followed to prepare the plant material.

We appreciate the reviewer’s concern as any lab study must struggle with how much of field conditions to replicate and how much conditions must be different based on fundamental limitations or the desire for experimental control. In our first draft, we can see that we did not provide a sufficient review of the literature to place our experimental design in context. To address this we now review the literature (see new Methods, Results, and Table 1) and show that our treatments bracket a range commonly done in winegrape lab studies.

Q3: I have to underline that the range of temperature selected for the experiment do not include limiting temperature for temperate plant development an expected optimum for growth ranging for 20 to 40 °C (see for example in attached figure from Criddle, R., Hansen, L., Smith, B., et al. (2009). Thermodynamic law for adaptation of plants to environmental temperatures. Pure and Applied Chemistry, 77(8), pp. 1425-1444. Retrieved 21 Dec. 2019, from doi:10.1351/pac200577081425).  
  
The low number of cultivar and plants able to flower are the consequences of the unphysiological temperature treatments given to the plants to complete the experiment within the same year of cutting. We have to diagnose that this treatments negatively affected the inflorescence differentiation in general and probably the plants able to differentiate had “weak” inflorescences maybe more prone to high temperature damage than normally differentiated inflorescence.

We agree with the reviewer that 20 to 40°C would be relevant treatments to consider and our chamber temperatures ranged from 17-40°C. Thus we believe we bracketed the relevant temperatures, while also attempting to maintain a difference in day/night temperatures. We do agree with the reviewer that our plants were forced outside of the usual timing, though this is commonly done with cuttings. We now stress this limitation in the discussion.

Additionally, we have added a literature review to put our experimental treatment more in context. While we understand the reviewer feels that our results are due how we forced the cuttings and our treatment temperatures, these temperatures are in line with other studies and our results do mirror findings on older cuttings. We now highlight this in the results and discussion, while also continuing to stress the caveats of our design.

Q4 Checklist

b. Is the quality of the figures and tables satisfactory?  
- No

f. Are the results presented correctly and interpreted in light of previous knowledge?  
- No

We now provide a new table reviewing other relevant literature, and would be happy to adjust our figures/tables further as requested. We hope our new literature review and changes to the discussion to stress the limitations of our design have improved our interpretation of the results.

Q5: The paper is formally well done but the experiment is very inconsistent.  
  
For what discussed in the point Q3, I have to conclude that:  
- the thermal range selected for the experiment does not include very limiting temperatures;  
- the vine management to complete the experiment was unphysiological with a general negative effect of inflorescence differentiation.

My suggestion is to not publish the paper.

We too were disappointed with the results of the chamber experiment, but we believe the study provides value in: (1) Supporting previous research on flower abortion and phenological responses to warming during flowering, and (2) highlighting the challenges of variety-rich studies by providing new data on correlations between lab and field phenology, showing the high-variation in lab flowering success across 50 varieties and providing some evidence for its reasons. We have now rewritten the abstract, introduction and discussion as well as added to the methods and results to better capture this.

Finally, we note the growing emphasis today on publishing all results of experiments. This includes null results, and results from replication studies, and embraces a general aim to fully publish results to prevent other researchers from attempting similar experiments without knowing the results and challenges of previous work (because that work goes unpublished). Given this, we believe publishing our results, with our related review of our methods and challenges, could advance the field by providing a template for future studies to consider. Based on our challenges, we have worked to clearly state our findings and recommendations for future work in the discussion.

**Comments on Revision (Response to Editor on 30 March 2020)**

Dear Dr. Brillante,

I cannot tell if we need to respond to the critiques of Reviewer 1 now (apologies, I am new to the interactive system), so I have provided a brief response below and can provide a fuller response if requested.

All the best,

Lizzie

We appreciate Reviewer 1’s concern about the limited number of varieties in each treatment and have tried to be clear about this limitation in our study at multiple points in the manuscript. For example, in the results, we state, “Because of the limited number of replicates per variety, we do not report variety-specific estimates and all statistics are done across varieties.”

We do not believe, however, that our study confounds variety and response to temperature. This would be an issue if we had only one type of variety in each temperature treatment (e.g., if the lowest temperature was all Tempranillo and the highest treatment was all Durif), but varieties were randomized as they were placed in treatments, thus we ended up with a random mix of varieties across treatments (with Verdelho and Sauvignon Blanc each present in four treatments). As we state in the results:

“Given the low number of plants that formed inflorescence, most varieties could be placed in only one or two temperature treatments (with very low or no replication per variety: chamber 1 (20**°**C) had one plant each of Cabernet Sauvignon, Durif, Sauvignon Blanc, and Verdelho. Chamber 2 (26**°**C) had one plant each of Durif, Pinot Gris, Sauvignon Blanc, and Verdelho. Chamber 3 (30**°**C) had three Durif plants, then one plant each of Gewürztraminer, Tempranillo, and Verdelho. Chamber 4 (34**°**C) had two Tempranillo plants, then one each of Dolcetto, Pinot Gris, Sauvignon Blanc, Syrah, and Verdelho. Chamber 5 (mean of 37**°**C) had two Tempranillo plants, and one each of Sauvignon Blanc, Verdelho, and Vinhão).”

Our results thus integrate across a diversity of varieties and can provide an overall estimate of the response to temperature (and, as our revision hopefully makes clear, our findings are generally in-line with the literature.) We would be happy to try to make this clearer in the text and address these concerns in more detail if requested.