nature portfolio

Peer Review File

Haplotype-based pangenomes reveal genetic variations and climate adaptations in moso bamboo populations

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Hou et al. have conducted a haplotype-resolved pan-genome analysis of 16 representative moso bamboo accessions, providing valuable insights into genetic diversity, population structure, and climate adaptation mechanisms in this ecologically and economically vital bamboo species. Specifically, the authors have assembled 16 high-quality genomes to construct a pan-genome of moso bamboo. To my knowledge, this is the first pan-genome for the subfamily Bambusoideae, thus providing extensive genomics resources for bamboo research. Additionally, the authors have assembled haplotype-resolved genomes to improve the moso bamboo reference and reveal actual variation levels through haplotypes. Based on these haplotypes, the authors constructed a haplotype pan-genome, analyzed allele-specific expression (ASE), and proposed that the ASE is related to environmental adaptation. Furthermore, genotype-environment association (GEA) analysis has identified variants associated with climate variables and predicted vulnerable moso bamboo populations. As moso bamboo's economic and ecological value rises amid worsening plastic pollution and climate change, these findings and datasets will critically inform molecular research on bamboo and beyond. While the work is commendable overall, there is room for improvement based on the following suggestions.

1. Line 37: To quantify the economic importance of the bamboo industry, it would be valuable to include figures related to the scale and development of this sector. Specifically, incorporating metrics such as annual output value, total area of bamboo forests, utilization rates, and contributions to rural incomes could effectively showcase the vital role of the bamboo industry in supporting livelihoods and national economic development. Adding such quantitative details would strengthen the paper by substantiating the significant economic impact of bamboo cultivation and products.

2. Line 94: The authors have assembled 16 haplotype moso bamboo genomes with meaningful results. To further characterize the quality of these assemblies, clarifying whether the reported N50 refers to contigs or scaffolds is needed. Moreover, with advancements in sequencing technologies, N50 values have increased substantially. However, directly comparing N50 across species is of limited value due to differences in genome size and chromosome numbers. It is therefore recommended that the authors calculate the maximum chromosome-level N50 for moso bamboo, as well as the proportion of the total length represented by the longest 24 contigs (the haploid chromosome number in moso bamboo) at the contig level. Reporting these metrics would better show the continuity and completeness achieved in these assemblies at both the chromosome and contig levels for this species. Specifying both chromosome-level N50 and the total proportion of the genome covered by the longest 24 contigs would thus more clearly convey the assembly quality in a species-appropriate context.

3. Line 112: Please provide the reasons for choosing the CYhap1 genome as the reference genome. Additionally, for subsequent molecular biology experiments and comparative genomics studies of moso bamboo, does the author still recommend selecting this genome as the reference genome?

4. Lines 123-124, 293-295: The authors have made an interesting observation that the majority of variations in moso bamboo stem from inter-haplotype differences. This insight could have broader implications for the study of other asexually reproducing plants. However, I have some concerns about the author's decision to continue using inter-haplotype variations for GEA analysis. It might be more appropriate to filter out these inter-haplotype variations before conducting GEA analysis.

5. Lines 144-146: It was recommended to add a figure to visualize this result.

6. Lines 169-171, 489: The authors have presented an interesting and novel result by constructing a haplotype pan-genome based on allelic genes. However, a more detailed description of the construction process of the haplotype pan-genome is needed, including the relevant thresholds and the rationale behind their selection. For instance, the authors mention that “Gene pairs within 40 kb were retained” in the method. It would be beneficial for the authors to explain why they chose the threshold of 40 kb. This additional information would provide greater clarity and understanding of the methodology used in the study.

7. Line 882: The authors have successfully constructed 16 genomes with high LAI values, which presents a unique opportunity for exploring LTRs in moso bamboo. Given this, incorporating analyses related to LTRs was recommended. It would be particularly interesting to investigate whether LTRs share inter-haplotype variations, similar to other observed variations. Additionally, considering that LTRs have been linked to structural variations, investigative this association in the context of moso bamboo could yield valuable insights.

8. Line 202: Please ensure uniformity in font sizes in figures, such as Figure 2b.

9. Line 288: Please add related method for identifying short variations in the methods section.

10. Line 445: It was recommended to add the related reference for the LAI method. Assessing genome assembly quality using the LTR Assembly Index (LAI) (https://doi.org/10.1093/nar/gky730).

11. Line 467: Please capitalize "DIAMOND" and provide the corresponding reference. Sensitive protein alignments at tree-of-life scale using DIAMOND (doi:10.1038/s41592-021-01101-x).

12. Line 475: Please add the related reference.

13. Line 678: It was recommended to cite the latest BUSCO reference. BUSCO: Assessing Genomic Data Quality and Beyond(https://doi.org/10.1002/cpz1.323).

14. Line 708: Please use the latest reference for the NCBI Nr database to ensure that the study adheres to the most up-to-date database and methodologies. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation (https://doi.org/10.1093/nar/gkv1189).

15. Line 710: Please use the latest reference for the UniProt database. UniProt: the Universal Protein Knowledgebase in 2023 (https://doi.org/10.1093/nar/gkac1052).

16. Line 714: The reference for the eggNOG database seems to be incorrect, should be: eggNOG 6.0: enabling comparative genomics across 12 535 organisms(https://doi.org/10.1093/nar/gkac1022).

17. Supplementary Table 11: It was recommended to fill in the table according to accession numbers.

18. Supplementary Table 19: Please correct the table and arrange environmental variables in numerical order.

19. Supplementary Fig. 12: Please add related descriptions.

Reviewer #2 (Remarks to the Author):

The manuscript by Hou et al. presents the creation of a haplotype-resolved pan-genome for moso bamboo. The authors conducted an in-depth analysis of this pan-genome, utilizing 427 re-sequenced samples to uncover significant genetic variations and allele-specific gene expressions linked to climatic adaptability. They leveraged these genomic insights alongside climatic data to attempt predictions of vulnerable moso bamboo populations, potentially aiding in assessing risks posed by environmental changes. While the data undeniably constitutes a valuable genomic resource for bamboo research, the analyses tend to be descriptive, and several conclusions are drawn without robust supporting evidence. Please find my detailed comments below:

1. The title 'haplotype-resolved pan-genomes' is misleading. Typically, 'haplotype-resolved' refers to a phased genome assembly of a single genome, applicable to both diploids and polyploids, rather than pan-genomes. The authors have actually constructed a graphical pan-genome by integrating structural variations from 16 haplotype-resolved genomes (i.e., 32 haplotypes), as described in their methodology.
2. Line 91, Supplementary Figure 1 merely depicts the geographic distribution of the 16 RMA samples. It does not clearly illustrate the claimed extensive genetic diversity. The rationale behind selecting these 16 RMAs is unclear—are there specific characteristics that warranted their selection?
3. Of the 16 RMAs, only three have been anchored to chromosomal-scale assemblies using Hi-C reads, leaving 13 at a contig-level assembly, which does not fully resolve haplotypes. I recommend that the term 'haplotype-resolved' be used cautiously and accurately, reflecting the actual level of assembly achieved.
4. Line 105 mentions 'an average of 54,815 protein-coding genes.' Is this figure inclusive of allelic genes? It would be informative to know the count of bi-allelic genes versus those present as a single allele.
5. Lines 123-124 note that inter-haplotypic variations exceed the genetic variations among different accessions. An explanation or hypothesis to account for this observation would be beneficial.
6. Lines 177-181 introduce the classification of double-allele, single-allele, and variable-allele gene sets, but the biological significance and relevance of these categories are not clear. What specific biological questions are these data intended to address?
7. Line 255-257 draws a conclusion that seems speculative. There isn't adequate evidence

provided to substantiate the claim that certain inter-haplotype variations lead to allele-specific expression (ASE).
8. Line 289 suggests that a population analysis of the 427 accessions should be conducted to understand their characteristics, population structure, and genetic diversity.
9. Line 296 queries the meaning of six climate variables (BIO1, BIO2, BIO5, etc.). Clarification on what these variables represent is necessary.
10. Line 340 mentions SSP585 and SSP126 without prior explanation. These should be defined upon their initial mention in the text.
11. Lines 354-355 hypothesize that greater genetic offsets under SSP585 compared to SSP126 may indicate increased risks to bamboo populations due to higher CO2 emissions. A more detailed explanation of this inference is required.
12. The Discussion section requires restructuring; it currently appears to be a recapitulation of the results rather than a thoughtful discussion.
13. Line 434 asks for clarification on what is referenced by Supplementary Table SX.
14. Line 474 has a missing citation that needs to be addressed.
15. Lines 487-493 discuss the identification of allelic genes solely through reciprocal alignments, which could result in numerous genes that are not part of any allelic pair. Some may be single-allelic genes, while others might represent haplotype-specific duplicated genes. The methodology for identifying and categorizing these genes should be elaborated upon.
16. Lines 519-526 detail allele-specific expression identification using HISAT2 with default parameters, which may not discriminate between highly identical allelic genes—thus unsuitable for calculating allele-specific expression accurately. A recommendation is made to select uniquely aligned RNA-seq reads to enhance this aspect of the study.

Reviewer #3 (Remarks to the Author):

This is an excellent MS with an impressive volume of data and results. I just have some suggestion in order to improve the understanding of the paper for the reader.
There is an understandable general trend in the MS towards the pangenome associated information and methods and an underplaying of the populational data. To make the populational results more understandable the MS need some improvements.

1- sample information
You mention that you have 16 representatives moso bamboo accessions with additional 427 resequenced accessions. However in methods section, sample collection, the information about this 427 samples is missing. Where they come from ? How many from each region?
Supplementary table 4 only have sequence information and only for 186 RNAseq samples.
2- In the methods section “Genome and transcriptome sequencing” there aren’t any information about the 427 resequenced samples

3- In the methods section “ Identification of climate-associated variants” is not clear how many samples you are using. The 16 you used for the pangenome or the 427 that you didn’t explain how you sequenced and where you get it from? The admixture plot and associated differentiation statistics is important for the reader to understand the pattern of distribution of the variation inside the species. The same for RDA results, the figure and associated information is missing. How much variation is explained by the 8 climatic variables?

4- line 587 "Calculation of genomic drift". Do you really want to say this?
A more detailed explanation about the genomic offset and the local, forward and reversed offset is needed, otherwise Figure 5e and f are very difficult to understand.

5- Is the Hi-C and RNAseq come form the same samples used for the pangenomes? The 186 samples used for RNAseq are only mention in the results and not on the methods. The origin of this samples have to be more clear in the methods section.

6- In the end of the introduction you already talk about you results!!. In the introduction, the explanation of the genomic offset and related measures as well as its potential and limitations are missing.

7- I missed the results of some analysis that are the base for the assembly evaluation, such as the K-mer analysis with Mercury.

8- Is not clear to me the rational of the distinction between the inter-haplotype and inter-accession measures, and why you based all your results in theses differences?
Line 128 - “Thus, these variants were in fact present between the haplotypes (inter-haplotype) of the two accessions simultaneously,
and not between the accessions (inter-accession)”.
Line 159 -”Inter-accession variations are absent between the reference haplotypes and present in other accessions. Inter-haplotype variations are present between the reference haplotypes”.

Thus that mean that when you compare with the reference genome you have more differences than when you compare each of two different accessions?
So all you inter-accessions results from multiple pairwise differences across accession and so you should have a distribution of values, one for each comparison?

I not sure If I understand properly the relevance of you analysis. But for me the relevant biological information come from comparing your set of genomes that correspond to the groups you identified with admixture. Accession of each K=3 group should have genomes more identical that when you compare among the 3 groups.

9 - line 133 - “suggesting that traditional methods of variant
identification overestimate heterozygosity in moso bamboo”. It deserve more elaboration in the discussion section.

10 – Line 288 Information about the origin and how you resequenced and analysed the 427 samples is missing.
Line 297 - Supplementary Fig 11 and Supplementary Fig 12 are missing in the supplementary material.

Line 302 – Don’t mention the all BIO variables here because you only used 6

Line 341 – how you select this three variables?

Figure 5b is difficult to read. Please label your populations and make the differences in the

RONA values more evident.
The use of the mean value buffers the differences across climate models.
Figure 5e and f are very complex and deserve a better discussion of the results. Some part of your results looks better explanation and discussion.
Are the results between RONA and GF congruent? Do they show the same variables?
How are you results hampered by the lack of sampling from the extreme southern and northern populations?

RESPONSE TO REVIEWERS' COMMENTS

Responses to the comments of Reviewer #1

Hou et al. have conducted a haplotype-resolved pan-genome analysis of 16 representative moso bamboo accessions, providing valuable insights into genetic diversity, population structure, and climate adaptation mechanisms in this ecologically and economically vital bamboo species. Specifically, the authors have assembled 16 high-quality genomes to construct a pan-genome of moso bamboo. To my knowledge, this is the first pan-genome for the subfamily Bambusoideae, thus providing extensive genomics resources for bamboo research. Additionally, the authors have assembled haplotype-resolved genomes to improve the moso bamboo reference and reveal actual variation levels through haplotypes. Based on these haplotypes, the authors constructed a haplotype pan-genome, analyzed allele-specific expression (ASE), and proposed that the ASE is related to environmental adaptation. Furthermore, genotype–environment association (GEA) analysis has identified variants associated with climate variables and predicted vulnerable moso bamboo populations. As moso bamboo's economic and ecological value rises amid worsening plastic pollution and climate change, these findings and datasets will critically inform molecular research on bamboo and beyond. While the work is commendable overall, there is room for improvement based on the following suggestions.

Response:

Thank you for your positive feedback and recognition of the significance and value of our work. Your insightful comments and suggestions have been invaluable in helping us improve the quality and clarity of our manuscript. In response to your suggestions, we made substantial revisions, as follows:

1. **Modifying imprecise statements throughout the revision to ensure accuracy and clarity**. We have carefully reviewed the manuscript and

rephrased any ambiguous or unclear sentences, providing additional context where necessary to improve readability and understanding.

2. **Incorporating additional details regarding the analytical methods employed in our study**. We have expanded the Methods section to provide a more comprehensive description of the sample collection, sequencing, and data analysis procedures. These revisions include providing more information on the distribution of the 427 resequenced accessions, as well as the specific bioinformatic tools and parameters used in pangenome construction, haplotype assembly, and ASE and GEA analyses.
3. **Correcting certain figures to ensure they accurately represent the data.** We have thoroughly checked all the figures and made necessary adjustments to improve their clarity and consistency with the results described in the revision.
4. **Updating the references and supplementary information.** We have ensured that all citations are accurate and up-to-date and that the supplementary materials are complete and properly referenced in the revision.

We believe that these revisions have significantly enhanced the quality and impact of our work, making it more accessible and informative for readers. By addressing your concerns and suggestions, we have strengthened the presentation and interpretation of our findings, highlighting the importance of our study in understanding the genomic diversity, population structure, and climate adaptation mechanisms of moso bamboo.

1. Line 37: To quantify the economic importance of the bamboo industry, it would be valuable to include figures related to the scale and development of this sector. Specifically, incorporating metrics such as annual output value, total area of bamboo forests, utilization rates, and contributions to rural incomes could effectively showcase the vital role of the bamboo industry in supporting livelihoods and national economic development. Adding such quantitative details would strengthen the paper by substantiating the significant economic impact of bamboo cultivation and products.

Response:

Thank you for the valuable suggestion. We have incorporated additional quantitative data in the revised manuscript to better illustrate the economic significance of the bamboo industry as follows:

Line 39–42:

"The China bamboo market, valued at $41.58 billion in 2020, is projected to surpass $138.63 billion by 2035[1,2], underscoring the increasing economic significance of the bamboo industry and its potential impact on the global market the increasing economic significance of the bamboo industry worldwide."

2. Line 94: The authors have assembled 16 haplotype moso bamboo genomes with meaningful results. To further characterize the quality of these assemblies, clarifying whether the reported N50 refers to contigs or scaffolds is needed. Moreover, with advancements in sequencing technologies, N50 values have increased substantially. However, directly comparing N50 across species is of limited value due to differences in genome size and chromosome numbers. It is therefore recommended that the author calculate the maximum chromosome-level N50 for moso bamboo, as well as the proportion of the total length represented by the longest 24 contigs (the haploid chromosome number in moso bamboo) at the contig level. Reporting these metrics would better show the continuity and completeness achieved in these assemblies at both the chromosome and contig levels for this species. Specifying both chromosome-level N50 and the total proportion of the genome covered by the longest 24 contigs would thus more clearly convey the assembly quality in a species-appropriate context.

Response:

Thank you for your valuable suggestions regarding the characterization of our assembly quality. To address this concern, we have made the following revisions:

1. **Clarification of N50 values and supplementary data.** To clarify, the N50 values mentioned in the original manuscript specifically refer to the contig-

level N50. In addition, we have supplemented the data with scaffold-level N50 values, which averaged 83 Mb across the 16 haplotype assemblies. Correspondingly, the average scaffold-level L50 value was 10 (please find the revised Supplementary Table 2 for detailed data).

2. **Genome coverage by longest contigs**. Following your recommendation, we determined the total proportion of the genome covered by the 24 longest contigs in each assembly. On average, the 24 longest contigs accounted for 76.25% of the total genome length (revised Supplementary Table 2). This metric provides a clear indication of the completeness and continuity of our assemblies at the contig level, considering the specific characteristics of the moso bamboo genome.

In summary, by reporting these additional metrics, we aim to present a clearer and more comprehensive picture of the assembly quality achieved in our study, accounting for the specific characteristics of the moso bamboo genome. We believe these revisions will enable readers to better appreciate the continuity and completeness of our assemblies at both the chromosome and contig levels in the context of this species.

3. Line 112: Please provide the reasons for choosing the CYhap1 genome as the reference genome. Additionally, for subsequent molecular biology experiments and comparative genomics studies of moso bamboo, does the author still recommend selecting this genome as the reference genome?

Response:

Thank you for raising this important question regarding the selection of the CYhap1 genome as the reference genome for moso bamboo. We appreciate the opportunity to provide a more detailed explanation of our decision and its implications for future research.

1. **CYhap1: A superior reference genome for moso bamboo.** The choice to select CYhap1 was based on its superior assembly quality compared to those

of other available assemblies. CYhap1 exhibited high gene continuity, as evidenced by its elevated contig N50 value. Moreover, CYhap1 exhibited superior phasing accuracy, with a low switch errors, suggesting a high level of haplotype precision. This information is crucial for accurately capturing the heterozygosity and allelic variations present in the moso bamboo genome. The combination of high contiguity and phasing accuracy makes CYhap1 an ideal reference genome for moso bamboo.

2. **Recommendation of CYhap1 for future moso bamboo research.** For future molecular biology experiments and comparative genomics studies involving moso bamboo, we recommend the continued use of CYhap1 as the reference genome. Among the currently available public reference genome assemblies for moso bamboo, CYhap1 stands out as the highest-quality genomic resource. Its superior assembly quality, as demonstrated by the aforementioned metrics, provides an excellent foundation for a wide range of genomic and molecular investigations in this species. The use of a consistent, high-quality reference genome across studies will facilitate the comparison and integration of results, ultimately advancing our understanding of moso bamboo biology.

4. Lines 123-124, 293-295: The authors have made an interesting observation that the majority of variations in moso bamboo stem from inter-haplotype differences. This insight could have broader implications for the study of other asexually reproducing plants. However, I have some concerns about the author's decision to continue using inter-haplotype variations for GEA analysis. It might be more appropriate to filter out these inter-haplotype variations before conducting GEA analysis.

Response:

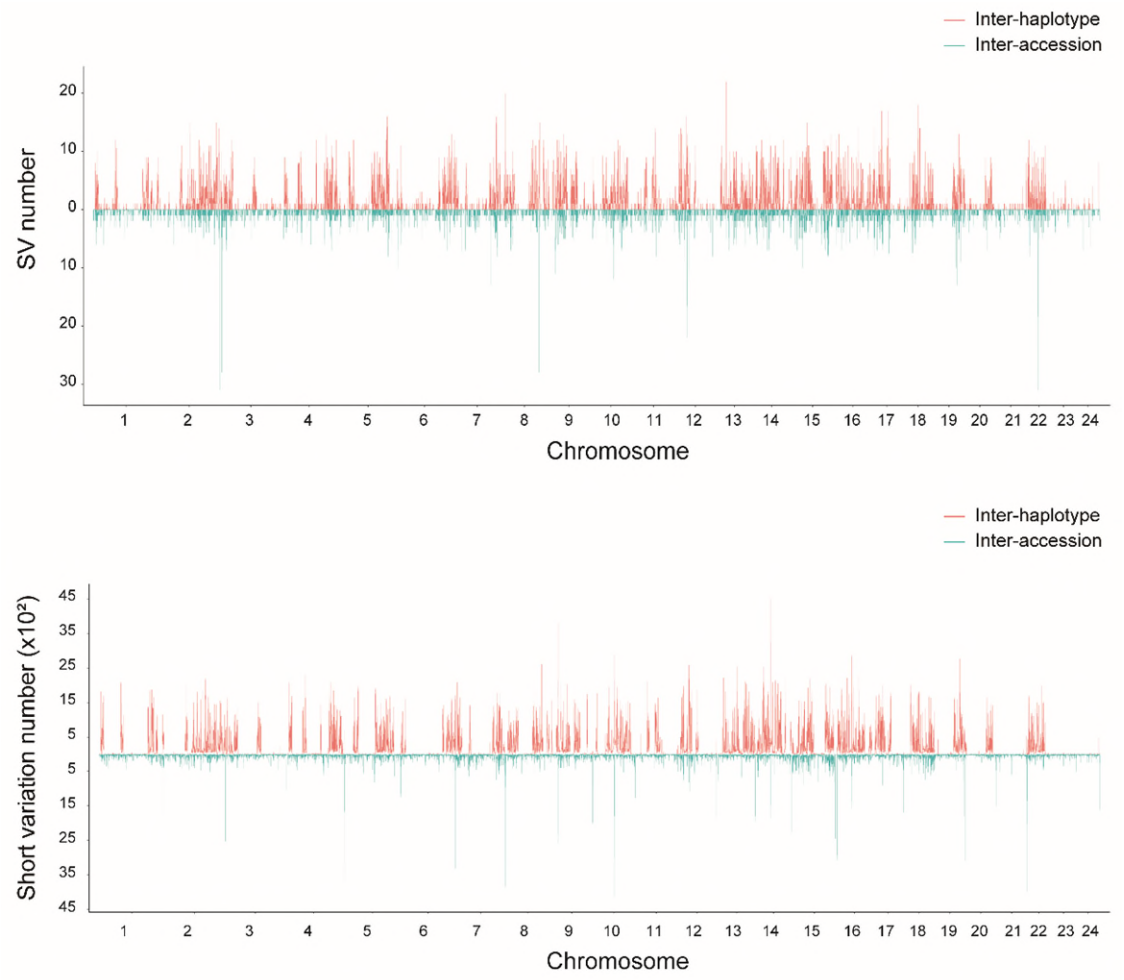
Thank you for the insightful observation regarding the potential implications of our findings on inter-haplotype variations in moso bamboo for the study of other asexually reproducing plants. We have carefully considered the reviewer's concern and made the following revisions to address this issue:

1. **Filtering strategy for the genome–environment association (GEA) analysis.** Considering the reviewer's concern about the use of inter-haplotype variations in the genome–environment association (GEA) analysis, we have carefully considered this issue and implemented a filtering strategy to address it, "Since most variations occurred between haplotypes rather than within accessions in the moso bamboo population, we filtered sites with minor genotype frequencies <0.05, leaving 1,467,461 SNPs, 103,955 InDels and 4,643 SVs for genome-wide identification of climate-associated variations." This approach effectively removes variations that are exclusively present between haplotypes, ensuring that the GEA analysis focuses on biologically meaningful variations.
2. **Rationale for the filtering approach.** For inter-haplotype variations, almost all individuals displayed a "0/1" genotype, indicating the presence of both alleles. By filtering out variations with minor allele frequencies (MAFs) < 0.05 and further removing variations with minor genotype frequencies (the frequency of the least common genotype) < 0.05, we can effectively exclude variations that are solely attributable to differences between haplotypes. This filtering strategy ensures that the GEA analysis captures variations that are more likely to be associated with environmental adaptations and other biologically relevant factors.
3. **Balancing the importance of inter-haplotype variations and the focus of the GEA analysis.** We believe that this approach strikes a balance between acknowledging the importance of inter-haplotype variations in moso bamboo and focusing the GEA analysis on variations that are more informative for understanding the species' adaptations to different environments. Moreover, we acknowledge that the study of inter-haplotype variations in asexually reproducing plants is an important area for future research, and we will perform further investigations into the implications of these variations for the evolution and adaptation of such species.

5. Lines 144-146: It was recommended to add a figure to visualize this result.

Response:

Thank you for the suggestion. To address this concern, we have added a new figure (revised Supplementary Fig. 7, please find below), which provides a clear visual representation of the frequency distribution of variations in moso bamboo.



Supplementary Fig. 7 Frequency distributions of both structural variations (SVs) and short variations (SNPs and InDels) across the moso bamboo genome. The upper panel depicts the distribution of SVs, while the lower panel illustrates the distribution of short variations. The x-axis indicates the chromosome numbers, and the y-axis indicates the number of variations on each chromosome. The blue line represents the frequency of inter-accession variations, which are differences between individual moso bamboo accessions, while the red line represents the frequency of inter-haplotype variations, which are differences between the two haplotypes within each accession.

6. Lines 169-171, 489: The authors have presented an interesting and novel result by constructing a haplotype pan-genome based on allelic genes. However, a more detailed description of the construction process of the haplotype pan-genome is needed, including the relevant thresholds and the rationale behind their selection. For instance, the authors mention that "Gene pairs within 40 kb were retained" in the method. It would be beneficial for the authors to explain why they chose the threshold of 40 kb. This additional information would provide greater clarity and understanding of the methodology used in the study.

Response:

Thank you for your interest in our novel approach for constructing a haplotype pangenome based on allelic genes and for the request for a more detailed description of the construction process. The 40 kb threshold was chosen because it approximates the average gene distance in the moso bamboo genome, which has a size of approximately 2 Gb and contains approximately 50,000 genes. This threshold allows for some positional flexibility, permitting allelic genes to be separated by up to one average gene interval. The rationale behind this choice is to account for potential structural variations and rearrangements that may occur between haplotypes while still maintaining a reasonable level of stringency in identifying allelic gene pairs. Then, the best reciprocal alignments were preserved as allelic pairs based on a combination of closer genomic distance and higher sequence identity. In summary, we have expanded the description of the pangenome construction process in the Methods section, including the relevant thresholds and the rationale behind their selection, as follows:

Line 607–622:

"The determination of alleles between any two haplotype assemblies was based on a combination of protein sequence similarity and relative position. Protein sequence similarity was calculated by aligning every haplotype genome using BLASTP v2.9.0+[3] (-evalue 1e-10). The relative positions between two genes from different genomes were

determined by aligning the assemblies using Minimap2 v2.24-r1122[4] (-ax asm5), and gene pairs within 40 kb (the average gene distance in the moso bamboo genome) were retained for further analysis. The 40 kb threshold was chosen because it approximates the average gene distance in the moso bamboo genome, which has a size of approximately 2 Gb and contains approximately 50,000 genes. This threshold allows for some positional flexibility, permitting alleles to be separated by up to one average gene interval. Based on the criteria of closer genomic distance and higher sequence identity, the best reciprocal alignments were preserved as allele pairs. Genes that met the similarity threshold for the other haplotype but were not among the best-aligned allele pairs were identified as haplotype-specific duplicated genes (Supplementary Table 28). To promote transparency and reproducibility, the script for this step has been made publicly available on GitHub (https://github.com/ZhaoGroupLab/moso-bamboo-pangenome)."

7. Line 882: The authors have successfully constructed 16 genomes with high LAI values, which presents a unique opportunity for exploring LTRs in moso bamboo. Given this, incorporating analyses related to LTRs was recommended. It would be particularly interesting to investigate whether LTRs share inter-haplotype variations, similar to other observed variations. Additionally, considering that LTRs have been linked to structural variations, investigative this association in the context of moso bamboo could yield valuable insights.

Response:

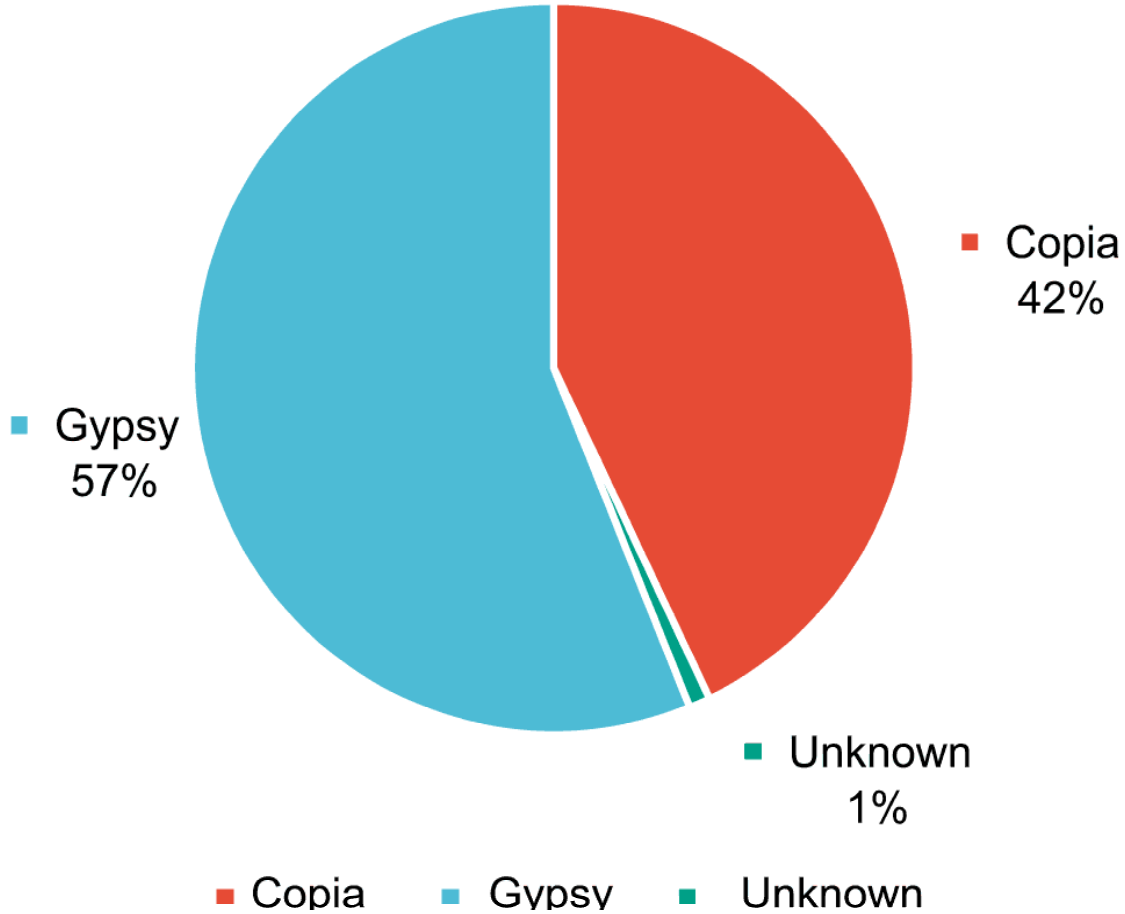
Thank you for the insightful suggestion. To address this concern, we performed an in-depth analysis of LTRs in the revised manuscript to investigate their distribution, inter-haplotype variations, and potential associations with structural variations (SVs). Our findings are as follows:

1. **Identification and classification of LTRs across moso bamboo genomes.** Our analysis revealed a total of 477,068 LTRs across the 16 moso bamboo genomes, with an average of 14,908 LTRs per genome. The identified LTRs

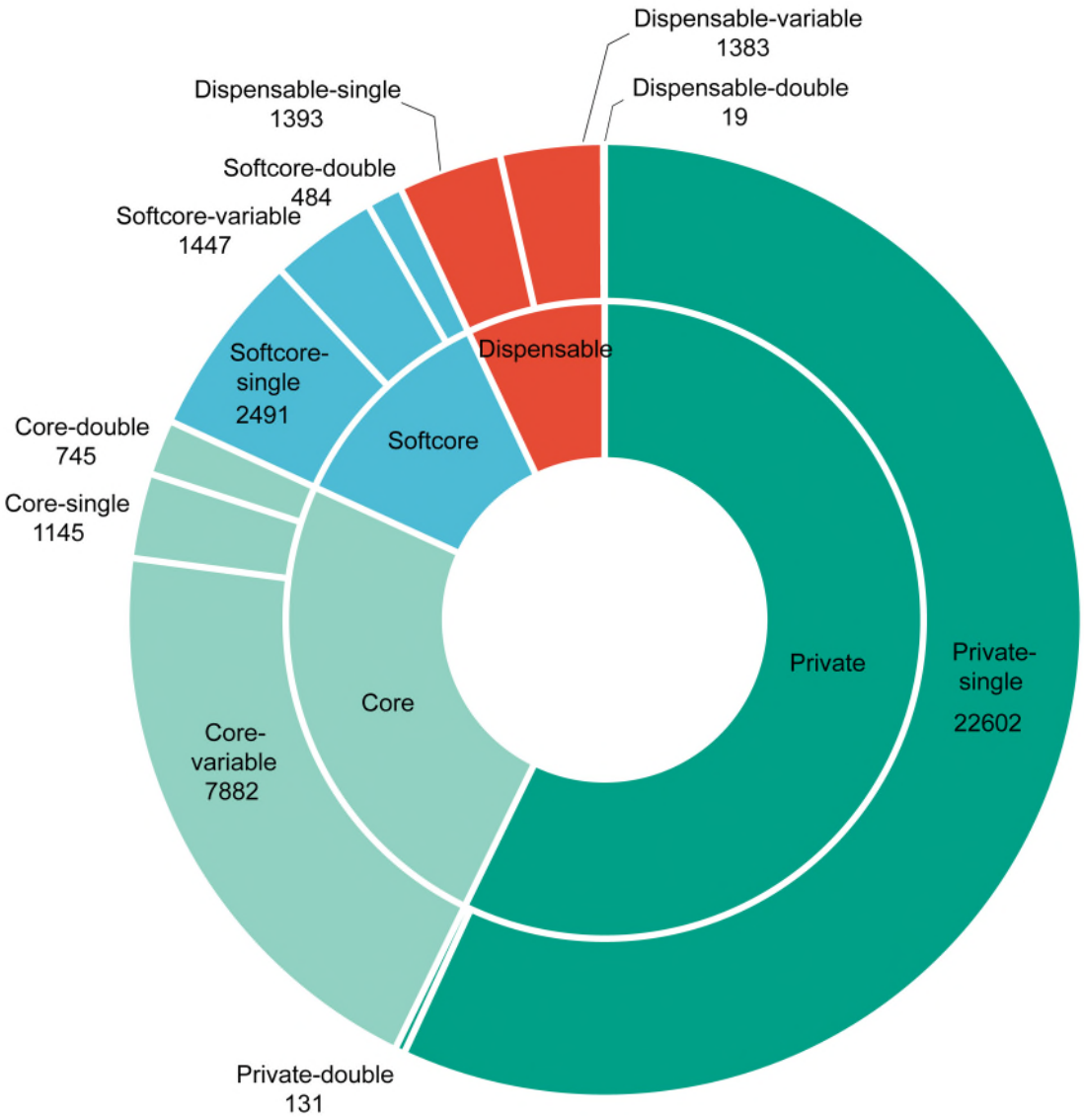
were classified into two major retrotransposon families, with approximately 42% belonging to the Copia family and 57% to the Gypsy family (revised Supplementary Fig. 24).

2. **Inter-haplotype variations in LTRs revealed by cluster analysis.** To investigate the potential inter-haplotype variations in LTRs, we conducted a cluster analysis of the identified LTRs. Interestingly, our results revealed that the "private-single" class, representing LTRs unique to individual haplotypes, constituted the highest proportion at 56.9% (revised Supplementary Fig. 25).
3. **Exploring the association between LTRs and structural variations.** Considering the potential association between LTRs and structural variations (SVs), we explored this relationship in the context of moso bamboo. Within a 5 kb window surrounding the identified LTRs, we detected a total of 5,565 SVs (revised Supplementary Table 29). This finding suggested that the presence or activity of LTRs may influence the formation of these SVs, providing valuable insights into the potential roles of LTRs in shaping structural variations in moso bamboo genomes.

The comprehensive analysis of LTRs in moso bamboo genomes, as presented in our revised manuscript, elucidates their distribution, inter-haplotype variations, and potential associations with structural variations. These findings contribute to a deeper understanding of the complex genetic landscape of moso bamboo and highlight the importance of considering repetitive elements, such as LTRs, in the study of genome evolution and diversity.



Supplementary Fig. 24 Pie chart illustrating the relative abundances of different LTR retrotransposon families in the moso bamboo genome.



Supplementary Fig. 25 Pie chart comparing of LTR retrotransposon types between the two haplotypes within each moso bamboo genome.

8. Line 202: Please ensure uniformity in font sizes in figures, such as Figure 2b.

Response:

Thank you for pointing out this issue. To address the oversight and ensure uniformity throughout the manuscript, we have thoroughly reviewed all the figures, including those in the main text and supplementary materials, to ensure uniformity in font size. Specifically, we have revised Figure 2b to maintain consistent font sizes across all text

elements within the figure. The updated Fig. 2b, presented below, showcases the standardized font sizes for improved consistency and legibility.

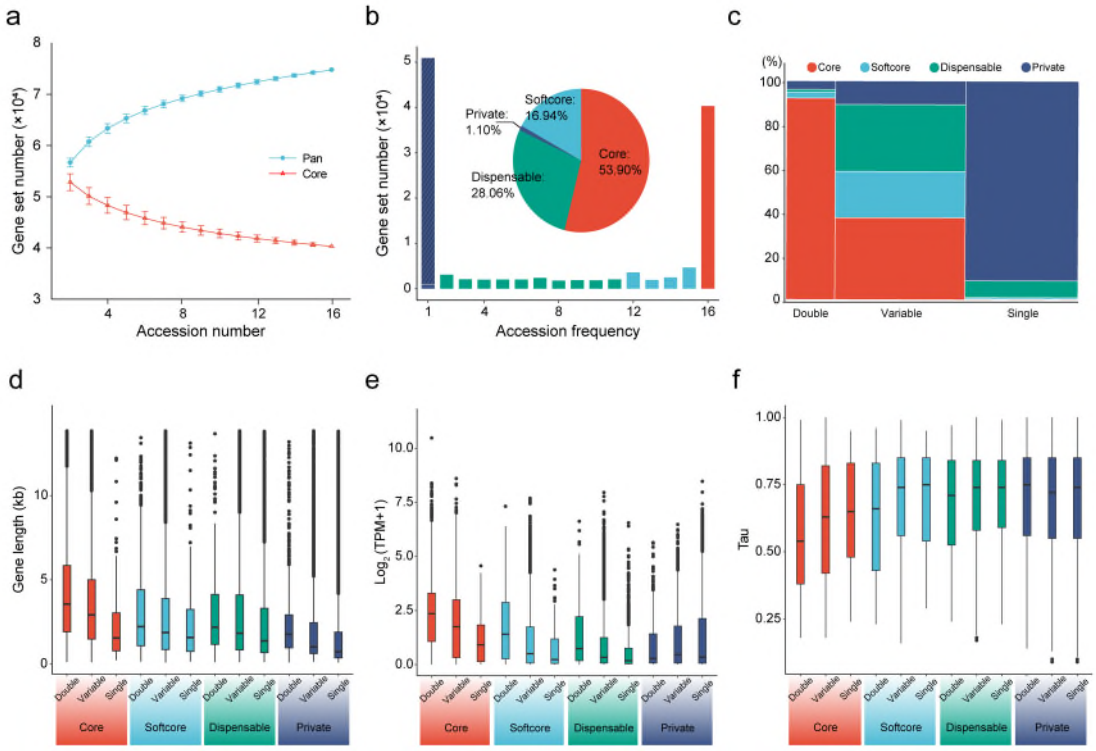


Fig. 2 Classification and characteristics of moso bamboo pangenome gene sets. a The number of gene sets in the pangenome (blue) and core genome (red) increased as a function of the number of moso bamboo accessions in the analysis (x-axis). **b** Compositions of the pangenome. The bar plots show the number of gene sets (y-axis) in each accession categorized by frequency (x-axis). The pie chart depicts the proportions of gene sets marked by each composition category: the core gene set (present in all accessions), softcore gene set (present in 12–15 accessions), dispensable gene set (present in 2–11 accessions), and private gene set (present in only one accession). The left block shows the number of unique gene sets (bottom) and the sum of unclustered genes in each genome (hatched area). **c** Distribution of gene sets in different groups based on gene frequency and alleles composition. The y-axis represents the four groups of moso bamboo divided based on gene frequency across accession (core, softcore, dispensable, and private), and the x-axis represents the 3 gene set groups categorized based on allele composition: double-allele sets (present in both haplotypes across accessions), single-allele sets (only detected in one haplotype across accessions),

and variable-allele sets (present in paired haplotypes of some accessions). Thus, all the gene sets were divided into 12 groups (core-double, core-variable, core-single, softcore-double, softcore-variable, softcore-single, dispensable-double, dispensable-variable, dispensable-single, private-double, private-variable, and private-single). The area of each group is proportional to the number of gene sets. **d, e, and f** Comparison of gene length (**d**), expression level (**e**), and tissue specificity index (Tau) (**f**) across the 12 gene set groups (x-axis). The y-axes show gene length in base pairs, log2(TPM+1) expression values, and the Tau specificity index, respectively. Statistical significance (*P*-value) is provided in Supplementary Tables 13–15.

9. Line 288: Please add related method for identifying short variations in the methods section.

Response:

Thank you for the valuable suggestion. To address this concern, we have incorporated a comprehensive overview of the steps involved in detecting and filtering short variations (SNPs and InDels), ensuring consistency with the previous project wherever possible. The revised Methods section now includes the following information:

Line 707–716:

"SNP and InDel calling based on resequenced reads

The raw sequencing reads were processed using the same pipeline as in our previous study to ensure consistency. Briefly, the filtered resequenced reads were aligned to the CY haplotype 1 reference genome using BWA v0.7.17[5] (-M). Aligned reads (BAM files) were sorted using SAMtools v1.9[6] (default parameters), and duplicates were removed using GATK v4.2.0[7] (default parameters). SNP and InDel calling was performed using the joint calling method within GATK. We obtained the genomic variant call format (GVCF) in ERC mode for each accession based on reads (-ERC GVCF --native-pair-hmm-threads 100). Then, we filtered SNPs directly based on

quality, removing variations with a quality score lower than 50 based on the quality score distribution."

10. Line 445: It was recommended to add the related reference for the LAI method. Assessing genome assembly quality using the LTR Assembly Index (LAI)(https://doi.org/10.1093/nar/gky730).

Response:

Thank you for this suggestion. We have added the following reference for the LAI (LTR Assembly Index) method to the revised manuscript (Reference 43).

"Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res*. **46**, e126(2018)."

11. Line 467: Please capitalize "DIAMOND" and provide the corresponding reference. Sensitive protein alignments at tree-of-life scale using DIAMOND (doi:10.1038/s41592-021-01101-x).

Response:

Thank you for pointing out this issue. We have capitalized "DIAMOND" throughout the text to ensure consistency and proper formatting of the software name. We have added the following reference to the revised manuscript (Reference 55).

"Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368(2021)."

12. Line 475: Please add the related reference.

Response:

Thank you for pointing out this issue. We have added the following reference to the revised manuscript (Reference 21).

"Zhang, X. *et al.* Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis*. *Nat. Genet.* **53**, 1250–1259(2021)."

13. Line 678: It was recommended to cite the latest BUSCO reference. BUSCO: Assessing Genomic Data Quality and Beyond (https://doi.org/10.1002/cpz1.323).

Response:

Thank you for this suggestion. We have updated the reference for BUSCO to the most recent version in the revised manuscript. The corrected reference is as follows (Reference 42):

"Manni, M., Berkeley, M. R., Seppey, M. & Zdobnov, E. M. BUSCO: assessing genomic data quality and beyond. *Curr. Protoc.* **1**, e323(2021)."

14. Line 708: Please use the latest reference for the NCBI Nr database to ensure that the study adheres to the most up-to-date database and methodologies. Reference sequence(RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation (https://doi.org/10.1093/nar/gkv1189).

Response:

Thank you for this suggestion. We have updated the reference for the NCBI Nr database to the latest version in the revision (Reference 56), as follows.

"O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745(2016)."

15. Line 710: Please use the latest reference for the UniProt database. UniProt: the Universal Protein Knowledgebase in 2023 (https://doi.org/10.1093/nar/gkac1052).

Response:

Thank you for this suggestion. We have updated the reference for the UniProt database to the latest version in the revision (Reference 57), as follows:

"UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531(2023)."

16. Line 714: The reference for the eggNOG database seems to be incorrect, should be:eggNOG 6.0: enabling comparative genomics across 12 535organisms(https://doi.org/10.1093/nar/gkac1022).

Response:

Thank you for pointing out this issue. We have updated the reference for the eggNOG database to the correct version in the revision (Reference 59), as follows:

"Hernández-Plaza, A. *et al.* eggNOG 6.0: enabling comparative genomics across 12 535 organisms. *Nucleic Acids Res.* **51**, D389–D394(2023)."

17. Supplementary Table 11: It was recommended to fill in the table according to accession numbers.

Response:

Thank you for the recommendation. We have updated the table, now designated Supplementary Table 12 (Supplementary Table 11 in the previous version of the manuscript), by including the related accession numbers for each entry.

"Supplementary Table 12. Gene membership in pangenome gene set groups."

18. Supplementary Table 19: Please correct the table and arrange environmental variables in numerical order.

Response:

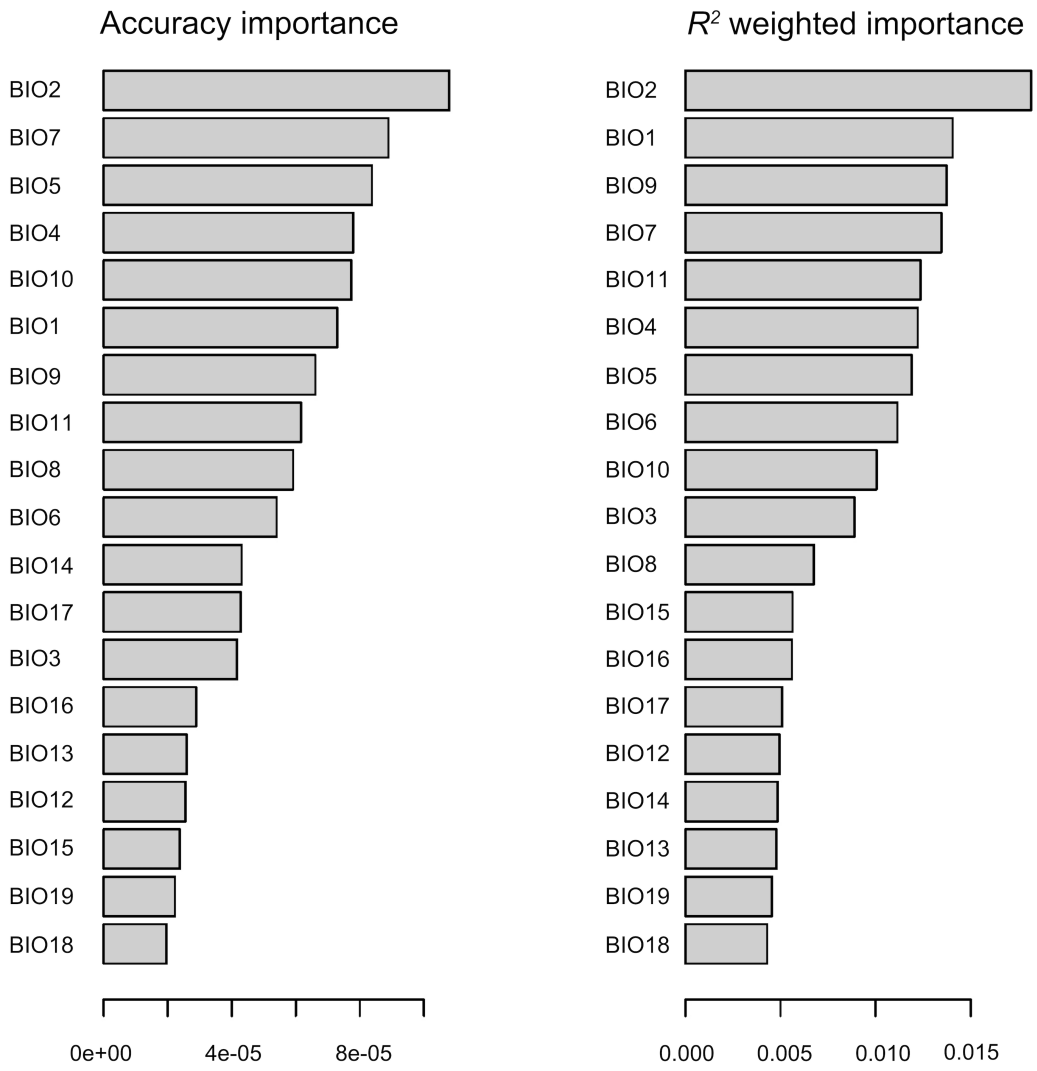
Thank you for the suggestion. We have revised the table, now designated Supplementary Table 21 (Supplementary Table 19 in the previous version of the manuscript), to ensure that the bioclimatic variables are presented in a clear and logical sequence.

"Supplementary Table 21. Nineteen bioclimatic variables across moso bamboo populations from the WorldClim database."

19. Supplementary Fig. 12: Please add related descriptions.

Response:

Thank you for this suggestion. We have added a detailed description of the figure, now designated Supplementary Fig. 15 (Supplementary Fig. 12 in the previous version of the manuscript), to provide context and facilitate the interpretation of the data presented.



Supplementary Fig. 15 Gradient Forest (GF) ranking of bioclimatic variables. The left panel shows the accuracy importance, and the right panel shows the R^2 weight importance.

Responses to the comments of Reviewer #2

The manuscript by Hou et al. presents the creation of a haplotype-resolved pan-genome for moso bamboo. The authors conducted an in-depth analysis of this pan-genome, utilizing 427 re-sequenced samples to uncover significant genetic variations and allele-specific gene expressions linked to climatic adaptability. They leveraged these genomic insights alongside climatic data to attempt predictions of vulnerable moso bamboo populations, potentially aiding in assessing risks posed by environmental changes. While the data undeniably constitutes a valuable genomic resource for bamboo research, the analyses tend to be descriptive, and several conclusions are drawn without robust supporting evidence.

Response:

Thank you for your acknowledgment of the value of our work and the constructive suggestions for improvement. We have carefully considered the concerns raised regarding the descriptive nature of some analyses and the lack of robust supporting evidence for certain conclusions. In response to the feedback, we have made substantial revisions to address these issues and strengthen the manuscript:

1. **Refinement of allele-specific expression (ASE) analysis**. To ensure the robustness of our ASE results, we reanalyzed the data using only uniquely aligned RNA-Seq reads. This was achieved by specifying the parameter "-k 1" in HISAT2, which ensures the inclusion of reads that uniquely align to the reference genome. This stringent filtering approach significantly reduces the influence of ambiguously mapped reads on ASE estimates, thereby enhancing the reliability of our findings on allele-specific gene expression patterns.
2. **Expansion of supporting data and analyses**. We have included additional supplementary data and analyses to further substantiate our claims and provide a more comprehensive picture of the genetic variations and their potential implications for climatic adaptability in moso bamboo. These include detailed information on the functional annotation of genes exhibiting significant ASE, as well as additional analyses exploring the relationships

between genetic variations and specific bioclimatic variables. These additions strengthen the evidence supporting our conclusions and provide a more robust foundation for our interpretations.

3. **The Discussion section was fully rewritten**. The Discussion section has undergone extensive revisions to provide a more in-depth and nuanced interpretation of our findings. We investigated the potential mechanisms underlying the observed genetic variations and their relevance to the climatic adaptability of moso bamboo. The revised Discussion section also places our results in the context of previous studies and highlights the novel insights gained from our work. This enhanced discussion provides a more comprehensive and balanced perspective on the implications of our findings.
4. **Enhancement and optimization of language and expression.** In collaboration with American Journal Experts (AJE), a subsidiary of Springer Nature, we have meticulously refined the language used throughout the revised manuscript and response letter, and an AJE editing certificate is provided below. Ambiguous or unsupported statements have been removed or clarified, and the narrative has been focused more sharply on key findings and their scientific implications. These language improvements have enhanced the clarity, precision, and scholarly rigor of the manuscript.

Response Fig. 1 AJE editing certificate.

1. The title 'haplotype-resolved pan-genomes' is misleading. Typically, 'haplotype-resolved' refers to a phased genome assembly of a single genome, applicable to both diploids and polyploids, rather than pan-genomes. The authors have actually constructed a graphical pan-genome by integrating structural variations from 16 haplotype-resolved genomes (i.e.,32 haplotypes), as described in their methodology.

Response:

Thank you for your suggestion regarding the potential ambiguity of the term “haplotype-resolved” in the context of our work. We agree that this term is more commonly associated with phased genome assemblies of individual genomes than with pangenomes. However, it is crucial to emphasize that our pangenome was indeed constructed based on haplotypes, as this is a key aspect of our method. To address this concern and improve the clarity of our manuscript, we have made the following modifications:

1. **Title revision**: We have revised the title to reflect the nature of our work more accurately. The revised title is: "Haplotype-based pangenomes reveal genetic variations and climate adaptations in moso bamboo populations".
2. **Revision consistency:** We have conducted a thorough review of the entire manuscript and have removed any instances where the description "haplotype-resolved pangenome" was used. This revision ensures consistency with the updated title and enhances the overall clarity of our work.

2. Line 91, Supplementary Figure 1 merely depicts the geographic distribution of the 16 RMA samples. It does not clearly illustrate the claimed extensive genetic diversity. The rationale behind selecting these 16 RMAs is unclear—are there specific characteristics that warranted their selection?

Response:

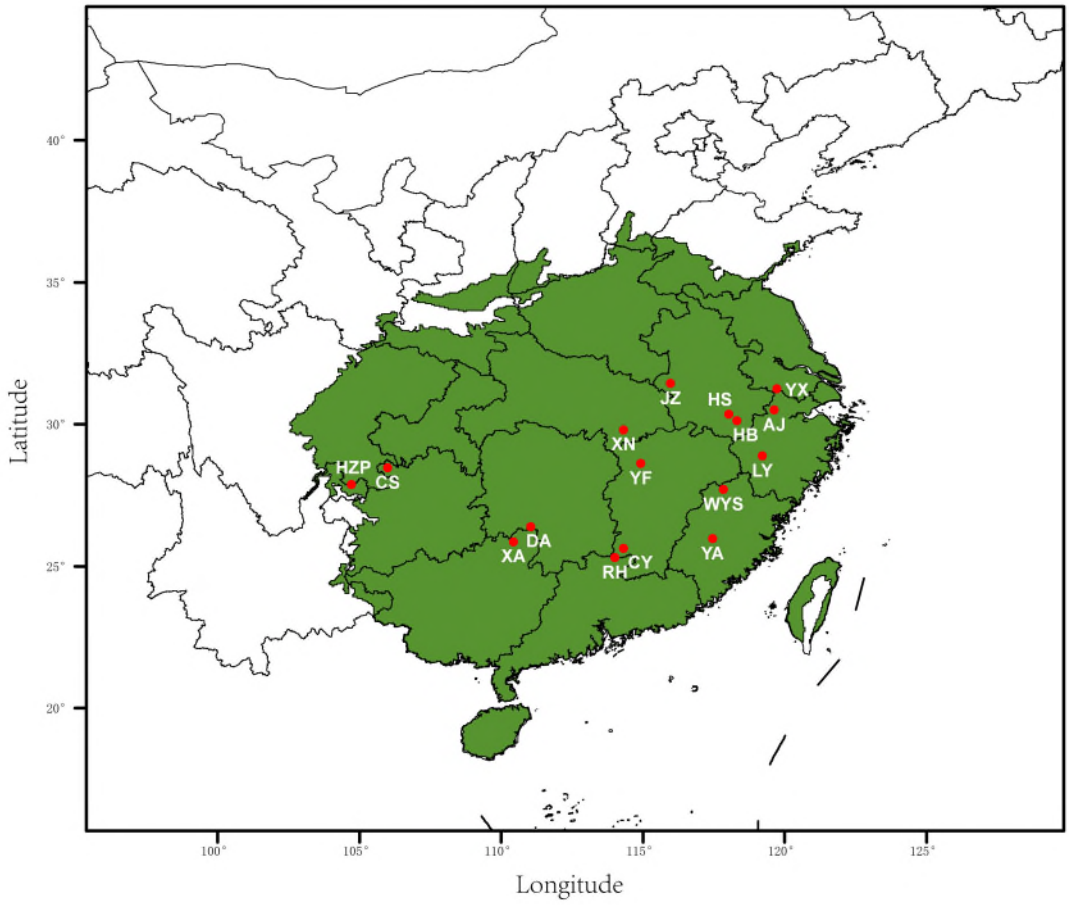
Thank you for raising this point. To address this concern, we implemented the modifications to improve the clarity and readability, as detailed below:

1. **Comprehensive sampling of major natural moso bamboo populations.** Our team has been dedicated to the investigation and exploration of bamboo germplasm resources for many years. Our team have published the first genome of the bamboo subfamily, i.e., the moso bamboo genome[8,9], conducted a comprehensive survey of moso bamboo germplasm resources in China[10], and published the first population resequencing study in the bamboo subfamily, focusing on moso bamboo population resequencing[11]. Through these efforts, we have preliminarily determined the distribution of natural moso bamboo populations and ensured that our sampling in this project has fully covered the relevant regions. Therefore, we would like to express that our sampling regions have already covered all the major natural distribution regions of moso bamboo in China, which we believe is sufficient for identifying the adaptive variations and predicting the risks for moso bamboo in these regions.

2. **Clarification of sampling rationale in methods section.** However, as you correctly pointed out, the necessary description in the manuscript to explain this rationale was not provided. Thus, we have added the related description and references to the sample collection in the Methods section, as follows:

Line 518–523:

"To optimize the representation of genetic and environmental diversity, we selected 16 representative moso bamboo accessions (RMAs) based on a previous phylogenetic study that identified the species' primary natural distribution in China[10,11] (Supplementary Fig. 1). Our sampling strategy aimed to capture the extensive genetic diversity present in moso bamboo by covering all its major habitats, ensuring a comprehensive representation of the populations."



Supplementary Fig. 1 Geographic locations of the 16 representative moso bamboo accessions across China. The specific coordinates of each accession are denoted by

red dots. The suitable habitat range of moso bamboo across China was delineated based on Global Biodiversity Information Facility (GBIF) data (green shaded area).

3. Of the 16 RMAs, only three have been anchored to chromosomal-scale assemblies using Hi-C reads, leaving 13 at a contig-level assembly, which does not fully resolve haplotypes. I recommend that the term 'haplotype-resolved' be used cautiously and accurately, reflecting the actual level of assembly achieved.

Response:

Thank you for your thoughtful suggestion regarding the use of the term "haplotype-resolved" in the manuscript. To address this concern and provide a clear and concise explanation, we have made the following revisions:

1. **Clarification of the assembly status of the 16 RMAs.** We acknowledge that our initial use of the term "haplotype-resolved" was inaccurate, as only three of the 16 RMAs were assembled with Hi-C reads, while the remaining 13 assemblies did not fully resolve the haplotypes. To ensure accurate and cautious use of the term "haplotype-resolved", we have carefully revised the related expressions throughout the revision and added the corresponding description to clarify the assembly status of the 16 RMAs (please find below).
2. **Clarification of the extent of haplotype assembly in the remaining 13 assemblies.** While the remaining 13 assemblies did not fully resolve the haplotypes, they still assembled the haplotypes to a certain extent, as evidenced by the results provided by haplotype assembly software (Hifiasm[12]) and its quality assessment indicators (switch errors). However, to avoid any potential confusion or overstatement, we have refrained from using the term "haplotype-resolved" to describe these assemblies.
3. **Minimal impact on subsequent haplotype-related analyses.** Although we did not perform haplotype-resolved assembly for all the accessions, this did not affect our subsequent analyses. The subsequent haplotype-related analyses primarily focused on the differences between the two haplotypes, such as

inter-haplotype variation, which represents the variation between two haplotypes, and allele-specific expression (ASE), which utilizes the expression differences between allele pairs. These analyses do not require a particular segment to be fully phased to a specific haplotype, as they rely on the identification of variations and expression differences between allele pairs.

Additionally, we have added corresponding descriptions for the 16 RMAs, as follows:

Line 110–115:

"These data allowed the construction of 16 high-quality assemblies (32 haplotype assemblies) with an average contig N50 length of 57.0 Mb (Table 1 and Supplementary Fig. 2). The average quality value (QV) of the final assembly was 64.26, with a k-mer completeness of 98.20% (Supplementary Fig. 3 and Supplementary Table 2). We observed an average switch errors of 5.44% for all assemblies (Table 1 and Supplementary Table 3)."

4. Line 105 mentions 'an average of 54,815 protein-coding genes.' Is this figure inclusive of allelic genes? It would be informative to know the count of bi-allelic genes versus those present as a single allele.

Response:

Thank you for your suggestion. According to your recommendation, we have calculated the number of biallelic and single-allele genes in the Results section and added the relevant results to the revised manuscript as follows:

Line 122–127:

"we predicted an average of 54,343 protein-coding gene models in all 32 haplotype assemblies (Table 1), with an average of 97.9% of them assigned putative functions (Supplementary Table 6). Among these gene models, an average of 92,506 genes were present as biallelic genes (corresponding to 46,253 allele pairs in each RMA), while

8,090 genes, on average, were present as single alleles in the haplotype assemblies (Supplementary Table 7)."

5. Lines 123-124 note that inter-haplotypic variations exceed the genetic variations among different accessions. An explanation or hypothesis to account for this observation would be beneficial.

Response:

Thank you for the suggestion. The distinction between inter-haplotype and inter-accession variations is crucial in our study because it allows us to capture the true extent of genetic diversity within and among moso bamboo accessions. By considering variations at both the haplotype and accession levels, we gained a more comprehensive understanding of the genomic heterogeneity present in this species. To address this concern, we elucidated two key aspects as follows:

1. **Quantifying genomic diversity: dissecting inter-haplotype and inter-accession variations.** Inter-haplotype variations refer to the differences between two haplotypes within a single accession, while inter-accession variations represent the differences between different accessions. Inter-haplotype variations were calculated by aligning the two haplotypes of each accession and identifying the variations between them, while inter-accession variations were calculated by aligning the genomes of different accessions and identifying the variations between them. Our analysis revealed that the numbers of inter-haplotype short variations (SNPs and InDels) and structural variations (SVs) were, on average, 10.4 times and 5.3 times greater, respectively, than those of inter-accession short variations and SVs.
2. **The evolutionary significance of haplotype-level variations in moso bamboo.** Our analyses revealed that inter-haplotype variation was substantially greater than inter-accession variation, suggesting that the primary source of genetic diversity in moso bamboo is the divergence between haplotypes within a single accession rather than the divergence between

different accessions. This finding has significant implications for our understanding of the evolutionary history and reproductive biology of moso bamboo. As a species that primarily reproduces asexually, moso bamboo accumulates genetic variations through rare somatic mutations within haplotypes rather than through meiotic recombination. The absence of meiosis and the long generation times of moso bamboo contributed to the maintenance of these haplotype-level variations over extended periods, leading to the observed pattern of higher inter-haplotype diversity than inter-accession diversity.

The details are provided below in the revised manuscript.

Line 474–484:

"Given the asexual reproduction of moso bamboo over extended periods and its 67-year flowering cycle[13,14], the primary source of variation is likely rare somatic mutations occurring within one haplotype. Asexual reproduction makes it difficult for variations accumulated in accessions to be transmitted, as the absence of meiosis prevents the exchange of genetic material between homologous chromosomes. We hypothesized that there would have been a difference between the two haplotypes in the common ancestor of moso bamboo populations in different regions and that the accumulation of somatic mutations in moso bamboo from different regions did not exceed the original difference between the two ancestral haplotypes. These factors have led to the phenomenon where quantitatively, inter-haplotype variations exceed the genetic variations among different accessions."

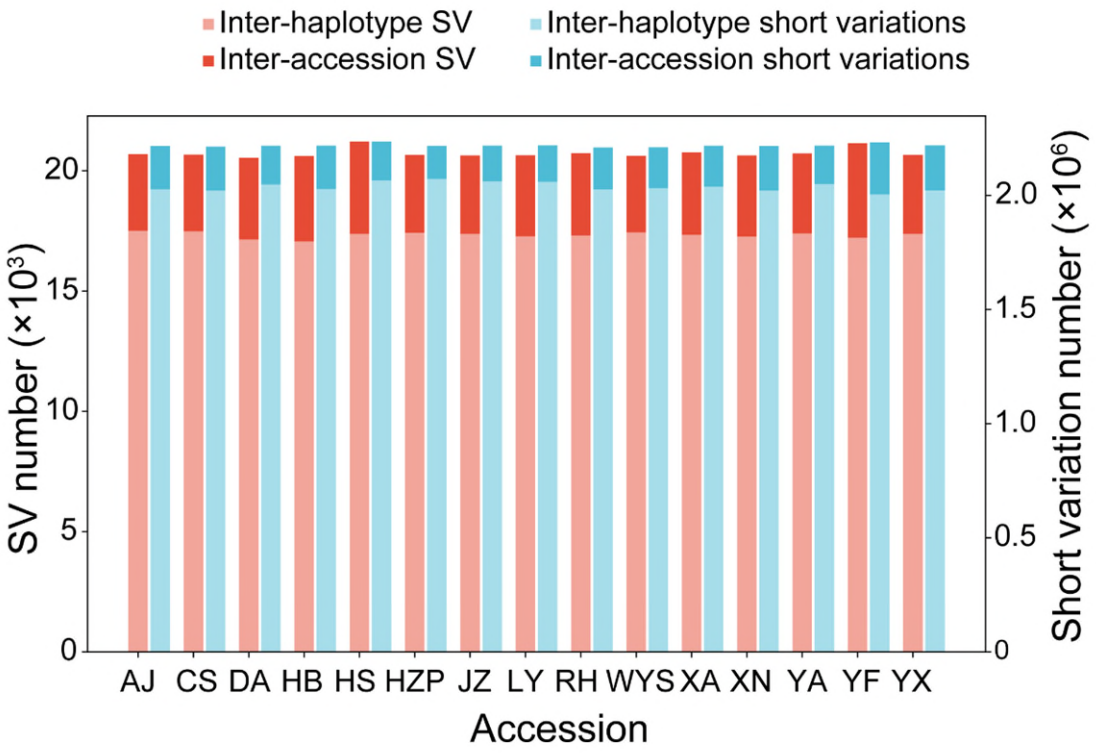
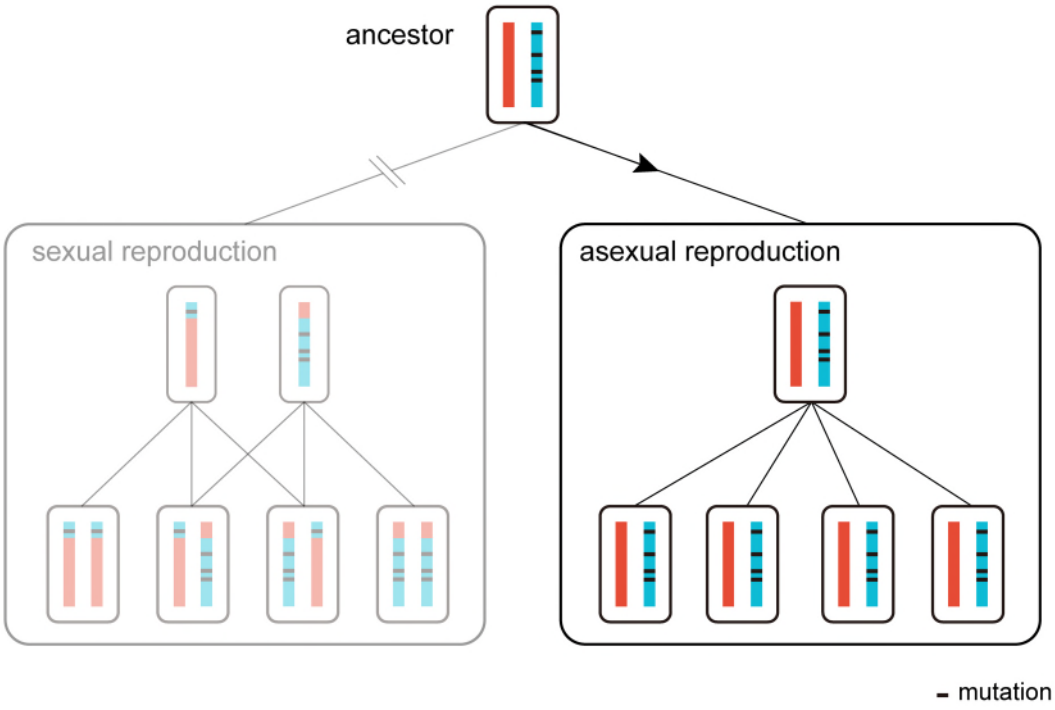


Fig. 1b Number of SVs (red) and short variations (SNPs and InDels, blue) categorized as either inter-accession (darker colors) or inter-haplotype (lighter colors). Inter-accession variations are absent between the reference haplotypes but present in other accessions. Inter-haplotype variations are present between the reference haplotypes. The x-axis represents accessions, and the y-axis shows the number of SVs/short variations.



Supplementary Fig. 26 Schematic diagrams illustrating sexual reproduction and asexual reproduction in moso bamboo. Although moso bamboo possesses the ability for both sexual (left) and asexual (right) reproduction, it predominantly relies on asexual reproduction through rhizome growth and vegetative propagation. This hypothesis, grounded in the reproductive biology of moso bamboo, provides a plausible explanation for the observed pattern of genetic variation. By considering the rarity of somatic mutations, the lack of meiotic recombination due to the predominance of asexual reproduction, and the potential for haplotype divergence in the ancestral population, we provided a schematic diagram for understanding the predominance of inter-haplotype variations in moso bamboo.

6. Lines 177-181 introduce the classification of double-allele, single-allele, and variable-allele gene sets, but the biological significance and relevance of these categories are not clear. What specific biological questions are these data intended to address?

Response:

Thank you for this suggestion. To address these concerns, we have carefully considered

each point and provided detailed responses accordingly.

1. **Clarification of the gene classification criteria used in our study.** We defined three gene sets based on their presence and distribution across haplotypes and accessions: double-allele, single-allele, and variable-allele sets. double-allele gene sets (the allele pair was detected in all accessions), single-allele gene sets (only one allele was detected in all accessions), and variable-allele gene sets (the allele pair was detected in some accessions).
2. **Challenges in defining specific functions for the three gene sets.** Our results indicate that double-allele gene sets account for a greater proportion (92.1%) of core gene sets than do variable- and single-allele gene sets, supporting the hypothesis that they may play more essential roles in moso bamboo biology. Conversely, the high degree of overlap between single-allele gene sets and accession-specific gene sets (91.3%) suggests that single-allele gene sets may be associated with functions specific to certain moso bamboo accessions from particular regions. Despite these observations, directly defining the specific functions of these gene sets remains challenging due to the vast number of genes involved in this genome-wide classification and the diverse range of functions they encompass. The complexity and breadth of the gene sets make it difficult to assign precise functional roles without further targeted investigations.
3. **Focus on a more specialized gene set.** To gain more targeted insights into the functional implications of allele-specific gene presence, we focused our attention on a more specialized gene set, the core-single gene set. This gene set refers to genes present in only one haplotype across all accessions, potentially representing a group of genes with unique allele-specific functions. By concentrating on this specific subset, we aimed to provide a more manageable and informative avenue for exploring the functional consequences of the presence of allele-specific genes in moso bamboo.

Based on the above information, we have made corresponding modifications in the

revised manuscript, as follows:

Line 224–233:

"Additionally, we focused on the core-single gene set, which represents genes present in all accessions but only in one haplotype assembly. Among the 47 core-single gene sets with known functions and an TPM greater than 1 in at least one RNA-Seq accession, we identified 27 gene sets whose functions were related to the environment adaptation (Supplementary Table 16). The functions of these gene sets include stress tolerance (e.g., the gene set GS0035370, encoding aldo-keto reductase 1 (AKR1))[15], disease resistance (e.g., the gene set GS0058418, encoding disease resistance protein RPM1)[16], and DNA damage repair (e.g., the gene set GS0062031, encoding dynamics of the (6-4) photolyase)[17]. These results suggested that some haplotype-specific genes may be involved in the environmental adaptation of moso bamboo."

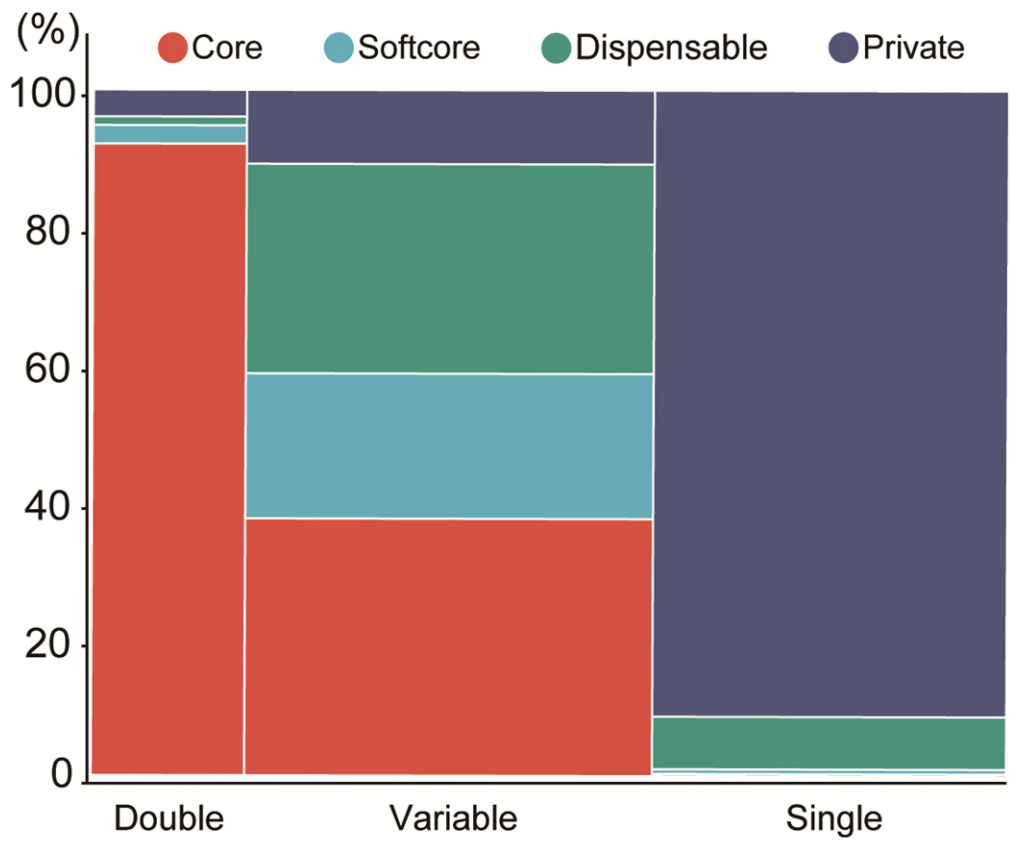


Fig. 2c Distribution of gene sets in different groups based on gene frequency and

alleles composition. The y-axis represents the four groups of moso bamboo divided based on gene frequency across accession (core, softcore, dispensable, and private), and the x-axis represents the 3 gene set groups categorized based on allele composition: double-allele sets (present in both haplotypes across accessions), single-allele sets (only detected in one haplotype across accessions), and variable-allele sets (present in paired haplotypes of some accessions). Thus, all the gene sets were divided into 12 groups (core-double, core-variable, core-single, softcore-double, softcore-variable, softcore-single, dispensable-double, dispensable-variable, dispensable-single, private-double, private-variable, and private-single). The area of each group is proportional to the number of gene sets.

7. Line 255-257 draws a conclusion that seems speculative. There isn't adequate evidence provided to substantiate the claim that certain inter-haplotype variations lead to allele-specific expression (ASE).

Response:

Thank you for your suggestion. As you noted, we agree that the evidence presented does not adequately substantiate the claim that certain inter-haplotype variations directly lead to ASE. Thus, we have revised the statement as follows:

Line 289–290:

"These results suggest that certain inter-haplotype variations could be associated with consistent ASE events."

8. Line 289 suggests that a population analysis of the 427 accessions should be conducted to understand their characteristics, population structure, and genetic diversity.

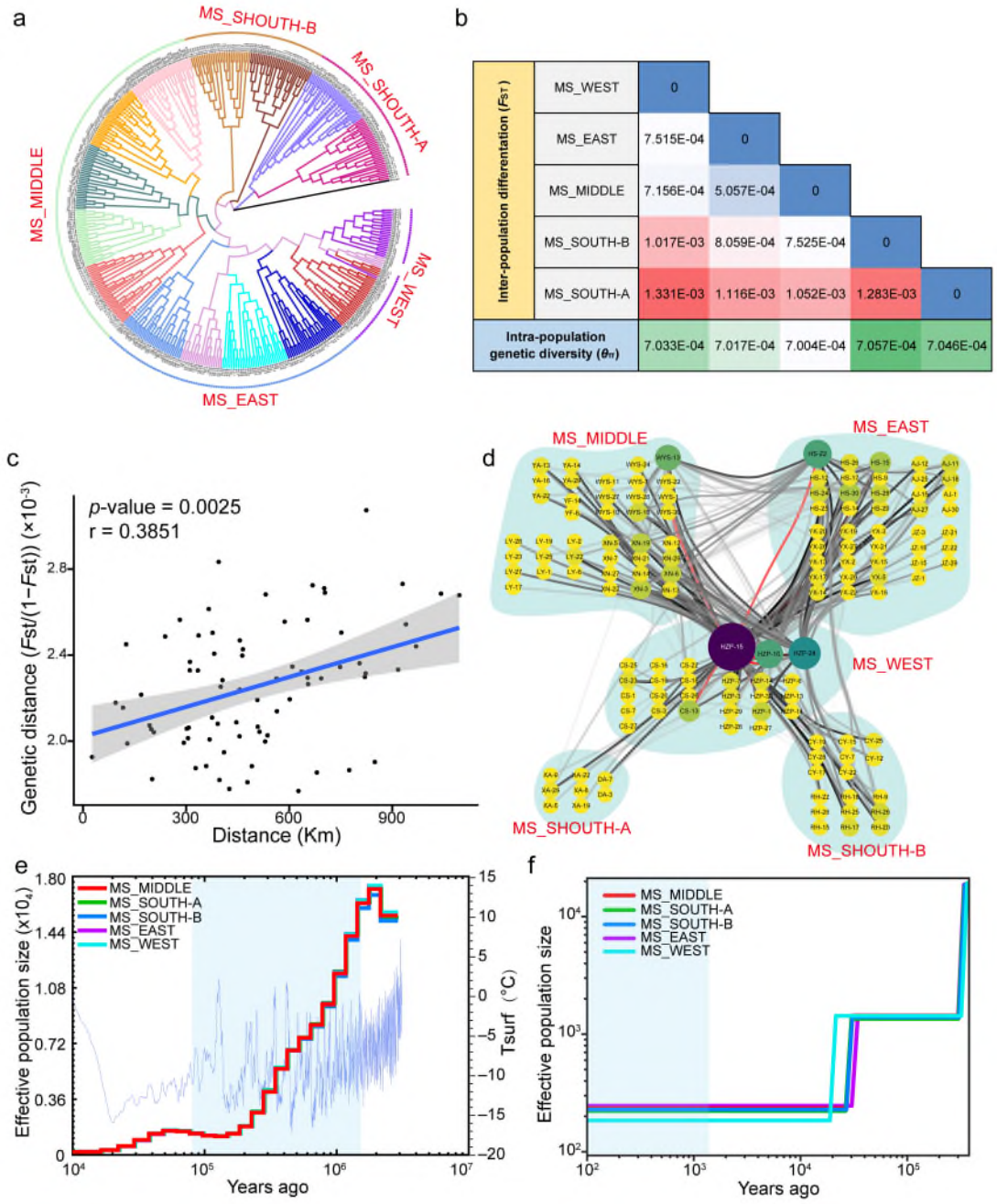
Response:

Thank you for this suggestion. We systematically performed and reported a comprehensive population analysis of these 427 accessions in our previous study[11]. To address this, we have made the following revisions:

1. **Summary of previous population analysis findings**

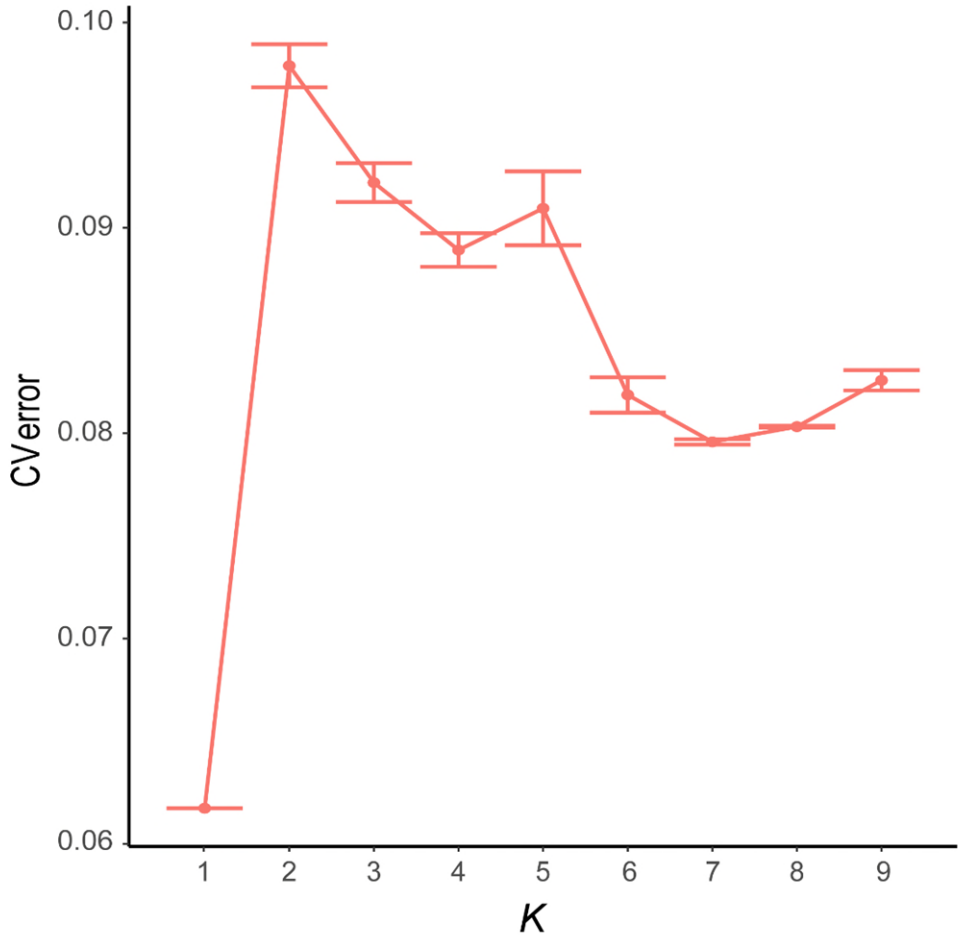
   - **ADMIXTURE analysis results.** The ADMIXTURE analysis with multiple random seeds revealed that the optimal model had a *K* value of 1, indicating rare and slight population differentiation. Based on the phylogenetic tree and geographic separation distances, the 427 accessions were divided into five phylogenetic groups.
   - **Population differentiation and isolation by distance.** Wright's *F* statistics (F_{ST}) reflected relatively low population differentiation. An isolation-by-distance (IBD) analysis demonstrated a significant positive correlation between geographic and genetic distance for the moso bamboo population.
   - **Demographic history and potential factors.** The inferred ancient population bottleneck and recent small population size without a rebound are possibly related to climate change and human activities.

2. **Rechecking the *K* value in the ADMIXTURE analysis.** Given the thorough analysis and reporting of these results, we did not repeat them in the previous version of the manuscript. However, because the reference genome update and gene–environment association (GEA) analyses require *K* values, we reperformed the ADMIXTURE analysis to confirm the accuracy of the population structure. The results consistently showed a *K* value of 1 (Supplementary Fig. 13), which was consistent with our previous findings.



Response Fig. 2 Overall population structure landscape and the inferred population demographic history (Zhao, H. et al. Nat. Commun. 2021)**. a** Rooted neighbor-joining phylogenetic tree of 427 moso bamboo individuals. The different line colors represent the fifteen geographical geographic areas, and the differently colored dotted lines nearby represent five groups that were empirically assigned in our study. **b** The genetic diversity ($\theta\pi$) and F_{ST} matrix of the five groups. The colors and numbers in the cells of the matrix represent the F_{ST} values. The numbers in the cells below the F_{ST} matrix represent the genetic diversity ($\theta\pi$). **c** Results of the Mantel test of the

relationship between geographical distance and genetic distance with MS_WEST excluded. The region in gray represents the 95% confidence intervals. **d** The connection of individuals with the lowest 1% pairwise genetic distances. The size and color of circles represent the degree of connectivity to a node. The lines in different colors indicated values of Hamming distance (genetic distance), with red indicating the shortest distance and for the others, darker colors indicate shorter distances and lighter colors indicate longer distances. **e** The demographic history of the fifteen geographic areas was inferred separately using the pairwise sequential Markovian coalescent (PSMC) method. The light blue line represents the historical surface temperatures, and the light blue shade indicates the bottleneck experienced during the last glacial period (LGP, 115,000-11,700 years ago). **f** The demographic history was inferred using SMC++. The LGP was shaded in light blue, and the reduction without a rebound in the effective population size during the last 2,000 years is shaded in light green. The results were scaled to real-time by assuming a generation time of 67 years and a mutation rate of 8.51×10^{-8} per generation.



Supplementary Fig. 13 Cross-validation (CV) error values for different values of *K* in an ADMIXTURE analysis.

9. Line 296 queries the meaning of six climate variables (BIO1, BIO2, BIO5, etc.). Clarification on what these variables represent is necessary.

Response:

Thank you for this suggestion. To address this concern, we have provided a brief description of each variable in the revised manuscript (please find below). Additionally, we have provided a comprehensive table (revised Supplementary Table 21) that includes the definitions and units for all 19 bioclimatic variables used in our study.

Line 334–339:

"Based on variable correlation and gradient forest (GF) ranking (see Methods)

(Supplementary Figs. 14–15), six bioclimatic variables were selected for RDA to further filter variations. These variables included annual mean temperature (BIO1), mean diurnal range (BIO2), max temperature of warmest month (BIO5), mean temperature of wettest quarter (BIO8), precipitation of driest month (BIO14), and precipitation seasonality (BIO15)."

Line 326–330:

"We applied both latent factor mixed models 2 (LFMM2) and redundancy analysis (RDA) to identify climate-associated genetic variations using 19 bioclimatic (BIO) variables from WorldClim, including 11 temperature-related variables (BIO1–BIO11) and 8 precipitation-related variables (BIO12–BIO19, Supplementary Table 21)."

"Supplementary Table 21. Nineteen bioclimatic variables across moso bamboo populations from WorldClim database."

10. Line 340 mentions SSP585 and SSP126 without prior explanation. These should be defined upon their initial mention in the text.

Response:

Thank you for this suggestion. We have added detailed explanations of the SSP585 and SSP126 scenarios to both the Results and Methods sections as follows:

In the Results section (Line 390–399):

"Four general circulation models (GCMs) were considered: the Australian Community Climate and Earth System Simulator Coupled Model version 2 (ACCESS-CM2)[18], the second generation CMCC Earth System Model (CMCC-ESM2)[19], the Goddard Institute for Space Studies Model E version 2.1 coupled with the GISS Ocean (GISS-E2-1-G)[20], and the Model for Interdisciplinary Research on Climate version 6 (MIROC6)[21], which participate in the World Climate Research Programme Coupled Model Intercomparison Project Phase 6 (WCRP CMIP6) under two shared

socioeconomic pathways (SSPs). The two SSPs included a low greenhouse gas emissions scenario (SSP126) and a high greenhouse gas emissions scenario (SSP585)."

In the Methods section (Line 760–771):

"Two shared socioeconomic pathways (SSP) were considered: a low-emissions scenario (SSP126) and a high-emissions scenario (SSP585), for two 20-year periods (2061–2080 for SSP126 and 2081–2100 for SSP585). The SSPs represent combinations of shared socioeconomic pathways and representative concentration pathways (RCPs). SSP126 is the abbreviation for the SSP1-RCP2.6 scenario. SSP1 (sustainability, taking the green road) assumes a gradual shift toward a more sustainable world, with emphasis on human well-being and reduced inequality. RCP2.6 represents one mitigation scenario leading to a very low forcing level[22,23]. Similarly, SSP585 is the shortest form for the SSP5-RCP8.5 scenario. SSP5 (fossil fuel development, taking the highway) assumes rapid economic growth driven by fossil fuels, with high energy demand and limited efforts to mitigate greenhouse gas emissions. RCP8.5 represents a very high-baseline emission scenario[22,23]."

11. Lines 354-355 hypothesize that greater genetic offsets under SSP585 compared to SSP126 may indicate increased risks to bamboo populations due to higher CO_2 emissions. A more detailed explanation of this inference is required.

Response:

Thank you for the suggestion regarding the clarification of the underlying assumptions and implications of the SSP126 and SSP585 scenarios. To address this concern, we have made the following revisions:

1. **Expanded explanation of the SSP126 and SSP585 scenarios.** We have provided a more detailed description of the SSP126 and SSP585 scenarios to better contextualize their potential impacts on moso bamboo populations. The combination of socioeconomic trends and climate policy in SSP126 results in substantial reductions in total agricultural land, while simultaneously dedicating large areas to bioenergy production and increasing forest area. In

contrast, SSP585 exhibits very high levels of fossil fuel use, up to a doubling of global food demand and up to a tripling of greenhouse gas emissions over the course of the century, marking the upper end of the emission scenario literature[24,25]. Consequently, SSP585 is associated with greater CO2 emissions than SSP126, which in turn leads to more drastic climate change.

2. **Detailed interpretation of findings.** We clarified that the divergent emission pathways of SSP126 and SSP585 are likely to have differential impacts on the adaptive capacity and vulnerability of moso bamboo populations. Compared with the more sustainable scenario of SSP126, the more extreme climate change associated with SSP585 may expose moso bamboo populations to greater adaptive challenges and potential risks.

Additionally, to provide a more accurate and nuanced interpretation of our findings, we have revised the statement as follows:

Line 422–427:

"Consistent with the RONA results, all the genomic offsets were greater under the SSP585 scenario than under the SSP126 scenario and the regions with high genomic offsets (brighter area) were also larger, suggesting that the more extreme climate change associated with fossil fuel development (SSP585) may expose moso bamboo populations to greater adaptive challenges and potential risks than under the more sustainable scenario (SSP126)."

12. The Discussion section requires restructuring; it currently appears to be are capitulation of the results rather than a thoughtful discussion.

Response:

Thank you for this suggestion. The Discussion section has been fully rewritten to concentrate on three critical aspects, thereby offering a more thoughtful and focused analysis of our findings.

1. **Harnessing the potential of haplotype-resolved genomes and pangenomes.**

Haplotype-resolved genomes and pangenomes are powerful tools for understanding genomic diversity, but their full potential remains untapped due to limitations in current analysis environments.

2. **Unraveling the complexities of inter-haplotype and inter-accession variations.** Inter-haplotype variations in moso bamboo exceed inter-accession variations, likely due to ancestral haplotype differences and rare somatic mutations, while traditional methods may overestimate heterozygosity.
3. **Haplotypes: Key players in moso bamboo environmental adaptation.** Our findings provide guidance for conservation strategies, such as habitat restoration and assisted migration, to mitigate the impacts of climate change.

The complete Discussion section is presented below:

Line 452–514:

"The introduction and application of haplotype-resolved genomes, graph-based pangenomes, and genus-level pangenomes have greatly enriched our understanding of the genomic diversity of species or taxa, providing powerful tools for revealing the genetic basis of important traits[26-28]. In this context, our study on moso bamboo (*Phyllostachys edulis*), an economically and ecologically vital nontimber resource, has made significant progress by employing the third-generation PacBio HiFi and Hi-C sequencing technologies to obtain the haplotype genomes of 16 representative moso bamboo accessions and construct a comprehensive pangenome. These genomic resources not only more comprehensively capture the heterogeneity of the moso bamboo genome but also provide valuable genetic information for a deeper understanding of moso bamboo adaptability to diverse environmental conditions. However, despite the significant advantages of haplotype genomes and pangenomes over traditional collapsed genomes, there are still challenges in their practical applications. Mainstream omics analysis workflows, such as transcriptomics and epigenomics, still predominantly rely on aligning sequencing data to linear reference genomes, failing to fully utilize the rich diversity information contained within

haplotype genomes and pangenomes. Therefore, while continuously improving the accuracy and completeness of genomic data, it is imperative to enhance the analytical framework and application environment of these high-quality genomic resources, thereby ensuring their optimal utilization.

To fully utilize these genomic resources, we integrated the haplotype genomes and the pangenome, revealing novel insights into the genomic architecture of moso bamboo. Our study revealed that the differences between the two moso bamboo haplotypes exceeded the differences between the two moso bamboo accessions. Given the asexual reproduction of moso bamboo over extended periods and its 67-year flowering cycle[13,14], the primary source of variation is likely rare somatic mutations occurring within one haplotype. Asexual reproduction makes it difficult for variations accumulated in accessions to be transmitted, as the absence of meiosis prevents the exchange of genetic material between homologous chromosomes. We hypothesized that there would have been a difference between the two haplotypes in the common ancestor of moso bamboo populations in different regions and that the accumulation of somatic mutations in moso bamboo from different regions did not exceed the original difference between the two ancestral haplotypes. These factors have led to the phenomenon where quantitatively, inter-haplotype variations exceed the genetic variations among different accessions. Additionally, we discovered that heterozygosity might be overestimated in traditional variation detection methods. When considering the haplotype genome, we found that universally heterozygous sites are also heterozygous in the reference genome and should not be regarded as variation sites between accessions. Filtering out these variations between haplotypes leads to a decrease in the detected heterozygosity, while also suggesting that genetic diversity is lower than originally estimated.

Due to its low genetic diversity, in-depth studies on the adaptability of moso bamboo to diverse environmental conditions are extremely important. We observed the core-single gene sets and allele-specific expression (ASE) phenomena closely related to environmental adaptability and identified two sets of climate-related heterozygous variation sites, which may imply that haplotypes play a significant role in the

environmental adaptation of moso bamboo. Our study also showed that under the high-emissions scenario SSP585, the moso bamboo population faces significant adaptive risks (Fig. 5), highlighting the importance of emission reduction measures for alleviating the pressure of climate change. Particularly in the northwestern region, we recommend prioritizing habitat restoration where the risk is most severe (Fig. 5c) and considering assisted migration for the northern population (Figs. 5e–f) while addressing potential competition risks. It is noteworthy that our samples contain only moso bamboo from all the major natural distribution regions of moso bamboo in China, missing some of the human-transplanted populations or extreme populations. Supplementing these populations, and even global moso bamboo accessions, could enable the identification of more variations adapted to extreme environments. For risk predictions like RONA and local offsets that do not involve migration, the absence of these samples is less impactful. However, for forward and reverse offset analyses, incorporating additional populations could uncover regions more conducive for moso bamboo cultivation and identify moso bamboo populations better suited for migration to extreme regions. Nevertheless, the application of genomic offset in conservation planning is still in its infancy, and empirical validation of its predictions is necessary to assess its practical utility[29,30]. This can be achieved through carefully designed experiments, such as common garden trials or controlled environment tests, which compare genomic offset predictions with realized fitness outcomes in populations exposed to environmental change[30,31].”

13. Line 434 asks for clarification on what is referenced by Supplementary Table SX.

Response:

Thank you for pointing out this oversight. To ensure clarity and accuracy, we have revised the related description as follows:

Line 548–549:

“For RNA-seq, tissues from three biological replicates per RMA were collected,

including leaves, stems, roots, and rhizomes (Supplementary Table 5)."

"Supplementary Table 5. Summary statistics of RNA-seq data from 186 samples."

14. Line 474 has a missing citation that needs to be addressed.

Response:

Thank you for pointing out the missing citations. We apologize for any inconvenience caused by this oversight and thank the reviewer for helping us improve the manuscript's accuracy and completeness. To rectify this oversight, we have made the following revisions:

1. **Addition of the missing citation.** We have added the related citation to the revised manuscript (please find below).
2. **Thorough review of the manuscript for other missing citations.** We have conducted a thorough review of the entire revision to ensure that all statements are properly supported by appropriate citations and that no other instances of missing citations are present.

"Zhang, X. et al. Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis*. *Nat. Genet.* **53**, 1250–1259(2021)."

15. Lines 487-493 discuss the identification of allelic genes solely through reciprocal alignments, which could result in numerous genes that are not part of any allelic pair. Some may be single-allelic genes, while others might represent haplotype-specific duplicated genes. The methodology for identifying and categorizing these genes should be elaborated upon.

Response:

Thank you for the constructive suggestion. To address this concern, we further elaborated on our methodology for identifying and categorizing these genes:

1. **Identification of haplotype-specific duplicated genes.** We identified haplotype-specific duplicated genes as those that exceeded the threshold of 80% identity and were located within 40 kb of a gene on the other haplotype but were not the best reciprocal match to form an allele pair. These genes are provided in the supplementary materials (revised Supplementary Table 28).
2. **Categorization of duplicated genes.** We realized that these duplicated genes had similar and closely located genes in the other haplotype, making it inappropriate to directly classify them as single-allele gene sets in the pangenome. To address this issue, we propose classifying gene sets containing these duplicated genes as variable gene sets. This classification acknowledges difficulty in definitively determining whether these genes are present in only one haplotype or in both haplotypes, given their similarity and close proximity to their counterparts. By categorizing them as variable gene sets, we can better represent the ambiguity associated with their haplotype-specific presence in the pangenome analysis.
3. **Reanalysis of all pangenome sections.** Implementing these modifications resulted in a decrease in the number of single-allele gene sets from 51,819 to 48,305. In addition, we recalculated all pangenome section, including the transcripts per kilobase million (TPM), gene length, and tissue specificity index (Tau) for each gene set group. The corresponding figures were updated to reflect these changes.

The reanalysis of pangenome section in Fig. 2 is presented below:

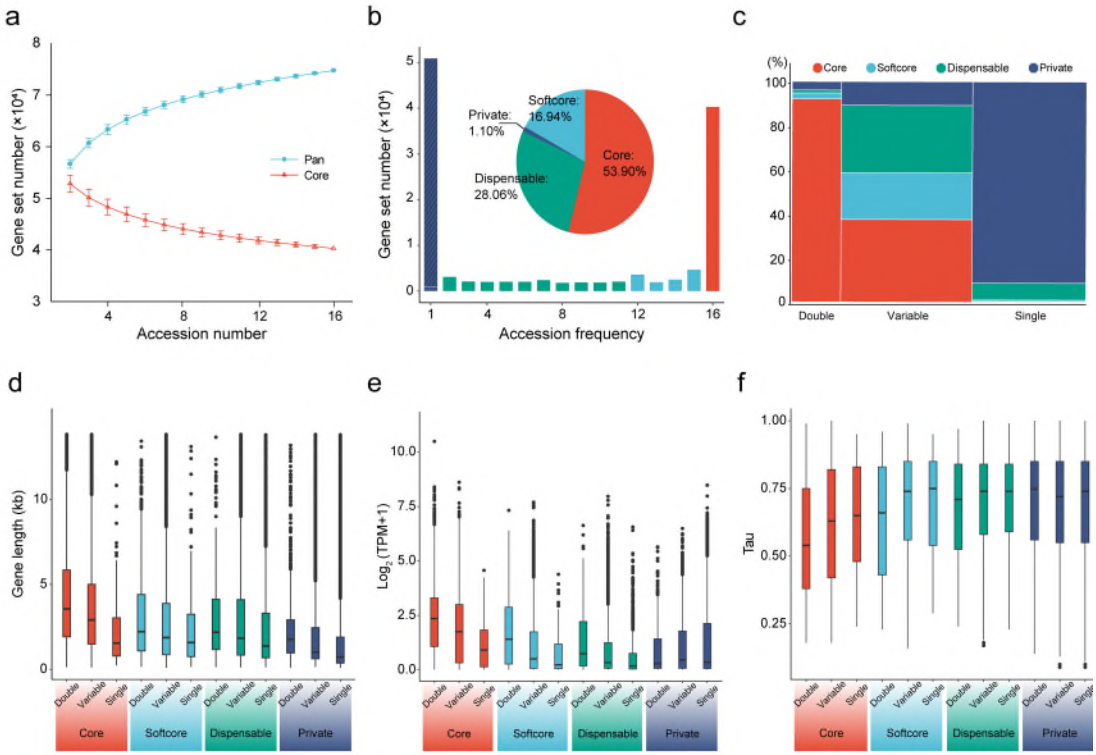


Fig. 2 Classification and characteristics of moso bamboo pangenome gene sets. a The number of gene sets in the pangenome (blue) and core genome (red) increased as a function of the number of moso bamboo accessions in the analysis (x-axis). **b** Compositions of the pangenome. The bar plots show the number of gene sets (y-axis) in each accession categorized by frequency (x-axis). The pie chart depicts the proportions of gene sets marked by each composition category: the core gene set (present in all accessions), softcore gene set (present in 12–15 accessions), dispensable gene set (present in 2–11 accessions), and private gene set (present in only one accession). The left block shows the number of unique gene sets (bottom) and the sum of unclustered genes in each genome (hatched area). **c** Distribution of gene sets in different groups based on gene frequency and alleles composition. The y-axis represents the four groups of moso bamboo divided based on gene frequency across accession (core, softcore, dispensable, and private), and the x-axis represents the 3 gene set groups categorized based on allele composition: double-allele sets (present in both haplotypes across accessions), single-allele sets (only detected in one haplotype across accessions), and variable-allele sets (present in paired haplotypes of some accessions). Thus, all the gene sets were divided into 12 groups (core-double, core-variable, core-

single, softcore-double, softcore-variable, softcore-single, dispensable-double, dispensable-variable, dispensable-single, private-double, private-variable, and private-single). The area of each group is proportional to the number of gene sets. **d, e, and f** Comparison of gene length (**d**), expression level (**e**), and tissue specificity index (Tau) (**f**) across the 12 gene set groups (x-axis). The y-axes show gene length in base pairs, $\log_2$(TPM+1) expression values, and the Tau specificity index, respectively. Statistical significance (*P*-value) is provided in Supplementary Tables 13–15.

Additionally, we have added a related description to the Methods section, as follows:

Line 618–620:

"Genes that met the similarity threshold for the other haplotype but were not among the best-aligned allele pairs were identified as haplotype-specific duplicated genes (Supplementary Table 28)."

Line 642–643:

"In addition, single-allele gene sets containing duplicated genes were reclassified as variable-allele gene sets."

16. Lines 519-526 detail allele-specific expression identification using HISAT2 with default parameters, which may not discriminate between highly identical allelic genes—thus unsuitable for calculating allele-specific expression accurately. A recommendation is made to select uniquely aligned RNA-seq reads to enhance this aspect of the study.

Response:

Thank you for the valuable suggestion. To address this concern, we have made the following revisions:

1. **Methodology revision:** We have revised our method by implementing the "-k 1" parameter in HISAT2, which identifies uniquely aligned RNA-seq reads based on the related reference[27]. This parameter ensures that only the best alignment is reported for each read, thereby improving the discrimination between highly identical allelic genes and enabling a more accurate calculation of allele-specific expression.
2. **Reanalysis of the ASE dataset:** Following this modification, we reanalyzed the entire ASE dataset, resulting in increases in the numbers of ASE genes (from 15,456 to 16,317) and gene sets (8,438 to 8,729). Despite these increases, the conclusions drawn from the updated analysis remain largely consistent with those of the previous analysis, further supporting the finding that ASEGs likely play key roles in environmental adaptation in moso bamboo.

Additionally, we have updated the relevant sections to reflect these changes as follows:

In the Methods section (Line 652–653):

"Reads were preprocessed with Trimmomatic v0.39 (default parameters) and aligned to the corresponding haplotype assemblies using HISAT2 v2.1.0 (-k 1)."

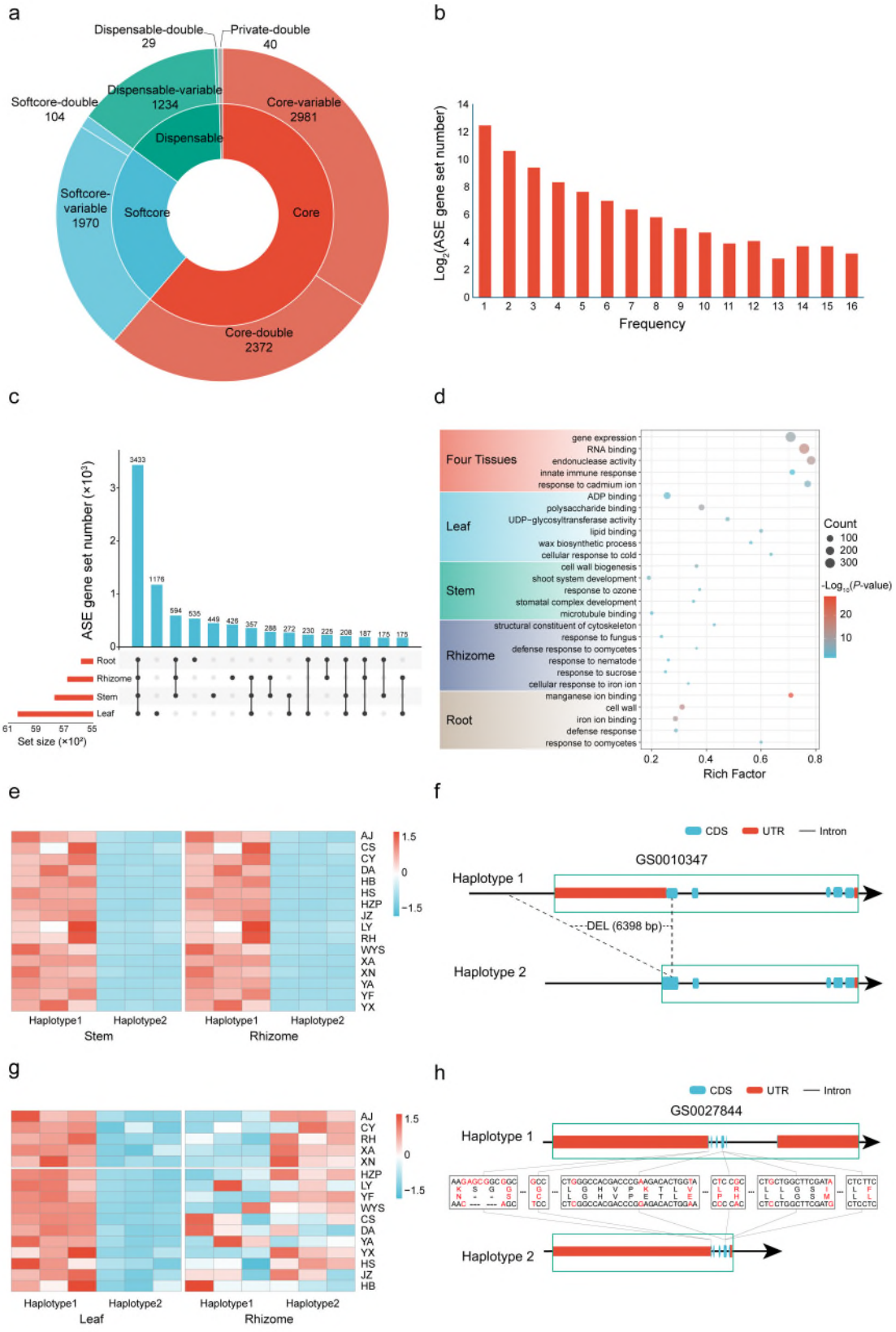


Fig. 3 Allele-specific expression related to environmental adaptation in moso bamboo. a Distribution of allele-specific expression gene sets (ASEGs) across the

moso bamboo pangenome categorized by accession (inner circle) and allele (outer circle). **b** Frequency distribution of ASEGs (y-axis) across all accessions (x-axis). **c** Tissue-specific distribution of ASEGs detected in all four tissues (leaf, stem, rhizome, and root) in the moso bamboo pangenome. **d** Gene Ontology (GO) enrichment analysis of ASEGs in all four tissues and exclusively in one tissue. The circle color represents the statistical significance (*P*-value), and the size represents the number of ASEGs. The rich factor is the ratio of ASEGs annotated to a given GO term over the total number of genes in that term. **e** An example of a consistent ASEG (the gene set GS0010347) exhibiting ASE patterns. The ASEG show high expression (red) in haplotype 1 and low expression (blue) in haplotype 2 in both the stem and rhizome. **f** Schematic representation of a 6,398-bp SV in GS0010347 that may induce ASE by altering protein sequences between haplotypes. The gene regions are green, the CDSs are blue, and the 5'/3' UTRs are red. **g** An example of an inconsistent ASEG (the gene set GS0027844) exhibiting tissue-specific patterns. For the five accessions (AJ, CY, RH, XA, and XN), the ASEGs showed higher expression in haplotype 1 compared to haplotype 2 in the leaf. Conversely, in the rhizome, the same ASEGs exhibited lower expression in haplotype 1 and higher expression in haplotype 2 relative to the leaf pattern. **h** Schematic representation of a DEL and several base substitutions were present between the CDSs of the two haplotypes in the gene set GS0027844. These sequence differences (red) changed the protein sequence between the two haplotypes.

Responses to the comments of Reviewer #3

This is an excellent MS with an impressive volume of data and results. I just have some suggestion in order to improve the understanding of the paper for the reader.

There is an understandable general trend in the MS towards the pangenome associated information and methods and an underplaying of the populational data. To make the populational results more understandable the MS need some improvements.

Response:

Thank you for your positive feedback on and constructive suggestions for our manuscript. We are pleased that you find the volume of data and results impressive. Your acknowledgment of the quality and significance of our work is highly encouraging. To address your concerns and make the population results more understandable, we have made the following improvements:

1. **Supplementing missing information.** We have supplemented the manuscript with previously missing figures and descriptions that are essential for supporting and illustrating our findings. These additions include information pertaining to the 427 resequenced accessions and the results from population analyses.
2. **Reanalyzing the ADMIXTURE and genotype–environmental associations.** We reanalyzed the ADMIXTURE analysis and found $K = 1$, which was consistent with a previous study. We also reperformed genotype–environment association (GEA) analysis and recalculated the genomic offset.
3. **Refining the narrative and interpretation.** We have rewritten the last parts of the Introduction and Discussion sections. The revised Introduction focuses on the research aims and questions. The revised Discussion section has undergone extensive revisions to provide a more in-depth and nuanced interpretation of our findings.
4. **Enhancement and optimization of language and expression.** In collaboration with American Journal Experts (AJE), a subsidiary of Springer Nature, we have meticulously refined the language used throughout the

revised manuscript and response letter, and an AJE editing certificate is provided below. Ambiguous or unsupported statements have been removed or clarified, and the narrative has been focused more sharply on key findings and their scientific implications. These language improvements have enhanced the clarity, precision, and scholarly rigor of the manuscript.

Response Fig. 1 AJE editing certificate.

1.sample information

You mention that you have 16 representatives moso bamboo accessions with additional 427 resequenced accessions. However in methods section, sample collection, the information about this 427 samples is missing. Where they come from? How many from each region? Supplementary table 4 only have sequence information and only for 186 RNAseq samples.

Response:

Thank you for this suggestion. We apologize for the missing information regarding the 427 resequenced accessions in the Methods section. To address this issue and improve the clarity of our work, we have made the following modifications:

1. **Clarifying the sources of the 427 resequenced accessions.** We obtained 427 resequenced accessions from our previous study[11], which extensively investigated natural moso bamboo populations across China. In that study, we identified major natural distribution regions of moso bamboo and conducted a comprehensive phylogenetic analysis to elucidate the evolutionary relationships among these populations. This analysis has already been reported in previous work.
2. **Utilizing prior research findings.** Building upon our team's prior research, we further utilized the 427 resequenced accessions generated in the previous project. By integrating these data with the newly constructed moso bamboo pangenome, we conducted a comprehensive association analysis to identify climate–related variations and gain insights into the environmental adaptability of moso bamboo.
3. **Supplemental method details.** Supplementary Table 27 provides detailed information about the 427 resequenced accessions, including their geographic position and the number of accessions from each region.

Additionally, we have added the following information to the sample collection section in the Methods section:

Line 525–528:

"Additionally, we used genetic information from 427 resequenced accessions obtained in our previous study[11] (Supplementary Table 27), which covered all the main natural distribution regions of moso bamboo in China, to enhance the genetic representation of our sampling in this study."

"Supplementary Table 27. List of the 427 resequenced accessions from a previous study."

2. In the methods section "Genome and transcriptome sequencing" there aren't any information about the 427 resequenced samples.

Response:

Thank you for the suggestion. To address this issue, we have added the following details to the related description in the Methods section:

Line 525–528:

"Additionally, we used genetic information from 427 resequenced accessions obtained in our previous study[11] (Supplementary Table 27), which covered all the main natural distribution regions of moso bamboo in China, to enhance the genetic representation of our sampling in this study."

3-1. In the methods section "Identification of climate-associated variants" is not clear how many samples you are using. The 16 you used for the pangenome or the 427 that you didn't explain how you sequenced and where you get it from?

Response:

Thank you for your suggestion. We have added the sample information to the Methods section as follows:

Line 724–727:

"The SVs identified from the graph-based pangenome were constructed using the 16 representative moso bamboo accessions, and the SNPs and InDels detected from the previously generated 427 moso bamboo resequenced accessions were used for the identification of climate-associated variations."

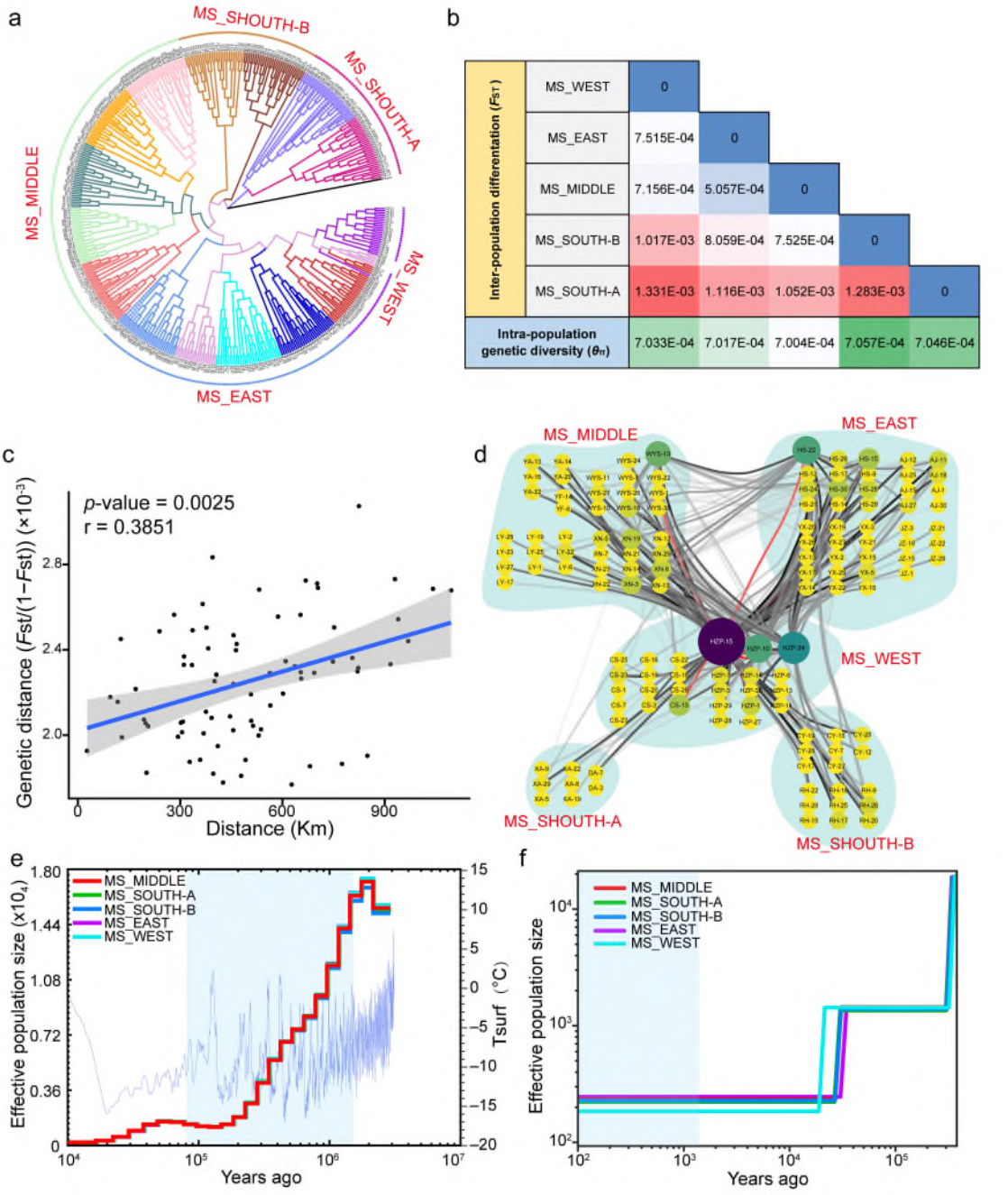
3-2. The admixture plot and associated differentiation statistics is important for the

reader to understand the pattern of distribution of the variation inside the species.

Response:

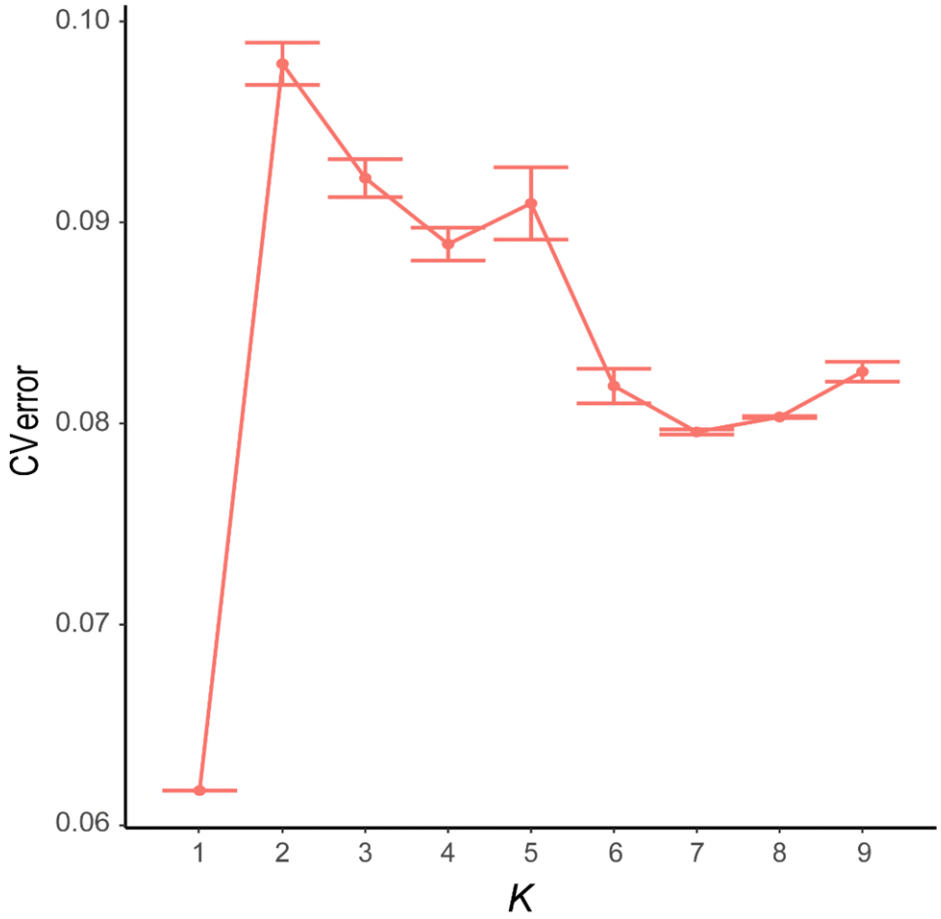
Thank you for this suggestion. We systematically performed and reported a comprehensive population analysis of these 427 accessions in our previous study (Zhao et al., Nat. Commun. 2021). To address this, we have made the following revisions:

1. **Summary of previous population analysis findings**
    - **ADMIXTURE analysis results. ADMIXTURE** analysis with multiple random seeds revealed that the optimal model had a *K* value of 1, indicating rare and slight population differentiation. Based on the phylogenetic tree and geographic separation distances, the 427 moso bamboo accessions were divided into five phylogenetic groups.
    - **Population differentiation and isolation by distance.** Wright's *F* statistics (F_{ST}) reflected relatively low population differentiation. An isolation-by-distance (IBD) analysis demonstrated a significant positive correlation between geographic distance and genetic distance for the moso bamboo population.
    - **Demographic history and potential factors.** The inferred ancient population bottleneck and recent small population size without a rebound are possibly related to climate change and human activities.
2. **Rechecking the *K* value.** Given the thorough analysis and reporting of these results, we did not repeat them in the previous version of the manuscript. However, because the reference genome update and gene–environment association (GEA) analyses require *K* values, we reanalyzed the admixture analysis to confirm the accuracy of the population structure. The results consistently showed a *K* value of 1 (Supplementary Fig. 13), which was consistent with our previous findings.



Response Fig. 2 Overall population structure landscape and the inferred population demographic history (Zhao, H. et al. Nat. Commun. 2021). a Rooted neighbor-joining phylogenetic tree of 427 moso bamboo individuals. The different line colors represent the fifteen geographical geographic areas, and the differently colored dotted lines nearby represent five groups that were empirically assigned in our study. b The genetic diversity ($\theta\pi$) and F_{ST} matrix of the five groups. The colors and numbers in the cells of the matrix represent the F_{ST} values. The numbers in the cells below the F_{ST} matrix represent the genetic diversity ($\theta\pi$). **c** Results of the Mantel test of the

relationship between geographical distance and genetic distance with MS_WEST excluded. The region in gray represents the 95% confidence intervals. **d** The connection of individuals with the lowest 1% pairwise genetic distances. The size and color of circles represent the degree of connectivity to a node. The lines in different colors indicated values of Hamming distance (genetic distance), with red indicating the shortest distance and for the others, darker colors indicate shorter distances and lighter colors indicate longer distances. **e** The demographic history of the fifteen geographic areas was inferred separately using the pairwise sequential Markovian coalescent (PSMC) method. The light blue line represents the historical surface temperatures, and the light blue shade indicates the bottleneck experienced during the last glacial period (LGP, 115,000-11,700 years ago). **f** The demographic history was inferred using SMC++. The LGP was shaded in light blue, and the reduction without a rebound in the effective population size during the last 2,000 years is shaded in light green. The results were scaled to real-time by assuming a generation time of 67 years and a mutation rate of 8.51×10^{-8} per generation.

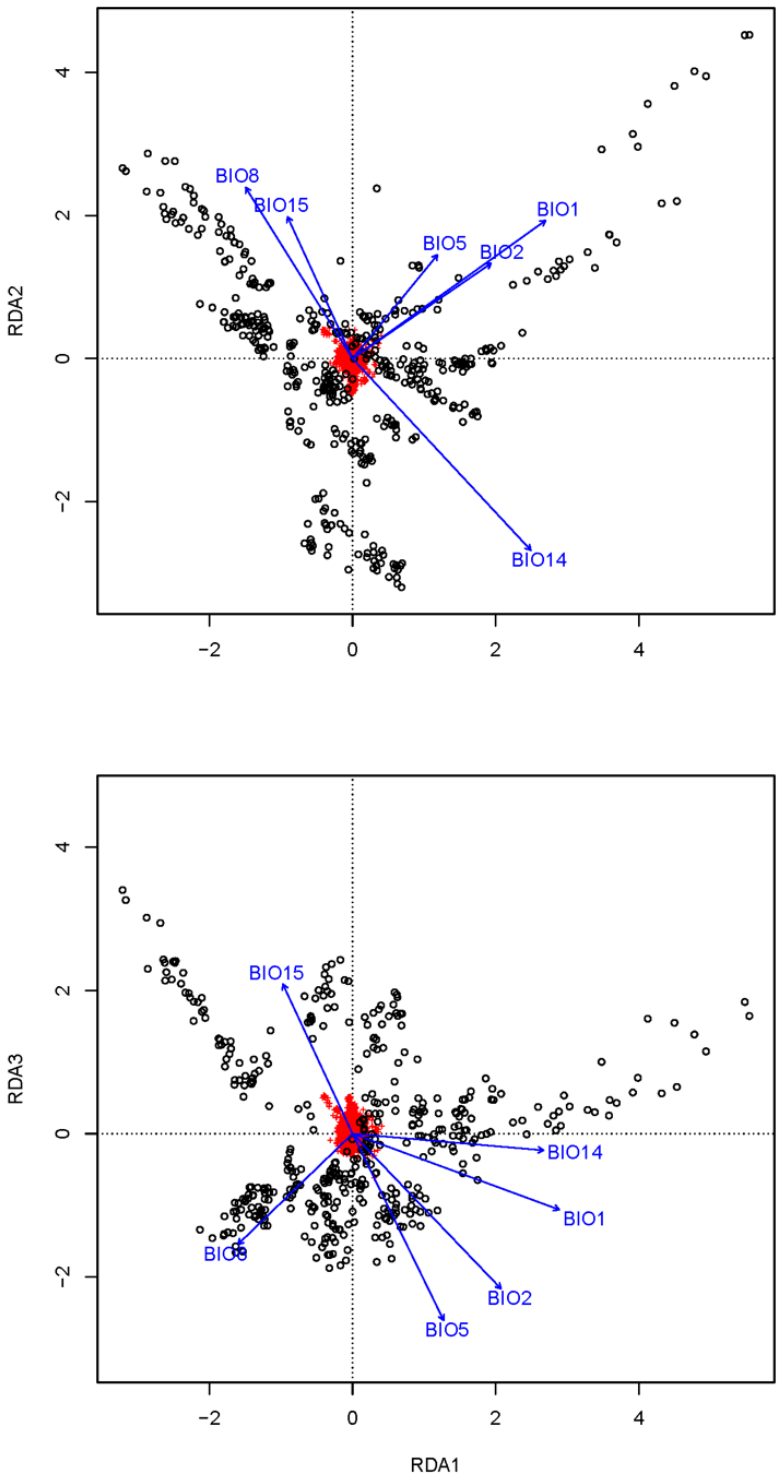


Supplementary Fig. 13 Cross-validation (CV) error values for different values of *K* in an ADMIXTURE analysis.

3-3. The same for RDA results, the figure and associated information is missing.

Response:

Thank you for this suggestion. To address this oversight, we have added the RDA figure in the revised manuscript as Supplementary Fig. 16 and provided the detailed results of the RDA in Supplementary Table 24, as follows:



Supplementary Fig. 16 Redundancy analysis (RDA) plot. The variations are represented by red dots (in the center of each plot), and each moso bamboo accession is depicted as black circles. The blue vectors represent the environmental variables. The upper represent the axes 1 and 2, and the lower represent axes 1 and 3.

“Supplementary Table 24. RDA information of candidate variations.”

3-4. How much variation is explained by the 8 climatic variables?

Response:

Thank you for this suggestion. We have added the related information to the Results section as follows:

Line 344–347:

"RDA revealed that the contribution of climate effects explained 35% of the genetic variation in the adaptive variations, which was substantially greater than that of geography (Supplementary Table 26)."

Supplementary Table 26. RDA models of inertia proportions of variations.

RDA models	Candidate Inertia	Proportion of candidate variations	Total Inertia	Proportion of total variations
F~ geno	136.71	0.13	1.31E+04	4.94E-03
F~ env	372.34	0.35	3.81E+04	1.44E-02
F~ geno|env	52.78	0.05	1.24E+04	4.68E-03
F~ env|geno	288.42	0.27	3.74E+04	1.41E-02
F~ env+geno	425.13	0.40	5.05E+04	1.90E-02
Total inertia	1,050.00	1.00	2.65E+06	1.00E+00

4. line 587 "Calculation of genomic drift". Do you really want to say this?

A more detailed explanation about the genomic offset and the local, forward and reversed offset is needed, otherwise Figure 5e and f are very difficult to understand.

Response:

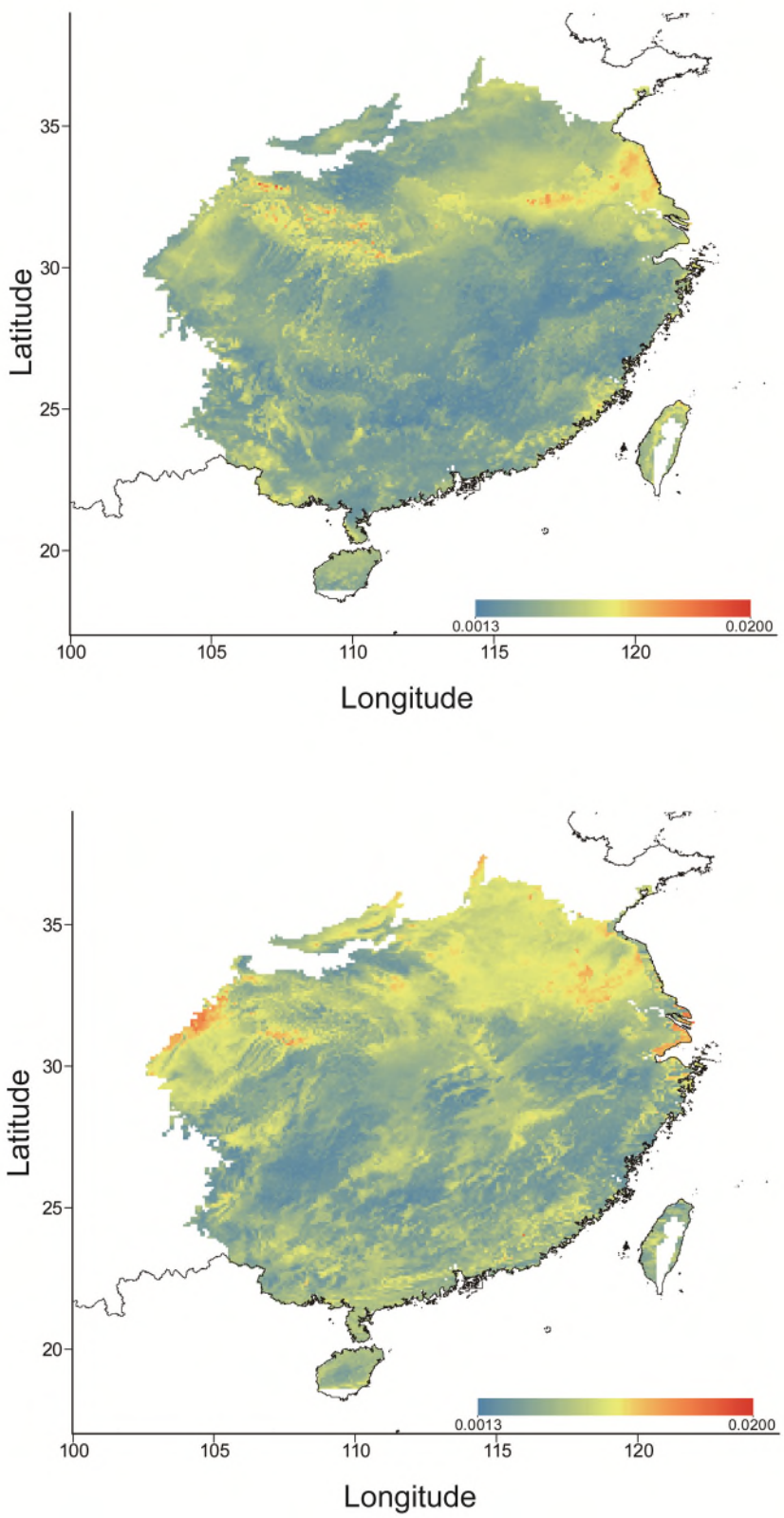
Thank you for pointing out this issue. We apologize for the confusion caused by the term "genomic drift", which should be corrected to "genomic offset". To address this concern, we have made the following modifications:

1. **Explaining the definition of genomic offset.** To improve the understanding of the genomic offset definition and its calculation, we have added an explanation to the Methods section, based on a previous reference[32].
2. **Interpreting and refining visualization of genomic offset results.** To

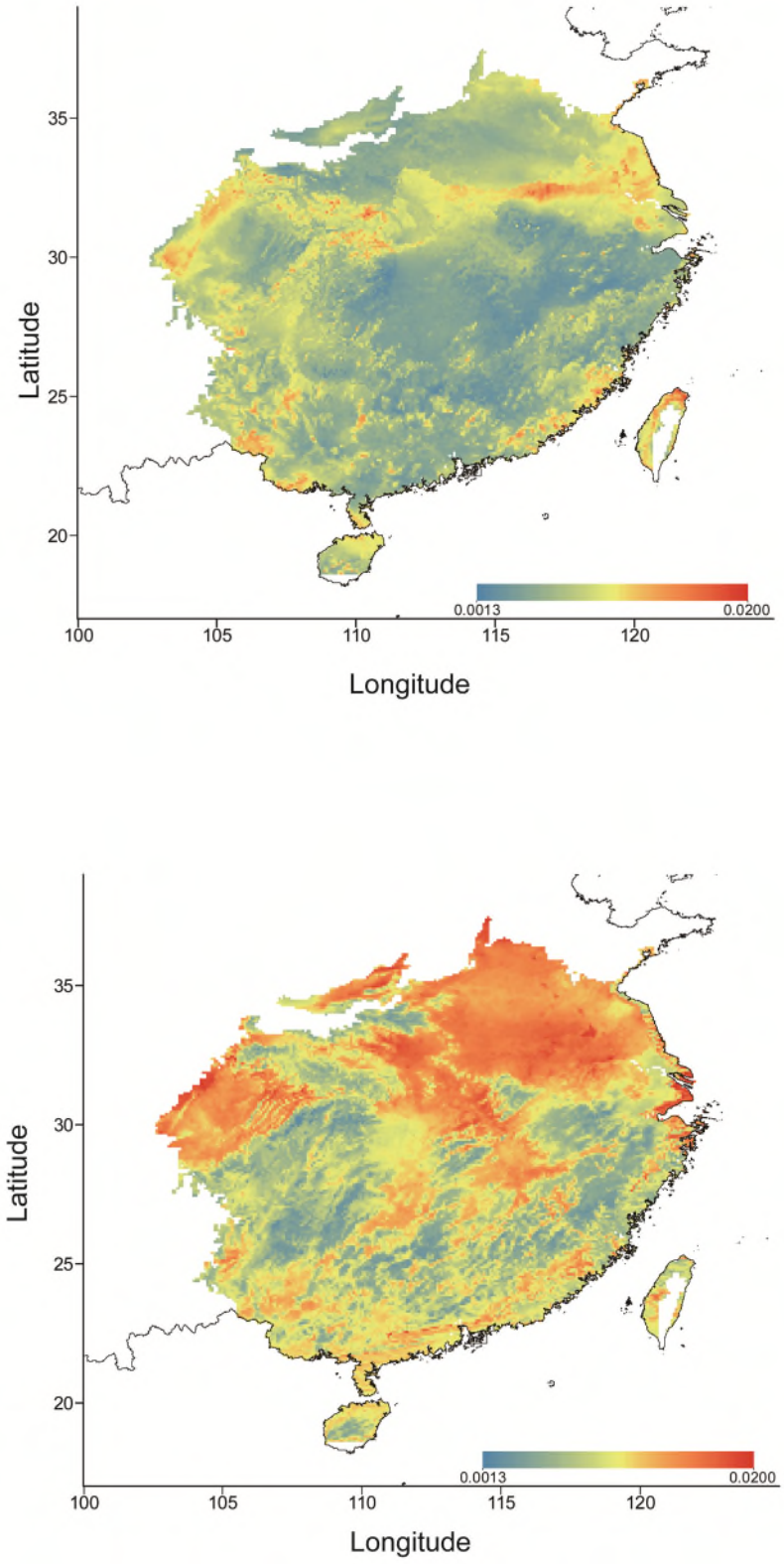
further aid in the understanding and interpretation of the genomic offset results, we have included individual genomic offset maps in the revised Supplementary Figs. 22–23 as follows:

In the Methods section (Line 776–786):

"Local offset, a measure of the vulnerability of a resident population to climate change, was calculated by estimating the predicted change in allele frequencies at climate-adaptive loci that was necessary for the population to adapt to local climate changes over time. In contrast, forward offset assumed that populations had unlimited migration ability. It was calculated by identifying the minimum predicted offset if propagules or alleles could move, through gene flow, to any suitable habitat within the range. Reverse offset represented the possibility that any population in the current range would be preadapted to a particular location in the future. Reverse offset was calculated by identifying the minimum offset between hypothetical populations within the current range in the future climate and populations in the current climate."



Supplementary Fig. 22 Map of the predicted forward genomic offset and reverse genomic offset averaged across four climate models across the distribution of moso bamboo under SSP126. The upper shows forward genomic offset, and the lower shows the reverse genomic offset.



Supplementary Fig. 23 Map of the predicted forward genomic offset and reverse genomic offset averaged across four climate models across the distribution of moso bamboo under SSP585. The upper shows forward genomic offset, and the lower shows the reverse genomic offset.

5. Is the Hi-C and RNAseq come from the same samples used for the pangenomes? The 186 samples used for RNAseq are only mentioned in the results and not on the methods. The origin of this samples have to be more clear in the methods section.

Response:

Thank you for your suggestion. To address this concern, we have revised the Methods section to provide a more detailed description of the sample collection process, as follows:

1. **Consistent sample source for genomic and transcriptomic data.** In our study, all samples used for HiFi sequencing, Hi-C analysis, and RNA-seq were collected from the same bamboo rhizome at each location. To ensure higher DNA quality, the HiFi and Hi-C samples were obtained from the same bamboo shoot.
2. **Supplementing information on RNA-seq samples.** We collected 186 RNA-seq samples from 3 or 4 tissues (young leaves, stems, roots, and rhizomes) of the same bamboo rhizome used for DNA sampling at each location. These RNA-seq data were used for annotation and allele-specific expression (ASE) analysis.

Additionally, we have rewritten the related descriptions in the revised Methods section as follows:

Line 528–531:

"In each region, moso bamboo shoots were collected for DNA extraction in April 2020. Concurrently, young leaves, stems, roots, and rhizomes were collected from the same moso bamboo rhizome as the RNA-Seq samples at each region."

6. In the end of the introduction you already talk about you results!!. In the introduction, the explanation of the genomic offset and related measures as well as its potential and

limitations are missing.

Response:

Thank you for this suggestion. To address these concerns, we have made the following changes:

1. **Focusing the Introduction on research aims and questions.** We have rewritten the final paragraph of the introduction to focus on the aims and key questions of our study, omitting any results, as follows:

Line 86–102:

"To address the pressing need for elucidating the genetic basis underlying the wide distribution and climate adaptation of moso bamboo, we aim to construct a haplotype-based pangenome for moso bamboo using PacBio HiFi and Hi-C sequencing strategies. By integrating comprehensive genomic datasets from 16 representative moso bamboo accessions (RMAs), we will characterize genome-wide genetic variations and allele-specific expression (ASE) at an unprecedented resolution. Furthermore, by harnessing the graph-based pangenome and high-resolution spatiotemporal climate data, we will identify genetic loci associated with local climate adaptation and quantify climate maladaptation risks across moso bamboo populations in China. Our research sought to address the following key questions: 1) What is the extent and pattern of haplotype-level genomic diversity of moso bamboo? 2) How does ASE contribute to the adaptive resilience of moso bamboo? 3) Which genomic variations underlie local climate adaptation in the moso bamboo populations? 4) How will projected climate change scenarios impact the climate maladaptation risks of moso bamboo populations? Addressing these questions will provide critical insights to inform evidence-based conservation and breeding strategies for safeguarding this ecologically and economically vital species in the face of rapid global climate change."

2. **Introducing the definition of genomic offset and its potential.** We have added

the definition and potential of genomic offset to the Introduction section as follows:

Line 59–66:

"Genomic offset represents the disruption of current genotype–climate relationships due to rapid shifts in climate[33].Genomic offset represents the disruption of current genotype–climate relationships due to rapid shifts in climate[34]. The advantage of genomic offset approaches lies in their ability to predict population responses and vulnerability to climate change from genomic data, serving as an alternative to common garden experiments[29,30]. Forward and reverse genomic offsets can also predict population maladaptation under migration scenarios[32]."

3. **Discussing the limitations of genomic offset.** We have provided the limitation of genomic offset in the Discussion section, as follows:

Line 510–514:

"Nevertheless, the application of genomic offset in conservation planning is still in its infancy, and empirical validation of its predictions is necessary to assess its practical utility[29,30]. This can be achieved through carefully designed experiments, such as common garden trials or controlled environment tests, which compare genomic offset predictions with realized fitness outcomes in populations exposed to environmental change[30,31]."

7. I missed the results of some analysis that are the base for the assembly evaluation, such as the K-mer analysis with Mercury.

Response:

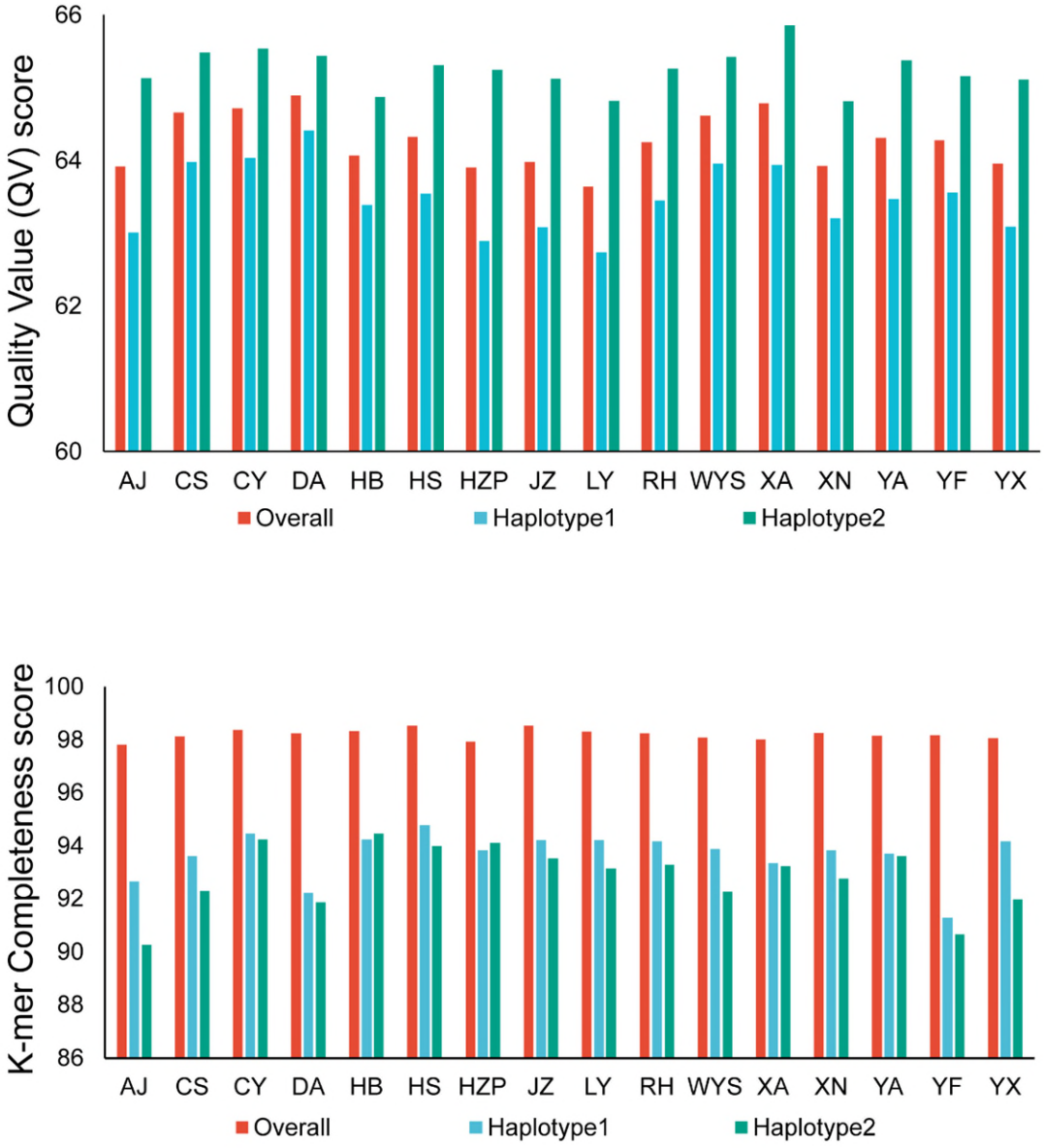
Thank you for the suggestion. To address this concern, we supplemented the k-mer analysis, including quality value (QV) and k-mer completeness using Merqury, and modified the Results and Methods sections accordingly.

In the Results section (Line 112–114):

“The average quality value (QV) of the final assembly was 64.26, with a k-mer completeness of 98.20% (Supplementary Fig. 3 and Supplementary Table 2).”

In the Methods section (Line 563–566):

“Meryl databases were generated with Meryl v1.41[35] (k = 20) for the raw sequencing reads and the assemblies. The quality value (QV) and k-mer completeness were then calculated using Merqury v1.3[35] (default parameters) by comparing the k-mer spectra of the assembly and the raw reads.”



Supplementary Fig. 3 K-mer assessment of the genome. The top panel displays the average Quality Value (QV), and the bottom panel illustrates the k-mer completeness.

8-1. Is not clear to me the rational of the distinction between the inter-haplotype and inter-accession measures, and why you based all your results in these differences?

Response:

Thank you for the suggestion. The distinction between inter-haplotype and inter-accession variations was crucial in our study because it allowed us to capture the true

extent of genetic diversity within and among moso bamboo accessions. By considering variations at both the haplotype and accession levels, we gained a more comprehensive understanding of the genomic heterogeneity present in this species. To address this concern, we elucidated four key aspects as follows:

1. **Quantifying genomic diversity: dissecting inter-haplotype and inter-accession variations.** Inter-haplotype variations refer to the differences between two haplotypes within a single accession, while inter-accession variations represent the divergence between different accessions. Inter-haplotype variations were calculated by aligning the two haplotypes of each accession and identifying the variations between them, while inter-haplotype variations were calculated by aligning the genomes of different accessions and identifying the variations between them. Our analysis revealed that the numbers of inter-haplotype short variations (SNPs and InDels) and structural variations (SVs) were, on average, 10.4 times and 5.3 times greater, respectively, than those of inter-accession short variations and SVs.
2. **The evolutionary significance of haplotype-level variations in moso bamboo.** Our analyses revealed that inter-haplotype variation was substantially greater than inter-accession variation, suggesting that the primary source of genetic diversity in moso bamboo is the divergence between haplotypes within a single accession rather than the divergence between different accessions. This finding has significant implications for our understanding of the evolutionary history and reproductive biology of moso bamboo. As a species that primarily reproduces asexually, moso bamboo accumulates genetic variations through rare somatic mutations within haplotypes rather than through meiotic recombination. The long generation times of moso bamboo contribute to the maintenance of these haplotype-level variations over extended periods, leading to the observed pattern of higher inter-haplotype diversity than inter-accession diversity.
3. **All analyses were based on haplotype-based genomes rather than the two variations.** The analyses in this study were all conducted using haplotype

assemblies. However, importantly, not all results were derived from inter-haplotype and inter-accession variations. These two types of genomic variations were first identified and characterized. Subsequently, pangenome and ASE analyses primarily utilized genome data and gene expression profiles, respectively. The GEA analysis integrated the graph-based pangenome and the previously published resequencing dataset of 427 accessions. The large sample size of the resequencing data provided robust statistical power to support the GEA analysis.

4. **Using haplotype-based data for comprehensive genomic analyses.** By accurately quantifying and distinguishing between these two types of variation, we were able to base our subsequent analyses on a more realistic representation of the genetic diversity present in moso bamboo. This approach was essential for the following aspects of our study:
    - **Pangenome construction:** The graph-based pangenome was built by integrating haplotype-based genomes, allowing us to capture the full spectrum of genetic variations, including those present at the haplotype level.
    - **Allele-specific expression analysis:** Accurate haplotype-based genomes are crucial for studying allele-specific expression patterns, which can be influenced by haplotype-specific variations.
    - **Genotype–environment association analysis:** By leveraging the graph-based pangenome and the 427 resequenced accessions, we identified climate-associated variations while accounting for the complex haplotype structure of moso bamboo.
    - **Population-level adaptations:** Distinguishing inter-haplotype and inter-accession variations enabled us to uncover the interplay between haplotype-level variations, individual-level differences, and population-level adaptations to environmental factors.

In summary, haplotype analyses were fundamental to our study because they allowed us to accurately capture the true genetic diversity of moso bamboo and its underlying evolutionary processes. By basing our analyses on these measures, we were able to gain insights into the genomic architecture, expression patterns, environmental adaptations, and population dynamics of this ecologically and economically important species.

8-2. Line 128 – "Thus, these variants were in fact present between the haplotypes (inter-haplotype) of the two accessions simultaneously, and not between the accessions (inter-accession)". Line 159 – "Inter-accession variations are absent between the reference haplotypes and present in other accessions. Inter-haplotype variations are present between the reference haplotypes". Thus that mean that when you compare with the reference genome you have more differences than when you compare each of two different accessions?

Response:

Yes. Our findings indeed suggest that when comparing the two haplotypes of the reference genome, we observe more differences than when comparing haplotypes from two different accessions. Specifically, our analysis revealed that the numbers of inter-haplotype short variations (SNPs and InDels) and structural variations (SVs) were, on average, 10.4 times and 5.3 times greater, respectively, than those of inter-accession short variations and SVs (Fig. 1b). These results indicate that the genetic differences between the two haplotypes within a single accession exceed the differences observed between distinct accessions. This finding underscores the importance of considering haplotype-level variations when studying the genetic diversity and evolutionary patterns of moso bamboo. By capturing and analyzing inter-haplotype variations, we can gain a more comprehensive understanding of the genomic heterogeneity within this species.

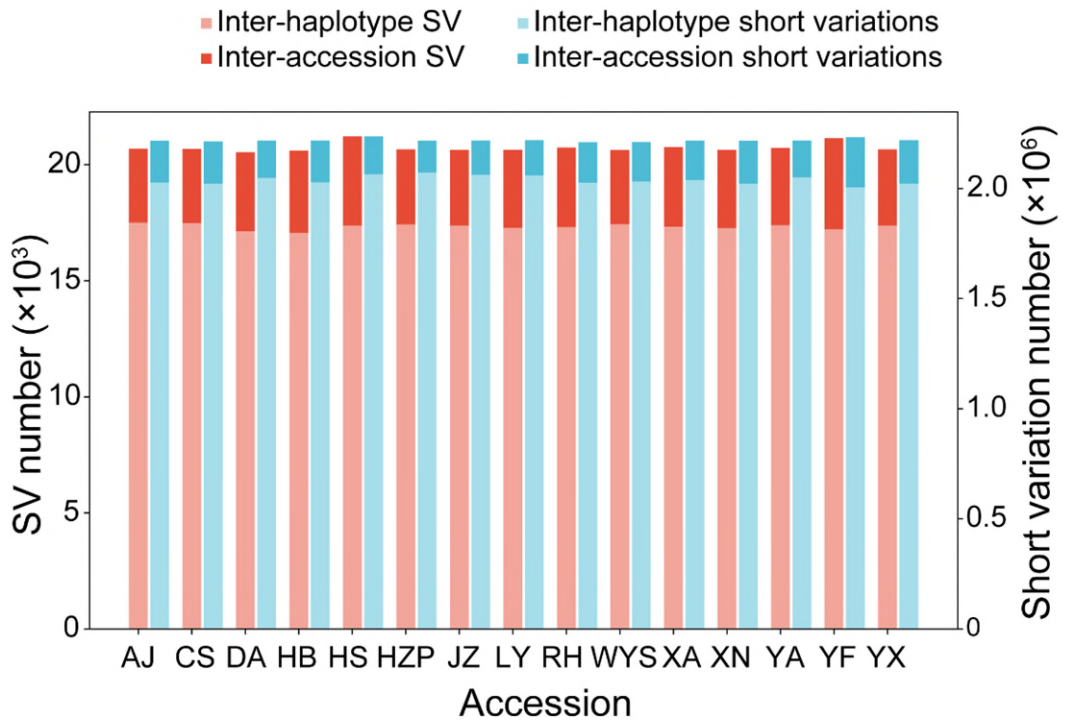


Fig. 1b Numbers of SVs (red) and short variations (SNPs and InDels, blue) categorized as either inter-accession (darker colors) or inter-haplotype (lighter colors). Inter-accession variations are absent between the reference haplotypes but present in other accessions. Inter-haplotype variations are present between the reference haplotypes. The x-axis represents accessions, and the y-axis shows the number of SVs/short variations.

8-3. So all you inter-accessions results from multiple pairwise differences across accession and so you should have a distribution of values, one for each comparison?

Response:

Thank you for this suggestion. According to your advice, we compared the variations in each accession with the variations between the haplotypes of the reference genome. This approach allowed us to assess the relative magnitude of inter-accession differences in the context of the haplotype-level diversity present in the reference genome. The values of these comparisons have been added to the revised Supplementary Table 30, which provides a comprehensive summary of the pairwise differences observed across accessions.

Supplementary Table 30. Comparison of variations between each accession and the reference genome haplotypes.

Accession	Short variation			SV		
	Same as reference (inter-haplotype)	Only in query (inter-accession)	Only in Reference (inter-accession)	Same as reference (inter-haplotype)	Only in query (inter-accession)	Only in Reference (inter-accession)
AJ	2,027,150	83,147	106,297	17,509	1,656	1,528
CS	2,021,790	78,516	113,659	17,486	1,634	1,551
DA	2,048,481	86,352	82,940	17,147	1,505	1,890
HB	2,028,758	83,939	105,328	17,071	1,576	1,966
HS	2,065,266	108,353	62,670	17,383	2,183	1,654
HZP	2,072,642	82,486	61,725	17,431	1,626	1,606
JZ	2,062,377	83,934	71,672	17,385	1,605	1,652
LY	2,060,083	85,260	73,755	17,282	1,609	1,755
RH	2,026,487	76,197	107,457	17,319	1,700	1,718
WYS	2,032,146	76,796	101,897	17,442	1,589	1,595
XA	2,039,072	83,367	94,910	17,337	1,722	1,700
XN	2,021,787	85,512	109,003	17,273	1,600	1,764
YA	2,050,853	83,801	83,331	17,402	1,683	1,635
YF	2,004,919	106,908	120,330	17,223	2,110	1,814
YX	2,022,490	81,956	112,175	17,386	1,622	1,651

8-4. I not sure If I understand properly the relevance of you analysis. But for me the relevant biological information come from comparing your set of genomes that correspond to the groups you identified with admixture. Accession of each K=3 group should have genomes more identical that when you compare among the 3 groups.

Response:

Thank you for this suggestion. To address these concerns, we have made the following modifications:

1. **Reanalysis of the population structure revealed a single population.** We reanalyzed the population structure using SNPs after removing linkage disequilibrium (LD). The updated analysis revealed a single population (K = 1), which is consistent with the conclusions of our previous study. We have added a description of this analysis to the Methods section.

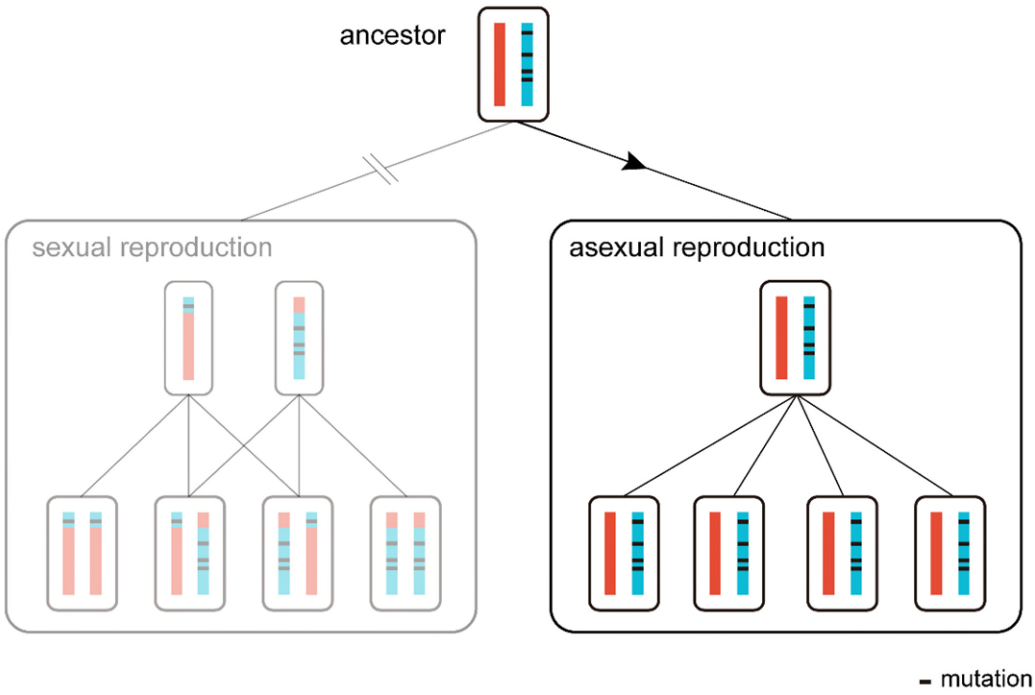
2. **Hypothesis explaining the predominance of inter-haplotype variations.** Our analysis demonstrated that, quantitatively, the variations between accessions did not exceed the variations between haplotypes within a single accession, and these inter-haplotype variations likely originated from the common ancestor of these samples. We have proposed the following hypothesis in the Discussion section to explain this observation.

In the Methods section (Line 717–721):

"To determine a pruned SNP set, we used PLINK v1.9[36] (-indep-pairphase 100 10 0.2). The resulting SNPs were then used to assess population structure using ADMIXTURE v1.3.0[37] (-cv -j4) for multiple repeats with different random seeds. The population structure showed $K = 1$ (Supplementary Fig 13), which was consistent with our previous findings[11]."

In the Discussion section (Line 474–484):

"Given the asexual reproduction of moso bamboo over extended periods and its 67-year flowering cycle[13,14], the primary source of variation is likely rare somatic mutations occurring within one haplotype. Asexual reproduction makes it difficult for variations accumulated in accessions to be transmitted, as the absence of meiosis prevents the exchange of genetic material between homologous chromosomes. We hypothesized that there would have been a difference between the two haplotypes in the common ancestor of moso bamboo populations in different regions and that the accumulation of somatic mutations in moso bamboo from different regions did not exceed the original difference between the two ancestral haplotypes. These factors have led to the phenomenon where quantitatively, inter-haplotype variations exceed the genetic variations among different accessions."



Supplementary Fig. 26 Schematic diagrams illustrating sexual reproduction and asexual reproduction in moso bamboo. Although moso bamboo possesses the ability for both sexual (left) and asexual (right) reproduction, it predominantly relies on asexual reproduction through rhizome growth and vegetative propagation. This hypothesis, grounded in the reproductive biology of moso bamboo, provides a plausible explanation for the observed pattern of genetic variation. By considering the rarity of somatic mutations, the lack of meiotic recombination due to the predominance of asexual reproduction, and the potential for haplotype divergence in the ancestral population, we provided a schematic diagram for understanding the predominance of inter-haplotype variations in moso bamboo.

9. line 133 - “suggesting that traditional methods of variant identification overestimate heterozygosity in moso bamboo”. It deserve more elaboration in the discussion section.

Response:

Thank you for the suggestion. We have removed the previous description and incorporated a revised elaboration in the Discussion section, as follows:

Line 485–490:

“Additionally, we discovered that heterozygosity might be overestimated in traditional variation detection methods. When considering the haplotype genome, we found that universally heterozygous sites are also heterozygous in the reference genome and should not be regarded as variation sites between accessions. Filtering out these variations between haplotypes leads to a decrease in the detected heterozygosity, while also suggesting that genetic diversity is lower than originally estimated.”

10-1. Line 288 Information about the origin and how you resequenced and analysed the 427 samples is missing.

Response:

Thank you for the suggestion. We have provided additional details regarding the origin of the 427 accessions and have described the resequencing and analysis processes in the revised Methods section, as follows:

Line 707–716:

“SNP and InDel calling based on resequenced reads

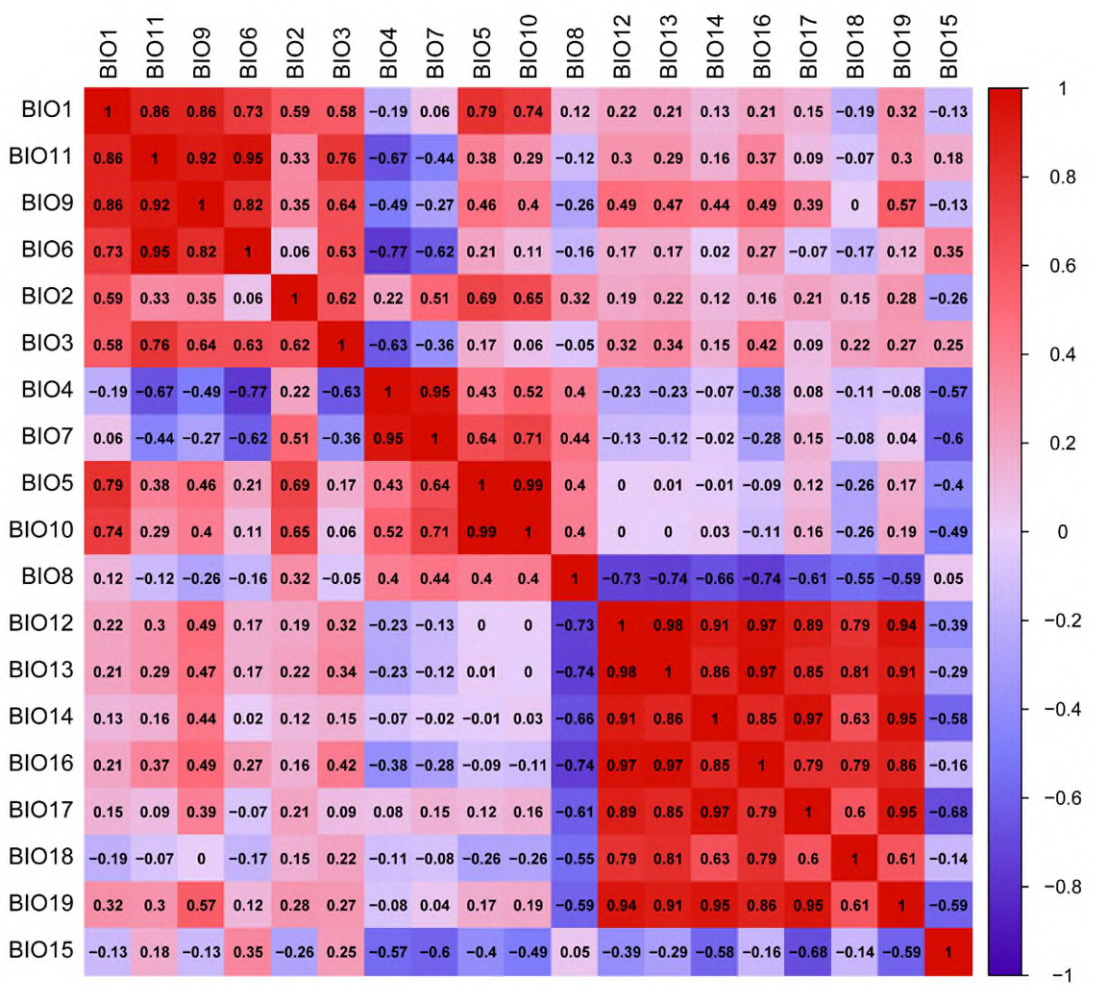
The raw sequencing reads were processed using the same pipeline as in our previous study to ensure consistency. Briefly, the filtered resequenced reads were aligned to the CY haplotype 1 reference genome using BWA v0.7.17[5] (-M). Aligned reads (BAM files) were sorted using SAMtools v1.9[6] (default parameters) and duplicates were removed using GATK v4.2.0[7] (default parameters). SNP and InDel calling was performed using the joint calling method within GATK. We obtained the genomic variant call format (GVCF) in ERC mode for each accession based on reads (-ERC GVCF --native-pair-hmm-threads 100). Then, we filtered SNPs directly based on quality, removing variations with a quality score lower than 50 based on the quality score distribution.”

10-2. Line 297 - Supplementary Fig 11 and Supplementary Fig 12 are missing in the

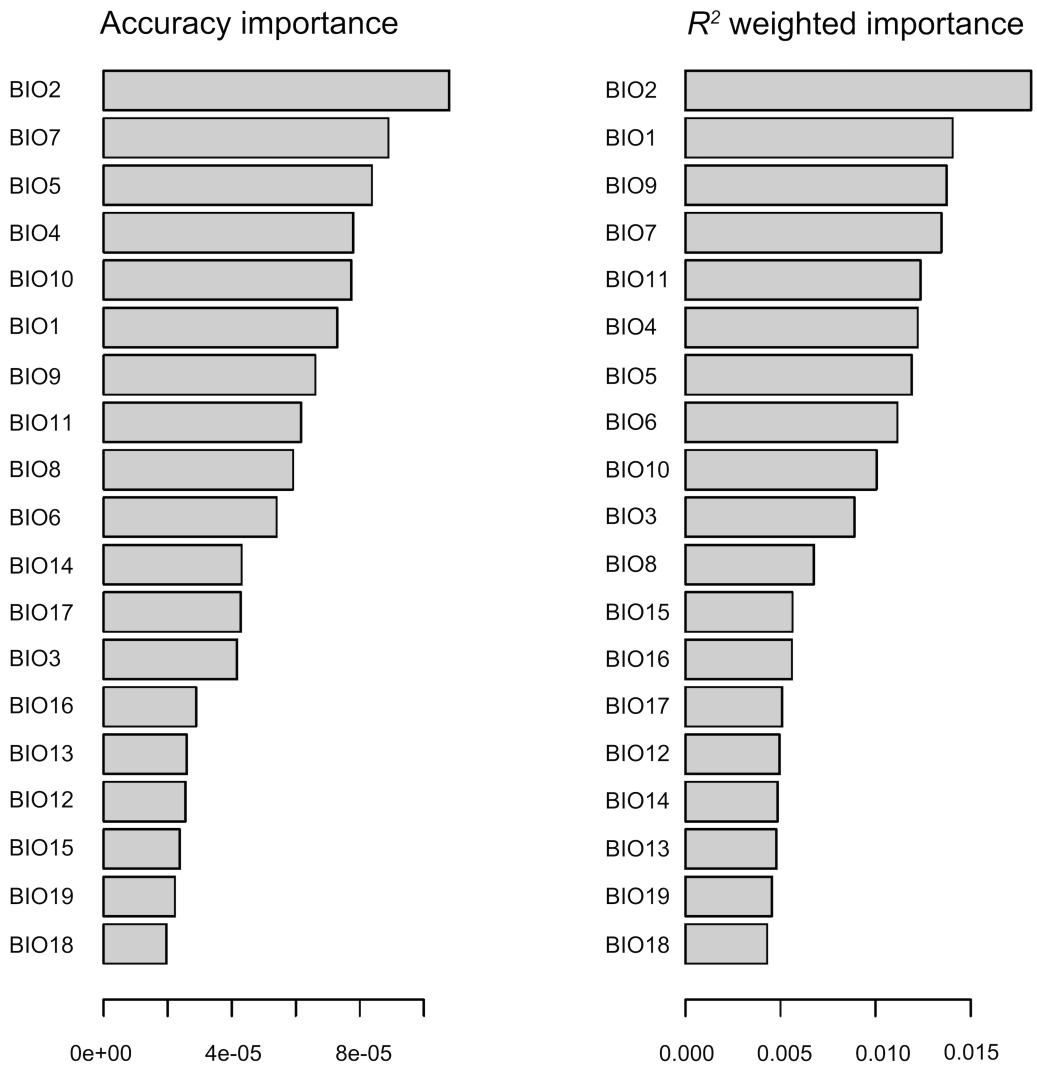
supplementary material.

Response:

Thank you for bringing this oversight to our attention. We apologize for the missing supplementary figures (revised Supplementary Fig. 14 and revised Supplementary Fig. 15) and have now included them in the revision, as follows:



Supplementary Fig. 14 Correlations of bioclimatic variables. Red represents a positive correlation, and blue represents a negative correlation.



Supplementary Fig. 15 Gradient Forest (GF) ranking of bioclimatic variables. The left panel shows the accuracy importance, and the right panel shows the R^2 weight importance.

10-3. Line 302 – Don't mention all BIO variables here because you only used 6.

Response:

Thank you for pointing out this inaccuracy. We have revised the description as follows:

Line 343–344:

"Additionally, compared with 123 variations related to precipitation, 996 variations were associated with temperature."

10-4. Line 341 – how you select this three variables?

Response:

Thank you for this suggestion. Previously, the three variables (BIO1, BIO5, and BIO10) were selected because, among the temperature-related variables, they exhibited relatively higher values. Additionally, the overall trend for temperature is projected to increase in the future, and these three bioclimatic variables were associated with high temperatures. However, we have reperformed the GEA analyses, including the subsequent RONA and genomic offset analyses. Compared to BIO1 (Annual Mean Temperature), BIO5 (Max Temperature of Warmest Month) and BIO10 (Mean Temperature of Warmest Quarter) are more directly related to high temperatures. Moreover, since the results of BIO5 and BIO10 are essentially consistent, we have chosen to focus the higher of the two, i.e., BIO5. Additionally, we have revised related description as follows:

Line 400–404:

"For all temperature-related variables, the RONA values were greater than the precipitation-related variables, and the differences between SSP585 and SSP126 were greater (Fig. 5a). For temperature, the overall trend indicated an increase in the future; therefore, we focused specifically on BIO5."

10-5. Figure 5b is difficult to read. Please label your populations and make the differences in the RONA values more evident. The use of the mean value buffers the differences across climate models.

Response:

Thank you for your valuable suggestion. To address these concerns, we implemented three modifications to improve the clarity and readability of the figures, as detailed below:

1. **Population labels**: We have clearly labeled each moso bamboo population represented in the figure, making it easier for readers to identify and

distinguish between them. This addition will facilitate a better understanding of the spatial distribution and potential risks faced by different populations.

2. **Differences in the RONA values**: To make the differences in the RONA values more evident, we adjusted the color scale and visual representation of the data. This enhancement will allow readers to more easily discern the variations in risk across populations and climate models, providing a clearer picture of the potential impacts of climate change on moso bamboo.
3. **Use of individual climate model values**: According to the reviewer's suggestion, we have now presented the RONA values for each individual climate model (ACCESS-CM2, CMCC-ESM2, GISS-E2-1-G, and MIROC6) instead of using the mean value across models. This change will provide a more detailed and transparent representation of the potential risks faced by moso bamboo populations under different climate change scenarios.

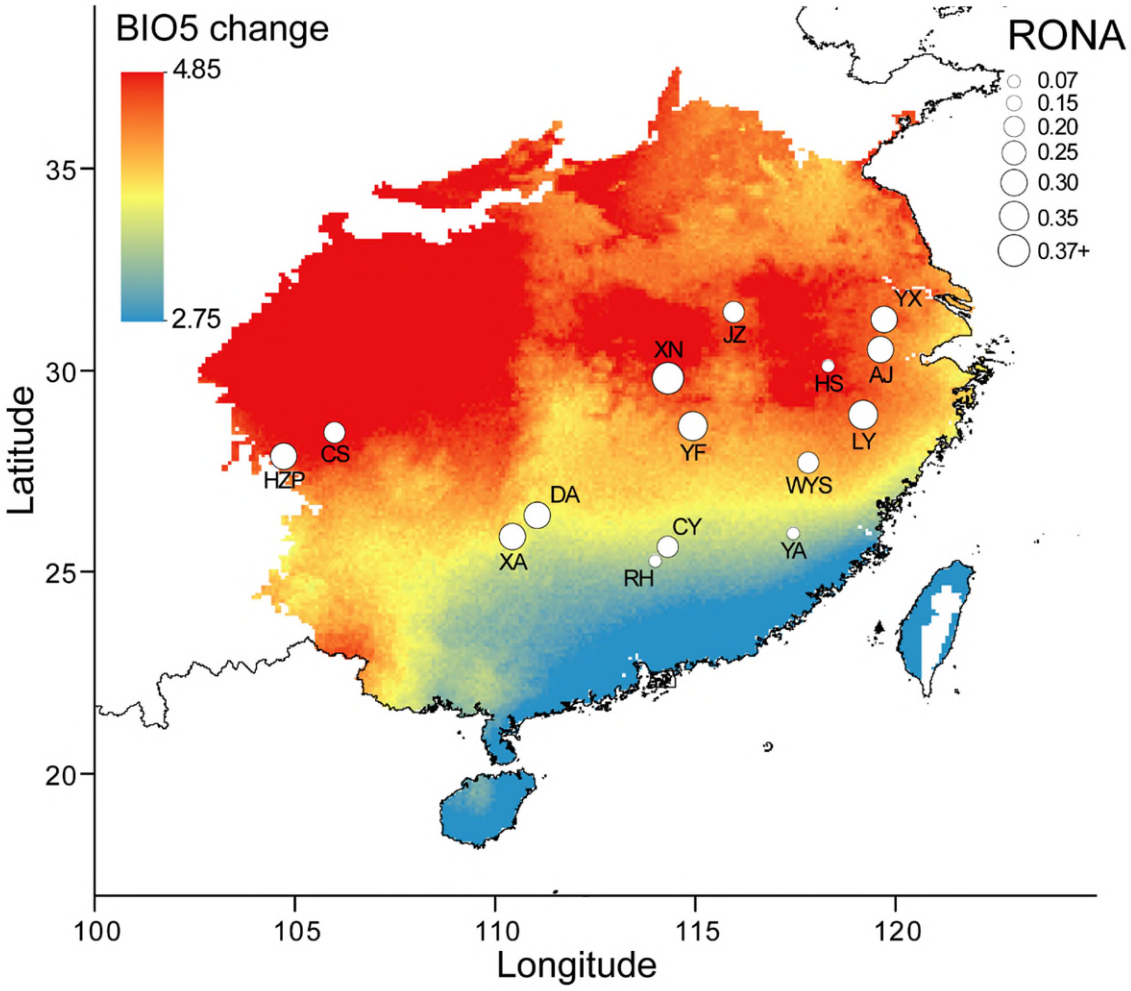


Fig. 5b Mean RONA estimates for the moso bamboo population under the high-emission scenario (SSP585) and the max temperature of warmest month (BIO5) for 2061–2080 based on four individual climate models (ACCESS-CM2, CMCC-ESM2, GISS-E2-1-G, and MIROC6). The map colors indicate projected climate changes in BIO5, with darker red indicating more substantial increases in temperature. The circle size represents the RONA values of different populations.

10-6 Figure 5e and f are very complex and deserve a better discussion of the results. Some part of your results looks better explanation and discussion.

Response:

Thank you for the suggestion. Figures 5e and 5f present a combinative visualization of three genomic offset (local offset, forward offset, and reverse offset) by mapping them

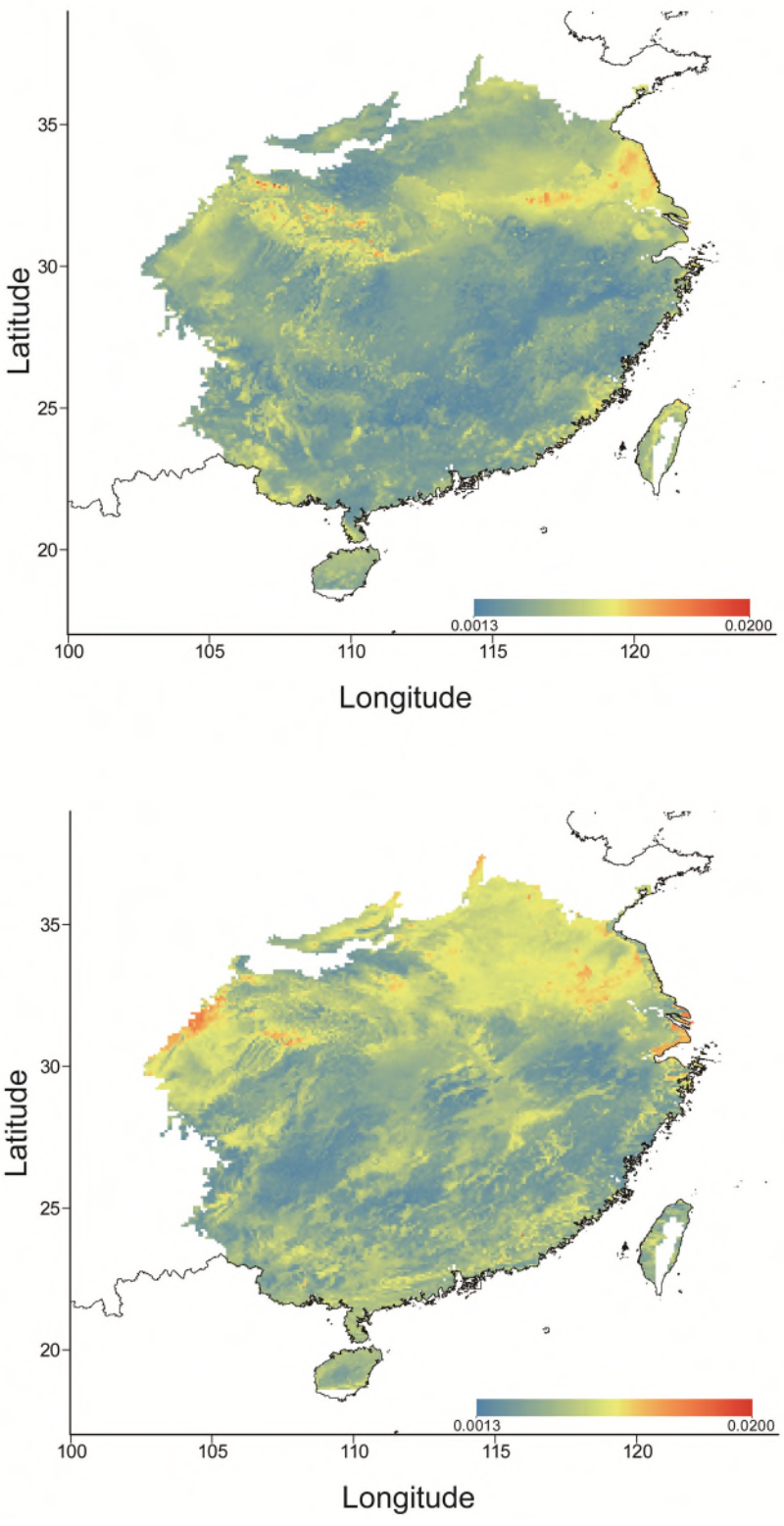
as red, green, and blue bands, respectively, in an RGB color space. This approach allows for a simultaneous comparison of the relative magnitudes of these offset measures across the geographic range of the species. In the figures, brighter cells (closer to white) have relatively greater genomic offset values, while darker cells (closer to black) have relatively lower values along each axis.

To enhance the interpretability of these complex figures, we have made the following modifications:

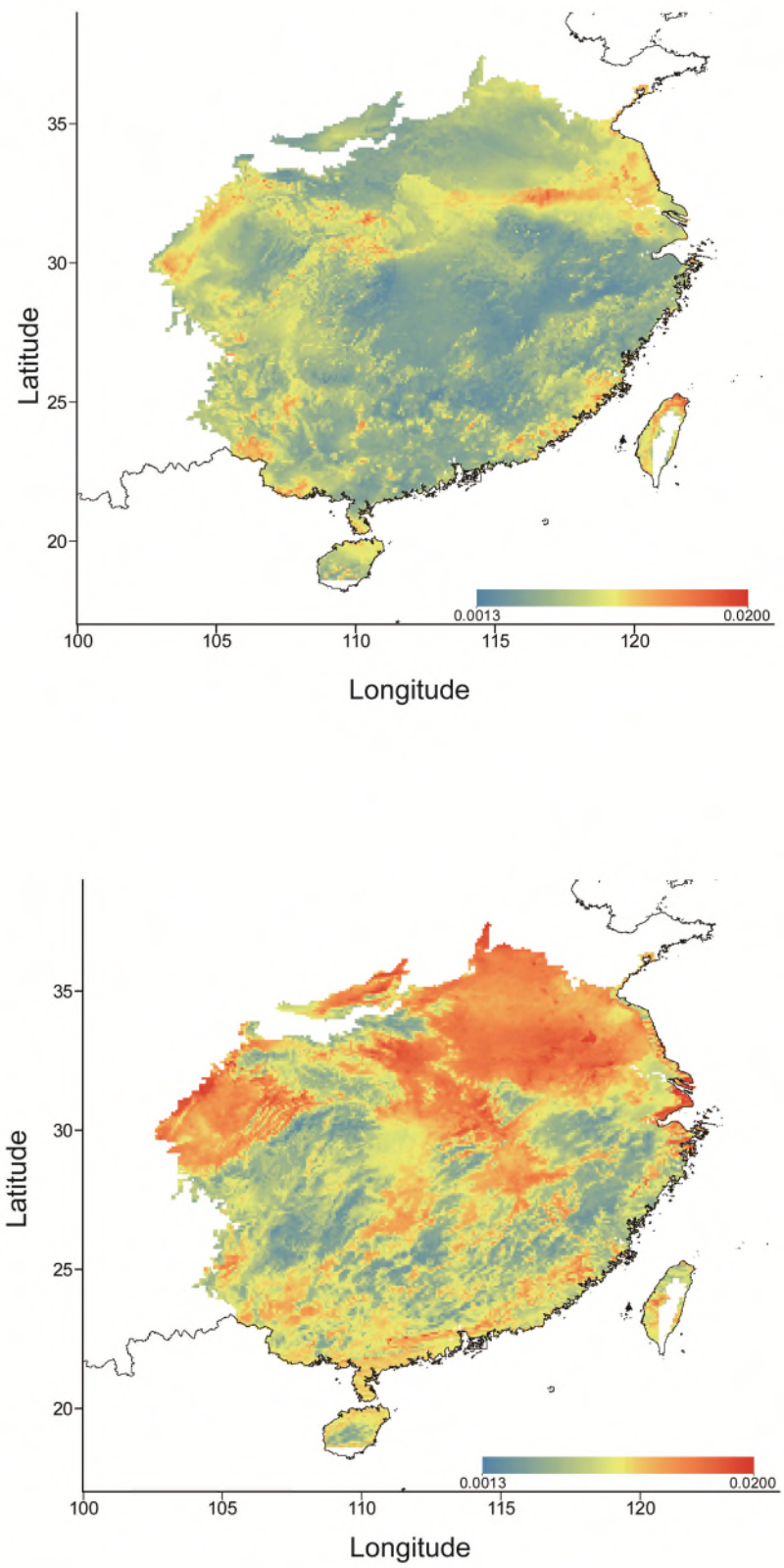
1. **Explaining the definition of genomic offset.** To improve the understanding of the genomic offset definition and its calculation, we have added an explanation to the Methods section, based on a previous reference[32].
2. **Interpreting and refining visualization of genomic offset results.** To further aid in the understanding and interpretation of the genomic offset results, we have included individual genomic offset maps in the revised Supplementary Figs. 22–23 as follows:

In the Methods section (Line 776–786):

"Local offset, a measure of the vulnerability of a resident population to climate change, was calculated by estimating the predicted change in allele frequencies at climate-adaptive loci that was necessary for the population to adapt to local climate changes over time. In contrast, forward offset assumed that populations had unlimited migration ability. It was calculated by identifying the minimum predicted offset if propagules or alleles could move, through gene flow, to any suitable habitat within the range. Reverse offset represented the possibility that any population in the current range would be preadapted to a particular location in the future. Reverse offset was calculated by identifying the minimum offset between hypothetical populations within the current range in the future climate and populations in the current climate."



Supplementary Fig. 22 Map of the predicted forward genomic offset and reverse genomic offset averaged across four climate models across the distribution of moso bamboo under SSP126. The upper shows forward genomic offset, and the lower shows the reverse genomic offset.



Supplementary Fig. 23 Map of the predicted forward genomic offset and reverse genomic offset averaged across four climate models across the distribution of moso bamboo under SSP585. The upper shows forward genomic offset, and the lower shows the reverse genomic offset.

Additionally, we have updated the Results section to provide a more comprehensive explanation of the patterns observed.

Line 411–430:

“In addition to the local genomic offset, we also calculated the forward and reverse genomic offsets (Figs. 5e–5f and Supplementary Figs. 22–23). Figs. 5e–5f showed a combinative visualization of three genomic offset (local offset, forward offset, and reverse offset) by mapping them as red, green, and blue bands, respectively, in an RGB color space. Brighter cells (closer to white) and darker cells (closer to black) presented relatively greater and lower values along each axis, respectively. Most of the northern regions appear brighter (Figs. 5e–f) indicated that they had relatively high offset values, suggesting that even with migration, they still face greater vulnerability compared to the southern regions. However, both the forward and reverse offsets were lower than the local offset (lower panels in Figs. 5e–f) in most of northern region, suggesting that assisted migration may to some extent enable adaptation to future climate change. Consistent with the RONA results, all the genomic offsets were greater under the SSP585 scenario than under the SSP126 scenario and the regions with high genomic offsets (brighter area) were also larger, suggesting that the more extreme climate change associated with fossil fuel development (SSP585) may expose moso bamboo populations to greater adaptive challenges and potential risks than under the more sustainable scenario (SSP126). As for moso bamboo in major natural distribution regions, they are still in a relatively safe position under the SSP126 scenario. However, under the SSP585 scenario, some major natural growth regions will face risks, especially the two westernmost natural growth regions (Fig. 5f).”

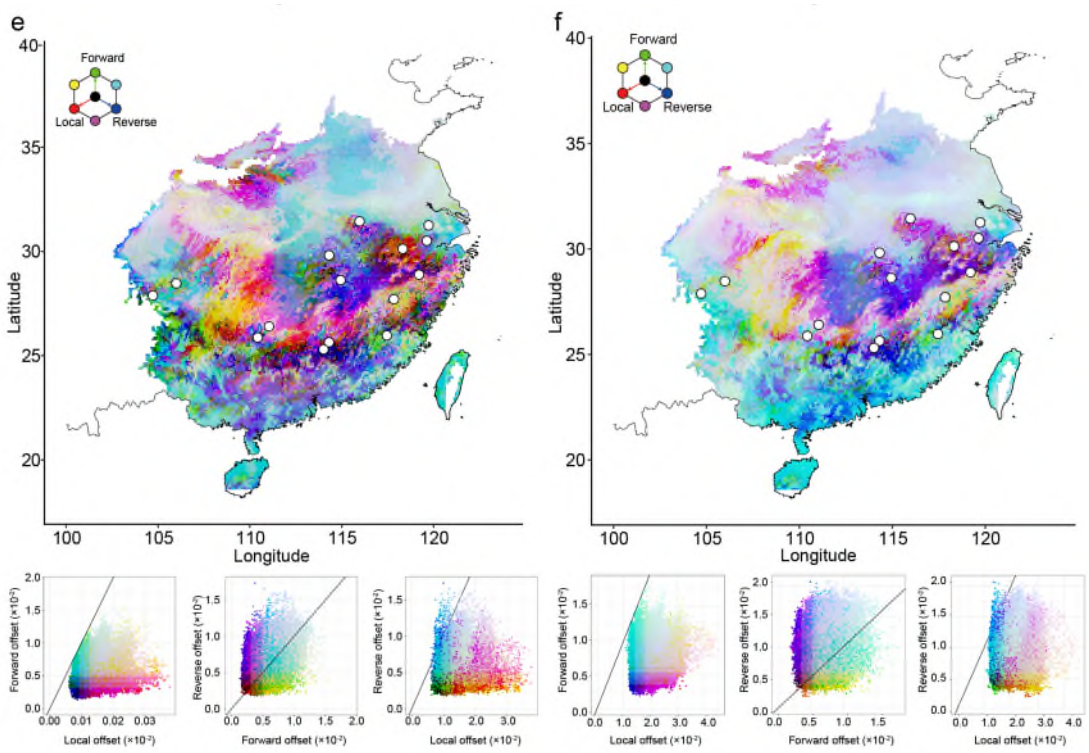


Fig. 5e–f RGB map showing local (red), forward (green), and reverse (blue) genomic offsets for SSP126 (e) and SSP585 (f), respectively. Brighter cells (closer to white) have relatively greater genomic offset values, and darker cells (closer to black) have relatively lower values along each axis. The lower panels are the bivariate scattergrams of e and f with 1:1 lines.

10-7 Are the results between RONA and GF congruent? Do they show the same variables?

Response:

Thank you for the suggestion. To address this concern, we structured our explanation around these key points, as detailed below:

1. **Complementary insights from RONA and GF analyses:** While RONA and GF analyses differ in their methodological approaches, they provide complementary insights into the potential vulnerability of moso bamboo populations under future climate change scenarios. RONA focuses on quantifying the risk of non-adaptedness for each individual bioclimatic variable, providing a detailed assessment of how specific aspects of the

climate may pose challenges for moso bamboo populations. In contrast, the GF analysis integrates the results across all 19 bioclimatic variables, offering a holistic measure of the overall risk level faced by each region based on the composite changes in future temperature and precipitation conditions.

2. **Congruence in vulnerability assessments:** Despite these methodological differences, both RONA and GF analyses consistently showed that vulnerability was greater under the high-emissions SSP585 scenario compared to the more sustainable SSP126 scenario. This congruence suggests that the more extreme climate change associated with SSP585 may expose moso bamboo populations to greater adaptive challenges and potential risks than the more moderate changes projected under SSP126.
3. **Importance of multiple analytical approaches:** These findings underscore the importance of considering multiple analytical approaches when assessing the potential impacts of climate change on species and ecosystems. By combining the detailed insights provided by RONA with the integrative perspective offered by GF, we can develop a more comprehensive understanding of the risks faced by moso bamboo populations and inform effective conservation and management strategies.

10-8 How are you results hampered by the lack of sampling from the extreme southern and northern populations?

Response:

Thank you for the suggestion. To address this concern, we implemented the modifications to improve the clarity and readability, as detailed below:

1. **Comprehensive sampling of major natural moso bamboo populations.** Our team has been dedicated to the investigation and exploration of bamboo germplasm resources for many years. Our team have published the first genome of the bamboo subfamily, i.e., the moso bamboo genome[8,9], conducted a comprehensive survey of moso bamboo germplasm resources in

China[10], and published the first population resequencing study in the bamboo subfamily, focusing on moso bamboo population resequencing[11]. Through these efforts, we have preliminarily determined the distribution of natural moso bamboo populations and ensured that our sampling in this project has fully covered the relevant regions. Therefore, we would like to express that our sampling regions have already covered all the major natural distribution regions of moso bamboo in China, which we believe is sufficient for identifying the adaptive variations and predicting the risks for moso bamboo in these regions.

2. **Future research on transplanted populations.** In the future, we plan to expand our research to include transplanted populations within China. This will further enhance our understanding of the genetic diversity and adaptive potential of moso bamboo across its entire range, including the extreme southern and northern populations.
3. **Acknowledging limitations.** As mentioned in the Discussion section, we have acknowledged the limitation posed by the lack of samples from some extreme populations. While our current sampling strategy provides a robust foundation for understanding the adaptive potential of moso bamboo in its major natural habitats, we recognize that incorporating data from these extreme populations could offer additional insights into the species' genetic diversity and adaptive capacity.

Line 501–509:

"It is noteworthy that our samples contain only moso bamboo from all the major natural distribution regions of moso bamboo in China, missing some of the human-transplanted populations or extreme populations. Supplementing these populations, and even global moso bamboo accessions, could enable the identification of more variations adapted to extreme environments. For risk predictions like RONA and local offsets that do not involve migration, the absence of these samples is less impactful. However, for forward

and reverse offset analyses, incorporating additional populations could uncover regions more conducive for moso bamboo cultivation and identify moso bamboo populations better suited for migration to extreme regions.”

References

1 China Forestry and Grassland Administration. *Development Plan for Forestry and Grassland Industry (2021–2025)*. (2019).

2 China Forestry and Grassland Administration. *et al. Opinions from Ten Departments on Accelerating the Innovative Development of the Bamboo Industry*. (2021).

3 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

4 Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574 (2021).

5 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

6 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

7 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

8 Peng, Z. *et al.* The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat. Genet.* **45**, 456–461 (2013).

9 Zhao, H. *et al.* Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*). *Gigascience* **7**, giy115 (2018).

10 Jiang, W. *et al.* Microsatellite markers revealed moderate genetic diversity and population differentiation of moso bamboo (*Phyllostachys edulis*)—a primarily asexual reproduction species in China. *Tree Genetics. Genomes* **13**, 130 (2017).

11 Zhao, H. *et al.* Analysis of 427 genomes reveals moso bamboo population structure and genetic basis of property traits. *Nat. Commun.* **12**, 5466 (2021).

12 Cheng, H. *et al.* Haplotype-resolved assembly of diploid genomes without parental data. *Nat. Biotechnol.* **40**, 1332–1335 (2022).

13 Isagi, Y. *et al.* Clonal structure and flowering traits of a bamboo [*Phyllostachys pubescens* (Mazel) Ohwi] stand grown from a simultaneous flowering as revealed by AFLP analysis. *Mol. Ecol.* **13**, 2017–2021 (2004).

14 Ma, P. *et al.* Negative correlation between rates of molecular evolution and flowering cycles in temperate woody bamboos revealed by plastid phylogenomics. *BMC Plant Biol.* **17**, 260 (2017).

15 Yu, J. *et al.* Analysis of aldo–keto reductase gene family and their responses to salt, drought, and abscisic acid stresses in *Medicago truncatula*. *Int. J. Mol. Sci.* **21**, 754 (2020).

16 Boyes, D. C., Nam, J. & Dangl, J. L. The *Arabidopsis thaliana* RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. *Proc. Natl. Acad. Sci. USA* **95**, 15849–15854 (1998).

17 Zhang, M., Wang, L. & Zhong, D. Photolyase: dynamics and mechanisms of repair of sun-induced DNA damage. *Photochem. Photobiol.* **93**, 78–92 (2017).

18 Bi, D. *et al.* Configuration and spin-up of ACCESS-CM2, the new generation Australian Community Climate and Earth System Simulator Coupled Model. *J.*

South. Hemisph. Earth Syst. Sci. **70**, 225–251 (2020).
19 Lovato, T. *et al.* CMIP6 simulations with the CMCC Earth System Model (CMCC-ESM2). *J. Adv. Model. Earth Syst.* **14**, e2021MS002814 (2022).
20 Kelley, M. *et al.* GISS-E2.1: configurations and climatology. *J. Adv. Model. Earth Syst.* **12**, e2019MS002025 (2020).
21 Kataoka, T. *et al.* Seasonal to decadal predictions with MIROC6: description and basic evaluation. *J. Adv. Model. Earth Syst.* **12**, e2019MS002035 (2020).
22 O'Neill, B. C. *et al.* The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.* **9**, 3461–3482 (2016).
23 Riahi, K. *et al.* The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Glob. Environ. Change* **42**, 153–168 (2017).
24 Hurtt, G. C. *et al.* Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6. *Geosci. Model Dev.* **13**, 5425–5464 (2020).
25 van Vuuren, D. P. *et al.* The representative concentration pathways: an overview. *Clim. Change* **109**, 5 (2011).
26 Shi, T. *et al.* The super-pangenome of *Populus* unveils genomic facets for its adaptation and diversification in widespread forest trees. *Mol. Plant* **17**, 725–746 (2024).
27 Sun, H. *et al.* Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nat. Genet.* **54**, 342–348 (2022).
28 Qin, P. *et al.* Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**, 3542–3558 (2021).
29 Gain, C. *et al.* A quantitative theory for genomic offset statistics. *Mol. Biol. Evol.* **40**, msad140 (2023).
30 Lotterhos, K. E. Interpretation issues with "genomic vulnerability" arise from conceptual issues in local adaptation and maladaptation. *Evol. Lett.*, qrae004 (2024).
31 Sang, Y. *et al.* Genomic insights into local adaptation and future climate-induced vulnerability of a keystone forest tree in East Asia. *Nat. Commun.* **13**, 6541 (2022).
32 Gougherty, A. V., Keller, S. R. & Fitzpatrick, M. C. Maladaptation, migration and extirpation fuel climate change risk in a forest tree species. *Nat. Clim. Change* **11**, 166–171 (2021).
33 Fitzpatrick, M. C. & Keller, S. R. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. *Ecol. Lett.* **18**, 1–16 (2015).
34 Chen, Y. *et al.* The combination of genomic offset and niche modelling provides insights into climate change-driven vulnerability. *Nat. Commun.* **13**, 4821 (2022).
35 Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
36 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

37 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have addressed all of my comments and the quality of the manuscript has been significantly improved.

Reviewer #2 (Remarks to the Author):

The manuscript has been extensively improved and I have no additional comments to add. Congratulations!

Reviewer #3 (Remarks to the Author):

This new version is much better than the previous one, again you have a very impressive volume of data, analysis and results. The respective justification and explanation of the results and the final discussion are now much better. The paper is now much easier to read and understand than the previous version.

I don't have any substantial comments to add but just minor remarks and two requests.

In this version I don ´t have access to Supplementary tables (or I cannot find it, only the excel file with the associated data). I only have access to a file with supplementary table 30 ! that is truncated!

Line 60-62 - repetition of the same sentence

Line 133 -The new red sentence is not clear

Line 339 - What you mean by "after integrating the LFMM2 and RDA results"

In the method section you put the option that you use for each command, in some cases it is easy to understand the meaning of the option. I give two examples: "The vg construct command (-a -S) was used to construct the initial graph"; "To determine a pruned SNP set, we used PLINK v1.982 (-indep-pairphase 100 10 0.2)."
The information is important but as it is just "noise". My suggestion is that you either explain it in the text or alternatively explain the option you use (other than the defaults) in a supplementary table or put a text in gitlab or zenodo repository (or add it to your GitHub). This is important for other scientists to be able to reproduce your results.
This information is particularly relevant for the PLINK, because from your previous version of the draft where you have an admixture K=3 to this version where you got a K=1, what seems to be different is the PLINK pruned. So, the rationale of the pruning process must be clearly explained.
You also have to explain your filtering rationale with vcftools, why do you use a MAF of 0.05? You

are excluding SNPs that appear in more than 20 samples/haplotypes. This type of SNPs can have a tendency to have a local/regional distribution and could have a relevant role for local adaptation.

Finally, I still have some difficulty in understanding your supplementary table 30 and your comparisons between haplotypes. Your Figure S26 is very useful. So, if you compare inter-haplotypes of the same accession, you should have a large difference as you describe but if you compare inter-accession, you should have a bimodal distribution. When you compare the red haplotypes inter-accession you have small differences, when you compare the blues, again you should have small differences, but when you compare the red and blue of different accession you should again have a large value. So, your inter-accession values depend on the type of haplotype you are making your comparison. What am I missing here?

RESPONSE TO REVIEWERS' COMMENTS

Responses to the comments of Reviewer #1

The authors have addressed all of my comments and the quality of the manuscript has been significantly improved.

Response:

Thank you for the positive assessment of our revised manuscript. Your insightful comments from the first round of revisions were instrumental in significantly improving our work, and we're pleased that our efforts to address your concerns have resulted in a substantially improved manuscript.

Responses to the comments of Reviewer #2

The manuscript has been extensively improved and I have no additional comments to add. Congratulations!

Response:

Thank you for the positive assessment of our revised manuscript. Your insightful comments from the first round of revisions were instrumental in significantly improving our work, and we're pleased that our efforts to address your concerns have resulted in a substantially improved manuscript.

Responses to the comments of Reviewer #3

This new version is much better than the previous one, again you have a very impressive volume of data, analysis and results. The respective justification and explanation of the results and the final discussion are now much better. The paper is now much easier to read and understand than the previous version.

Response:

Thank you for the positive assessment of our revised manuscript. Your insightful

comments from the first round of revisions were instrumental in significantly improving our work, and we're pleased that our efforts to address your concerns have resulted in a substantially improved manuscript.

1. I don't have any substantial comments to add but just minor remarks and two requests. In this version I don´t have access to Supplementary tables (or I cannot find it, only the excel file with the associated data). I only have access to a file with supplementary table 30! that is truncated!

Response:

We sincerely apologize for the inconvenience caused by the missing supplementary tables. We have now thoroughly reviewed and included all supplementary tables, ensuring they are complete and accessible. Specifically:

- Supplementary Table 5 (Supplementary Table 30 in previous) has been checked and uploaded to include the full data.
- We have verified the accessibility of all supplementary materials and added to the submission system.

We understand the importance of these materials for a comprehensive review of our work and regret any difficulties this may have caused. We are committed to providing all necessary information to support our research findings.

2. Line 60-62 - repetition of the same sentence

Response:

Thank you for pointing out this oversight. We have removed the duplicated sentence and thoroughly checked the entire revision to ensure no similar issues remain.

3. Line 133 -The new red sentence is not clear

Response:

Thank you for the suggestion. We have revised the sentence as follows:

Line 131–133:

"To comprehensively characterize the genetic variations across the 16 RMAs, we selected CYhap1 as the reference genome due to its superior quality compared to the other accessions, and successfully resolved its haplotypes."

4. Line 339 - What you mean by "after integrating the LFMM2 and RDA results"

Response:

Thank you for your suggestion. By "integrating the LFMM2 and RDA results," we mean that we identified and retained only those genetic variations that were simultaneously detected by both methods (LFMM2 and RDA) as being associated with environmental factors. This conservative approach was taken to:

- Increase the reliability of our findings by leveraging the strengths of both methods
- Minimize false positives by focusing on the most consistent signals
- Prioritize the most robust genotype–environment associations

To clarify this in the revision, we have revised the description, as follows:

Line 284–285:

"After retaining the variations identified by both LFMM2 and RDA methods (Supplementary Fig. 18 and Supplementary Data 14–15),"

5. In the method section you put the option that you use for each command, in some cases it is easy to understand the meaning of the option. I give two examples: "The vg construct command (-a -S) was used to construct the initial graph"; "To determine a pruned SNP set, we used PLINK v1.982 (-indep-pairphase 100 10 0.2)."

The information is important but as it is just "noise". My suggestion is that you either explain it in the text or alternatively explain the option you use (other than the defaults) in a supplementary table or put a text in gitlab or zenodo repository (or add it to your GitHub). This is important for other scientists to be able to reproduce your results.

Response:

Thank you for your insightful suggestion. We fully agree that while the presentation of

command options are crucial for reproducibility, they may inadvertently disrupt the flow of the main text for readers. To address this concern, we have made the following revisions:

- **Removal of detailed command options.** We have removed the detailed command options from the Methods section in the main text to improve readability and maintain a clear focus on the key methodological aspects of our study.
- **Creation of supplementary table for command options.** We have created a new Supplementary Table (Supplementary Table 13) that lists all the command options used in our analyses. This table includes the software name, version, and the specific options used for each command, along with brief explanations of their functions. Additionally, in the Code availability section, we now direct readers to Supplementary Table 13 for detailed information on the command options employed in our study. This approach allows readers to easily locate and reference this information without detracting from the main narrative of our manuscript.

Supplementary Table 13. Software packages, versions, and parameters used in this study.

Software	Version	Parameters	Function
Genome assembly			
CCS algorithm	v6.2.0	default parameters	Generate CCS reads
Hifiasm	v0.16.1-r375	default parameters	Genome assembling
BUSCO	v5.4.3	-m genome	Genome evaluation
LTR_retriever	v2.9.0	default parameters	LTR prediction
Meryl	v1.41	k = 20	Generate Meryl databases
Merqury	v1.3	default parameters	Genome evaluation
Genome annotation			
RepeatModeler	v2.0.3	-LTRStruct	de novo repeat library
RepeatMasker	v4.1.2-p1	default parameters	de novo repeat library
GeneWise	v2.4.1	-max_gene_length 23707 -segmentSize 1000000 -overlapSize 100000 -	Gene prediction

		coverage_ratio 0.4 -evalue 1e-9	
Trimmomatic	v0.38	default parameters	RNA-seq quality control
HISAT2	v2.1.0	-min-intronlen 20 -max-intronlen 20000 -dta -score-min L,0.0,-0.4	Mapping to genome
AUGUSTUS	v3.4.0	default parameters	ab initio prediction
BUSCO	v5.4.3	default parameters	Annotation evaluation
DIAMOND		--sensitive --max-target-seqs 200 --evalue 1e-5	Mapping to database
Switch errors calculation			
pbmm2	v1.5.0	--preset HIFI --sort	Mapping to genome
DeepVariant	v1.4.0	--model_type PACBIO	Variation detection
Whatshap	v1.1	--ignore-read-groups	SNP phasing
nucmer	v4.0.0rc1	-mum -l 1000 -c 200 -g 200	Genome alignment
show-snps	v4.0.0rc1	default parameters	SNP calling
Bwa mem	v0.7.17	default parameters	Mapping to genome
DeepVariant	v1.4.0	-SP5M	Variation detection (Hi-C)
Allele identification			
BLASTP	v2.9.0+	-evalue 1e-10	Protein alignment
Minimap2	v2.24-r1122	-ax asm5	Genome alignment
Allele-specific expression			
HISAT2	v2.1.0	-k 1	Mapping to genome
StringTie	v1.3.5	default parameters	TPM calculation
DESeq2	v1.34.0	default parameters	Differential expression analysis
SNP and InDel identification			
nucmer	v4.0.0rc1	--mum -l 100 -c 200 -g 200	Genome alignment
show-snps	v4.0.0rc1	-Clr	SNP calling
svim-asm	v1.0.2	--min_sv_size 1 --min_sv_size 50	InDel calling
pbmm2	v1.5.0	--preset HIFI --sort	Mapping to genome
DeepVariant	v1.4.063	--model_type PACBIO	Variation calling
Structural variation identification			
Minimap2	v2.24-r1122	--paf-no-hit -ax map-hifi	Mapping to genome
cuteSV	v1.0.13	-l 50 -s 5 --max_cluster_bias_INS 1000 --diff_ratio_merging_INS 0.9 --max_cluster_bias_DEL 1000 --	SV calling

		diff_ratio_merging_DEL 0.5	
Ngmlr	v0.2.7	default parameters	Mapping to genome
sniffles	v1.0.12	-s 5 -l 50	SV calling
pbmm2	v1.5.0	--preset HIFI	Mapping to genome
pbsv	v2.8.0	default parameters	SV calling
Minimap2	v2.24-r1122	--paf-no-hit -a -x asm5 --cs	Mapping to genome
svim-asm	v1.0.2	--min_sv_size 50	SV calling
Nucmer	v4.0.0rc1	--maxmatch -l 100 -c 500	Mapping to genome
Assemblytics	v1.2.1	default parameters	SV calling
Graph-based pangenome construction			
vg construct	v1.38	-a -S	Construct vg from a reference
vg gbwt	v1.38	-P	Manipulate GBWTs
vg snarls	v1.38	default parameters	Generate snarls from graph
vg index	v1.38	default parameters	Generate vg index
vg giraffe	v1.38	default parameters	Short read mapping to vg
SNP and InDel calling based on resequenced reads			
BWA	v0.7.17	-M	Mapping to genome
GATK GVCF	v4.2.0	-ERC GVCF --native-pair-hmm-threads 100	SNP calling
PLINK	v1.9	-indep-pairphase 100 10 0.2	LD pruning
ADMIXTURE	v1.3.0	-cv -j4	Assess population structure
Identification of climate-associated variations			
vcftools	v0.1.13	--minDP 3 --max-missing 0.2 --maf 0.05	Filter SNP
LEA	v3.6.0	K = 1	LFMM2 analysis
Calculation of RONA and genomic offset			
pyRona	v0.36	-P 0.01	RONA analysis

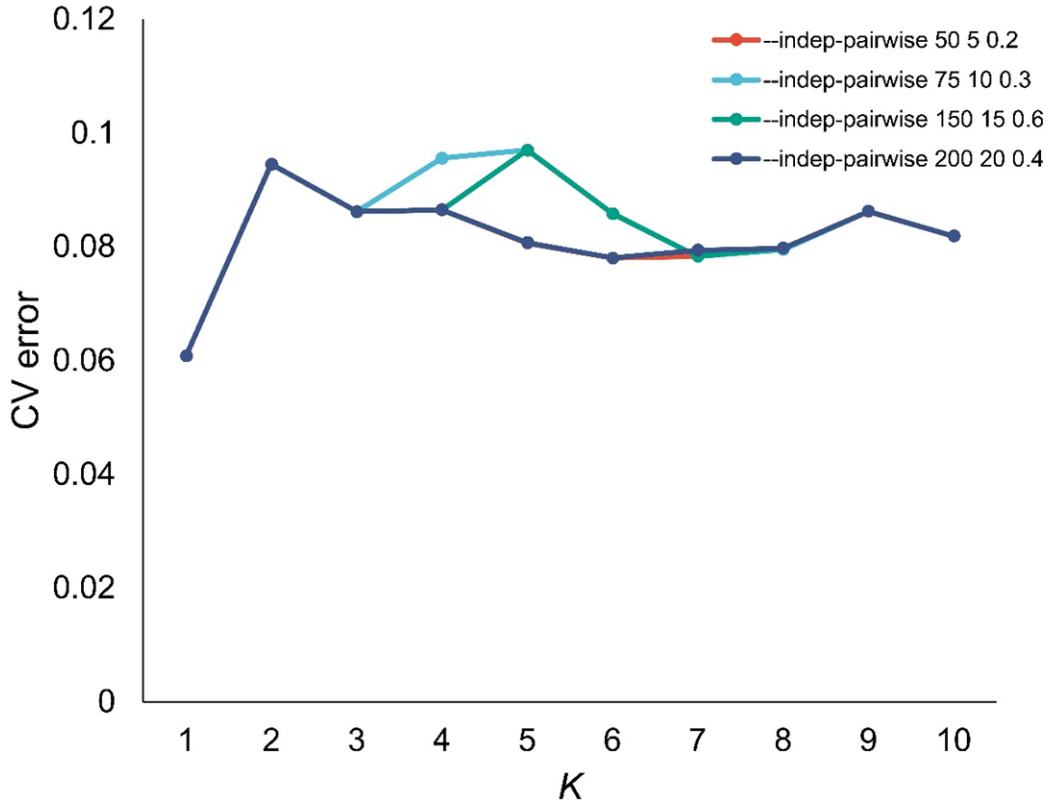
6. This information is particularly relevant for the PLINK, because from your previous version of the draft where you have an admixture K=3 to this version where you got a K=1, what seems to be different is the PLINK pruned. So, the rationale of the pruning process must be clearly explained.

Response:

Thank you for raising this important suggestion regarding the change in the admixture *K* value between the previous and current versions of our manuscript. We appreciate the opportunity to clarify the rationale behind our PLINK pruning process and its

impact on the results, as follows:

- **Refinement of dataset for admixture analysis.** In the first version, we performed linkage disequilibrium (LD) pruning and admixture analysis on a combined file containing single nucleotide polymorphisms (SNPs), insertions and deletions (InDels), and structural variations (SVs). However, upon further consideration and in light of recommendations from the literature[1], we realized that this approach was suboptimal due to the inherent differences between these variant types. To ensure the accuracy of our admixture analysis, we decided to focus solely on the most abundant variant type, SNPs.
- **Optimal *K* value and consistency with prior findings.** By refining our dataset to include only SNPs, we obtained an optimal $K = 1$ in the admixture analysis. This result is consistent with our previous findings and aligns with the known low genetic diversity of moso bamboo. The change in the *K* value can be attributed to both the PLINK filtering parameters and the refinement of our dataset.
- **Robustness testing with different PLINK pruning thresholds.** To validate the robustness of our results, we tested several different PLINK pruning thresholds and calculated the corresponding *K* values (Response Fig. 2). Consistently, the results showed an optimal $K = 1$ across these different filtering parameters, further supporting the reliability of our findings.



Response Fig. 2 Cross-validation (CV) error values for different values of K in different PLINK parameters. To validate the robustness of our results, we tested four different PLINK pruning thresholds and calculated the corresponding *K* values. Consistently, the results showed an optimal $K = 1$ across these different filtering parameters, further supporting the reliability of our findings.

7. You also have to explain your filtering rationale with vcftools, why do you use a MAF of 0.05? You are excluding SNPs that appear in more than 20 samples/haplotypes. This type of SNPs can have a tendency to have a local/regional distribution and could have a relevant role for local adaptation.

Response:

Thank you for raising this important suggestion regarding our filtering rationale with vcftools and the potential impact on locally adapted variants. We would like to address your concerns as follows:

- **Prioritizing common variants and minimizing false positives.** Our primary focus was on analyzing common variations and removing rare variants, which

tend to have a high false positive rate and can introduce noise into the analysis. By setting a minor allele frequency (MAF) threshold of 0.05, we aimed to prioritize variants that are more likely to be biologically relevant and minimize the inclusion of potential sequencing errors or low–confidence calls. This threshold is commonly used for filtering SNPs in genome–environment association (GEA) studies and other association analyses[2,3].

- **Retention of locally adapted variants.** We would like to clarify that we retained variants that appeared in more than 20 samples, rather than excluding them. Given that our sample size for each region is greater than 21, even if a variant is specific to a particular region, it would still be captured in our dataset. Therefore, we do not expect to miss out on locally adapted variants due to our filtering criteria.

We hope this explanation clarifies our rationale for the chosen filtering parameters and addresses your concerns regarding the potential impact on locally adapted variants. We believe these additions will enhance the clarity and reproducibility of our work.

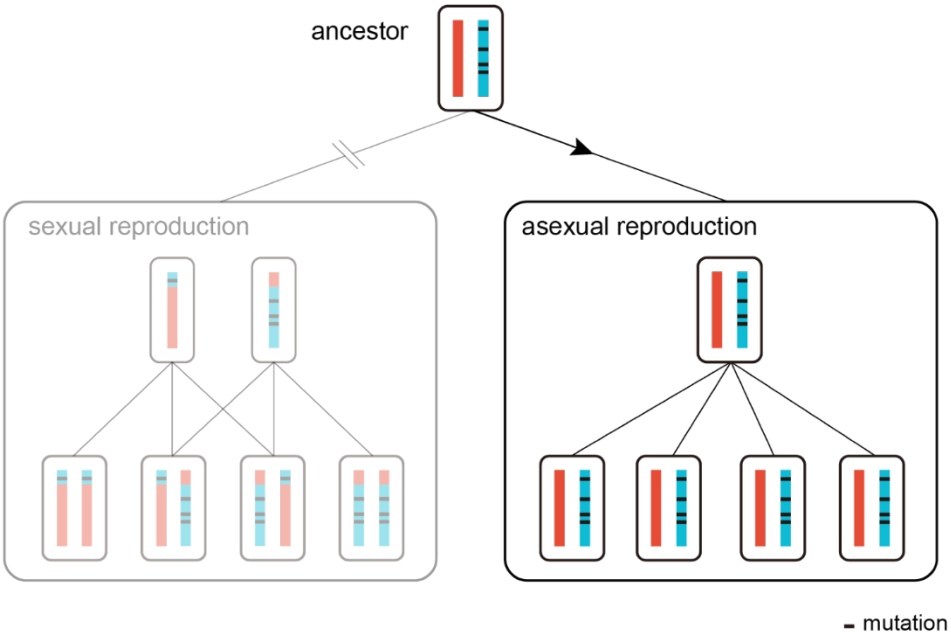
8. Finally, I still have some difficulty in understanding your supplementary table 30 and your comparisons between haplotypes. Your Figure S26 is very useful. So, if you compare inter-haplotypes of the same accession, you should have a large difference as you describe but if you compare inter-accession, you should have a bimodal distribution. When you compare the red haplotypes inter-accession you have small differences, when you compare the blues, again you should have small differences, but when you compare the red and blue of different accession you should again have a large value. So, your inter-accession values depend on the type of haplotype you are making your comparison. What am I missing here?

Response:

Thank you for your insightful question about the comparisons between haplotypes. We appreciate your feedback on the usefulness of Supplementary Fig. 26 in understanding the haplotype comparisons. Allow us to clarify the points you raised, as follows:

- **Limitations of haplotype assembly.** It is important to note that only three out of the 16 representative moso bamboo accessions (RMAs) were assembled with Hi-C reads. The remaining 13 assemblies did not fully resolve the haplotypes across all chromosomes. To avoid confusion, we have removed the related descriptions of "haplotype-resolved" throughout the text.
- **Focus on intra-accession haplotype differences.** Our primary focus was on the differences between the two haplotypes within the same accession (blue and red of the same accession). To overcome the issue of incomplete phasing, we treated the two haplotypes of a sample as a whole when comparing them to the reference genome.
- **Inter-accession haplotype comparisons.** In response to your question about comparing different haplotypes between different samples, we further analyzed the SNPs among the three well-phased accessions (CY, HZP, HB). We found that the variations have a bimodal distribution as you mentioned. The same haplotypes (same color) from different accessions were relatively small, while the variations between different haplotypes (different colors) were larger (Supplementary Table 7). This finding is consistent with our previous conclusion that the variation between moso bamboo haplotypes is greater than the variation between samples.

We hope this explanation clarifies the limitations of our haplotype assembly and the comparisons made in our study. We have revised the manuscript to ensure that these points are clearly communicated and to avoid any confusion regarding the extent of haplotype resolution.



Supplementary Fig. 26 Schematic diagrams illustrating sexual reproduction and asexual reproduction in moso bamboo. Although moso bamboo possesses the ability for both sexual (left) and asexual (right) reproduction, it predominantly relies on asexual reproduction through rhizome growth and vegetative propagation. This hypothesis, grounded in the understanding of the reproductive biology of moso bamboo, provides a plausible explanation for the observed pattern of genetic variation. Considering the rarity of somatic mutations, the lack of meiotic recombination due to the predominance of asexual reproduction, and the potential for haplotype divergence in the ancestral population, we provided a schematic diagram for understanding the predominance of inter-haplotype variations in moso bamboo.

Supplementary Table 7. Number of SNP between three haplotype-resolved accessions.

Assembly 1	Assembly 2	SNP number
CY hap1	HZP hap1	785,577
CY hap1	HZP hap2	2,792,068
CY hap2	HZP hap2	787,969
CY hap2	HZP hap1	2,794,863
CY hap1	HB hap1	433,970
CY hap1	HB hap2	3,085,151

CY hap2	HB hap2	468,651
CY hap2	HB hap1	3,137,431
HZP hap1	HB hap1	683,455
HZP hap1	HB hap2	2,907,917
HZP hap2	HB hap2	667,226
HZP hap2	HB hap1	2,883,394

* The number of SNPs was obtained directly from genome alignment.

References

1. Sang, Y. *et al.* Genomic insights into local adaptation and future climate-induced vulnerability of a keystone forest tree in East Asia. *Nat. Commun.* **13**, 6541(2022).
2. Zhao, W. *et al.* Effects of landscapes and range expansion on population structure and local adaptation. *New Phytol.* **228**, 330–343(2020).
3. Qin, P. *et al.* Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**, 3542–3558(2021).