26-Sep-2018   
  
Dr. Daniel J. Kliebenstein   
University of California   
Department of Plant Sciences   
One Shields Avenue   
Davis, California 95616   
  
  
Dear Dan:   
  
Thank you for choosing to send your manuscript entitled "Crop domestication and pathogen virulence: Interactions of tomato and Botrytis genetic diversity" for consideration at The Plant Cell. Your submission has been evaluated by members of the editorial board as well as expert reviewers in your field, and we regret to inform you that we are not able to recommend publication of this manuscript in its present form. However, during the post-review consultation session, we agreed that this is potentially important work that would be appropriate for publication in TPC, and that we would welcome a resubmission if the major points raised by the reviewers can be addressed. This would be treated as a new submission, but we would attempt to use the same reviewers. Nevertheless, reviewers will be asked to assess as a new manuscript (i.e. are the claims fully supported by the data; do the results presented move the field forward), and not only whether previous reviewer comments have been addressed.   
  
As you will read below, both reviewers provided very careful evaluations of this work, and offered multiple suggestions on how the manuscript can be improved. Reviewer 1's comments mostly focus on technical issues concerning your GWAS analyses, seeking clarification on exactly how these were performed, and requesting clear justification for the choice of methods, and when more than one method was used, some discussion of how the results differed. You should be able to address these concerns with some relatively modest rewriting. Reviewer 2 had more significant concerns, which mostly focused on whether the present data adequately support your conclusion that domestication has impacted disease resistance to Botrytis cinerea. This reviewer offers several suggestions on how to more thoroughly test this question. In the post-review discussion, the consensus was that domestication appears to have had a very minor impact, if any, on disease resistance, thus we suggest you tone down this conclusion, unless additional data can be provided to support it.   
  
We also recommend that you integrate data from Zhang et al 2017 (Plant Cell. 29(11):2727-2752) on Botrytis virulence loci with the lesion data in this manuscript in order to assess which Botrytis loci affect virulence on both species and which are host-specific virulence loci. Is there a a correlation between tomato and Arabidopsis across the 97 isolates in terms of virulence? Either a positive or a negative result would be interesting.   
  
It will be important to convince the editors and reviewers that the study adds significant new understanding of mechanisms or processes and that the major claims made are fully justified by the data presented.   
  
Note also that supplemental materials should be restricted to large datasets and tables, presentation of replicates, and validation of reagents, methods, or genotypes. Any data that are used to support the main claims must be in the main manuscript. Supplemental figure legends must indicate what figure in the main manuscript is supported by the supplemental data presented.   
  
We thank you for your interest in and support of The Plant Cell. We wish you good luck with your research and we look forward to seeing future submissions of your work.   
  
On behalf of the editorial board,   
  
Roger Innes, Senior Editor   
Sabeeha Merchant, Editor-in-Chief   
  
The Plant Cell   
  
---------------------------------------------------------------------------- Reviewer comments:   
  
Reviewer #1 (Comments for the Author):   
  
Review of Soltis et al. "Crop domestication and pathogen virulence: interactions of tomato and Botrytis genetic diversity"   
  
This ms describes a study on the genetic basis of pathogen virulence and host defense in an interaction between Botrytis cinera and tomato. The authors demonstrate that there is abundant genetic variation for pathogen virulence and that this variation differs among different genotypes of the host plant. My background is mainly in quantitative genetics and not plant-pathogen interactions, so I will focus my review on the GWA study and associated analyses.   
  
Over all I find the analyses to be adequately done and the results seem intuitively believable to me. The manuscript is by design rather complex (pathogen lines x host lines x domestication), so my comments below are mainly intended to increase clarity of the ms and to point out some lack of details in the various analyses. This will (hopefully) make the ms simpler to read and make it easier for the reader to digest the results.   
  
The overall data on lesion length was analyses using a general linear model (as described on lines 711-724) and the results are presented in Table 1. First, it is not clear to me from the M&M section what terms were considered fixed or random in the model. From reading lines 712-714 it seems that all variables were considered as fixed effects, yet the model was supposedly analysed using lme4, which is a package specifically designed for analysing mixed models (i.e. models including both fixed and random effects). In fact the lmer function in the lme4 packages require at least one random effect to even run. So I would like to see these issues - which (if any) terms were random and which were treated as fixed? Also, how were the % genetic variance calculated for the traits included in the model? This is never explicitly stated in the ms.

<Response>

The data sets used for the GWA analyses differ as they rely on mapping the pathogen sequencing data versus two alternative reference assemblies (T4 and B05.10). Why two different reference genomes are used is not really motivated in the paper. Judging by the SNP numbers presented, about 10% of the SNPs are missing when called against the B05.10 reference compared to T4. What fraction of SNPs are shared between the two reference genomes and do frequencies of SNPs common between the two reference genomes differ? The reason I'm asking is simply to get a feeling for how different the datasets are for the two reference genomes and if it is motivated to include moth analyses in the paper. Having two data sets called against different reference genomes certainly adds a level of complexity that might or might not be needed to an already complex ms. Right now the ms lacks justification for keeping this added complexity.

<Response>

The authors also perform the GWA analyses using two different statistical methods, ridge-regression using bigRR and the linear-mixed model approach as implemented in GEMMA. Since bigRR doesn't provide p-values the authors use a permutation approach and then also implement this for the GEMMA analyses. I assume the latter is used to keep the p-value calculations consistent across methods? If so, this could be stated in a sentence in the M&M section for clarity.

<Response>  
  
GEMMA do provide p-values for association tests and it would be interesting to see how these relate to the p-values from the permutation approach. Do the number of significant SNPs differ when selected based on the permutation approach or selected based on (multiple-test corrected) p-values from GEMMA? It might be worth to just simply mention any such differences (or lack of) in the text.

<Response>  
  
  
Also, like with the two reference genomes it is not clear whether the results are different enough using GEMMA or bigRR to justify including them both in the paper. It would be possible to select one method and add the other to the Supplementary materials if they give essentially the same results. If both methods are kept, it would be good for the authors to more clearly justify why having both analyses is valuable. What insights do the two GWA analyses methods provide that they don't provide alone?

<Response>  
  
  
  
Minor comments:   
  
Line 309: "..showed statistically similar variation" - that's not how hypothesis testing works! Lack of significance is not evidence in favour of the null hypothesis. I would simply drop the work "statistically" in this sentence.   
  
Lines 365 and 368: These two lines both state that two isolates were found to be more virulent on domestic tomato lines. This seems a bit repetitive to me as the two lines are only one sentence apart.   
  
  
Figure 4b) is rather hard to read. Would it be possible to highlight the points that overlap with the vertical lines, to make them clearer? Right now they are hard to see the points when they overlap with the dashed lines.   
  
Line 467: I'm not sure the phrase "SNP calling between hosts" is a good choice of words here. "SNP calling" is (at least to me) something entirely different from the process of identifying significant SNPs form the GWA analyses (which I think is what the authors intend to say)   
  
  
  
  
  
  
  
  
Reviewer #2 (Comments for the Author):   
  
This manuscript presents a very in-depth, quantitative, analysis of how pathogen and host genotypes influence virulence in the Botrytis-tomato pathosystem. The results show that a very large number of Botrytis genetic loci affect variation in virulence on tomato, and similarly, that tomato genetic variation affects susceptibility to Botrytis. No major loci emerged as dominating virulence or susceptibility respectively. The authors however did highlight small but statistically significant differences in susceptibility between wild and domesticated accessions. Based on the scale of the study and the comprehensiveness of the quantitative analysis, this paper has the potential to be landmark in the study of generalist pathogens. Although, the lack of major loci on which to focus follow-up reductionist studies might be disappointing to some readers, this should not detract from the impact of the paper.   
  
Although I am quite enthusiastic about this paper, there are nevertheless several substantial issues that need to be addressed, and there are some places where I thought the paper could be further strengthened.   
  
Major issues.   
  
1. The authors should include more comprehensive description of the Botrytis data set, especially the genetic structure of the collection (e.g. a neighbor-joining tree). Zhang et al 2017 (Plant Cell. 29(11):2727-2752; mis-cited in the references) only gives a table of isolate origins. Are all isolates genetically distinct members of a pan-mictic population, or do some represent clonal clades? What is the distribution of allele frequencies in the collection across all polymorphic sites? This information is important for assessing the pathogen component of the paper.

<Response>

2. The title and overall focus of the manuscript are on the effect of domestication. Yet Table 1 shows that domestication accounts for only 0.8% of the total variation in lesion area, and 3.5% of the genetic variation. Another 0.8%/3.7% was attributable to domestication:isolate interaction, but was not statistically significant. Furthermore, only six each wild and domesticated tomato accessions were included. Although statistically significant for this particular set of tomato accessions, I question whether this effect is large enough, and the conclusion robust enough, to support the focus of the paper. How can we know that the 3.5% difference is not a spurious artifact of the choice of accessions, especially given the much larger amount of variation attributable to individual plant genotype?

<Response>

The following procedure would be informative (but would not address my wider concern): If two bins of plant accessions were created, each with three of the wild and three of the domesticated accessions chosen at random, and the procedure repeated, say, 100 times, in what percentage of permutations would the two bins explain 3.5% or more of the genetic variation in pathogen virulence.

<Response>

3. Lines 333-355. The authors used a Wilcoxon signed-rank test as an alternative to test if the rank of B. cinerea isolate-induced lesion size significantly changes between pairs of tomato genotypes. However, no details of the test are given. For example what was the input for the test? Was it the raw lesion sizes, the model-adjusted lesion sizes, or the ranks of the lesion sizes on the individual accessions. Since the test begins by calculating the actual differences between paired input values, before ranking the absolute values of the differences, this question is important, especially when a large number of the input values are closely bunched.

<Response>

Furthermore, given that 77.8% of the variation is attributable to non-genetic sources, and was a high statistically significant source of variation, surely the author's procedure is simply affirming this source of variation in the data set. This would mean that the data presented in Table 2 is the result of experimental variation and not genetic variation in the host accessions.

<Response>

4. Lines 356-372. I have less concerns about the application of the Wilcoxon signed-rank test to the mean lesion areas on domesticated versus wild accession, since these represent means of 12 measurements each. Also the analysis is backed by the follow-up ANOVA tests. However, I do remain concerned whether the 2 isolates (out of 95) that showed as having significantly different virulence on domesticated versus wild accessions, truly are responding to domestication. Given the large amount of genetic variation in the plant accessions overall, is it possible that the two isolates are responding to genetic factors in the plant pools that are unrelated to domestication. If the plant pools were permuted as described in Comment #2, in how many permutations would 2/95 or more isolates show significant differences between the two pools. The FDR adjustment used by the authors does not account for the genetics of the plant pools; it only accounts for experimental variation in the lesion area measurements. In light of these concerns the statement "this B. cinerea population contains two highly domestication-sensitive isolates which are more virulent on domesticated tomato" seems a substantial over-statement.

<Response>

Related to the above, if the two isolates reported as "highly domestication-sensitive" are eliminated from the data set, is there any significant signal of domestication sensitivity left in the remaining set of 93 isolates? What happens to the 3.5% of genetic variation attributable to domestication?

<Response>

5. Lines 717-720; 738-740. It is typical to treat terms such as experiment, block, individual plant, leaf, and leaflet as random effects in linear models. (A fixed effect is one I could reproduce in a new experiment, e.g. isolate and plant accession; a random one is an effect I could not reproduce, e.g. experiment and block). What is the justification for treating them as fixed effects. The authors state that "significance of individual terms in the model did not change" but they do not provide documentation of that point. In particular, they do not show us how the estimated percentages of the variance accounted for by the different terms changes. Table 1 should contain the results from a conventional analysis (experiment and block as random effects) rather than an unconventional analysis. And in the ANOVA analysis of isolates sensitive to domestication, are the same 2 isolates identified if experiment is treated as a random effect (line 363).

<Response>

6. Zhang et al 2017 presents data on lesion sizes on Arabidopsis for all these isolates. Including data on the correlation between lesion sizes on Arabidopsis and tomato would provide a fuller picture of the genetic underpinnings of virulence in Botrytis. It would be especially informative know if any of the SNPs presented in Figure 4 are also significantly associated with variation in virulence against Arabidopsis.

<Response>

7. Surprisingly, the results from Zhang et al are not discussed in this paper at all, either in the Introduction or the Discussion. How does the finding that "... the JA and SA signaling pathways functioned to constrain/canalize the range of virulence in the pathogen population, but the underlying transcriptomic response was highly plastic. These data showed that plants utilize major defense hormone pathways to buffer disease resistance..." impact the observations presented here. Does this mean that much of the potential variation in pathogen virulence encoded by its genetic diversity is actually masked by the ability of the plant defense system to adapt to a variety of virulence mechanisms?

<Response>  
  
Lesser points   
  
8. Lines 138 - 143 "In addition to SNP diversity, the genomic sequencing showed that B. cinerea has a high level of recombination and genomic admixture, as if it were a randomly intermating population. As such, a collection of B. cinerea isolates contains genetic variation in a wide range of virulence mechanisms, offering the potential to challenge the host with a blend of diverse virulence mechanisms."   
Where is this documented? No reference is given and the documentation is not found in Zhang et al (2017).   
  
9. Lines 207-208; 214-216. Authors should reference Zhang et al (2017).   
  
10. Line 219-221; 236 and following. What is the justification for using lesion area directly in the linear model. The square root of the lesion area, which is a measure of the linear rate of progression of the lesion margin, would seem more likely to be directly explainable by a linear model. Did the authors compare the model fit between the area and the sqrt of the area?   
  
11. Lines 291-293. "we identified a significantly greater (18%) resistance of wild tomato in comparison to domesticated tomato across the population of B. cinerea isolates (Figure 2 and 3, Table 1)." What are the actual lesion sizes and the standard errors on the lesion sizes. These data are not given in any of the figures or table referenced. Also, how does the 18% number change if the two "highly domestication-sensitive" isolates are removed?   
  
12. Line 405. Here and elsewhere, the authors refer to a 99.9% effect size threshold obtained from 1000 permutations. In fact, a 99.9% threshold cannot be accurately determined from 1000 permutations as it represents 1/1000 and is likely to be idiosyncratic. Although such a threshold is very likely more stringent than the 99% threshold, the authors should be cautious about how they described this threshold.   
10,000 permutations would be required to more accurately determine a 99.9% threshold.   
  
13. Lines 408-411. "The ridge regression approach (bigRR) identified from 1,284 to 25,421 SNPs within B. cinerea that were significantly associated with altered virulence on the 12 different host genotypes" Do the authors mean the numbers of SNPs identified varied from host genotypes? If so, the sentence should be reworded to make this a little clearer. Similar issue on lines414 and 415.   
  
14. Lines 440-441 and Figure 6. If SNP block 5-11 represents the 5'UTR and 13-26 the body of the gene, then isn't the arrow in Fig 6a indicating the direction and startpoint of transcription in the wrong place (3' end instead of 5' end)?   
  
15. Lines 427-447. In addition to the pectinesterase gene, two other glycosyl hydrolases (Bcin14g00850, GH28, polygalacturonases) and (Bcin14g00650, GH\_31, alpha-glucosidases, alpha-galactosidases) seem of obvious interest, especially the polygalacturonase. Were these examined in more detail? If so, why not?   
  
16. Lines 484-485. "This had a high degree of overlap between the wild phenotype and domesticated phenotype." What does "This" refer to. Need to clarify.   
  
17. Line 661. The references contain an incomplete citation with an incomplete title. The correct citation appears to be "Zhang W, Corwin JA, Copeland D, Feusier J, Eshbaugh R, Chen F, Atwell S, Kliebenstein DJ. Plant Cell. 2017 Nov;29(11):2727-2752. doi: 10.1105/tpc.17.00348. Epub 2017 Oct 17. Plastic Transcriptomes Stabilize Immunity to Pathogen Diversity: The Jasmonic Acid and Salicylic Acid Networks within the Arabidopsis/Botrytis Pathosystem.   
  
18. Line 856. "plant phenotype" should read "plant accession". Also, the color coding should be given.