Metabolome-scale Genome-wide Association Studies Reveal Chemical Diversity and Genetic Control of Maize Specialized Metabolites

Shaoqun Zhou, Karl A. Kremling, Nonoy Bandillo, Annett Richter, Ying K. Zhang, Kevin R. Ahern, Alexander B. Artyukhin, Joshua X. Hui, Gordon C. Younkin, Frank C. Schroeder, Edward S. Buckler, and Georg Jander

Plant Cell. Advance Publication March 28, 2019; doi: 10.1105/tpc.18.00772

Corresponding author: Georg Jander gj32@cornell.edu

Review timeline:

TPC2018-00772-LSB	Submission received:	Oct. 15, 2018
	1st Decision:	Nov. 16. 2018 *revision requested*
TPC2018-00772-LSB1	1st Revision received:	Jan. 31, 2019
	2nd Decision:	Feb. 25, 2019 *revision requested*
TPC2018-00772-LSB2	2nd Revision received:	March 5, 2019
	3rd Decision:	March 7, 2019 *acceptance pending, sent to science editor*
	Final acceptance:	March 27, 2019
	Advance publication:	March 28, 2019

REPORT: (The report shows the major requests for revision and author responses. Minor comments for revision and miscellaneous correspondence are not included. The original format may not be reflected in this compilation, but the reviewer comments and author responses are not edited, except to correct minor typographical or spelling errors that could be a source of ambiguity.)

TPC2018-00772-LSB 1st Editorial decision – *revision requested* **Nov. 16, 2018**

Reviewing editor comments and **author responses**:

Point 1. How were metabolite identities assigned and information on ones that are validated using standards should be presented? There are also other concerns about metabolite identification.

RESPONSE: Benzoxazinoids were identified based on known masses, experiments conducted with purified standards for prior publications, known benzoxazinoid profiles of 25 maize inbred lines that are included in the GWAS panel, and several years of experience in the identification and analysis of maize benzoxazinoids. Phenylpropanoid hydroxycitric acid esters were identified based on high-resolution mass spectrometry and NMR.

To validate the claims that were previously supported only by UV absorbance profile-based identification, we have now included an MS2 analysis of whole leaf extract of B73 seedlings. We queried 94 spectra associated with major peaks on the total ion chromatogram in the ReSpect for phytochemical database and acquired probable identification of 19 of them in addition to 6 benzoxazinoid compounds we have previously identified. The resulting raw MS2 files are uploaded alongside with all other LC-MS files in Cyverse (doi.org/10.25739/9dsj-kw33), and the representative MS2 spectra used for identifications are included as supplemental data.

We recognize that UV absorbance profile-based identification is coarse and agree with the reviewers that the retention time window-metabolite class association is imperfect. However, as shown in Figures 2 and 3 and acknowledged by reviewer #2, this approach does provide an efficient proxy to test hypotheses and provide biological insights when performing metabolome and genome-scale analyses.

Point 2. There were noted discrepancies in the genotype numbers, and other Materials and Methods information needs to be addressed.

RESPONSE: These inconsistencies arose from failed germination, sample loss during processing, and failed LC-MS runs. Hence, not all genotypes were analyzed for both tissue types under both ionization modes. The actual numbers of usable spectra are described in lines 130-133 of the revised manuscript. All of these MS spectra have

also been deposited in Cyverse, doi.org/10.25739/9dsj-kw33. Rather than throwing out any genotype that did not have the "complete" set of LC-MS data, we included whatever was available to maximize the statistical power for GWAS, which was done for each tissue type-ionization mode combination independently.

Point 3. How will the data be distributed and made available?

RESPONSE: Excel file summaries of large LC-MS and GWAS data sets are included as supplementary material with this submission. Raw LC-MS result files and GWAS mapping data have been deposited in Cyverse, doi.org/10.25739/9dsj-kw33. This database is freely accessible to the public.

Point 4. The heritability calculation using solely a single genotype was of concern to several reviewers.

RESPONSE: We agree that calculating heritability based on only one genotype does not provide a full picture of maize metabolism. However, this calculation also addresses Reviewer 1, Point #5, which asks about the variance of metabolites in the repeated measurements of B73 versus variance in the population as a whole. Given the available time and resources, it was not feasible to collect and assay multiple samples from each maize inbred line. We have added text to the Discussion section (lines 452-461) to explain the limitations of our broad sense heritability analysis.

---------- Reviewer comments:

[Reviewer comments shown below along with author responses]

TPC2018-00772-LSB1 1st Revision received Jan. 31, 2019

Reviewer comments and **author responses**:

Reviewer #1:

In this manuscript, Zhou and collaborators performed a non-targeted metabolomics analysis in leaf tip and leaf base tissues of young seedlings of the so-called Goodman panel, a set of 286 maize inbred landraces for which numerous studies have been published, including a recent RNA-Seq paper on different plant tissues. The authors performed an analysis of the genetic architecture and genome-wide properties of the metabolome. They then showed how this dataset can be used to perform GWAS analysis and were able to find already known hits and new ones. Overall, I think this is a very valuable dataset for researchers interested in maize and the metabolomics community in particular and the plant biology community at large.

The paper would benefit from a reorganization of the content to first go over the metabolic genome-wide architecture and then provide examples of specific metabolites. More detail should be given about how the identification of compounds was performed and on how the data is going to be shared with the rest of the community to make it useful and easily reused by others. See below my comments and suggestions.

Point 1. No description is provided on how the mass features were assigned to compounds. Did the authors use standards? Was the identification based on previously published mass spectra? This should be included in the Materials and Methods section.

RESPONSE: We identified benzoxazinoids based on known retention times and fragmentation patterns that we are quite familiar with from prior research with maize mutant lines and purified standards. These prior publications have been cited in the Methods section, lines 640-641. In the revised manuscript, we provide additional NMR data that were used to identify the phenylpropanoid hydroxycitric acid esters (Supplemental Figure 6 and Cyverse, doi.org/10.25739/9dsj-kw3). Other compound identities were tentatively assigned by comparing the MS2 data to ReSpect for Phytochemical online (Sawada et al., 2012), as described in lines 631-635.

Point 2. Related to the previous point. The authors claim (and I agree with them) that the data here will be very valuable to the community and highlight several possible uses. Can the authors comment on how are they going to make their data available? Will it be deposited e.g. in maizegdb where it can be easily queried, searched? Will mass spectra will be provided so searches against mass spectra can be run?

RESPONSE: This project has generated two types of large data sets: GWAS mapping data and LC-MS data. Excel summary files are provided in the supplementary data with this manuscript. For the majority of users, these supplementary data files will be sufficient to link a mass feature to a specific gene or, conversely, determine whether a specific gene is associated with a metabolite QTL. For more specialized readers, the full genetic mapping data and raw MS data files have been deposited in Cyverse, doi.org/10.25739/9dsj-kw33. Unfortunately, MaizeGDB is not currently set up to take MS and GWAS data.

Reviewer #2:

In this study the authors measured 3,991 mass features that were taken from the base and the tip of the leaf of a 282 association panel. Using this data, they characterized the maize leaf metabolic diversity across the different genotypes. Interestingly, the data shows that the maize varieties were predominantly differentiated by benzoxazinoid. GWAS of these mass features identified candidate genes and hot spots of often related masses. Confirmation of these results was done by comparison to known metabolic genes as well as by QTL mapping. The authors suggest that the data generated could be used for public data mining and linking metabolites with biosynthetic and regulatory genes. Several metabolomics GWAS papers have previously been published; however this study presents several novelties and interesting approaches:

1) This study characterizes specific leaf zones from maize seedlings, which represent distinctive differentiation phases that are known to effect specialized metabolism and therefore helps to overcome the signal dilution and confounding effects caused by analyzing a whole tissue. In some cases, the authors showed the relevance of their findings by comparing them to publicly available gene expression data. This approach demonstrated nicely that there could be different regulatory genes effecting different metabolites or similar metabolites within these leaf zones, highlighting the complexity of the metabolism on one hand, but also the limitation of GWAS of metabolic traits on the other hand. This work suggests that in order to drill down the metabolic regulation, we need to go into sub-tissue levels.

2) This study demonstrates how correlation network analysis and GWAS can help structural and functional assignment of many unknown metabolites. Although this is not a new concept, its implementation in this study highlights the efficiency of this approach.

3) This study provides interesting insights into the maize metabolite general genetic architecture, class distributions, occurrence and regulation across the different subpopulation. The authors' approach to segregate the mass features by UV profile was demonstrated to be efficient in characterizing patterns as well as annotation.

4) This study performed mGWAS using very new SNP data, which helps increase the resolution of the mGWAS.

Point 1. In figure A: the authors show, using PCA, that there was a separation between the leaf zones but not the subpopulations. No explanation was offered, even though this is unexpected. Is this a pattern we might expect in general? Is this unique to maize? How does that fit with the finding of many relatively rare metabolites throughout the panel?

RESPONSE: If most of the relatively uncommon metabolites are randomly distributed in the analyzed population, we would expect that PCA would not allow differentiation by sub-population. We have provided additional discussion of this observation in the revised Discussion section (lines 469-471). As shown in Figure 2, there are few metabolites other than those co-eluting with known benzoxazinoids that vary in a subpopulation-specific manner.

Point 2. It is unclear from this manuscript whether there are any previous findings showing if flavonoids are only expressed in the leaf tip vs. leaf base. No discussion was offered or reference in the Introduction.

RESPONSE: We provided some references regarding developmental regulation of flavonoids and other maize metabolites in the revised Discussion section (lines 465-466).

Point 3. Confirmation of novel genes was not done at the transgenic level for new candidate genes but they were confirmed using data from QTL mapping. This is a legitimate approach; however, it would have been interesting to

conduct such an experiment for at least one of the candidate genes underlining the hotspots or at the very least suggest in the discussion what we would expect if we to do this.

RESPONSE: We mention the possibility of transgenic approaches or transient gene expression experiments in the revised Discussion (lines 528-530). It was not possible to conduct transgenic maize experiments in the two-month time period that we were given to revise this manuscript.

Point 4. Figure 2A: although the range of flavonoid shows higher amounts of significant metabolites, so does the benzoxazinoid range and the outer range. This figure might need further clarification to support the flavonoid claim.

RESPONSE: Yes, it is true that metabolites/mass features scattered across all intervals of the chromatogram show significant variation by tissue type. However, as shown in the top panel of Figure 2D, the density of this distribution is significantly higher in the time interval where known flavonoids eluted. This concept is explained in more detail in the revised text.

Point 5. The samples used for the metabolic analysis were composed of pooled leaves from two seedlings. No discussion was offered, even though it is customary that more tissues and more repeats are usually sampled when performing metabolic analysis

RESPONSE: In the current study, the replicates for GWAS are not the individual plant lines, but rather the alleles at each locus. Since all loci are bi-allelic (non-biallelic ones were filtered out), that means each allele is sampled an average of 141 times in the entire Goodman diversity panel (fewer since we did not obtain data for all lines). There are thousands of genotyped maize inbred lines available. The resources that were available for growing maize and running mass spectrometry assays limited our experiments. If resources were available to analyze additional samples, more information would be gained by analyzing additional inbred lines, rather than running replicated samples of the current set of inbred lines. We have added a paragraph to the Discussion section explaining this (lines 420-429).

Reviewer #3:

The work by the authors constructs non-targeted metabolic profiling and a follow-up mGWAS analysis of a maize population. Although some progress in maize metabolism has been described in the work, the overall quality and novelty need substantial improvement. For example, despite profiling thousands of metabolic features, the authors only tentatively annotated a few metabolites without any attempt to identify any of them. A number of conclusions drawn by the authors have been reported in previous studies, while others are not convincing. Moreover, in addition to the reported genes, the authors need to confirm the function of newly assigned genes underlying the metabolic traits of interest. The data presentation and manuscript readability need substantial improvement.

Point 1. The descriptions of the protocol for plant growth and the extraction of metabolic samples are unclear.

RESPONSE: We have provided more details in the Methods section regarding sample growth and extraction.

Point 2. The authors should provide a detailed list of plant materials. For example, the number of maize materials mentioned in Supplemental Dataset 1 does not correspond to the 264 mentioned in the manuscript (lines 49).

RESPONSE: Due to the lack of seed germination for some maize inbred lines, losses during sample processing, and occasional low-quality UPLC-MS runs, full spectra were obtained for the following: leaf tips, negative ionization (221 inbred lines and 17 B73 control); leaf tips, positive ionization (258 inbred lines and 25 B73 control); leaf bases, negative ionization (220 inbred lines and 22 B73 control); and leaf bases positive ionization (223 inbred lines and 22 B73 control). (lines 127-132 in the revised manuscript).

Point 3. The authors claimed that they used the same conditions to obtain metabolic materials corresponding to previously published transcriptomes. However, the overall batch-to-batch difference in light, temperature, water, and fertilizer control has to be considered before comparison. They should obtain transcriptome data for these 264 inbred lines directly, or use previous materials for metabolic analysis.

RESPONSE: In the two-month time period that was provided for the revision of this manuscript, we were not able to repeat these experiments and measure both metabolite content and gene expression in the same plant samples.

The use of pre-existing transcriptome data for comparison to new experiments is a common practice in plant research, including research that has been published in Plant Cell. For instance, many Arabidopsis researchers make use of gene expression databases such as Genevestigator for comparisons to their newer data sets. We have to assume that there is at least some relevance in comparing previously published gene expression data with subsequent experiments. Otherwise, why even bother publishing transcriptome data? In our experiments, we have gone further than most other researchers in trying to replicate the exact growth conditions that were used for the earlier transcriptome studies.

We mention in the revised Discussion that there is like to be variation due to the fact that the transcriptome data and metabolome data do not come from exactly the same plants (lines 430-439).

Point 4. For large-scale population, the authors should quickly freeze the samples (such as putting them into liquid nitrogen). All samples are vacuum-dried and then weighed to avoid metabolite degradation.

RESPONSE: As described in the Methods section, samples were harvested, weighed, immediately frozen in liquid nitrogen, and stored at -80°C until analysis by mass spectrometry. We chose to weigh and analyze frozen leaf samples rather than samples that had been vacuum-dried. There is no inherent benefit to freeze-drying samples before extracting them for analysis by mass spectrometry. In fact, some metabolites get lost during the freeze-drying process.

Point 5. In addition, is there only one biological repeat in the study?

RESPONSE: Yes, there is only one biological repeat of each maize inbred line. As described above in the response to Reviewer 2, replicates in this analysis are the individual alleles at each locus in the genome. This is also explained in lines 420-429 of the revised manuscript.

Point 6. The authors obtained all the metabolic data using high resolution mass spectrometry (Thermo QE), however, the structure of metabolites were assigned using only the UV absorption chromatograms. Metabolite annotation should be performed using MS2 data and the identification should be carried out for the "key metabolites" described in the study. The authors also need to provide all raw mass data, including the m/z values, the RT, and the fragmentation patterns (MS/MS) used in this study. In addition, in 2D-NMR spectroscopy, the authors should supplement the map information of HSQC spectra and HMBC spectra.

RESPONSE: We provide the MS2 data for key metabolites mentioned in this study as Supplemental Data 4. All raw mass data have been deposited in the Cyverse data repository, doi.org/10.25739/9dsj-kw33. The HSQC spectra and HMBC spectra have been deposited in Cyverse (doi.org/10.25739/9dsj-kw33) for the more specialized readers who want to look at them. Our current supplemental table should be sufficient for an audience with a basic structural chemistry background. Having static pictures of the spectra doesn't really help an advanced audience since the signal-to-noise ratios cannot be customized and presented well in this format.

Point 7. In the Methods section, each extract was analyzed for positive and negative ion patterns. Why are the number of samples in Supplemental Dataset 1 and 2 inconsistent?

RESPONSE: The inconsistency in the number of samples results from the fact that no usable data were obtained from some chromatograms. We have modified the manuscript to state how many maize inbred lines were analyzed in the positive and negative ionization modes, respectively, for the leaf tip and leaf base samples. (lines 127-132)

Point 8. There are a large number of repetitive features in positive and negative ions mode. The authors need to should describe in detail the criteria and methods for filtering the dataset in terms of the removal of redundant features within each mode and also between the two modes.

RESPONSE: We could not find an automated way to confidently assign repetitive features across ionization modes, which were analyzed separately through the XCMS-CAMERA pipeline. So, rather than removing data, we decided to take a more conservative approach and leave all results in the analysis. The specific parameters that we used are described in the Supplementary Methods.

Point 9. Line137-138. I am not sure about this conclusion for the fact that there are a large number of other metabolites that co-elute with these metabolites. They may either not have UV absorbance or with their UV spectra covered by the metabolites with high abundance, such as flavonoids as phenylpropanoids. Why not perform

metabolite identification and/or annotation when doing this analysis?

RESPONSE: Irrespective of the presence of other metabolites in that part of the chromatogram, absorbance data make it clear that the leaf tips have a higher abundance of flavonoids than the leaf bases (Figure 3). We have included an MS2 analysis of whole leaf extract of B73 seedlings in the revised manuscript to address this issue. However, as the reviewer is probably aware, in the absence of purified standards, it is not possible to assign molecular structures based solely on mass spectrometry data. For this reason, 90% or more of all metabolites in any plant species are structurally unknown. It was not possible to solve this problem in the two months that we were given to revise this manuscript.

Point 10. Line143-145. It may be reasonable the certain types of metabolites have the same or similar accumulation patterns. However, considering the co-elution of metabolites of different chemical category, it is therefore better to carry out accumulation pattern analysis for each specific type of metabolite. It is difficult the assign the chromatogram to a specific mass feature in complex samples, especially for those of low abundance.

RESPONSE: Given the uncertainties of assigning identities of individual metabolites, it is better to do a global analysis as we have presented here. We and many others have found that metabolically related metabolites have similar retention times (yes, of course there are exceptions, but we are looking at overall patterns). Therefore, the observation that different parts of the chromatogram differentiate either tissue type or genotype is an indication that there are overall patterns in the data set. In particular, this allows us to say that flavonoids are different between leaf tip and leaf base and benzoxazinoids are different between tropical and temperate maize. Such conclusions could not be derived from individual metabolites and instead depend on a broader analysis of the relevant data.

Point 11. The paragraph starts from Line 169. This finding has been reported previously by a number of studies.

RESPONSE: We have changed this paragraph to indicate that co-regulation of metabolically related metabolites is commonly observed in plants (lines 194-196 in the revised manuscript). The purpose of our analysis was to look at this phenomenon on a global scale in maize, rather than investigating individual pathways or sets of metabolites as is more commonly done.

Point 12. Line 182-184. There are two problems with this conclusion: 1) although structurally related metabolites may co-elute, a large number of structurally unrelated metabolites co-elute due to the complexity of the (plant) metabolome. 2) Structurally related metabolites may not necessarily co-elute. Metabolites modified by their glycosylation and aromatic acylation will result in significant decreased and increased retention in RP-HPLC.

RESPONSE: We agree with this statement. Yes, of course there is variation. However, there are overall patterns that are nevertheless apparent from these data. The two situations that are described (presence of unrelated metabolites and modifications that change the retention time) would add noise to the data and decrease the chance of seeing the overall patterns that we mention. We have added Supplemental Figure 2, which further confirms the absorption patterns characteristic of phenylpropanoids, benzoxazinoids, and flavonoids. We should also point out that this is not a new observation. Plenty of other scientists have noted that particular classes of metabolites, e.g. flavonoids, have similar retention times.

If a QTL affects the abundance of multiple mass features, these mass features tend to have similar retention times (Figure 8). This is also consistent with the hypothesis that metabolically related metabolites have similar retention times in our data set.

Point 13. Line 227-234. Direct evidence is needed to make the paper stronger. This is also true for the conclusion as described from Line 320-339.

RESPONSE: Lines 227-234. Although GRMZM2G108309, a gene model encoding a predicted protein phosphatase 2C family protein, is the only annotated gene located near this genetic mapping interval in B73 and other published maize genomes, we are not yet certain that this is actually the causative locus. In particular, we would have to assume that a 3' regulatory region is affecting the function of this gene. Since B73 is missing about one third of the genes of the maize pan-genome, it is quite possible that there are additional genes present in this mapping area in other inbred lines. Effectively, we could be mapping the absence of a gene rather than natural variation in a gene that is present in sequence maize genomes. Therefore, it is not yet

reasonable to try to obtain direct evidence for the enzymatic function of GRMZM2G108309. It also is not feasible to sequence additional maize genomes in the two-month revision time provided for this manuscript. Other groups are funded to do large-scale sequencing of multiple maize genomes. Therefore, we intend to take a "genome waiting" approach to determine whether other candidate genes are downstream of GRMZM2G108309 and could account for the observed benzoxazinoid phenotype. This rationale is explained in the revised Discussion section (lines 503-517).

Lines 320-339. The genetic mapping data are quite unequivocal in showing that SNPs in the coding region of GRMZM2G063909 are the cause of genetic variation in the metabolite abundance. We agree that cloning an in vitro enzyme activity assays are necessary to confirm the function of GRMZM2G063909 in the synthesis of phenylpropanoid hydroxycitric acid esters in maize. This is clearly and interesting chemical reaction and, to the best of our knowledge, has not been described previously. However, this is beyond the scope of the current project and it was not possible to complete these experiments in the two-month revision time for this manuscript.

TPC2018-00772-LSB1 2nd Editorial decision – *revision requested* Feb. 25, 2019

There are a few editorial clarifications requested by one of the reviewers that identify some issues that readers may be confused by. This round of revision will not go back out for review but handled internally.

TPC2018-00772-LSB2 2nd Revision received March 5, 2019

Reviewer comments and **author responses**:

Reviewer #1:

The authors have addressed most of the comments made by the reviewers in the previous submission. I do not have major concerns just some minor comments/suggestions:

Point 1. Line 140: I am not that surprised about this result. The genetics populations defined by Flint-García are still very diverse within them, and there are many compounds that are unique to a few lines or not. A highly replicated experiment using one representative line of each subpopulation would probably be well separated in a PCA by genetics.

RESPONSE: This comment is similar to the explanation in the Discussion section (Lines 477-480 in the revised manuscript).

Point 2. Line 22 (Fig 6B): It is confusing to me how the average H2 of those features that have an H2>0.2 can be (much) lower than 0.2. See right panel of figure 6B, bin of mass features present in 0.1-0.2 genotypes. Also, for consistency, it would be better to express this values in % like in the panel above. Fig 6C: Again, for consistency, label X axis as Genotypes Present In.

RESPONSE: We presume the reviewer means "those features that have an H2<0.2", i.e. the white bars that represent the averages for all metabolites. This would be the white bars for 0-10% and 10-20% in the right panel of Figure 6B. We checked our calculations and they are correct. As we mention in the text of our manuscript, the heritability calculations are somewhat skewed because a relatively small number of B73 plants was used to estimate environmental variation, and a much larger number of plants was used to estimate genetic variation. This explains the negative average heritability in 0-10% column in the left panel of Figure 6B and the very small values in the right panel for 0-10% and 10-20%. In the right panel, some negative values caused the averages to be lower than 0.2. The gray bars, which do show heritability for "those features that have an H2>0.2" are all considerably higher than 0.2, with average heritability ranging from 0.55 to 0.92. The Figure 6B X-axis has been changed to percentages. The Figure 6C X-axis labels have been changed to "% Genotypes Present In"

Point 3. Line 227. While I agree with this reasoning, running blank injections would allow the authors to check if the less common features are present in blank injections due to background signal of material used/solvents/column etc.

RESPONSE: Blank injections were run at the beginning of each batch and between every 60 runs within the batch.

Thus, we were able to rule out background signals. This procedural detail was omitted in the previous versions of the manuscript and was added to this revision (Lines 600-602).

Point 4. Line 427: I would agree with this unless you want to try to map rare metabolites or when the heritability is low. In that scenario, increasing the number of replicas per line rather than the number of lines would be preferable I think. The authors even mention in line 246-247 that it is not possible to map metabolites that are present in less than 10% of the lines given the size of the population. Increasing the number of replicates per line would help there. This is assuming that the rare metabolite would also be rare in the new added lines.

RESPONSE: We do not agree with this line of reasoning. Having replicates of the individual lines will not increase the statistical power of the GWAS analysis or the number of recombination breakpoints that allow genetic mapping. We agree that running replicate samples of the same lines would increase accuracy. However, we would gain greater benefits by analyzing an equivalent number of genetically independent maize inbred lines.

As an example, if a rare allele is present in not 20 inbred lines, we probably would not be able to map it. If we measured our original mapping population in triplicate, we would measure the phenotypic effects of that rare allele 60 times, but we would not gain much additional mapping resolution. On the other hand, if we tripled the size of our mapping population, we would not only measure the phenotypic effects of the rare allele 60 times but would also get improved mapping resolution from 40 genetically independent maize inbred lines. Yes, of course this assumes that the rare alleles would also be rare in a larger sample set of maize inbred lines. That is something that would need to be determined empirically.

Point 5. Line 436-438: I understand why the authors added this comment based on the previous revision of the manuscript. In my opinion, even if the plants had not been grown in the same greenhouse and conditions, it would have been fine to compare and combine their data with that of Kremling et al.

RESPONSE: We agree with the reviewer's comment. However, since this point was a concern for another reviewer and perhaps also some future readers of our manuscript, we prefer to leave this section of the text as it is.

Reviewer #2:

In this study the authors characterized the metabolic diversity and genetic architecture of the non-targeted metabolome measured from either the tip or base of leaves belonging to the Goodman diversity panel. The authors used extensive metabolic and SNP data, which helped uncover novel information about the overall metabolic diversity of maize leaves as well as its genetic architecture. Interestingly, the authors showed that only a small fraction of the metabolites are detected in the majority of the lines and that most mass features were found in less than half of the lines. Using a unique approach, they have also discovered that the leaf tip and base are differentiated by flavonoid while the subpopulations are mainly differentiated by benzoxazinoid. In addition the data set produced in this study was also used to show that metabolically related structures tend to be co-regulated. Although this is not a novel concept, the data set generated can help with the identification of unknown compounds in the future. The GWAS performed on the data uncovered novel associations that can be utilized by the community as well as novel insights on the metabolome genetic architecture. The study shows that the genetic architectures of different leaf zones, which represent different differentiation phases do not completely overlap. The analysis also identified a few genomic hotspots of the two leaf zones that were not completely overlapping. The efficacy of the GWAS was nicely demonstrated by comparing the found QTLs to known QTLs and cross referencing the findings with expression data and RIL data. Overall this study uncovered novel insights on the maize metabolome and provided an important community resource for future discoveries. The manuscript is well presented and demonstrates nice integration of several approaches to provide better understanding of the genetic basis of the maize metabolome

Point 1. On line 237, the title claims that specialized metabolites are not strongly influenced by occurrence, but this is not discussed in the text, even though the data is shown in figure 7.

RESPONSE: In the Results section, we focused the presentation of figure 7B on the only clear pattern we could identify (Lines 273-278). An additional sentence has been added to clarify that we could not make further interpretations based on the current level of analysis (Lines 274-275). These points were further mentioned in the Discussion section (Lines 474-483).

TPC2018-00772-LSB2 3rd Editorial decision – *acceptance pending* **March 7, 2019**

We are pleased to inform you that your paper entitled "Metabolome-scale genome-wide association studies reveal chemical diversity and genetic control of maize specialized metabolites" has been accepted for publication in The Plant Cell, pending a final minor editorial review by journal staff. At this stage, your manuscript will be evaluated by a Science Editor with respect to scientific content presentation, compliance with journal policies, and presentation for a broad readership.

Final acceptance from Science Editor **March 27, 2019**