

Dear Editor,

Thank you for the opportunity to re-submit our manuscript now entitled "Trait plasticity is more important than genetic variation in determining species richness of associated communities", to *Journal of Ecology*. We appreciated hearing that the referees and Editors felt our previous manuscript had much potential; however, we agree that its presentation was long and muddled. My coauthors and I have invested a substantial amount of effort in revising the manuscript, taking into account the major and minor comments raised by the referees and Editors. Below, we have provided a point-by-point response (bold font) that illustrates how we have addressed these comments in the revised manuscript. We hope that you find that this revision has resolved the previous problems and now provides a coherent and concise message that is of interest to the broad readership of *Journal of Ecology*.

Sincerely and on behalf of my co-authors,  
Matt Barbour

-----  
Associate Editor  
Handling Editor Comments for Authors:

I concur with the 3 reviewers that the paper presents a very nice experiment, but that the results are hard to integrate. There are several points that need major special attention including the use of Type III ANOVA when you test interactions (R1), the effectiveness of the aphid treatment, given its low numbers (R2, but also mentioned by others) and the clarity and integration of the paper (R2 & 3). The nature of the critics raised needed exceeds those of a typical "major revisions". However, I would like to leave the door open to resubmitting a new paper with the same data if the authors are willing to substantially re-write the ms to incorporate reviewers suggestions and make a more focused paper.

**We appreciate that the Associate Editor recognized the quality of our experiment in the prior manuscript despite its convoluted presentation. In our revision, we have tried to address the key points brought up by the Associate Editor and the three referees. In particular:**

- **The prior version of the manuscript used Type II ANOVA when testing interactions, which both the Associate Editor and Reviewer 1 pointed out as questionable. In the revised version, we used an approach that is analogous to using Type III ANOVA when testing interactions, as advocated by the Associate Editor and Reviewer 1. Specifically, we centered and scaled all input variables (mean=0, SD=1) and used Bayesian linear-mixed models to estimate both main effects and statistical interactions. With the design of our experiment and the used of centered/scaled input variables, we can interpret both main effects and statistical interactions within the same statistical model (Schielzeth 2010, *Methods in Ecology & Evolution*).**
- **The effectiveness of the aphid treatment was questioned by the Associate Editor and all three reviewers in the prior version of the manuscript. In the revised version, we now make it clear to the reader that our aphid treatment was effective (Lines 140-141, 318-320, Fig. A1, A3), although the effects were generally weak.**
- **As noted by the Associate Editor and Reviewer 2 & 3, the prior version of the manuscript lacked clarity and integrated results. Addressing this issue required us to revise all of the analyses and rewrite the entire manuscript. We discuss how we address these specific**

**issues in our responses to Reviewer 2 & 3. By incorporating their feedback, we feel our revised manuscript is now clear and coherent.**

-----

Referees' comments:

Reviewer: 1

#### COMMENTS FOR THE AUTHOR

I have read the manuscript 'Partitioning plant genetic and environmental drivers of above and belowground community assembly'. It reads well and includes a lot of detail on the complex interactions that we see in natural communities. Certainly, we are missing studies on the relative importance of genetic and environmental factors. However, I feel that the final message is somewhat that the traits you studied elsewhere in a different study (plant chemical variation) are likely the ones that explain more variation than those you measured here (mostly plant growth traits). By using the older data as a proxy you make the case for this (and I am aware that often not everything can be analysed), but it does leave the message of this paper missing the final link.

**We agree that using the older leaf chemistry data as a proxy resulted in the previous manuscript lacking a clear message. Identifying this message required a complete revision of our analyses and refocusing of the manuscript. As part of this, we now restrict our analyses to the trait data collected during this experiment and ignore the older leaf chemistry data. At the same time, we still make clear to the reader how unmeasured traits may influence the interpretation of our results (Lines 412-424).**

#### Major comments

1. I would move the hypotheses hidden in the methods 'study site' section to the end of the introduction, probably after the main research questions. Here you would also have to extend the introduction to include the background of the system, i.e. the section in the methods before the hypotheses.

**As suggested, we moved the hypotheses in the 'Study Site' section to the end of the Introduction.**

2. In the experimental setup, (L164) were aphids added at each new time point (this is how I understand it), and were then aphids removed from aphid-absent plants (if there were any)? Then if there were only 0.05-7 aphids on average on plants, does this mean there was little reproduction? How often were aphids replenished? I wonder if aphids didn't 'accept' certain genotypes, how might this impact the power of the aphid environment analyses?

**Yes, aphids were added at each new time point and, at the same time, aphids were removed from aphid-absent plants if there were any. We make this clear on Lines 134-137.**

**Yes, aphid reproduction was variable and generally low. We did, however, still detect variation in aphid abundance among plant genotypes in the aphid treatment (Fig. A1) as well as a G x Aphid**

**effect on ant abundance (Fig. A1) and trichome density (Fig. A3). These results indicate that our aphid treatment was effective.**

3. For the analyses, the use of type II SS is ok for unbalanced data sets but I think not so suitable if there are significant interactions in the model, since it is more focused on the order (or non-order) of main effects. Please can you justify the use of these, and do the models differ when you use type III SS? i.e. how robust is the data to the different types.

**Good point. In our revised analyses, we used Bayesian linear-mixed models, so technically the type of sum-of-squares is no longer an issue. But, we acknowledge the general point that we need to be cautious when we want to interpret both main effects and statistical interactions. To do this, we centered and scaled all of our input variables (mean=0, SD=1), which enabled us to interpret both main effects and statistical interactions in our models (Shielzeth 2010, Methods in Ecology & Evolution).**

4. There is some inconsistency in the interpretation of the wind experiment, with respect to the relative effect of genotype and environment. For example, in the results (L492) you say the abiotic is more important yet later at L632 you say that genotype is more important in both, then L658 again you say the environment dominated. I think this partially just needs a clarification in the text on what you are looking at, since wind has perhaps a stronger effect on plant traits, height and shoot length (but not shoot count), yet less effect on foliar arthropods, for richness and abundance (but not rarefied richness). Yet, the abstract again suggests that wind is more important than genotype for foliar arthropods and bacteria. Please clarify in the text at each stage again what it is that you are discussing.

**We have completely rewritten the manuscript and believe this specific issue is no longer relevant. Still, we acknowledge the reviewer's general point and have sought to make consistent interpretations throughout the manuscript though.**

5. In the discussion, from L672 and the next few paragraphs, you mention a couple of times 'the only other genotype-by-environment experiment/study..' but then have multiple citations across the text. The importance of each of these studies in relation to yours and others needs to be clearer, what is unique to that study that has not been done before, and then how is your progressing from these.

**In the revised version of the Discussion, we have attempted to provide better context and comparison of our work to others.**

6. The aboveground and belowground communities have been studied rather independently. I would expect feedback from each, as plants can change their rhizosphere community in response to herbivore feeding. Also, aphids can respond to changes in the rhizobacteria/fungal community belowground. How might feedback from these communities have effected one another? I'd like to see some discussion on this.

**Good point. On Lines 367-373 of the Discussion, we now explicitly discuss how feedbacks between above and belowground communities could be modified by GxE effects on above and belowground plant traits.**

Minor comments

1. Figure 1A needs explanation somewhere of the black and white dots, I assumed this to be aphid +/- treatments.

**This figure is no longer in the manuscript, but we have made sure to use figure legends to avoid confusion about symbols.**

2. Figure 2A, can you give a bit more explanation (maybe in the text at L486), on what this is actually showing and how this demonstrates the interaction is driven by genotype J.

**This figure is no longer included in the manuscript.**

3. Figure 4. What do you mean the marginally significant effects are transparent? Clarify this.

**This figure is no longer included in the manuscript.**

4. Methods L261 and 264, add in that you also included all 2-way interactions to make the methods fully clear

**On Lines 256-257, we now make clear that we also included all 2-way interactions.**

5. Results. L419, just re-mention which species name is the ant and which is the aphid (I had to go back to look after reading through the methods and then starting the results section)

**We believe this specific issue is no longer relevant for the revised manuscript.**

6. Results. L 516, change Fig. 1C to Fig. 2C, and L524 Fig. 1D to Fig. 2D

**We have double-checked that all figure references are accurate in the revised manuscript.**

7. Results. L519, Table A6, could you add the functional group of each OTU/species to the table?

**This specific issue is no longer relevant since we did not include any Tables in the revised manuscript.**

8. Results. L570. Change Table 11 to Table 1

**This specific issue is no longer relevant since we did not include any Tables in the revised manuscript.**

9. Discussion. L714. Compared to what other studies? Give some citations here.

**This specific issue is no longer relevant for the revised version of the Discussion.**

-----

Reviewer: 2

COMMENTS FOR THE AUTHOR

I like that the authors manipulated two sources of environmental variation (wind, aphids) and replicated this at multiple sites. These aspects make the study very complete. Here are a few general comments and then more specific ones for each section.

1. I would suggest reducing/simplifying the statistical analyses and the results section is too long. Perhaps you could focus only on effects on the associate community and then the SEM to tease a part genetic and direct and indirect effects of the environment. Maybe the section on treatment effects on traits could be reduced substantially or maybe left as part of the methods (referring to supplementary material for these results). Could also delete some analyses (e.g. wind effects on soil characteristics) that did not yield important findings. At this moment, it is difficult to grasp the main findings of the results section so I recommend simplifying and focusing on the core analyses and key results.

**As suggested we have simplified our statistical analyses and focused on the key results. In particular, we have focused on quantifying the indirect effects of different sources of trait variation (G+E+GxE) on species richness. We feel this has resulted in a more concise and coherent presentation of our results.**

2. The effect of site could be viewed as another source of environmental effects on plants and their associated faunas. Perhaps the relationship the site effect and the treatment effects could be addressed up front. The design clearly has advantages in this sense, and maybe these should be explicitly stated and capitalized on, i.e. the relationship between treatment and site effects as environmental influences on plant associates.

**Thank you for this suggestion. Our revised analysis now explicitly quantifies the effect of site (blocks and plots) as another source of environmental effects on plants and their associated faunas. This suggestion actually resulted in a complete reframing of the manuscript. We realized that by including the indirect effects of site that trait plasticity was more important than genetic variation in determining species richness of associated communities.**

3. Can direct vs. indirect effects of the environment be effectively separated based on the current design and analyses? Does the lack of plant trait effects mean that direct effects must be acting or that residual variation not explained by traits are necessarily due to direct effects of the environment? Or is it that other unmeasured plant traits could be driving effects on the community but were not evaluated (i.e. indirect effects were not adequately tested)? I can't see how one could determine one scenario vs. the other, so perhaps some more thought has to be put into the a priori expectations of what is being tested as well as the interpretation of the results.

**Great point. For this reason, we re-focused our analyses on quantifying the effects of intraspecific trait variation. Thus, we no longer make any attempt to partition the direct effects of the environment. This limits the scope of our study to comparing the relative importance of trait plasticity vs. genetic variation in the traits we measured. However, we think that this provided a much needed focus and clarity to the manuscript, while still being able to test a novel question.**

4. I did not understand how these SEM models were fed (see specifics further down). This is a key part of the study which relates plant genetics, plant traits, and environment to the associate community and I would like to know more about the basics of this analysis. See some specifics on this ahead.

**This specific issue is no longer relevant since we do not use SEM in the revised manuscript.**

\*Side comment: the lead author has used the term community “assembly” in a previous study, but I wonder if this could be a bit misleading since assembly dynamics, in the strict sense, were not addressed. Maybe this is just a matter of opinion and a more loose use of this term is fine. For example, I was imagining a different terminology, such as “effects on community structure”, “structuring of associate communities”, etc. that didn’t imply assembly dynamics per se (stuff like priority effects and other temporal dynamics of community properties).

**We no longer use the term community “assembly” in the manuscript.**

#### Introduction

This section emphasizes the importance of addressing connections and feedbacks between above- and below-ground communities associated with plants. However, the paper ends up delivering an isolated assessment of effects on each component. I wouldn’t necessarily suggest adding more analyses to explore these connections (the paper is already heavy on the analytical side), but still, this part of the intro could generate some expectation that these dynamics will be evaluated. Perhaps just modify the language a bit to tailor to limits of the present work. Also, perhaps the term “phenotypic plasticity” could be narrowed down to “plant-mediated effects on associated faunas” (here and throughout the paper). This is a bit more specific and relates to plant induced responses influencing associates, and you could relate this parenthetically as a form of plasticity, to connect to a broader context.

**We have modified the Introduction so we no longer generate expectations that we will test feedbacks between above- and below-ground communities associated with plants.**

**We have opted to use the term “trait plasticity” throughout the manuscript. While the revised manuscript focuses on quantifying the effects of trait plasticity on associated communities, we do not feel that narrowing down our language to “plant-mediated effects on associated faunas” is necessary. We feel our use of trait plasticity is clear in the revised manuscript and useful for connecting our research to others studying plasticity.**

L. 43-46: Direct and indirect effects can be (and have been) evaluated in controlled experiments, I think the issue is that they’re conducted in a single site. The key point is that having site replication (e.g. ideally along an environmental gradient) is important in order to address the context dependency or spatial variation in such direct/indirect dynamics.

**Quantifying direct and indirect effects of biotic/abiotic factors on community structure requires explicit measurements of plant traits, which has rarely been done (restricted to GxE studies of soil nutrients). Because of this, the relative importance of trait plasticity vs. genetic variation across natural environmental gradients remains unclear. While we agree that having site replication is important, our experiments were not designed to test the context dependency of GxE effects across sites (i.e. GxE).**

L. 92: H estimates are environment-dependent, so this begets to some extent the logic of this sentence. Maybe talk more broadly about magnitude or degree of genetic variation varying among traits?

**Good point. We have modified this sentence accordingly (Lines 78-79).**

L. 95: Wording for q. 1 is not clear to me, particularly the ending "... and other environmental factors". Also, might be good to specify in q. 3 that this refers to plant genetic variation and plant plasticity as bottom-up sources of variation in associated faunas.

**We agree that the wording for Q1 in the previous manuscript was not clear. Our revised manuscript focuses on answering a single question: "what is the relative importance of trait plasticity vs. genetic variation in structuring plant-associated communities?" (Line 73-74).**

Methods

What was the spatial scale (area over which gardens were distributed) for the aphid and wind experiments?

**The ant-aphid experiment was conducted at 5 ant mounds distributed across an area of 300 m by 50 m. The wind experiment was conducted at 10 natural willow stands distributed across an area of 400 m by 40 m. We include this spatial information on Lines 122 and 144.**

L. 126-29: The distinction between these three alternative dynamics don't seem clear cut. For example, aphids may directly influence other herbivores or indirectly do so via changes in plant traits. Perhaps divide these dynamics into direct (e.g. competitive) and indirect (plant-mediated or ant-mediated) effects of aphids on the associate community. Also, might be helpful to relate this back to biotic effects of the environment (rather than genetics) on plant phenotypes.

**We have now refocused our analysis to quantify indirect effects via changes in plant traits. We make this distinction clear now in the context of the ant-aphid experiment on Lines 81-86.**

L. 144-46: Again, perhaps thinking about direct and indirect effects of this environmental source on plant-associated communities (or some other division) is a cleaner (and simpler) way of thinking of pathways of potential effects (and within each of these two categories provide examples such as those described).

**As above, we make this distinction clear for the wind experiment on Lines 87-90.**

L. 222-224: It seems somewhat problematic that phenolic compounds were not quantified for the actual plants used in the experiment. Heritability values are environment-dependent this justification may be weak. Particularly the last line "and therefore unlikely to be strongly influenced by environmental variation" is questionable since phenolic compounds (and other groups) are highly inducible (e.g. by abiotic factors such as light availability, not only herbivory). As the authors mention, there's also the potential for G x site and G x treatment effects that are not addressed by using a different set of plants to quantify defenses. I'm not sure if these limitations defeat the purpose of including phenolics in the paper, particularly given that genetic and environmental sources of plant phenotypic variation are the specific subject under study.

**We agree that using the phenolic chemistry data for this study has major limitations, which is why we no longer use these data in the revised manuscript. Thus, our inferences about the relative importance of trait plasticity vs. genetic variation are limited to the traits we measured in this study.**

L. 260: Any rationale/justification for including genotype (and its interaction) as a fixed rather than random effect?

**We now model genotype (G) and G x Environment as random effects. We were able to do this because our revised analyses use Bayesian methods to estimate effects. Before, we were observing high correlations between G and GxE effects, which is why we chose to model them as fixed effects previously (detailed reasoning is given here: <http://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>, “Singular models: ... correlations estimated as +/- 1).**

L. 263-64: was this a repeated-measures GLMM?

**Yes it was. We now analyze years separately since different traits were measured in different years.**

L. 273: it is not clear from this description that the analysis of wind effects was based on a split-plot design. Was the wind x G interaction assessed? This model structure appears different than the model for the aphid treatment but shouldn't they be the same (except for the year effect)?

**The model structure for the wind and ant-aphid experiment is basically the same. We have tried to make this more clear on Lines 251-257.**

L. 358: So these models included significant in the models above, where they were treated as fixed effects? This is confusing to me, both using subsets of significant terms and re-running models (sort of like a non-declared stepwise procedure) as well as treating factors as fixed in one analysis and then as random in another. Are these analyses somewhat redundant relative to the SEM analyses in the next section. Both seem to be quantifying the relative importance of genetic and environmental (plant-based and non plant-based) variation.

**Good point. We removed this entire section since it was no longer needed after re-focusing our analyses.**

L. 359-61: Can you separate direct from plant-based (indirect) environmental effects in this analysis?

**No we cannot. This is one of the major reasons why we modified our previous analyses. By focusing on the indirect effects of the environment via the traits we measured, we can explicitly compare the effects of trait plasticity vs. genetic variation on associated communities.**

L. 369: This section is key, but I was not able to understand how genetic and environmental effects were quantified (is this related to the variance components estimated from analyses described in the previous paragraph?). I would like to know what specific data on genetic and environmental effects were inputted in the SEMs. This section provides technical information on the statistical features of the analysis, but I can't understand the basics of how the effects were estimated and the type of data that was fed to these models.

**This specific issue is no longer relevant since we do not use SEM in the revised manuscript.**

Results



L. 424-25: Were aphid abundances kept roughly the same among plants in the aphid treatment? Five aphids were inoculated per plant, and based on the genotype means for aphid abundance (0.05 to 7 aphids) it seems like aphid numbers later in the experiment were similar to the number initially inoculated. If this is the case, how can genetic variation in ant abundance (mediated by aphids) be evaluated within the aphid treatment if aphid numbers were relatively constant among plants and genotypes and similar to initial numbers?

**Aphid abundances were not kept at similar abundances in the aphid treatment. Although the absolute differences were small, the amount of replication in our experiment enabled us to have sufficient power to demonstrate that there were clear differences in the relative abundance of aphids among genotypes in the aphid treatment (Fig. A1). This effect on aphids resulted in a G x Aphid effect on ant abundance (Fig. A1) and on leaf trichome density (Fig. A3).**

L. 492: were there G x wind effects on abundance, richness?

**We did not observe G x Wind effects on arthropod richness (Fig. A5). In order to maintain a clear and coherent manuscript, we restricted our revised analyses of community responses to species richness.**

-----  
Reviewer: 3

#### COMMENTS FOR THE AUTHOR

Barbour et al. present a comprehensive study on the partitioning of variation found among willow genotypes. Using two different large field experiments, the authors test how a number of variables affect the arthropod community on willows aboveground, and the community of microorganisms belowground. The variables of interest include plant genotype, variation in several plant traits, presence of a keystone herbivore (aphids), and an abiotic condition (wind exposure). I am very impressed by the size and quality of the study, and in my opinion each experiment represents the state-of-the-art for field experiments on variation partitioning. Unfortunately, while each experiment is strong individually, the combination of the two the main weakness of this study in my opinion.

The two experiments use the same set of plant genotypes, are carried out in the same geographic area at the same time, and use largely the same statistical toolset. However, there is relatively little integration of the results from the two experiments. As an example, plant traits of the willow genotypes are measured in both experiments, but are never directly compared between experiments. Among the few places where results are compared side-by-side are table 1 and figure 1a-d, but these comparisons raise more questions than they answer. Table 1 presumably is meant to compare relative contributions of genetic and environmental effects in the two experiments, but since most values are missing for the ant-aphid experiment the only meaningful comparison is for the two values of arthropod abundance. It is therefore unclear what message the authors want to convey with this table. Additionally, it highlights that the two experiments were carried out in different years, raising the question whether these (presumably unitless?) net effect sizes can still be compared, and how large the year-to-year variation would be within experiment. Are the values in Figure 1 from different years as well, or were they all measured in the same year? Are the values in panels a-b and c-d meant to be compared directly? If so, the y-axes should be on the same scale.

While the two experiments share many similarities, there are important differences, including total duration (years with data) and traits measured in different years. The two experiments are presented sequentially in each section of the manuscript, and the authors attempt to maintain a logical structure by the frequent use of subheadings. Nonetheless, I often found this back-to-back presentation confusing, especially since the important differences between experiments are generally only mentioned once in the methods. It would be helpful if they were mentioned again when a new result is presented for which these differences are important. In addition, and given the lack of integration of the two parts, the sequential presentation often feels like a listing of results and does not make for very attractive reading.

Overall, the sheer number of information contained in these experiments makes it difficult to follow the author's arguments, and it doesn't help that the manuscript currently reads more like two stories that have been spliced together. In the current state I doubt whether the combination of the two experiments provides a benefit beyond the sum of the manuscript's parts. Given the high quality of the two experiments, it would perhaps be justified to split this story into two, and submit two streamlined manuscripts for back-to-back publication. However, if the authors decide against this approach I would strongly encourage them to attempt to integrate the results more, starting with the points that I criticized above. In that case I think it should also be possible to significantly shorten the manuscript, particularly by taking better advantage of the similarities between the experiments. In several places a section in the methods or results for the ant-aphid experiment is followed by a very similar section on the wind experiment. By reducing some of the subdivisions and condensing similar parts where possible, I think redundancies could be significantly reduced, and the manuscript be made easier to follow throughout. I acknowledge that this will not be an easy task, but one way or another simplification and streamlining of the manuscript will be necessary for it to be read by a broad audience.

**We gave substantial consideration to splitting the manuscript into two, but in the end decided to keep the results in one integrated manuscript. We have rewritten the entire manuscript bearing Reviewer 3's comments in mind. In particular, we made a concerted effort to present all results in standardized units that could be easily compared across experiments, years, and community responses. We also present the results together rather than with repeated back-to-back presentations. By incorporating Reviewer 3's comments, we were able to reduce many redundancies present in the previous version. We hope we have now created an integrated manuscript that is better together than as two separate streamlined manuscripts.**

In addition to this major issue I only have a few more minor comments that I list below.

L164: Was there any colonization by aphids of plants in the control plants? If so, how were these plants treated? The authors claim later in the manuscript that aphid effect on arthropod communities were independent of aphid numbers (L588), suggesting some induced change in the plant triggered by even few aphids. Presumably this could present a problem if controls were colonized by aphids even for short amounts of time?

**Aphids that colonized control plants were removed, which we now make clear on Lines 136-137. The reviewer brings up a valid point, but aphid colonization of control plants was extremely rare. Thus, we do not expect this to have had a major influence on our results.**

L165: The timings of the two experiments are quite complicated. Presumably all data for the ant-aphid experiment is from 2012, but for the wind experiment it isn't always clear if data is from 2012, 2013, or

averaged across years. Personally, I would appreciate a supplementary figure detailing the timings of the experiments in graphical form.

**We have made every attempt to clarify the timing of the two experiments in the Methods and Results sections. We decided not to include a supplementary figure detailing the timings because we felt that the revised manuscript is now much clearer.**

L203: I believe the denominator of the equation should be the 'fresh mass' (the higher weight) for the result to be a fraction.

**No, the denominator should be 'dry mass'. Leaf water content should not be reported as a fraction, which can actually weaken the power of analyses (Munns and PrometheusWiki Contributors 2010, <http://prometheuswiki.org/tiki-pagehistory.php?page=Plant%20water%20content%20and%20relative%20water%20content&preview=6>).**

L420/426: Language in which results are reported should reflect statistical significance. 'Had little effect' with a p-value of 0.46, 'had no effect' with a p-value of 0.195 is counterintuitive.

**Good point. We have revised our language to coincide with the effect sizes reported in the revised manuscript. We report the median estimated effect size along with 50% and 95% credible intervals (percentile intervals of posterior estimates) throughout the Results section.**

L425: I wonder if average numbers are useful for quantification of ant visits, as this depends on all censuses across which numbers are averaged to be identical in all conditions – assuming for example that ant activity followed diurnal patterns but plants were sampled randomly throughout the day, zero values could be meaningless. I'm not sure if it is a better solution, but I would at least try to compare results with the sum of observed ants as well.

**These reported averages were across experimental units. As with each arthropod species, we used the maximum abundance observed on a plant over the summer as the response variables (Lines 196-205).**

L429/L449: 'would be mediated by plant traits' seems incorrect, as 'mediate' implies that plant traits are a separate actor, rather than one aspect of plant genetic variation. 'would be caused' or 'would be due to' would be better. Check usage of 'mediate' throughout.

**We avoid using 'mediate' in the revised manuscript when referring to effects due to plant traits.**

L438: This is a surprisingly strong effect of aphid induction on trichomes, which aren't usually a very plastic trait. I think this could be highlighted more in the discussion, and I would be interested to learn the authors interpretation of this result.

**While we agree this is an interesting result, it is only of minor importance in the revised version of the manuscript. Thus, while we mention it in the supplementary material (Fig. A3), we don't believe it warrants inclusion in the Discussion.**

L451-461: Two sets of results are presented for the wind experiment, the second set described as the end of the experiment. It is unclear to me what these time points correspond to, and this should be explained better (see also my comment regarding the unclear timings in this experiment).

**In the revised manuscript, we have made sure to clearly report the timing of results in the wind experiment.**

L474-475: The genetic units of willow are sometimes referred to as genotypes, and sometimes as clones. For consistency and to avoid confusion the same name should be used throughout.

**We now consistently refer to the genetic units of willow as 'genotypes' throughout the manuscript.**

L570-577: The reporting of the PCA results here is very repetitive and could be significantly condensed.

**We have condensed the reporting of the PCA results to avoid repetition in the revised manuscript (Lines 294-302).**