nature portfolio

Peer Review File

Integrative multi-omics analysis reveals genetic and heterotic contributions to male fertility and yield in potato

Corresponding Author: Professor Chunzhi Zhang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Because of the tetrasomic feature, genetic analysis of cultivated potato is difficult to carried out for mapping genes and studying their genetic effects. In this manuscript, the authors constructed a F2 population by crossing two diploid homozygous potatoes and generates transcriptomics and metabolomics data to discover genes involved in heterosis in potato genome. They identified a PME gene playing important role in male-fertility, contributing to the formation of heterosis in potato. Multi-omics association analyses using F2 population of hybrid diploid potato is a significant novel idea. The knowledge obtained from this work, making potato as a seed-prorogated diploid crop with hybrid vigor, may greatly contribute to the world-wide potato industry. I have multiple lines of suggestions that hope the authors to consider for further improvement of this manuscript.

Major comments.

1. Line 160-162: The authors classified all the QTLs to local and distant eQTLs. Does the definition of “local and distant” correspond to cis and trans”, respectively? To my knowledge, cis and trans are more commonly used to describe the action of a QTL on target gene. Also, the author should mention explicitly, how trans and cis are defined. It’s not simply based on distance between gene and QTL. Additionally, this statement “A distant eQTL hotspot could regulate the expression of hundreds of genes” is not absolutely correct. A cis/distant QTL usually refers to a transcription factor or other type of regulators that influence expression level of downstream genes. A cis/local QTL usually refers to genomic regions, e.g. promoters, enhancers or other regulatory SV/InDel elements that can also influence numerous genes located in the same genomic region. Anyway, I suggest the authors to clarify the definition of local/cis and distant/trans based on possible function.

2. Line 167-168: “highly correlated metabolites suggest that they might be controlled by the same regulator in this population”. This statement is also not absolute. Many metabolites are synthesized in a cascade fashion in one pathways. If one metabolic compound located in the upstream stage of the pathway is influenced, it will cause co-change of their downstream metabolic products. As a matter of fact, it’s better to do a dimension reduction of a set of clustered metabolites with high correlation using PCA or other algorithms, and then use PC1 or PC2 to represent the expression pattern of a cluster of metabolites to do GWAS or QTL mapping. The authors may take a look at the MODAS paper by Liu S et al. 2022.

3. The authors utilize GAME9 gene, an AP2/ERF transcription factor, as one example to illustrate their success in identifying a major regulator that influence the pathway of solanine biosynthesis. I am wondering whether the expression level of GAME9 genes and the clustered solanine metabolites are significantly different in the two parental inbred-line. This information should be included in Fig. 3, as it may perfectly explain why gene expression and metabolite are segregated in the F2 population. Additionally, I also wonder whether there is a genomic variation, SV or large InDel or PAV or haplotype, can be found in the neighborhood of GAME9 genes. It would be a good example to show that genomic variation as a cis-acting QTL influencing gene regulatory network and metabolic network. If possible, if the authors possess wild ancestor genotype data of potato and different potato germplasm, it will be also good to check whether this gene is subject to domestication or adaptive selection based on nucleotide diversity. These results will add more story for this work. The current content of this work is too descriptive and lack biological stories.

4. If I understand correctly, the authors may compute a MPH (mid-parent heterosis) value for each F2 hybrid based on the

comparison with the phenotype of parental inbred line, using yield-related trait. Then the MPH value, which may represent the heterosis degree, can be used as a trait to do QTL mapping or GWAS. I am wondering if the authors can use one or a few representative yield-related traits to do such an analysis, and see if the heterotic QTL can be also found by this method.

5. In the section of “QTL mapping of the multi-omics traits”, Line 115, the authors mentioned that “We mapped several qualitative traits to known loci using bulked-segregant analysis (BSA)”; In the Method part line 428, the authors mentioned that “we conducted transcriptome sequencing of developing tubers of 204 F2 individuals (80 days after transplanting)”. I am wondering how the 204 F2 individuals is selected. Were they selected based on the BSA result, and what qualitative traits they used for BSA? Did different traits have different subgroups of individuals with contrasting traits? All of these information is important but not sufficiently provided in the Method part. It’s the same question regarding the selection of 215 F2 individuals (120 days after transplanting) for metabolome profiling.

6. Another problem is that, transcriptome and metabolome profiling utilized different developmental stages, which are 80 and 120 days after transplanting. I am not familiar with potatoes, but either transcriptome and metabolome should be changed quite a lot in 40 days’ development. The authors may have a reason to used different developmental stages, and it’s better include the explanation in Discussion part. Moreover, were the metabolome and transcriptome of parental inbred lines profiled in the same stage? I didn’t see this information was mentioned in anywhere.

7. Data analysis of the entire paper was based on computing Pearson correlation between gene expression, metabolite abundance and phenotype traits. Then, the relationship with “significant” P values was detected to build WGNCA network or co-localize QTLs. As we all know, high-dimensional omics data including transcriptomic data, metabolomic data and imaging data are extremely complex and heterogeneous. Therefore, I believe quite big proportion of the detected “significant relationship” are false-positive and perhaps, actual biologically meaningful relationship of among eQTL, mQTL and various traits were not detected. Since this is not a methodology paper, and the datasets are original and the authors offer good examples with experimental validation on gene function, I am not picky on the algorithms or methods they used in this work. However, I believe these datasets are quite insufficiently mined.

Minor comments

1. Fig S1a, X axis better indicates the chromosome numbers;
2. Fig S1c, Y axis is loge, or log2, or log10?
3. Line 489, “Detection of pollen viability and pollen tube gemination assay”, it should be tube germination? Also line 497.

Reviewer #2

(Remarks to the Author)
This paper reports two interesting results from the genetic and phenotypic characterization of the diploid F2 potato population with 1,064 individuals derived from the cross of the two highly homozygous inbred lines. One is identification of the yield-related QTLs in the potato lines; the other is identification of the PME gene that involves pollen formation in the potato lines, providing an insight into the mechanisms underlying heterosis related to yield and male fertility. The results could be interesting to the colleagues who work in the fields of plant science and crop breeding.

The main concerns:
1.Concerning the yield-related QTLs:
(1) What are the main genes-encoded functions that may contribute to the heterosis of yield?
(2) What are the molecular mechanisms that may result in the yield-related heterosis in the hybrids, at the levels of protein functions or at the levels of gene expression or both aspects?
(3) Is there any molecular genetic evidence for the yield-related heterosis from the gene mutations (natural or artificial)?

2. Concerning the identification of the PME involved in the male-fertility heterosis:
(1) Genetic complementation of the gene-edited mutants is required to confirm that the pollen-defective phenotype actually was caused by the PME-edited mutations although multiple PME-edited alleles were obtained due to the possibility of mistargeting.
(2) In generation of the gene-edited mutant plants via tissue culture, the tissue-chimeric gene-edited plants could be often found. Therefore, what about the gene-editing rates in the PME-edited mutant plants used in the assays for pollen development?

Minor revision
Writing of the manuscript needs to be further improved.
1. The vague expression of many sentences needs to be revised, for examples:
(1) "…… should be……";
(2) "Fortunately,……";
(3) "Our findings will accelerate……";
and more.

2.The title of the manuscript does not actually express the main interesting achievement of the study, which might be not good for attracting the reading interest of audiences.

Reviewer #3

(Remarks to the Author)
Li. Et al. The genetic and heterotic analyses of an elite diploid potato hybrid.

Summary. This manuscript reports on the construction and genetic analyses of a diploid F2 potato population with 1,064 individuals. The F2 was derived from the cross of two homozygous inbred lines. The authors report that they investigated 20,929 traits generated from 26 multi-omics datasets and identified 32,073 QTLs in these materials. Using gene expression data, they constructed a de novo systems-genetics network in potato, which can be used for gene discovery. They also used these materials to study the genetic basis of heterosis for two traits - yield and male fertility. Overall, they found positive heterotic effects for yield-related QTLs and negative heterotic effects of metabolite QTLs (mQTLs), which they suggested contribute to significant yield heterosis in hybrid potato. Additionally, they identified a PME gene with a dominance heterotic effect that plays an important role in male-fertility in potato. The authors conclude that this study provides significant genetic resources for the potato community and may facilitate the application of heterosis in diploid potato breeding.

Overall Review – This manuscript is well written and relatively easy to follow but I honestly have had a hard time reviewing it. This is mostly due to the extreme scope of the materials presented. To be fair, even though I have worked in potato breeding and genomics and I have fairly significant experience in identifying QTL connected to important traits in potato and using them for marker-assisted breeding in an applied variety development program , them methods and techniques presented in this paper are very new to me. I do have major concerns on the total number of traits reported. Is it really possible to investigate 20,929 traits and identify 32,073 QTL for those traits and make any realistic sense of this volume of data? I am very skeptical of this approach, but I will not entirely discount it. However, I do not recommend accepting this publication as it is currently presented, as it needs major revisions and more focus.

Specific Comments
Results
1. How do you evaluate 20,929 traits. My experience tells me that this is impossible to do at any scale with a high degree of precision and repeatability. Before accepting this I need to know more about the specific traits and how they are connected.
2. I worry about large-scale projects like the one presented here as they have potential to make sweeping statements based on datasets that are not well defined nor well controlled. For example, we revealed the basis of potato heterosis....for what trait and what is the basis. How specifically is heterosis measured and what about epistatic effects? Is this heterosis too?
3. Results Line 92 - please add more data on filtering of the SNPs. Were all of the SNP's used? We have several high-quality reference genomes available. Which were used? I'm thinking DM was one of them as reported in cite 23.
4. Results Line 124 - How do you protect from false discovery with so many QTL?
5. Results Line 125 – one year pollen viability…. or pollen viability during 1 year of the experiments. I read this sentence as pollen viability after storage for 1 year. Reword to make this more clear.
6. Line 135 – "Data generated by hyperspectral imaging helped to detect 5,850 pQTLs of tuber reflectance" I feel like the authors are trying to impress us with huge numbers of QTL but I'd encourage then to focus on important QTL that contribute to deeper understanding rather that blowing the reader away with huge numbers of QTL that we don't understand. Can this be done with a few key traits with a good explanation as to why the traits were selected?
7. Lines 247 – 248 – These are not a very high % explanation. They may be significant but if these are the greatest effects I wonder how important they really are?
8. Line 267 – Is it correct to say that all eQTL are genes? These are expression profiles, right? Is all expression regulated by a gene or genes? Aren't there other methods of regulation?
9. Line 278 - A cautionary note about metabolites, though I am by no means a metabolite expert. But I worry about how the authors seem to be relating the metabolites to genes per se. Based on my discussions with food scientists and plant biochemists, many metabolites are secondary products or breakdown products that have their own energetics and breakdown products are often formed from natural molecular rearrangements due to the cellular environment, etc. Have the authors considered this perspective?
10. Line 289 – "These results further support our findings in metabolites heterosis and provide insights into heterosis of clonally-propagated corps." Based on my observations above I'm not sure that I can make this leap of faith in this statement at this point.
11. Line 300 - A pectin methylesterase contributes to male fertility heterosis – I recommend that this should be a separate paper standing on its own with more robust analyses. I note this work has also be elaborated on in citation 23.

Discussion

1. Overall - Given the sweeping amount of research conducted the discussion section is very short and lacks depth or analysis/interpretation.
2. Lines 378 – 386 - This is a very disjointed section and I don't see good justification for many of the statements made. For example. Tuber dry matter serves as the nutrition provided for asexual reproduction. I'm not even sure what the authors are trying to say here? Is this from an evolutionary perspective? An energetics point of view?

Methods

Collection of phenome data – The authors need to describe the collection of their phenomic data more precisely. I went to the rice paper cited in 62 - 64 to see what was done with rice and there is no comparison to what is reported herein. Please elaborate or…preferably, focus on key traits that make a difference in potato production.

I prefer to let others comment on QTL mapping strategy (e.g., Is CIM the appropriate methods for these studies?) and on metabolomic and eQTL research.

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I have read the revised manuscript and the authors' response to my comments. All of my concerns have been fully addressed. I have no further questions then.

Reviewer #2

(Remarks to the Author)
In this revision, the authors have not address any of all the question points concerned from the last review, but it is understandable for the difficulty to address all the question points due to the current limitation of biotechnology and genetic assays for potato materials based on the explanation from the authors. No more new question from this review.

Reviewer #3

(Remarks to the Author)
This manuscript has been improved considerably from the original submitted. The authors clearly have taken the reviewers comments and suggestion constructively and they have submitted a much-improved manuscript. Thank you.

Here are a two suggested edits based on my most recent review.

Line 84. Importantly, we revealed the genetic basis of heterosis in for male fertility and 21 tuber yield traits in this elite hybrid potato cross (or population?) either population or cross is ok by me.

Line 375 This QTL database will provide useful genetic markers for molecular breeding and
gene discovery in potato.

Discussion
The discussion has improved considerably…However, I feel that it does not end with a good summary. I suggest that the authors add one more paragraph summarizing their work. It can pull from the abstract. But in reading through the discussion I felt like a good summary is still needed to end an otherwise good discussion. Maybe the authors can highlight some of the aspect of this work that they mentioned in the responses to reviewers along the lines of one of the major goals of this research was to provide new genetic resources and phenomic and metabolomic resources for the potato community and to contribute to our collective understanding of heterosis in potato. Something along this line, which you have done in the abstract already, but it would be good to finish strong in the discussion.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.
The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

Author Responses (AR): We sincerely thank the Editor for handling our submission as well as the four reviewers for their constructive advice, which helped us substantially improve our manuscript. In light of the reviewers' concerns, we have performed additional analyses and have extensively revised the manuscript. Our point-by-point responses to all reviewer comments are listed below (in blue font).

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Because of the tetrasomic feature, genetic analysis of cultivated potato is difficult to carried out for mapping genes and studying their genetic effects. In this manuscript, the authors constructed a F2 population by crossing two diploid homozygous potatoes and generates transcriptomics and metabolomics data to discover genes involved in heterosis in potato genome. They identified a PME gene playing important role in male -fertility, contributing to the formation of heterosis in potato. Multi-omics association analyses using F2 population of hybrid diploid potato is a significant novel idea. The knowledge obtained from this work, making potato as a seed-prorogated diploid crop with hybrid vigor, may greatly contribute to the world-wide potato industry. I have multiple lines of suggestions that hope the authors to consider for further improvement of this manuscript.

AR-1: Thank you for these positive comments and your suggestions to improve our manuscript.

Major comments.

1. Line 160-162: The authors classified all the QTLs to local and distant eQTLs. Does the definition of "local and distant" correspond to cis and trans", respectively? To my knowledge, cis and trans are more commonly used to describe the action of a QTL on target gene. Also, the author should mention explicitly, how trans and cis are defined.

It's not simply based on distance between gene and QTL. Additionally, this statement "A distant eQTL hotspot could regulate the expression of hundreds of genes" is not absolutely correct. A cis/distant QTL usually refers to a transcription factor or other type of regulators that influence expression level of downstream genes. A cis/local QTL usually refers to genomic regions, e.g. promoters, enhancers or other regulatory SV/InDel elements that can also influence numerous genes located in the same genomic region. Anyway, I suggest the authors clarify the definition of local/cis and distant/trans based on possible function.

AR-2: We defined local eQTLs as those that were within 100 kb of their target regulated genes, which would be equivalent to *cis*-eQTLs in genome-wide association studies (GWAS). In this study, we identified eQTLs using a segregating F_2 population, considering each bin (2.2 kb – 3.8 Mb) (rather than each SNP, as in GWAS) as a genetic marker. We therefore used the terms local/distant rather than *cis/trans*. Moreover, we have made no inference about whether these eQTLs act as *cis*-regulating elements or *trans*-acting factors. We have clearly stated this in the Methods section (lines 504–505).

Our analysis was primarily inspired by a review article (*Nature Reviews Genetics*, 2006; https://doi.org/10.1038/nrg1964) and a research article (*Molecular Plant*, 2017; https://doi.org/10.1016/j.molp.2017.12.011), based on the following comments: "*because mapping studies reveal the locations of QTLs, an expression QTL can be immediately classified as 'local' (near the genomic location of the gene encoding the transcript) or 'distant' (elsewhere in the genome). Because mapping studies do not reveal the underlying molecular nature of QTLs, we prefer the strictly positional terms local and distant to the commonly used terms cis- and trans-linking, which have implicit mechanistic connotations*" from the review and "*Consequently, cis-regulatory mutations are usually detected as local eQTL in linkage populations*" from the research article.

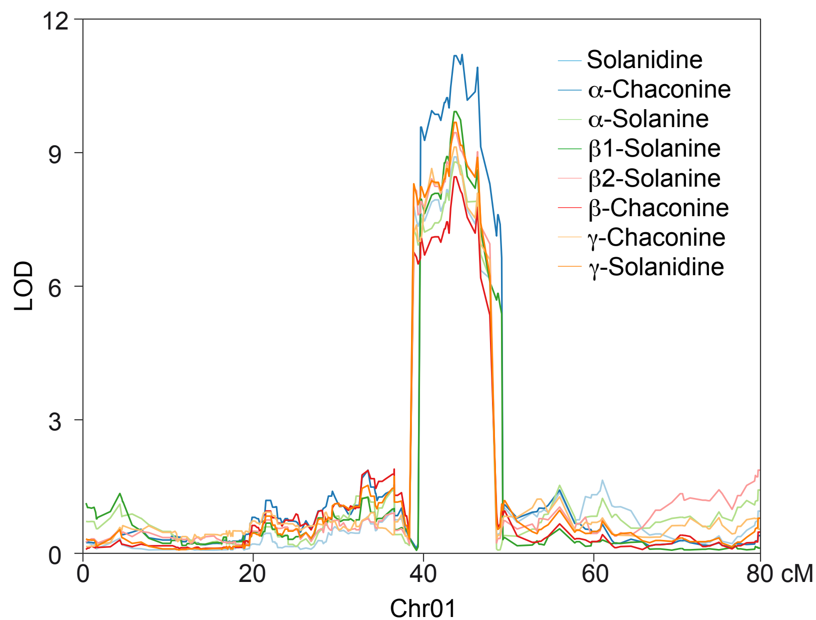
We fully understand your concern about the function of *cis*/*trans* eQTLs, in which a *cis*/*trans* eQTL is a *cis*/*trans*-acting element/factor. In this study, we classified the

eQTLs by their distance from their target genes to partially represent how they act, in line with research in tomato[1], peach[2] and *Brassica napus*[3]. "Local/distant" thus better represent their mode of action than "*cis/trans*". In addition, to avoid any misunderstandings, we have deleted the sentence "A distant eQTL hotspot could regulate the expression of hundreds of genes" (line 163)

2. Line 167-168: "highly correlated metabolites suggest that they might be controlled by the same regulator in this population". This statement is also not absolute. Many metabolites are synthesized in a cascade fashion in one pathway. If one metabolic compound located in the upstream stage of the pathway is influenced, it will cause co-change of their downstream metabolic products. As a matter of fact, it's better to do a dimension reduction of a set of clustered metabolites with high correlation using PCA or other algorithms, and then use PC1 or PC2 to represent the expression pattern of a cluster of metabolites to do GWAS or QTL mapping. The authors may take a look at the MODAS paper by Liu S et al. 2022.

AR-3: Thank you for this suggestion. We have changed the sentence to now read "highly correlated abundance of some metabolites suggests that they might be controlled by the same regulator or affected by a change in the abundance of upstream metabolites from a single pathway" (please see line 169).

We then conducted QTL mapping following the suggested dimensionality reduction for the highly correlated solanine content using MODAS (the parameters -pca). We repeatedly identified the same genomic region by this method (Response Fig. 1). We have added these results in the manuscript on lines 200–203 and Supplementary Fig. 6a.



Response Fig. 1. QTL mapping for solanine content following dimensionality reduction using MODAS.

3. The authors utilize GAME9 gene, an AP2/ERF transcription factor, as one example to illustrate their success in identifying a major regulator that influence the pathway of solanine biosynthesis. I am wondering whether the expression level of GAME9 genes and the clustered solanine metabolites are significantly different in the two parental inbred-line. This information should be included in Fig. 3, as it may perfectly explain why gene expression and metabolite are segregated in the F2 population. Additionally, I also wonder whether there is a genomic variation, SV or large InDel or PAV or haplotype, can be found in the neighborhood of GAME9 genes. It would be a good example to show that genomic variation as a cis-acting QTL influencing gene regulatory network and metabolic network. If possible, if the authors possess wild ancestor genotype data of potato and different potato germplasm, it will be also good to check whether this gene is subject to domestication or adaptive selection based on nucleotide diversity. These results will add more story for this work. The current content of this work is too descriptive and lack biological stories.

AR-4: Indeed, the expression levels of *GAME9* and the solanine contents significantly differ between the two parental lines (Fig. S6c, e and manuscript lines 214–222).

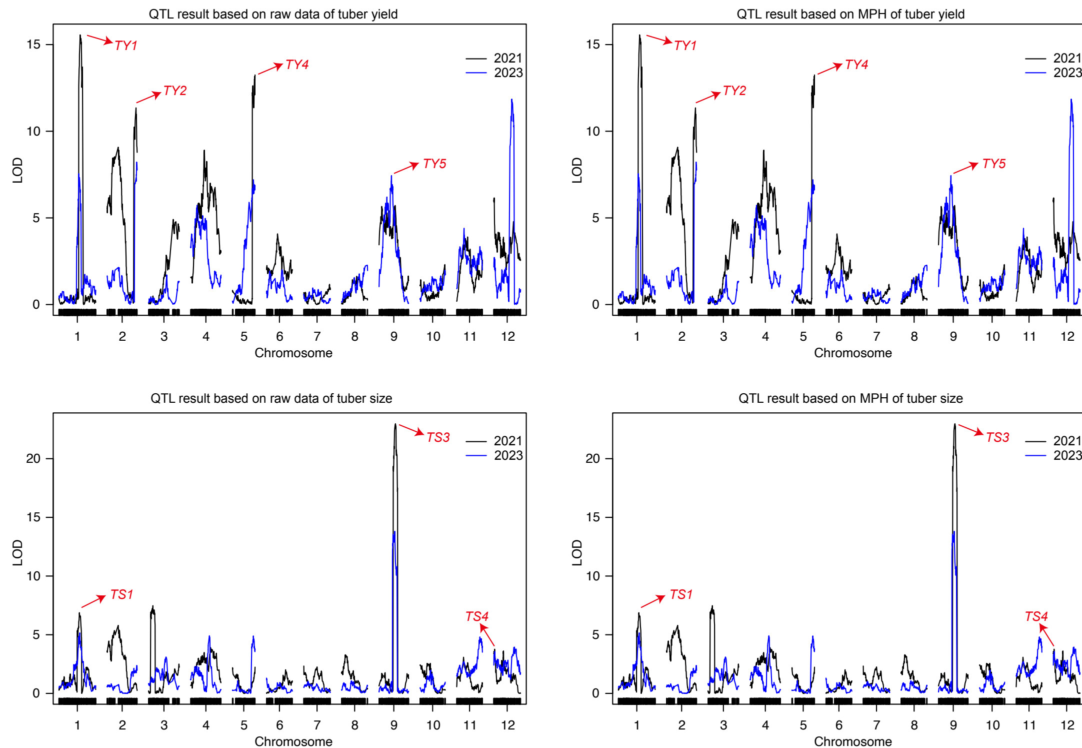
To explore possible causative *cis*-acting elements, we compared the promoter sequence (2000 bp upstream of the ATG) of *GAME9* in the two parental lines (Response Fig. 2). We identified an 11-bp insertion/deletion (InDel) polymorphism located 121 bp upstream of the ATG. This 11-bp insertion disrupts an *activation sequence-1* (*as-1*) element in the E4-63 parent (characterized by low *GAME9* expression levels). The *as-1* element has been reported to activate gene expression in plants[4-6]. We therefore hypothesize that this 11-bp InDel may explain the altered gene expression levels seen for *GAME9*. These results have been added to the manuscript (lines 216–220 and Supplementary Fig. 6d).

As you stated above, solanine content is a domestication trait in potato tubers[7]. In a previous study, Hardigan et al. found that the *GAME9* locus was located within a domestication-selection sweep[8]. We have added this information in the manuscript (lines 222–224).

Response Fig. 2. DNA sequence alignment of the 2000-bp promoter region of *GAME9* in the two parental lines. The *as-1* element is indicated as a red box.

4. If I understand correctly, the authors may compute a MPH (mid-parent heterosis) value for each F2 hybrid based on the comparison with the phenotype of parental inbred line, using yield-related trait. Then the MPH value, which may represent the heterosis degree, can be used as a trait to do QTL mapping or GWAS. I am wondering if the authors can use one or a few representative yield-related traits to do such an analysis, and see if the heterotic QTL can be also found by this method.

AR-5: Thank you for this suggestion. To test your idea, we have selected tuber yield and tuber size to conduct such an MPH-based QTL mapping analysis, as these two traits were mapped to several common QTLs in 2021 and 2023. The LOD profiles using the raw values and their MPH-transformed values were identical for both tuber yield and tuber size in both years. Thus, we identified the same heterotic QTLs mentioned in the manuscript, including the dominant QTLs *TY2*, *TY4*, *TY5*, *TS3*, *TS4* and the overdominant QTLs *TY1* and *TS1* (Response Fig. 3). Since the MPH calculation is essentially a set of arithmetic operations (addition, subtraction and division) of the raw data, which are all linear transformations, the LOD scores did not change. Thus, performing QTL mapping with raw data or their MPH-transformed values leads to the same results.



Response Fig. 3 QTL mapping for tuber yield and tuber size using raw data and MPH-transformed values.

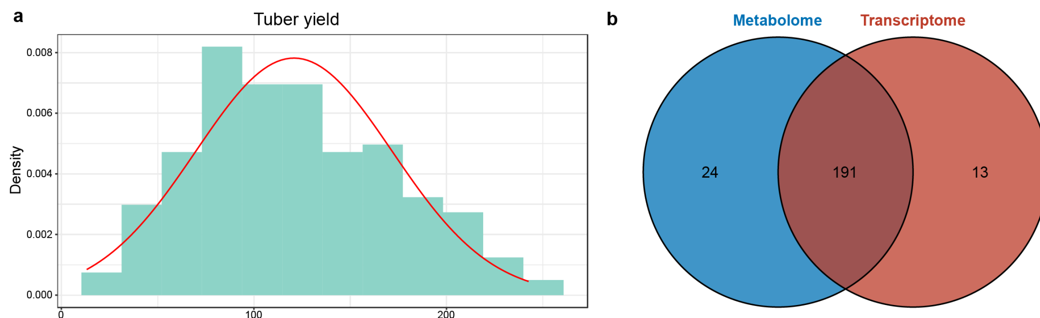
5. In the section of "QTL mapping of the multi-omics traits", Line 115, the authors mentioned that "We mapped several qualitative traits to known loci using bulked-segregant analysis (BSA)"; In the Method part line 428, the authors mentioned that "we conducted transcriptome sequencing of developing tubers of 204 F2 individuals (80 days after transplanting)". I am wondering how the 204 F2 individuals is selected. Were they selected based on the BSA result, and what qualitative traits they used for BSA? n? All of these information is important but not sufficiently provided in the Method part. It's the same question regarding the selection of 215 F2 individuals (120 days after transplanting) for metabolome profiling.

AR-6: We have added details about sampling in the Methods sections; please see lines 455–456, 471 and 477–480.

As shown in Supplementary Fig. 2, the traits used for BSA are qualitative traits, namely, tuber shape, tuber flesh color, tuber bud color, and yellow leaf. We used tuber

shape as an example here to describe the analysis pipeline. We randomly selected F_2 plants with long (51 samples) or round (48 samples) tubers to generate two contrasting pools, which we used for BSA. We selected different sets of plants as a function of the trait of interest among the 1,064 F_2 individuals. BSA was used to confirm the reliability of our population for gene discovery, which was independent of the transcriptome and metabolome data. We have listed the traits and the individuals used for BSA (please see revised Supplementary Table 1).

Regarding our sampling strategy, we planted all 1,064 F_2 individuals for evaluation of their pollen viability and yield-related traits; we then randomly selected 204 of them for transcriptome analysis. The yield data for these 204 individuals followed a normal distribution (Response Fig. 4a). When performing combined transcriptome and metabolome analyses, we performed metabolome analysis on 215 individuals, of which 191 individuals are shared with the transcriptome dataset, adding an additional 24 individuals selected randomly for metabolome analysis, making 215 individuals (Response Fig. 4b).



Response Fig. 4 Distribution of tuber yield for 204 randomly selected F_2 individuals used for RNA-seq (a) and sampling strategy of transcriptome and metabolome (b).

6. Another problem is that, transcriptome and metabolome profiling utilized different developmental stages, which are 80 and 120 days after transplanting. I am not familiar with potatoes, but either transcriptome and metabolome should be changed quite a lot in 40 days' development. The authors may have a reason to used different developmental stages, and it's better include the explanation in Discussion part.

Moreover, were the metabolome and transcriptome of parental inbred lines profiled in the same stage? I didn't see this information was mentioned in anywhere.

AR-7: The time at which we collected the samples for the RNA-seq and metabolome analyses was an integral part of our experimental design; we chose to collect the samples for RNA-seq analysis before the samples for metabolome analysis, considering that any changes in gene expression would lead to later changes in the accumulation and composition of metabolites in tubers at a final mature stage. This strategy has been demonstrated to be effective in previous studies[1-3,9] and allows better correlation between gene expression profiles and metabolite accumulation. Additionally, based on our breeding experience, the tubers of the parental lines typically reach full maturity in approximately 100 days. However, in our F_2 population, the growth period varied, with most materials maturing at around 100 days, while some still exhibited vigorous aboveground growth at later dates. To ensure accurate yield and metabolite data and to standardize the sampling time, we extended the period to 120 days for a uniform harvest. In addition, 60–90 days after transplantation is when tubers rapidly expand in the parental lines (as observed during breeding), a stage that is characterized by increases in gene expression levels and metabolite accumulation, making it an optimal time for transcriptome sampling. For example, in a study of the pear fruit metabolome, researchers collected samples for RNA-seq 15 days before the samples used for metabolome analysis due to the large changes happening in the fruits in the 2 weeks before reaching fully maturity[2]. Tuber development is slower than that of fruits, requiring about 2–3 months from initiation to full maturity. Therefore, we chose to collect the RNA-seq samples 40 days before full maturity to ensure uniform sampling. We have added these details in the Discussion section (lines 377–382). The time of sample collection for the transcriptome and metabolome data of the parental lines was the same as that for the F_2 population. The data for the parental lines and their F_1 hybrid have been previously reported[10], so we did not collect new samples from these genotypes in this study. We have added this information in the Methods section (line 479).

7. Data analysis of the entire paper was based on computing Pearson correlation between gene expression, metabolite abundance and phenotype traits. Then, the relationship with "significant" P values was detected to build WGNCA network or co-localize QTLs. As we all know, high-dimensional omics data including transcriptomic data, metabolomic data and imaging data are extremely complex and heterogeneous. Therefore, I believe quite big proportion of the detected "significant relationship" are false-positive and perhaps, actual biologically meaningful relationship of among eQTL, mQTL and various traits were not detected. Since this is not a methodology paper, and the datasets are original and the authors offer good examples with experimental validation on gene function, I am not picky on the algorithms or methods they used in this work. However, I believe these datasets are quite insufficiently mined.

AR-8: Indeed, multi-omics data are complex. The issue of false positives is also one that we were concerned about in this work, and which we tried to avoid. We have taken the following measures to ensure the reliability of our findings and provide a useful database for potato researchers.

(1) Filtering of gene expression data: We excluded genes with low or no expression ($FPKM \leq 1$) in > 90% of all F_2 tubers to decrease the false positive rate (FDR).

(2) Strict FDR controls for WGCNA: We rigorously applied an $FDR < 0.001$ for the WGCNA. To facilitate data usage by other researchers, we have uploaded the correlation analysis data, including *p*-values, *q*-values, and correlation coefficients, allowing for further filtering as needed (Supplementary Table 4).

(3) Integration with QTL data: Only gene expression data and yield traits/metabolite abundance mapping to the same locus were integrated with the WGCNA data. By constructing the triple relationships (gene–bin–trait) and integrating them with the QTL data, we decreased the number of correlations to 3,499 (please note that the previous data were based on eQTLs using raw FPKM; the revised data are based on eQTLs using qq-normalized FPKM values). Furthermore, to lower the FDR for QTL detection, we

only considered LOD values >5 in the triple relationship analysis. In similar studies in tomatoes, there were 232,934 correlations and 13,361 triple relationships[1]; in peach, 18,052 significant locus–trait associations and 12,691 expression–metabolite correlations were detected[2], yet these studies effectively pinpointed target genes. In this study, we effectively identified *GAME9* as a candidate gene that regulates solanine accumulation in tubers.

We acknowledge that WGCNA may include false positives; other multi-omics analysis tools, such as MODAS or other software based on machine learning, may also be applicable to the datasets produced in this study. Therefore, we have uploaded all raw data and triple relationship results for usage by other researchers (Supplementary Table 5). Many other traits can also be mined using these datasets. In this study, we mainly focused on traits related to fertility and tuber yield. These data will provide potato researchers with invaluable resources that are currently lacking in potato breeding due to their tetrasomic genomes.

Minor comments

1. Fig S1a, X axis better indicates the chromosome numbers;

AR-9: Now changed.

2. Fig S1c, Y axis is loge, or log2, or log10?

AR-10: The axis is on a log10 scale. Now specified.

3. Line 489, "Detection of pollen viability and pollen tube gemination assay", it should be tube germination? Also line 497.

AR-11: Now changed, thank you.

Reviewer #2 (Remarks to the Author):

This paper reports two interesting results from the genetic and phenotypic characterization of the diploid F2 potato population with 1,064 individuals derived from the cross of the two highly homozygous inbred lines. One is identification of the yield-related QTLs in the potato lines; the other is identification of the *PME* gene that involves pollen formation in the potato lines, providing an insight into the mechanisms underlying heterosis related to yield and male fertility. The results could be interesting to the colleagues who work in the fields of plant science and crop breeding.

AR-11: Thank you for your positive comments.

The main concerns:

1.Concerning the yield-related QTLs:

(1) What are the main genes-encoded functions that may contribute to the heterosis of yield?

AR-12: Unlike in the well-studied crops rice and maize, genetic and molecular studies in potato are lagging behind; in fact, genes related to high yield have yet to be cloned through forward genetics. Thus, knowledge about genes associated with tuber yield is limited. Yu et al reported that heterogeneous expression of the human RNA demethylase gene *FTO* in potato caused an ~50% increase in yield[11]. However, *FTO* is neither a potato gene nor a heterotic gene. Genes associated with photoperiod signaling such as *CYCLING DOF FACTOR 1* (*CDF1*) and *CONSTANS-LIKE1* (*COL1*) in potato can affect tuber yield by controlling tuberization time[12,13]. However, these genes showed no variations between the parental lines.

In terms of heterosis, causative genes of yield in rice[14] and maize[15] within identified QTLs can be directly picked out because these genes have already been reported, allowing researchers to continue studying their contributions to heterosis. However,

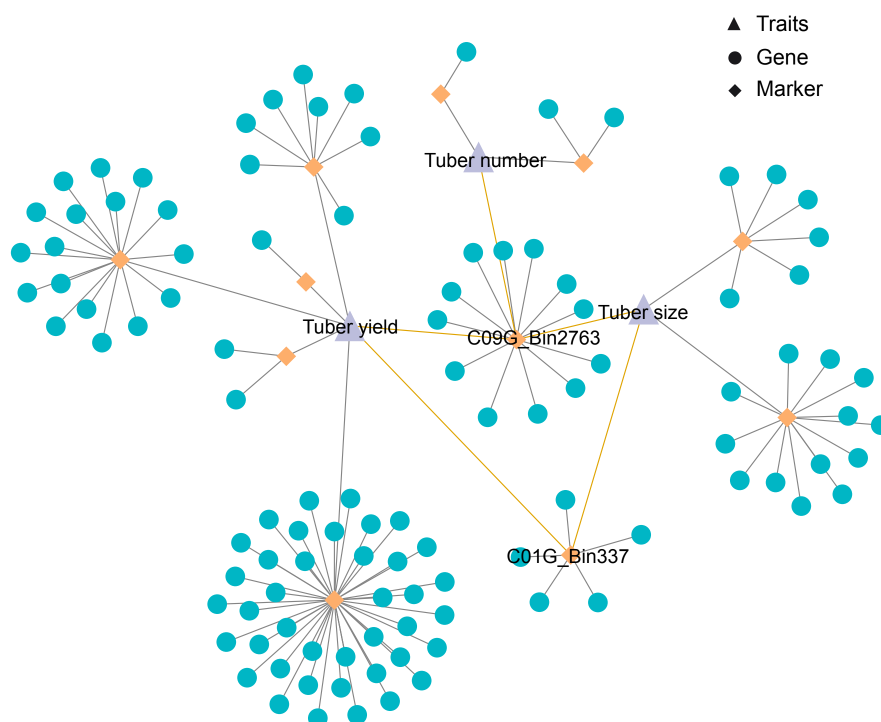
this is very difficult to do in potatoes at present. The lack of highly homozygous inbred lines hampers gene discovery in potato. In this study, the parental lines used to generate the F_2 population come from two potato subspecies, exhibiting high genetic diversity. We have identified 24 QTLs associated with yield traits, but determining the underlying candidate genes will require fine-mapping and transgenic confirmation.

Since yield traits are controlled by multiple genes, we are currently developing introgression lines (using the female parent A6-26 as the recurrent parent and the male parent E4-63 as the donor). In the future, we hope to transform the multi-QTLs into single QTL and facilitate the gene discovery.

(2) What are the molecular mechanisms that may result in the yield-related heterosis in the hybrids, at the levels of protein functions or at the levels of gene expression or both aspects?

AR-13: As mentioned in the above response, it is difficult to identify the causative genes behind yield-related heterosis in potato. Nevertheless, we have explored this question based on available data. By integrating the transcriptome, QTL mapping, and yield data, we have constructed triple relationships between three yield-related traits (tuber yield, tuber number and tuber size). As shown in Response Fig. 5, tuber yield, tuber size and tuber number are controlled by bin C09G_Bin2763, while QTLs for both tuber size and tuber yield locate within bin C01G_Bin337. Furthermore, from WGCNA, we identified 64 genes whose expression levels are significantly correlated with different yield components, of which only St_E4-63_C02G003152 was correlated with both tuber yield and tuber size. However, this gene is only involved in QTLs for tuber yield (both years) and was not detected in any QTL related to tuber size. More evidence is needed to ascertain its function on yield; thus, we did not mention this gene in the current manuscript. C09G_Bin2763 is an interesting genomic region, as its contribution to yield appears to be achieved by increasing tuber size while decreasing tuber number. This finding also aligns with our understanding of the negative correlation between tuber size and tuber number.

In general, considering the complexity of yield-related traits and the collective evidence garnered in other crops, we believe that yield heterosis in potato is affected by both protein functions (sequence variation) and gene expression, as in other crops[16]. In potato, we previously reported that gene expression complementation contributes to yield heterosis (complementation of allele-specific expression gene in the parental lines in hybrid)[10]. Additionally, sequence variation also affects yield heterosis. In rice, a non-functional allele of *Heading date 3a* (*Hd3a*) associated with rice heterosis had a 2-bp substitution, resulting in a delay of flowering under long-day conditions[14]. In potato, amino acid changes in StCDF1 can also regulate yield via promoting tuberization under long-day conditions[12] (there is no sequence variation of the *StCDF1* gene in the parental lines). Thus, we speculate that sequence variation can also affect yield heterosis by regulating yield-related traits in potato.



Response Fig. 5. Diagram of triple relationships for three yield-related traits.

(3) Is there any molecular genetic evidence for the yield-related heterosis from the gene mutations (natural or artificial)?

AR-14: As mentioned above, it is currently difficult to analyze yield-related heterosis at the single-gene level. Additionally, cultivated potatoes are highly heterozygous and almost completely self-incompatible, which hampers the discovery of natural mutations and the development of artificial mutagenized populations. Highly homozygous diploid potatoes were recently developed[17], but research on diploid potatoes is still in its infancy. We hope that after challenges such as self-incompatibility and the complexity of tetraploid genetics are addressed, potato research will progress rapidly, as in other crops. Overall, considering these issues in potato and in yield research, we have expanded the Discussion section to reflect these issues. Please see lines 389–393.

2. Concerning the identification of the *PME* involved in the male-fertility heterosis:

(1) Genetic complementation of the gene-edited mutants is required to confirm that the pollen-defective phenotype actually was caused by the *PME*-edited mutations although multiple *PME*-edited alleles were obtained due to the possibility of mistargeting.

AR-15: We fully understand your concerns about possible off-target effects during gene editing. However, complementation of the gene-edited mutants would be challenging in potatoes. Transformation efficiency is highly genotype dependent, with only a few diploid lines being transformable. The diploid clone 01-58 that we used for transformation has a highly heterozygous genome. Genetic complementation require self-pollination of the generated gene-edited mutants to eliminate the influence of the *Cas9* gene (driven by the cauliflower mosaic virus 35S promoter, conferring a constitutive level of Cas9 activity) by segregating it out. However, selfed progeny will show severe inbreeding depression, such as decreased growth vigor and male fertility, which will affect the phenotypic evaluation of any complementation lines. In previous potato studies, there have been no precedents for complementation of a gene-edited mutant with the wild-type copy of the edited gene. Given these constraints, performing complementation with gene-edited mutants is infeasible at the moment.

To address your concern about possible off-target effects, we have conducted whole-genome sequencing (WGS) with 60× coverage (~50 Gb) for the wild-type clone 01-58 and three transgenic gene-edited lines. We used established methodologies from published studies to identify and evaluate off-target effects[18] (*Nature Biotechnology*, 2021). First, using Cas-OFFinder[19], we identified 131 potential off-target sites across the genome based on the sgRNA (revised Supplementary Table 7). We then compared the sequence at all of these sites in the three transgenic plants to that of the WT clone 01-58. Although we employed relatively relaxed screening criteria for off-target site identification (allowing up to 5 mismatched bases out of the 20 for the Cas9 target site), we detected no off-target effects at any of the possible 131 sites in the three transgenic plants, with editing only occurring at the desired target site (the *PME* gene). These results indicate that the decrease in pollen viability is most likely caused by the mutation of the *PME* gene. We have incorporated these findings into the Results section (lines 340–345. The WGS data for 01-58 and the three transgenic plants have been uploaded to NCBI with BioProject accession number PRJNA878602 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA878602).

(2) In generation of the gene-edited mutant plants via tissue culture, the tissue-chimeric gene-edited plants could be often found. Therefore, what about the gene-editing rates in the PME-edited mutant plants used in the assays for pollen development?

AR-16: As explained above, the diploid clone 01-58 is highly heterozygous, which required us to conduct all analyses with T_0 plants showing editing rates of 100%. We previously developed a gene editing vector for efficient GFPuv-based fluorescence screening in potato[20]. Combined with the kanamycin screening, we can efficiently identify positive transgenic plants. By phenotyping and genotyping 10 T_0 plants, we successfully found three plants with a near 100% editing rate, as shown in the manuscript (although the mutation types are different in flower cells, the editing rates were 99.1%, 97.6% and 99.65% for KO-1, KO-2 and KO-3, respectively). The editing rate was identified by Hi-TOM[21], which has been widely used in mutation identification.

The Hi-TOM sequencing data for 01-58 and the three transgenic plants have been uploaded to NCBI with BioProject accession number PRJNA878602 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA878602).

Minor revision

Writing of the manuscript needs to be further improved.

1. The vague expression of many sentences needs to be revised, for examples:
(1) "…… should be……";
(2) "Fortunately,……";
(3) "Our findings will accelerate……";
and more.

AR-17: We carefully reviewed the entire manuscript and revised these statements accordingly.

2.The title of the manuscript does not actually express the main interesting achievement of the study, which might be not good for attracting the reading interest of audiences.

AR-18: We have changed the title to "Integrative multi-omics analysis reveals genetic and heterotic contributions to male fertility and tuber yield in an elite diploid potato hybrid".

Reviewer #3 (Remarks to the Author):

Li. Et al. The genetic and heterotic analyses of an elite diploid potato hybrid.

Summary. This manuscript reports on the construction and genetic analyses of a diploid F2 potato population with 1,064 individuals. The F2 was derived from the cross of two homozygous inbred lines. The authors report that they investigated 20,929 traits generated from 26 multi-omics datasets and identified 32,073 QTLs in these materials. Using gene expression data, they constructed a de novo systems-genetics network in

potato, which can be used for gene discovery. They also used these materials to study the genetic basis of heterosis for two traits - yield and male fertility. Overall, they found positive heterotic effects for yield-related QTLs and negative heterotic effects of metabolite QTLs (mQTLs), which they suggested contribute to significant yield heterosis in hybrid potato. Additionally, they identified a PME gene with a dominance heterotic effect that plays an important role in male -fertility in potato. The authors conclude that this study provides significant genetic resources for the potato community and may facilitate the application of heterosis in diploid potato breeding.

Overall Review – This manuscript is well written and relatively easy to follow but I honestly have had a hard time reviewing it. This is mostly due to the extreme scope of the materials presented. To be fair, even though I have worked in potato breeding and genomics and I have fairly significant experience in identifying QTL connected to important traits in potato and using them for marker-assisted breeding in an applied variety development program, them methods and techniques presented in this paper are very new to me. I do have major concerns on the total number of traits reported. Is it really possible to investigate 20,929 traits and identify 32,073 QTL for those traits and make any realistic sense of this volume of data? I am very skeptical of this approach, but I will not entirely discount it. However, I do not recommend accepting this publication as it is currently presented, as it needs major revisions and more focus.

AR-19: Thank you for your constructive advice to improve our manuscript. Below is a point-by-point response.

Specific Comments

Results

1. How do you evaluate 20,929 traits. My experience tells me that this is impossible to do at any scale with a high degree of precision and repeatability. Before accepting this I need to know more about the specific traits and how they are connected.

AR-20: These 20,929 traits comprise 19,166 expressed genes in tubers (each expressed gene is considered a trait), 679 metabolites detected in tubers and 1,084 phenotypic values. The 19,166 expressed gene are based on RNA-seq data. The 679 tuber metabolites were generated by ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS). The 1,084-phenotypic values primarily consist of yield-related traits and above- and under-ground traits (such as shoot morphology–related and tuber structure–related traits) collected using high-precision cameras (please also see AR-25; we have revised the hyperspectral and RGB data as you suggested). Detailed phenotypic data are listed in Supplementary Table 1. Six yield and male fertility–related traits (plant height, flowering time, tuber yield, tuber number, tuber size and pollen viability), as well as four qualitative traits (tuber shape, tuber flesh color, purple tuber bud and yellow leaf) used for BSA-seq were collected manually. Other phenotypic data were generated using high-precision cameras, a method widely employed in crop research, as they ensure the accuracy and reliability of data collection. We have added more detailed explanations for these datasets in Supplementary Table 1. WGCNA was used to connect the expression data with tuber-related traits (metabolites and yield-related traits). These methods have been used in other crops[1-3].

2. I worry about large-scale projects like the one presented here as they have potential to make sweeping statements based on datasets that are not well defined nor well controlled. For example, “we revealed the basis of potato heterosis”....for what trait and what is the basis. How specifically is heterosis measured and what about epistatic effects? Is this heterosis too?

AR-21: We previously described heterosis for male fertility and tuber yield in hybrid potatoes[10]. Thus, in this study, we explored the overall genetic pattern of gene expression and metabolite content and mainly focused on the yield-/male fertility–related heterosis, then further identified dominance and overdominance QTLs. The method to evaluate heterosis is explained in the Methods section (lines 527–533) and has been employed for similar studies in many crops[14,15,22]. The large-scale detection

of epistasis remains challenging in heterosis studies due to the many potential epistatic interactions across the genome.[15] Thus, we only looked for epistasis for four yield-related traits in the two-year experiments (plant height, tuber yield, tuber number and tuber size) using IciMapping software[23]. However, we identified no epistatic interactions between the 24 yield-related QTLs in both years. Thus, we did not report these results in this study. We have now added this information in the Discussion section; please see lines 416–423.

We understand your concern about the scale of the study and of our statements. The dominant/overdominant effects identified in this F_2 population may not be applicable to all potato populations. Therefore, we have modified phrases like "the basis of potato heterosis" to, for example, "the basis of [yield/male fertility/metabolites] heterosis in the elite hybrid potato" to more clearly restrict our findings to this F_2 population rather than all potato germplasms. Please see lines 34, 84, 186, 188, 297, 388.

3. Results Line 92 - please add more data on filtering of the SNPs. Were all of the SNPs used? We have several high-quality reference genomes available. Which were used? I'm thinking DM was one of them as reported in cite 23.

AR-22: We have added more details about SNP calling in the Methods section, please see lines 436–441.

Indeed, DM has a high-quality genome. There are some structural variations between DM and the two parental lines used in this study[24], which would lead to the misidentification of some SNPs when using DM as the reference. Thus, we used the genome of the male parent E4-63, which is also of high quality and has been used for genome-assisted breeding and pan-genome analyses, for comparison to the female parent for identifying all relevant SNPs[17,24]. The information of reference genome used in this study is given on line 435. We have previously used the E4-63 genome to clone a gene, *Dissected Leaf 1*, regulating leaf shape[25].

4. Results Line 124 - How do you protect from false discovery with so many QTL?

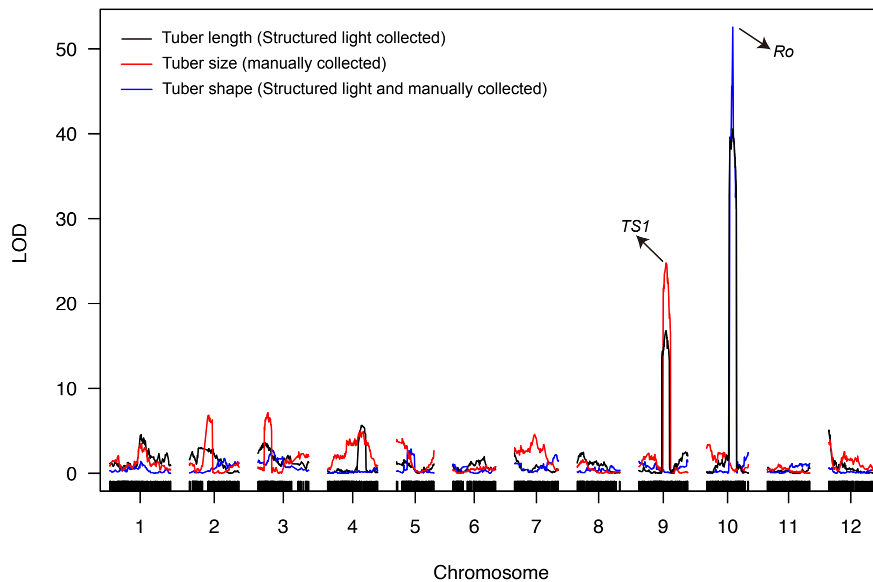
AR-23: The large number of QTLs identified in our study is due to the extensive range of traits we evaluated. The vast majority of them are expressed genes in tubers (19,166 genes out of 20,929 traits; please refer to AR-20). Similar studies using tomato introgression lines identified 889 QTLs for 74 metabolites[26] (1,264 QTLs for 679 metabolites in our study) and 32,204 eQTLs for 11,992 genes[27] (24,371 eQTLs for 19,166 genes in our study). This study not only explores heterotic effects in the hybrid potato population we developed, but also provides a database for potato researchers. Thus, accuracy was one of our concerns during the collection of all phenotypes. We have implemented several measures to ensure the reliability of our phenotypes and QTL mapping, as follows.

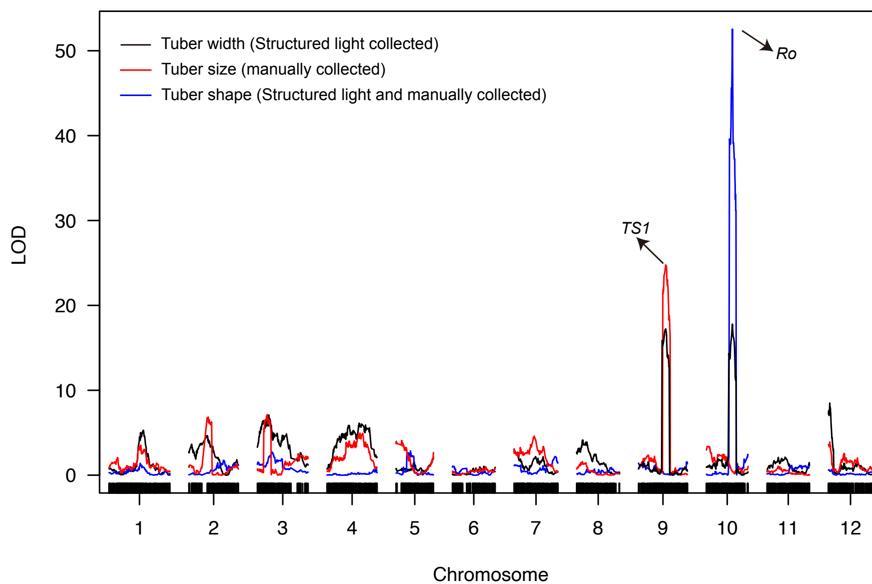
1. Accurate measurement of traits: As mentioned above, despite the large number of traits, the measurements in this study were accurate. Traits such as tuber yield, tuber size, tuber number and plant height were collected over two years (three replicates for each F_2 individual and each year). Other traits were measured using RNA-seq or high-precision cameras, ensuring high accuracy. The identification of QTLs for each trait was an independent event, and reliable trait assessment ensured the accuracy of our QTL identifications.

2. High LOD threshold: To further enhance the accuracy of QTL identification, we adopted a high LOD threshold. While many studies typically use LOD values ranging from 2 to 3[28-32], we raised the LOD threshold to 3.5 in this study to filter out potential false-positive loci. Additionally, in the analysis of triple relationships, we included only QTLs with LOD values >5 in constructing the network, further ensuring the reliability of our results.

On the basis of these results, we rapidly identified the transcription factor gene *GAME9* related to solanine accumulation. The successful cloning of the *PME* gene also demonstrated the quality of our datasets. Another example is the mapping of genes

related to tuber shape and size. We developed a pipeline to construct a 3D model of tubers using the structure light modeling method and extracted tuber traits such as tuber length and tuber width. Tuber length and tuber width are correlated with both tuber size and tuber shape. Interestingly, QTLs for tuber length and tuber width co-localize with QTLs related to tuber shape and tuber size (Response Fig. 6). Among these QTLs, one overlapped with the *Ro* gene on chromosome 10 regulating tuber shape that was previously reported[33]. This is a cross-validation between traits collected through imaging and those obtained through manual surveys.





Response Fig. 6. QTL maps showing QTLs for tuber length and width co-localizing with those related to tuber size and shape.

5. Results Line 125 – one year pollen viability…. or pollen viability during 1 year of the experiments. I read this sentence as pollen viability after storage for 1 year. Reword to make this more clear.

AR-24: Thank you for pointing out this. We have revised this sentence to "We collected data on pollen viability (male fertility-related) and flowering time (yield-related) during the 2021 trial and obtained data for four other yield-related traits (plant height, tuber yield, tuber number and tuber size) during the trials in 2021 and 2023". Please see lines 125–128.

6. Line 135 – "Data generated by hyperspectral imaging helped to detect 5,850 pQTLs of tuber reflectance" I feel like the authors are trying to impress us with huge numbers of QTL but I'd encourage then to focus on important QTL that contribute to deeper understanding rather that blowing the reader away with huge numbers of QTL that we don't understand. Can this be done with a few key traits with a good explanation as to why the traits were selected?

AR-25: Hyperspectral technology has been widely used for detecting dry matter and metabolites in crop grains, with its primary application being the prediction of the content of these substances. Genetic mapping of hyperspectral data is not its main application direction (our intention was to provide as comprehensive a dataset as possible for other potato researchers). We have revised this section to emphasize its application in predicting tuber dry matter and metabolite content and removed the QTL detection. The specific modifications are as follows.

To more accurately evaluate tuber reflectance, we divided the near-infrared light and the visible light values into detailed segments based on wavelength. Specifically, near-infrared light (1,000–1,700 nm) was divided into 172 segments, while visible light

(400–1,000 nm) was divided into 314 segments. The details about the wavelengths were added to revised Supplementary Table 3. The previous manuscript performed QTL mapping for total reflectance (488 traits [172+314 wavelengths and two tuber surface areas under near-infrared or visible light]) and average reflectance (total reflectance/tuber surface area, 486 traits [172+314 wavelengths]) at all wavelengths. In the revised manuscript, we now only present the raw data for total reflectance (488 traits) and have removed the QTL mapping results for the traits related to total (488) and average reflectance (486). We mainly focused on the role of hyperspectral data in predicting dry matter and metabolite content. Furthermore, using stepwise linear regression and 10-fold cross-validation, we identified the key wavelengths according to their predictive effectiveness on dry matter and different metabolites (please see revised Supplementary Table 3). We have also revised the manuscript; please refer to lines 136–138, 145–146 and revised Supplementary Table 3.

7. Lines 247 – 248 – These are not a very high % explanation. They may be significant but if these are the greatest effects I wonder how important they really are?

AR-26: Crop yield is a very complex trait to which many minor- and medium-effect loci and several major loci contribute in a given population. In this study, the parental lines of the F_2 population come from two potato subspecies: *S. tuberosum* Group Stenotomum and *S. tuberosum* Group Phureja. There are nearly five million SNPs between the two parents, resulting in a high genomic and phenotypic diversity in the F_2 population. The explanation of *TY1* is relatively not high in the F_2 population because of the other six minor QTLs. The total explanation of the seven *TY* QTLs is 25.4% and 24.4% in 2021 and 2023, respectively.

8. Line 267 – Is it correct to say that all eQTL are genes? These are expression profiles, right? Is all expression regulated by a gene or genes? Aren’t there other methods of regulation?

AR-27: We apologize for this misunderstanding. eQTLs are loci regulating the expression of a target gene. In the context of this study, we consider that all eQTLs regulate the expression of a single gene, but not all eQTLs are genes. We have changed "all eQTLs in a gene" to "all eQTLs associated with the expression of a gene" (line 277). The eQTLs may be part of the promoter sequence or gene(s) encoding transcription factors that affect the expression of the target gene. Please refer to AR-2 for more details about *cis*/*trans* regulation.

9. Line 278 - A cautionary note about metabolites, though I am by no means a metabolite expert. But I worry about how the authors seem to be relating the metabolites to genes per se. Based on my discussions with food scientists and plant biochemists, many metabolites are secondary products or breakdown products that have their own energetics and breakdown products are often formed from natural molecular rearrangements due to the cellular environment, etc. Have the authors considered this perspective?

AR-28: The accumulation of metabolites requires the expression of genes encoding related biosynthetic enzymes, and molecular rearrangements also require enzymatic catalysis, which is a product of gene expression. In some rare cases, high temperature or light exposure can also cause changes in metabolites, but all 1,064 F_2 individuals are grown in the same environment and the environment has the same impact on all plants. Numerous articles have reported the link between metabolites and gene expression[1-3,27], and the methodologies are well established. In our study, we determined the composition and contents of metabolites from mature tubers collected in the same environment for all F_2 individuals. Metabolite accumulation in mature tubers is relatively stable. Although molecular rearrangements occur, if the relative genes are segregated in the F2 population (either expression level or protein sequence variation), we can also identify relevant QTLs based on their products.

10. Line 289 – "These results further support our findings in metabolites heterosis and provide insights into heterosis of clonally-propagated corps." Based on my

observations above I'm not sure that I can make this leap of faith in this statement at this point.

AR-29: Thank you for pointing out this. We have deleted the statement "provide insights into heterosis of clonally-propagated crops" to limit our conclusions to this hybrid potato population. Please also see AR-32.

11. Line 300 - A pectin methylesterase contributes to male fertility heterosis – I recommend that this should be a separate paper standing on its own with more robust analyses. I note this work has also be elaborated on in citation 23.

AR-30: The *PME* gene is in this study and the *FBA1* gene in citation 23 (https://doi.org/10.1016/j.cell.2021.06.006) are two different genes. The *FBA1* gene encodes a bHLH transcription factor (located on chromosome 02) and regulates stamen development, while the *PME* gene encodes a pectin methylesterase (located on chromosome 07) and regulates pollen viability.

During breeding, we also observed differences in male fertility and seed number between the two parental lines, which are two key traits for hybrid breeding in diploid potato. We chose to map the genetic loci behind pollen viability, leading to the cloning of the main-effect gene *PME* and evaluated its contribution to fertility heterosis.

Discussion

1. Overall - Given the sweeping amount of research conducted the discussion section is very short and lacks depth or analysis/interpretation.

AR-31: We have expanded the text in the Discussion section about yield heterosis and some insights for in-depth studies in the future. Please see lines 368–372, 377–395 and 416–423.

2. Lines 378 – 386 - This is a very disjointed section and I don't see good justification for many of the statements made. For example. Tuber dry matter serves as the nutrition provided for asexual reproduction. I'm not even sure what the authors are trying to say here? Is this from an evolutionary perspective? An energetics point of view?

AR-32: We meant this statement in the context of resource/energy utilization. We think that the hybrid allocates more resources to the production of dry matter (starch, protein) that is critical for tuber storage and germination. Studies in maize have found that the heterosis effect of metabolites in kernels (sink) is opposite to that in leaves (source), with most metabolites content showing negative heterosis in kernels[34]. Similarly, in *Arabidopsis thaliana*, researchers discovered that the amount of biomass and the concentrations of most metabolites were negatively correlated[35] and further found a more efficient utilization of resources and a corresponding advantage[36]. Similar observations have been reported in other studies[37,38]. Overall, research on metabolic heterosis and the relationship between metabolites and dry matter/biomass is insufficient. Our findings in potato tubers are consistent with existing reports. We have revised this part in the Discussion to make our points more clearly (please see lines 407–414).

Methods

Collection of phenome data – The authors need to describe the collection of their phenomic data more precisely. I went to the rice paper cited in 62-64 to see what was done with rice and there is no comparison to what is reported herein. Please elaborate or…preferably, focus on key traits that make a difference in potato production.

AR-33: We have generated two Supplementary Files: Supplementary File 1 provides a detailed explanation of our data collection pipeline, and Supplementary File 2 describes the meanings and calculation methods of each phenotype.

Additionally, plant RGB imaging and hyperspectral technologies are relatively new technologies that have been used to detect crop traits for only 4–5 years. Indeed, some data collected from RGB-based images have not yet been linked to clear agronomic traits, which is common in plant phenomics[39] (The latest research; https://doi.org/10.1111/nph.19942). Our intention was to provide a dataset that could serve as a potential reference for other potato researchers. According to your suggestions, we have filtered the phenotypes generated by RGB imaging, removing those that were not clearly defined (Response Table 1) and retaining only certain phenotypes. Furthermore, we have added an "Explanation" column in Supplementary Table 1 to explain the meaning of each RGB trait.

Response Table 1. Traits omitted from the RGB datasets

Omitted traits	Full name
SH	Height of the bounding rectangle
THR	The ratio of total projected area and hull area
FDIC	Fractal dimension with image cropping
F1	Relative frequencies
F2	Relative frequencies
F3	Relative frequencies
F4	Relative frequencies
F5	Relative frequencies
F6	Relative frequencies
T1-G	Correlation
T2-G	Advantages of the small gradient
T3-G	Advantages of the large gradient
T4-G	Energy
T5-G	Intensity inhomogeneity
T6-G	Gradient inhomogeneity
T7-G	Mean gray
T8-G	Mean gradient
T9-G	Gray entropy
T10-G	Gradient entropy
T11-G	Entropy of mixing
T12-G	Differential moment
T13-G	Deficit score
T14-G	Gray mean variance
T15-G	Gradient mean variance
T1-H	Correlation

T2-H	Advantages of the small gradient
T3-H	Advantages of the large gradient
T4-H	Energy
T5-H	Intensity inhomogeneity
T6-H	Gradient inhomogeneity
T7-H	Mean gray
T8-H	Mean gradient
T9-H	Gray entropy
T10-H	Gradient entropy
T11-H	Entropy of mixing
T12-H	Differential moment
T13-H	Deficit score
T14-H	Gray mean variance
T15-H	Gradient mean variance

I prefer to let others comment on QTL mapping strategy (e.g., Is CIM the appropriate methods for these studies?) and on metabolomic and eQTL research.

AR-34: CIM has been effective in mapping QTLs and studying heterosis using F_2 populations in rice[14] and maize[15]. The methods used for metabolome analysis and eQTL mapping have also been successfully used in previous studies[1-3].

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

AR-35: Thank you for reviewing our manuscript.

Reference:

1. Zhu, G. *et al.* Rewiring of the fruit metabolome in tomato breeding. *Cell* **172**, 249-261 (2018).
2. Cao, K. *et al.* Combined nature and human selections reshaped peach fruit metabolome. *Genome Biol.* **23**, 146 (2022).
3. Li, L. *et al.* Characterization of novel loci controlling seed oil content in Brassica napus by

marker metabolite-based multi-omics analysis. *Genome Biol.* **24**, 141 (2023).

4. Strompen, G., Grüner, R. & Pfitzner, U.M. An as-1-like motif controls the level of expression of the gene for the pathogenesis-related protein 1a from tobacco. *Plant Mol. Biol.* **37**, 871-883 (1998).
5. Garretón, V., Carpinelli, J., Jordana, X. & Holuigue, L. The as-1 Promoter Element Is an Oxidative Stress-Responsive Element and Salicylic Acid Activates It via Oxidative Species. *Plant Physiol.* **130**, 1516-1526 (2002).
6. Krawczyk, S., Thurow, C., Niggeweg, R. & Gatz, C. Analysis of the spacing between the two palindromes of activation sequence-1 with respect to binding to different TGA factors and transcriptional activation potential. *Nucleic Acids Res.* **30**, 775-781 (2002).
7. Johns, T. & Alonso, J.G. Glycoalkaloid change during the domestication of the potato, Solanum Section Petota. *Euphytica* **50**, 203-210 (1990).
8. Hardigan, M.A. *et al.* Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proc. Natl Acad. Sci. USA* **114**, E9999-E10008 (2017).
9. Zhang, Z. *et al.* Multiomics Analysis Reveals the Chemical and Genetic Bases of Pigmented Potato Tuber. *J. Agric. Food Chem.* **71**, 16402-16416 (2023).
10. Li, D. *et al.* The multi-omics basis of potato heterosis. *J. Integr. Plant Biol.* **64**, 671-687 (2022).
11. Yu, Q. *et al.* RNA demethylation increases the yield and biomass of rice and potato plants in field trials. *Nat. Biotechnol.* **39**, 1581-1588 (2021).
12. Kloosterman, B. *et al.* Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature* **495**, 246-50 (2013).
13. Abelenda, J.A., Cruz-Oro, E., Franco-Zorrilla, J.M. & Prat, S. Potato StCONSTANS-like1 Suppresses Storage Organ Formation by Directly Activating the FT-like StSP5G Repressor. *Curr. Biol.* **26**, 872-81 (2016).
14. Huang, X. *et al.* Genomic architecture of heterosis for yield traits in rice. *Nature* **537**, 629-633 (2016).
15. Liu, H. *et al.* Genome-wide identification and analysis of heterotic loci in three maize hybrids. *Plant Biotechnol. J.* **18**, 185-194 (2020).
16. Chen, Z.J. Genomic and epigenetic insights into the molecular bases of heterosis. *Nat. Rev. Genet.* **14**, 471-482 (2013).
17. Zhang, C. *et al.* Genome design of hybrid potato. *Cell* **184**, 3873-3883 (2021).
18. Jin, S. *et al.* Genome-wide specificity of prime editors in plants. *Nat. Biotechnol.* **39**, 1292-1299 (2021).
19. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473-1475 (2014).
20. Du, J. *et al.* The Modification of Gene Editing Vector for Efficient GFPuv Fluorescence Screening and Its Application in Potato Genetic Transformation. *Scientia Agricultura Sinica* **56**, 2223-2236 (2023).
21. Liu, Q. *et al.* Hi-TOM: a platform for high-throughput tracking of mutations induced by CRISPR/Cas systems. *Science China Life Sciences* **62**, 1-7 (2019).
22. Zhou, G. *et al.* Genetic composition of yield heterosis in an elite rice hybrid. *Proc. Natl Acad. Sci. USA* **109**, 15847-15852 (2012).

23. Meng, L., Li, H.H., Zhang, L.Y. & Wang, J.K. QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* **3**, 269-283 (2015).
24. Tang, D. *et al.* Genome evolution and diversity of wild and cultivated potatoes. *Nature* **606**, 535-541 (2022).
25. Li, D. *et al.* Dissected Leaf 1 encodes an MYB transcription factor that controls leaf morphology in potato. *Theor. Appl. Genet.* **136**, 183 (2023).
26. Schauer, N. *et al.* Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat. Biotechnol.* **24**, 447-454 (2006).
27. Szymański, J. *et al.* Analysis of wild tomato introgression lines elucidates the genetic basis of transcriptome and metabolome variation underlying fruit traits and pathogen response. *Nat. Genet.* **52**, 1111-1121 (2020).
28. Chen, J. *et al.* Construction of a high-density genetic map and QTLs mapping for sugars and acids in grape berries. *BMC Plant Biol.* **15**, 28 (2015).
29. He, D.-H. *et al.* QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of Gossypium hirsutum × Gossypium barbadense. *Euphytica* **153**, 181-197 (2007).
30. Dugo, M.L. *et al.* Genetic mapping of QTLs controlling horticultural traits in diploid roses. *Theor. Appl. Genet.* **111**, 511-520 (2005).
31. Sun, Z. *et al.* Minor-effect QTL for heading date detected in crosses between indica rice cultivar Teqing and near isogenic lines of IR24. *The Crop Journal* **6**, 291-298 (2018).
32. Wang, J. *et al.* QTL mapping of grain length in rice (Oryza sativa L.) using chromosome segment substitution lines. *Genet. Res.* **88**, 93-104 (2006).
33. Van Eck, H.J. *et al.* Multiple alleles for tuber shape in diploid potato detected by qualitative and quantitative genetic analysis using RFLPs. *Genetics* **137**, 303-309 (1994).
34. Romisch-Margl, L. *et al.* Heterotic patterns of sugar and amino acid components in developing maize kernels. *Theor Appl Genet* **120**, 369-81 (2010).
35. Meyer, R.C. *et al.* The metabolic signature related to high plant growth rate in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **104**, 4759-4764 (2007).
36. Meyer, R.C. *et al.* Heterosis manifestation during early Arabidopsis seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids. *Plant J* **71**, 669-83 (2012).
37. Bajpai, P.K., Reichelt, M., Augustine, R., Gershenzon, J. & Bisht, N.C. Heterotic patterns of primary and secondary metabolites in the oilseed crop Brassica juncea. *Heredity (Edinb)* **123**, 318-336 (2019).
38. Kleessen, S. *et al.* Metabolic efficiency underpins performance trade-offs in growth of Arabidopsis thaliana. *Nat. Commun.* **5**, 3537 (2014).
39. Zhang, Z. *et al.* Integrating high-throughput phenotyping and genome-wide association studies for enhanced drought resistance and yield prediction in wheat. *New Phytol.*

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

I have read the revised manuscript and the authors' response to my comments. All of my concerns have been fully addressed. I have no further questions then.

Response: Thank you.

Reviewer #2 (Remarks to the Author):

In this revision, the authors have not address any of all the question points concerned from the last review, but it is understandable for the difficulty to address all the question points due to the current limitation of biotechnology and genetic assays for potato materials based on the explanation from the authors. No more new question from this review.

Response: Thank you for your understanding.

Reviewer #3 (Remarks to the Author):

This manuscript has been improved considerably from the original submitted. The authors clearly have taken the reviewers comments and suggestion constructively and they have submitted a much-improved manuscript. Thank you.

Here are a two suggested edits based on my most recent review.

Line 84. Importantly, we revealed the genetic basis of heterosis in for male fertility and 21 tuber yield traits in this elite hybrid potato cross (or population?) either population or cross is ok by me.

Response: Thank you for this suggestion. We have revised "elite hybrid potato" to "elite hybrid potato cross".

Line 375 This QTL database will provide useful genetic markers for molecular breeding and gene discovery in potato.

Response: We have revised "gene discovery of potato" to "gene discovery in potato".

Discussion

The discussion has improved considerably…However, I feel that it does not end with a good summary. I suggest that the authors add one more paragraph summarizing their work. It can pull from the abstract. But in reading through the discussion I felt like a good summary is still needed to end an otherwise good discussion. Maybe the authors can highlight some of the aspect of this work that they mentioned in the responses to reviewers along the lines of one of the major goals of this research was to provide new genetic resources and phenomic and metabolomic resources for the potato community and to contribute to our collective understanding of heterosis in potato. Something along this line, which you have done in the abstract already, but it would be good to finish strong in the discussion.

Response: Thank you for this suggestion. We have added a paragraph to summarize our work, as follows: "Overall, molecular breeding and functional genomics study in diploid potatoes are still in its infancy. This study provides valuable genetic and phenotypic resources for the potato community. By integrating multi-omics data to construct a systems genetics network, along with studies on heterosis, our findings in this work contribute to gene discovery and enhance our understanding of the genetic basis of heterosis in potatoes."