nature portfolio

Peer Review File

A pangenome reference of wild and cultivated rice

Corresponding Author: Professor Bin Han

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

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

Please note that the attachment of Referee 3 for the first version is uploaded at the last of the TPR file.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
The manuscript by Guo et al. reported: 1) the generation of massive high quality de novo genome assemblies, which included 129 Oryza. rufipogon (wild rice) and 16 O. sativa (cultivated rice) accessions; 2) characterization of genome evolutionary patterns with these assemblies, particular on transposon elements; 3) insights and a new model of domestication origins and process for Asian cultivated rice.

Due to the domestication bottleneck, only a fraction of the genetic diversity present in wild progenitors was transferred to cultivated varieties. Wild rice thus represents a critical genetic reservoir to further improve cultivated rice. These unprecedent wild rice genome assemblies serve as invaluable resources to assist the community in identifying such missed beneficial genes/alleles. Just from this standpoint, this manuscript represents a significant milestone for crop genomics.

Meanwhile, it may be prudent to add "draft" in the title, given that only two accessions (Gla4 and W1943) were assembled at the chromosome level through a combination of PacBio HiFi and Hi-C techniques (Line 141-143), while other accessions were assembled to the contig level and subsequently anchored to pseudo-chromosomes (Line 601-602). Echoes the stages of “Draft” to “Finished” then “Complete” in the single genome era, adding “draft” follows the precedent set by “A draft human pangenome reference” (2023) and further reinforces this standard for future pangenome studies. Importantly, this simple term opens the possibility for a "Finished" and "Complete" pangenome reference in future.

The analyses and results of genome assemblies, evaluation, and genome evolution sections are sound and exciting. One potential issue is the differentially expressed alleles (DEA) part. Two out of three standards used for identifying DEA (Line 672-675), I believe, were wrong in the current written forms. For Equation 1 (Line 672) a division between two positive numbers should not yield a negative result, therefore, ‘TPMp-gene / TPMa-gene < -2’ is always false and won’t do anything for DEA identification. The correct one should be ‘TPMp-gene / TPMa-gene < ½’. Equation 3 (Line 675) also is unsuitable for identifying DEA (My guess ‘|TPMp-gene - TPMa-gene| >=10’ would be the intended equation). However, based on a lot of unexpected typos in Figures (see below), it is hard to determine if both equations were also just typed wrong or if these they were indeed used in analyses. If the latter is the case, it would be necessary to redo the analyses using the corrected equations and update the results accordingly.

Another potential issue is related to Fig. 2a, which showed japonica is so different than other groups. Since all these variants were identified by mapping to Nipponbare (a japonica variety), a question is how much the reference bias contributed to this phenomenon. Using an outgroup accession, such as O. meridionalis, as the reference probably is a potential solution.

The assembled pangenome enables extensive analyses on the genome evolution. The authors revealed that Or-IIIa has a greater mean and variance in the genome size than japonica, which was mainly attributed to the Gypsy families. Or-IIIa and japonica have different outcome directions from illegitimate recombination among Gypsy retrotransposons: japonica tends to reduce Gypsy copies while Or-IIIa tends to increase. These initial characterizations likely stimulate further investigations of mechanisms on this fascinating genome phenomenon. For instance, an early study (10.1038/s41467-018-07974-5) conducted GWAS for inferred copy number variation from short reads with cultivated rice. Given the large variation between wild and cultivated rice (such as RIRE3, Os001 and RIRE2 shown in sFig 13a) estimated from assemblies, conducting similar GWAS may lead interesting discoveries.

The question of how O. sativa was domesticated from wild progenitor is being actively debated. Early findings utilizing low-density SNP (Ref19) from the lab eventually led to the hypothesis of “a single domestication event with multiple origins”. However, an alternative hypothesis was formulated as “multiple domestications” based on either the same genomics dataset in the Ref19 (Ref23) or new dataset (Ref24). This study generated novel insights with a higher resolution of phylogenetic relationships among wild rice, such as previously characterized Or-I can be further clustered into two subtype Or-Ia and Or-Ib; proposed a new cluster of cultivated rice, intro-indica, derived from crossing between indica and aus; then intro-indica crossing with japonica derived basmati rice. A refined “single domestication event with multiple origins” model was then proposed.

My major comment pertains to the identification of selective sweep regions. These selective regions established the foundation of further characterization between wild and cultivated rice. In this analysis, only Or-IIIa and Or-1b were used as the ancestral group (Line 375 & 964). While it is acceptable that Or-Ia was excluded because of extensive gene flow between Or-Ia and indica (Line 372-373), the exclusion of other groups, Or-IIIb, Or-II, and Or-Unspecific, seems unjustified. Including these groups should increase the power of detecting the same selective sweep regions. In fact, this strategy (all wild rice groups) was used in the Ref19 from the lab. Unless the authors had other reasons, which were unspecified in this version, it would be better to use the same strategy of including all other groups (Or-Ia can be excluded) in the analysis.

It may be good to combine all the sequenced accessions in this study and Ref24, which supported another hypothesis, to check the relationships among these different subtypes of wild rice (A sFigure with a phylogenetic tree would serve the purpose). The main reason is that Ref24 separated perennial and annual wild rice into two subgroups (O. rufipogon and O. nivara, respectively), while based on Ref19, the authors probably still treated these two as one O. rufipogon (If so, it is better to reiterate this point in this manuscript). Even though the geographic distributions of these sampled wild rice accessions in both studies are highly overlapped, the varied nomenclatures challenge readers attempting to make connections and comparisons. For instance, if the Niv2 group in Ref24, progenitor of indica, is clustered together with Or-Ia in this study, it would clearly address a potential concern that the author’s ‘single domestication event’ model might be due to the fact that Niv2 was not included in the sample. This extra simple analysis might provide valuable insights into the ongoing debate.

Some minor points:
Fig 2b was cited to support the “a propensity for enrichment in intergenic repetitive regions”. However, reading this message from Fig 2b is not straightforward (the circular plot is too clouded). Why not use one chromosome as a representative, coupling with Pearson correlations between densities among these elements, to show the trend.

Fig 2g. Both x and y are related to -log(P), but the significance threshold was only plotted for y, which may generate unnecessary confusions. I would recommend extending x-axis to 0 and add the threshold for x too, which follows the practice in the cited Ref40. Furthermore, modify Line “we designated these” (Line 733) to avoid the confusion of a new category being defined in this manuscript (Ref40 actually defined this category).

Fig 4a & sFig 17. Clearly define the last panel of blue and orange whether each group is based on one representative accession or the whole group.

For sFig7, it might be more informative to reorganize the accessions based on the presence/absence of the RLK gene and the size of the syntenic segment.

Line 201 “Acknowledging the propensity of RGAs to form clusters in the genome” can be removed as the analysis results didn’t align well with this assumption.

Line 457-467 and the related Fig. 5 introduced SNP categories as japonica/indica genetic bottleneck. I would suggest finding alternative terms for such category as “bottleneck” has special meaning in domestication related literature.

Duplicated statements in Line 500-503.

Line 512 to 518. The statement about LABA1 may require clarification for a better comprehension.

Line 540 indicated some new functional polymorphisms were identified in known domestication genes. It would greatly enhance clarity to depict the locations and consequences for the newly identified polymorphisms for genes in sFig 21. For example, is the A:T mutation in OsC1 located in CDS and altered an amino acid?

Check Line 750-751 is the intended statement.

Although the main text is well written and clean, a lot of unexpected typos occur in figure legends and labels: “Sourth” in most of Figures and sFigures when geographic regions were mentioned; “Hetrozygous allele” in Fig 1b&f, sFig 4, & sFig 24; “domsticattion gene” for Fig 4e; “State of demoestication” in Fig 4c, sFig 18 & sFig 22; “misiing allele” in Fig 5c; “phynotype” in Fig 5d; “maped” in sFig 10; “acitivity” and “oxidorreductase” in sFig 13; “America North” in sFig 15; “WastAsia” in sFig 18 & sFig 22; “Accesstions” in sFig 21; “MIT gene” in sFig 24; “Comparation” in sTable 3 & sTable 4; “out study” in sTable 8; “Grap-pangenome” in sTable 9. These typos did not impact the analyses and results but impair the manuscript quality.

Referee #2

(Remarks to the Author)

The manuscript presents a comprehensive analysis of a pangenome reference of 149 wild and cultivated rice, providing valuable insights into rice genetics and evolution. The investigation into the origin and domestication of Asian cultivated rice, including the identification of introgression events and selection patterns, adds significant value to the field. The manuscript's findings are likely to interest a wide audience, ranging from geneticists and plant biologists to researchers studying crop domestication and evolution.

However, I have the following major concerns:

1: The manuscript provides an important resource for the discovery of elite natural variations, and studies on the functional genomics, and evolutionary biology. However, the results in this version do not sufficiently highlight the importance of this resource. For example, in the identification of potential elite alleles present in wild species, the authors briefly examined the resistance gene analogs using this remarkable genetic resource but did not draw substantive conclusions or provide detailed examples, leaving this section feeling empty.

2: Considering that many functional genomics research teams worldwide are not familiar with big data analysis, it would significantly enhance the paper's citations and impact if the manuscript could provide a convenient entry point for accessing related data for the resource, such as genome sequence variation queries, alignment, distribution ranges, or differentiation sites between indica and japonica rice.

3: Regarding heterozygous loci, the heterozygosity rate in wild rice species is high. However, I did not see relevant information on the heterozygosity rates of wild species. Although the authors paid attention to heterozygous loci during genome assemblies, this consideration was not evident in the subsequent analyses, such as genetic variation, gene flow, population structure and introgression, selective sweep, and the genomic variation information used in the analysis of rice origin.

4: I suggest that the author reconsider the importance of each section and remove or shorten some standard descriptions or contents without new findings. For example, the graph-based pan-genome did not provide any substantive content. Graph-based pan-genome analysis is only valuable if it yields new insights, so I suggest removing it. The content regarding gene-based pan-genomes and transposable elements mostly consists of conventional data descriptions and should be substantially shortened, as they did not yield new conclusions or exciting content.

5: Many descriptions in the results did not explicitly mention the datasets used. For example, lines 441-442: "Our analysis identified 442,855,122 SNPs and 13,853 PAVs differentiating indica from japonica." While I checked the methods section, I found that the authors used 90 indica and 36 japonica cultivars. This should be briefly mentioned in the results because different datasets may yield different results and conclusions. Therefore, the authors should also explain why they chose these datasets for each analysis when a new dataset was used.

6: The readability of the paper needs improvement. The paper had a number of long sentences and paragraphs. For example, the first paragraph of "Extensive variation and graph-based pan-genome" was too long; readers may find it challenging to read such a lengthy paragraph.

Referee #3

(Remarks to the Author)
Comments on the MS “A pangenome reference of wild and cultivated rice” by Gu et al.
This MS reported an effort to construct a pan-genome of Oryza rufipogon (OR), the presumed progenitor of Asian cultivated rice (Oryza sativa, or OS). This study had two key results. The first was the construction of the OR pan-genome comprising a total of 69,531 genes with 13,728 wild rice specific genes based on chromosome-level assemblies of 129 genetically diverse OR accessions and 16 diverse OS varieties. The second was to provide “strong” evidence in supporting their early hypothesis of single origin of OS, i.e. all Asian cultivated rice varieties today was originated from a single domestication event, based on simple population genetic analyses using their OR data combining with 128 OS genome data in the public domain. Although this study generated useful genomic information of the OR population interesting to the public rice scientific community, the total amount, quality and novelty of the reported work were far from the high standard required for publication in Nature. In particular, their conclusions on the OR pan genome and the single origin (domestication) hypothesis of OS were apparently based on incorrect comparisons and misinterpretation of the results. Thus, the MS is not acceptable for publication in Nature. Specific comments are given as follows:
1. The average 24.2x genome sequencing coverage for the 133 OR genomes was insufficient to ensure the claimed high-quality assemblies of most OR genomes, given the fact that most OR genomes contain highly heterozygous loci (58.9%). The relatively low quality of the genome assemblies was also evidenced by the unreasonably high heterozygous loci (34.3%) of the 9 OS genomes (Supplementary table 5), which are expected to be low given the self-pollinated nature of the OS accessions, including Nipponbare.
2. Because pan genome is a property of population or species, the comparison between the OR and OS genomes in pan-genome parameters of this study was invalid given the extremely small population size (16) of the OS samples. For example, the authors constructed the OR pan genome consisting of 69,531 genes, which was much bigger than the OS pan genome from the 16 OS accessions. However, Zhang et al. (2022) reported that the OS pan genome obtained from 105 OS accessions consists of 70,624 protein-coding genes, which is equivalent to the OR pan-genome reported in this study. Regarding the 13,728 wild rice specific genes, the reviewer did not see any evidence how the authors made the comparison between the OR pan-genome with the previously reported OS pan-genome to determine the 13,728 wild rice specific genes.

According to Zhang et al. (2022), there are, on average, ~0.2% private genes in each OS accession. If so, most of the 13,728 wild rice specific genes could be well attributable to the small sample size of the OS accessions used in this study. This was obvious in their statement “We observed a greater abundance and diversity of RGAs in O. rufipogon, averaging 1,710, compared to around 1,664 RGAs in the 16 OS accessions”. This difference by 46 (~2.7%) in the number of RGAs, is insignificant considering the possible large variation in RGA number among different accessions within OS or among different accessions within an OS population such as indica or japonica.

3. In addition to its economic importance as the staple food feeding most Asian people, OS is well-known for its wide geographical distribution, subspecific and population differentiation, and extremely rich genetic/genomic diversity within populations. Thus, the origin (domestication) of OS has been an issue of long debate, with two major hypotheses: the multiple origins of OS and single origin of OS. The fundamental difference between the two primary hypotheses is that OS was domesticated in different geographic locations at different times in the former hypothesis, while domestication of OS was presumed to have occurred only once in one location in the latter. In the reviewer’s opinion, any conclusion on this issue has to reconcile with all evidence from genetic/genomic studies and archeological discoveries. The authors were apparently trying to obtain evidence in supporting the hypothesis of single origin, i.e. “OS was first domesticated as proto-japonica cultivars from OR population III (Or-III) in China and indica cultivars were subsequently domesticated as proto-japonica cultivars, which were then spread southward and eastwards in Asia and crossed with local OR population group I (Or-I)”. Unfortunately, the authors did not perform sufficient population genetic analyses to support or reject any of the hypotheses, and their conclusions were based on the misinterpretation of their results with the following three major problems.

3.1 First, one most obvious expectation from their hypothesis of single origin was a severe genetic bottleneck and dramatic reduced genetic diversity genomewide in OS populations. With the inclusion of additional 128 OS accessions from the public domain in their data analyses, both OR and OS showed clear population differentiation of geographic correspondence (Fig. 3a). The phylogenetic analyses resolved the OR accessions into three well-differentiated populations, OrIII classified with OS japonica accessions, OrI with OS indica accessions and OrII in between, consistent with three previous reports based on SNPs from a low-coverage (~1x) sequencing of the much large OR and OS populations containing all OR and OS accessions used in this study. Compared to the OR populations, the OS populations (indica, japonica, aus and basmati) showed greater population differentiation but slightly reduced within population diversity (Fig. 3c). These results were consistent with virtually all previous reports of similar studies. The observed reduction of genetic diversity in japonica, when compared with its presumed ancestor, OrIIIa, was ~57%, indicating presence of a significant level of genetic diversity within population japonica, consistent with that reported by Wang et al. (2022). More severe genetic bottlenecks would have been expected for populations indica, aus and basmati, if they were derived from a single domestication event of japonica. Thus, there is no evidence for severe genetic bottlenecks (greatly reduced diversity genomewide) observed in all OS populations when compared with their most closely related OR populations of the same geographic origin.

3.2 Secondly, the authors’ inferences on the separation history of different OR and OS populations (Fig. 4) were largely reached based on their analytic results of the OR and OS genomic sequencing data using the Multiple Markovian Coalescent (MSMC2) method. However, according to Mather et al. (2020), the MSMC2 method developed to infer past population sizes and divergence time of related lineages has two assumptions, the neutrality of genome sequence variation and random mating of the studied population. When the assumptions are violated, the parameters of the demographic history inference obtained by MSMC2 could be incorrect and misleading. In particular, when MSMC2 is used for highly structured populations such as OR and OS, the results are likely to be extremely misleading. Moreover, the impact of population structure on the results of MSMC2 cannot be assessed directly from the observed coalescence times. “Concerningly, this problem appears to grow worse as larger numbers of genomes are used for MSMC2, so the problem cannot necessarily be overcome by adding more data” (Mather et al. 2020). Now, both the OR and OS accessions used in this study are highly structured. OS is known to be a highly self-pollinated species and the OR accessions used in this study are not random-mating populations (though OR accessions have higher outcrossing rates than OS accessions). This was clearly seen from their well-differentiated population structure. Most importantly, the OR and OS accessions from different geographic origins are known to adapt to a diverse range of environments, each carrying large numbers of genes/alleles that had gone through various kinds of selection during evolution. Thus, both underlying assumptions of the MSMC2 method are severely violated in this study. Thus, the authors are required to explain how this had affected their results and conclusions.

3.3 Thirdly, it is well-known that results from population genetic analyses are extremely sensitive to sampling, and the sampling problem in this study were reflected in at least two aspects: (1) the sequenced OR and OS accessions do not cover the whole spectrum of OR and OS geographic distributions, because the sampled OR accessions were collected in relatively recent time and most and real ‘ancestor’ OR populations got lost as their habitats were largely occupied by OS varieties over time worldwide; and (2) the few selected “domestication genes” for detailed analyses by the authors only account for a very small portion of rice genes that had gone through varied types of selection during the process of domestication. In the former case, for example, the OR and OS accessions used in this study failed to include those (japonica landraces and OrIII accessions) from the foothills of the Himalayas (North India, Nepal and North Myanmar). In the latter cases, the sampled 11 “domestication-genes” used in gene PAV analysis (Fig. 4d) were too few to reveal the whole history of domestication. According to a recent report using much larger population sizes of OR and OS populations, Wang et al. (2022) reported a total of 14,202 and 5,163 domestication selected genes in OS subspecies japonica and indica, respectively, with 3,115 shared between the two subspecies. Of these, only a very small proportion of (3.4 % in japonica and 1.9 % in indica) of the domestication selected genes were attributable to genetic bottlenecks that either showed uniform low polymorphism or no variation in gene CDS regions. They found that selection in japonica and indica during domestication tended to act on different genes in the same pathways, on different functional alleles of the same loci, or on different members of the same gene families, indicating that japonicas and indicas may have compensatory regulation and functional genetic networks affecting the same phenotypes for their adaptations to different environments and their responses to artificial selection during the process of domestication, indicating that genetic bottleneck played little role during

domestication of both japonica and indica. These results led them to the proposition of the multiple origins (domestication) of OS. Unfortunately, the authors of this study failed to perform appropriate analyses which would allow them to test the major hypotheses regarding the OS origin.
4. In addition to above comments, the authors have to provide answers to the following two questions to support their hypothesis of single origin of OS:
4.1. First, it is well-known that the subspecific (japonica-indica) differentiation of OS involves a large portion of their genome constitutions, as indicated in the MS "We identified a total of 855,121 indica–japonica differentiated SNPs (Supplementary Table 16) based on a very stringent standard of 0.9 frequency difference. We also applied the same standard to identify 13,853 indica-japonica differentiated PAVs between 26 indica cultivars and 13 japonica cultivars". These results were consistent with the results of the 3,000 rice genome project (Wang et al. 2018; Zhang et al. 2021) that the indica–japonica subspecific differentiation represents the maximum direction of within OS diversity, and indica and japonica each contains many (hundreds of) subspecific genes/gene families, and thousands of subspecific alleles (functional haplotypes) at >13,000 gene loci (Zhang et al. 2021). It is also known for the presence of high levels of reproductive barriers (hybrid sterility and hybrid breakdown) between japonica and indica. The mountainous areas along the Himalayas including Southwest China are known as the center of diversity for OS, where various ecotypes of OS, ranging from typical indica to temperate japonica and upland (japonica) ecotype, are distributed across different longitudes and show tremendous phenotypic and genetic diversity. However, the recorded history of OS cultivation in Southwest China was <3,000 years. Then, the authors have to explain what genetic mechanism(s) could have generated such high levels of genetic diversity, population differentiation and reproductive barriers in >3,000 years if all OS accessions were originated from single ancestor not adapted to the diverse environments along the Himalayas. The authors proposed that the hybridization and introgression from population Orl into "ancient japonica" to explain how indica and aus were derived (Fig. 4e). However, this explanation is implausible because of the severe hybrid sterility and hybrid breakdown between japonica and Orla would have blocked the massive gene flow between them. The alternative explanation would be that indica and aus accessions of South Asia were more likely to have been directly domesticated from the local Orla and Or1b at ancient times independent from the domestication events in China, while the observed introgression between indica and Osla and between aus and Oslb interpretation was more likely to have occurred after domestication. Thus, the authors are required to provide evidence to reject the alternative explanation.
4.2 Secondly, the earliest archeological evidence in China (lower Yangtze River) and India (lower Ganges River) has traced the cultivation of OS back to ~12,000 years and ~9,000 years ago, respectively (Fuller et al. 2010; Gross and Zhao 2014; Ma et al. 2016; Zuo et al. 2017). Thus, the authors have to provide evidence for the ancient human migration between the two countries that occurred at least 9,000 years ago if their hypothesis of single origin for OS holds true.

In summary, the results provided by the authors on their single origin (domestication) hypothesis of OS does not reconcile with all evidence required to support their hypothesis, which led the reviewer to the conclusion that the MS is not suitable for publication in Nature based on above comments. Nevertheless, the reviewer is hoping that these comments would be helpful for the authors in their future efforts to revise their data analyses and interpretation of results.

Referee #4

(Remarks to the Author)
In this manuscript, Professor Han and colleagues provide an updated pangenome of wild and cultivated rice. The authors added 145 chromosome-level assemblies from a 129 genetically diverse Oryza rufipogon and 16 diverse O. sativa varieties. They integrated this data with previously published data to provide an important new pan-genome resource, for example, they identified resistance genes in wild rice not found in cultivated rice. Overall, the study is well-executed, and I have no technical objections to the methods used or the results presented.

Previous rice pan-genomes or super pan-genomes focused on cultivated varieties (papers cited include Zhao et al. 2018, Qin et al. 2021, Shang et al. 2022, Zhang et al. 2022, Zhou et al. 2023, Huang et al. 2021, Wu et al. 2023; however missing citation for Wang J et al. 2023 Genome Biology 24: article 19). The authors claim that an important scholarly gap is that previous rice pan-genomes focused on cultivated rice. However, looking at the literature cited were some O. rufipogon and other rice species included in previous pan-genomes (e.g., Zhao et al. 2018, Zhou et al. 2023, Wu et al. 2023). As a reviewer, the first question that I have is: What significant new insights were gained by adding additional O. rufipogon genomes?

There is an ongoing debate amongst scholars about the multiple or single origins of domesticated rice (e.g., Civan et al. 2015, Huang and Han 2015, Choi and Purugganan 2018, Gutaker et al. 2020, Xu and Sun 2021, Jing et al. 2023). Not being a specialist in rice domestication, I read Jing et al. 2023 (Multiple domestications of Asian rice. Nature Plants 9, 1221-1235). I found that paper to be an excellent overview of the issues surrounding how complicated the issues are in this debate. For example, there is no consensus as to whether the Asian rice progenitors (O. rufipogon and O. nivara) were two distinct species and whether these species are geographically and genetically subdivided. Or how inferring the number of origins or domestication "events" is deeply complicated by introgression among rice subpopulations. So, my second question that I have is: What are the significant new insights gained on the origin(s) or domestication of rice not previously reported?

In summary, the manuscript is a well-executed assembly of an updated rice pan-genome that should be an important resource for rice biology. However, as written, the insights provided seem incremental relative to previously published work (e.g., Jing et al. 2023). Perhaps what could be emphasized in a revision is how adding the new data (e.g., "outgroups") dramatically changes how we understand rice origins and domestication.

Version 2:

Reviewer comments:

Referee #1

(Remarks to the Author)
I am impressed by the extra data, analyses, and reasoning of this revision. Overall, the manuscript is significantly improved, but a few minor issues still remain.

Hi-C reads were added for 30 accessions to verify the quality of PacBio assemblies. sTable 5 and sFig. 3 showed the congruence between assemblies by PacBio only and PacBio + Hi-C, demonstrating the high quality of these assemblies. The current title is reasonable. Meanwhile, sFig. 3 showed that most chromosome pairs had an overhang at the right end, which means the discrepancies consistently occurred in the end of a chromosome. Please confirm this is not the artifact of plotting.

The authors further compiled a 549-accession new dataset by including resequencing data from Jing et al. (Ref 25). The results from this new dataset not only addressed the key concern of sampling issues but also connected both studies: the subgroups defined by two studies were aligned (such as Or-IIIb ~ Niv1 and Or-II ~ Ruf2 + Ruf1b). The discrepancy between this study and Jing et al. centered on the definitions of domestication and adaptation/ improvement. As stated in the manuscript (Line 549-553), this study represented the viewpoint of domestication as a process only involving selecting several obvious morphological traits and improvement/adaptation as a subsequence process, while another viewpoint is improvement/adaptation being a part of domestication. From the authors' point of view, the proposed model was supported by the results. Meanwhile, I have a related question based on Line 556-559, which mentions "accurate haplotype analysis of LABA1 in this study". Was this accuracy due to the graph-based pangenome (indicated in Response), instead of a single reference genome, was used for calling variants?

I appreciate the effort of conducting GWAS to identify potential loci/genes associated with the copy number variation of Gypsy families, which are main contributors to the genome evolution between wild and domesticated rice. Although some families showed nice GWAS signals, I agreed that exploring the mechanisms should be followed up in future studies.

In addressing the comment for Fig. 2a about the reference bias, the author conducted comprehensive analyses and confirmed the result in Fig. 2a was biased but other evolutionary analyses were unaffected regardless which reference genome was used. However, despite confirming the bias, the original Fig. 2a and the corresponding statements were retained in this revision. In my opinion, these parts should be removed from manuscripts.

About the third DEA equation, the manuscript indicated three standards (Line 718), implying that any gene that meets one of these equations is DEA, but the Response indicated Equation 3 is a filter, or an extra condition for Equation 1 and/or 2. The equation 3 is good as a filter, but not a good standard for classifying DEA (for instance, p-gene reads = a-gene reads = 5). Please confirm which is the case and revise the manuscript accordingly.

The new sFig. 25 compared Pi and Fst among different datasets to support that the applied standard of Or-IIIa + Or-Ib vs Os is reasonable. However, panel a and b are very similar. Please confirm that removing 13 Or-Ia accessions didn't change the estimates that much.

Fig. 3b and Fig. 4c showed the results from Admixture with different K groups. A common practice is separating each group by, for example, a vertical line. However, this was not followed for most K groups. Furthermore, Line 383-385 cited Fig. 3b about "Intro-indica", but the "Intro-indica" cluster was not labeled in Fig. 3b.

A typo of "Transaction" in sFig. 15.

Referee #2

(Remarks to the Author)
I am satisfied with the responses and revisions to my concerns. However, I still have several issues that need addressing:

1: I would like clarification on the criteria used to define a novel sequence, which resulted in a 3.87G sequence being absent from the Nipponbare reference genome. This criterion was not detailed in the methods section.

2: Regarding the domestication debate, the author suggested that the previous study utilized a gene set that included traits related to post-domestication (or genetic improvement), potentially leading to the conclusion of multiple domestications. The authors also emphasized that the initial stages of rice domestication primarily focused on key agronomic traits, such as plant architecture, hull color, seed shattering, awn length, pericarp color, and grain size.

This raises questions about the criteria for determining domestication traits. For instance, why are hull color and pericarp

color considered key domestication syndromes at the initial stage? It's difficult to imagine why people would prioritize hull color and pericarp color during the initial stages of rice domestication. While grain size, which is often associated with genetic improvement, is not?

Additionally, the authors concluded that O. rufipogon and O. nivara are not genetically or geographically distinct enough to warrant classification as separate species. However, I could not find information on the genetic and geographical differences between O. nivara and O. rufipogon in the results section. Could you please clarify or add this information?

3:Most of the first paragraph in the discussion repeats content from the results or introduction. It would be more effective to shorten this paragraph and merge it with the second paragraph to create a more cohesive discussion.

Referee #3

(Remarks to the Author)
Comments on the revised MS of 2024-03-04666B

1. The authors did not understand my question on the invalid comparison between the pan genomes of OS and OR. Although the sampled 16 OS accessions did not affect the pan genome estimation of the sampled OR accessions, the authors should use the pan genome of OS accessions of equivalent sample size and similar geographic distribution for their comparison and claim. In other words, the authors should compare all OR pan genes they discovered with all previously published OS pan genes in order to be sure those 13,728 wild rice "novel" genes are indeed OR specific, even though they are absent in the 16 sampled OS accessions. In the authors' response "Following your suggestion, we have compared our identified wild rice-specific genes with three published rice pan-genomes (Qin et al. 2021; Zhang et al. 2022; Shang et al. 2022) (Supplementary Table 10 in the revised manuscript). As expected, the number of genes exclusive to wild rice is different in this broader comparison. However, the results demonstrate that most of the wild rice-specific genes we identified (63%-82%) remain unique, even when compared against a larger set of cultivated rice genomes. These are truly important and novel findings." If so, the authors' claim for the discovering "13,728 wild rice-specific genes in their abstract does not hold true because they should include the results into the revised MS, stating only at most ~63% of the 13,728 genes are likely OR specific. Again, when the authors did not obtain an accurate estimate on the number and diversity of RGAs in OS (given the 16 OS accessions they used), their claim that OR population has more PGAs and higher diversity at PGAs than OS remains invalid, until an OS population of equivalent size and geographic distribution is used in their analyses.
2. The reviewer does not understand how the authors reached the speculation "These cultivars have been identified to be domesticated through crossing the proto-japonica type cultivar (with multi-domestication traits, such as non-shattering seed habit, from prostrate to erect growth, transition from black to straw-white seed hull, etc) with diverse wild rice O. rufipogon populations for adapting various environments (Gutaker et al. 2020)", given the fact that japonica cultivars or japonica type of ORs are known to have very low outcrossing rates, and "the modern breeding process" involving hybridization and artificial selection of rice has a short history < 130 years.

3. Unfortunately, the authors did not answer to my question regarding how the serious violation of the assumptions "the neutrality and random mating" for software MSMC2 influences their conclusions. In fact, the authors should use computer simulation to address how the breeding systems and different types of selection affect their results. They did not even try to test the alternative hypothesis of the multiple origins using their data.

4. In the authors' response, "I would like to clarify that our model does not propose a single ancestor for all types of Asian cultivated rice. "Instead, we suggest that different groups of O. sativa just shared a common domestication event, with their respective ancestors diverging and adapting to distinct ecological environment before this event." This resulted in significant genetic differentiation and diversity among these groups of O. rufipogon. Subsequent artificial selection and genetic improvements further enhanced this diversity, tailoring rice varieties to meet diverse farming and culinary needs (Jiang et al. 2022). Therefore, indica, through shares key domestication loci with japonica, shows larger diversity and higher genetic differentiation from japonica. The domestication process of wild to cultivated rice was gradual; japonica experienced intense artificial selection and significant genetic bottlenecks, whereas indica could have rapidly incorporated domestication genes through gene flow, preserving its diversity. This suggests that indica did not experience a separate, de novo domestication, challenging the multiple domestication hypothesis." To the reviewer, the authors' arguments are in conflict to their conclusion "We also provided strong evidence to support the initial domestication of all Asian cultivated rice occurred only once" because this conclusion would imply all Asian cultivated rice has a single japonica type of OR ancestor. Unfortunately, the authors did not seem to understand the most recent multi-origin hypothesis of OS which has two key elements: 1) the Asian cultivated rice has multiple ancestors, ancient japonica types in different places of China, ancient japonica types in South and southeast Asia, ancient indica types in China, India and Southeast Asia; 2) the domestication events may have occurred multiple times in different places of China, South and Southeast Asia at different times, while the OR populations had become well differentiated into different populations in different geographic regions (clearly indicated from all previous and the present studies).

5. The authors' response to the reviewer's question on the consistency between the genetic/genomic evidence and the archaeological evidence was extremely weak. The earliest histories of rice cultivation in China and India were completely independent according to the archaeological evidence, and both could be traced back to the time that far preceded to the earliest recorded interaction event(s) between the two ancient civilizations. Thus, to the reviewer, all these arguments provided no evidence regarding direct relationships between domesticated rice types (indica, Aus, Basmati and japonica) in

South Asia and those domesticated japonica in China, nor any direct relationship between the japonicas and indicas in China.

In summary, the reviewer is not convinced by the authors' responses and arguments to the questions. Given the presence of large numbers of previous studies on rice population structure and genomic variation. The population size of 129 OR accessions does not warrant the high quality of the OR pan genome and their comparison between the pan genomes of OS and OR remains invalid. In particular, their selection of "domestication genes" and their interpretations on the analyzed results were inappropriate and highly biased. Their selection of references is highly biased in favor of their conclusions. They failed to test the alternative hypothesis of the multiple origins (domestications) of rice using their data. Taking together, the reviewer feel strongly that the MS does not meet the high standard for publication in Nature.

Referee #4

(Remarks to the Author)
This is a revised manuscript by Professor Han and colleagues. I (Reviewer 4) had two simple questions about the original manuscript:
(1) What significant new insights were gained by adding additional O. rufipogon genomes?
(2) What are the significant new insights gained on the origin(s) or domestication of rice not previously reported?

In their rebuttal to question 1, the authors replied that they constructed high-quality pan-genomes, created important genetic resources, and enhanced evolutionary studies. Specifically, they "solidified our 'single domestication event' model. The precise variations dataset have enabled detailed classifications, revealed the genetic evolutionary path of new subpopulations, and facilitated the analysis of the origins of indica-japonica differentiation."

In their rebuttal to question 2, the authors replied that they provided a new understanding of the phylogenetic relationships of wild rice, identified the ancestors of each cultivated rice group, refined the evolutionary paths of two O. sativa subgroups, and reinforced the hypothesis of a single domestication event for Asian cultivated rice.

I'm still not sure where on the spectrum these two replies are between an incremental improvement and an advance worthy of publication in Nature, so I bow to the wisdom of other reviewers who had methodological issues with the manuscript or who are experts in rice domestication.

Version 3:

Reviewer comments:

Referee #1

(Remarks to the Author)
My only comment is about keeping Fig. 2a (and the new sFig. 14a). Variation of polymorphism counts across different subgroups doesn't provide significant insights. And it suffers strong reference bias. Unless I missed something, removing Fig. 2a, sFig. 14a, and Line 273-278 won't impact the subsequent analyses.

In terms of statistic test, the error bars in Fig. 2d and sFig. 34 were not defined.

sFig. 15, the p-value of inversions in genes was 1.21e-02, which didn't pass the threshold of 0.01. Maybe it is better to annotate this was not significant while others were significant.

Referee #2

(Remarks to the Author)
I am satisfied with the responses and revisions to my concerns, and don't have any more concerns.

Referee #3

(Remarks to the Author)
Remarks: In summary, the reviewer is not convinced by the authors'responses and arguments to the questions. Given the presence of large numbers of previous studies on rice population structure and genomic variation. The population size of 129 OR accessions does not warrant the high quality of the OR pan genome and their comparison between the pan genomes of OS and OR remains invalid. In particular, their selection of "domestication genes" and their interpretations on the analyzed results were inappropriate and highly biased. Their selection of references is highly biased in favor of their conclusions. They failed to test the alternative hypothesis of the multiple origins (domestications) of rice using their data. Taking together, the reviewer feel strongly that the MS does not meet the high standard for publication in Nature.
Response: We would like to say that our study is fully based on the analysis of original genomic data, which were generated from diverse accessions carefully collected by Japanese and Chinese scientists. We hope that we have fully addressed these concerns.

The reviewer sees the considerable efforts from the authors in addressing my comments. However, the reviewer is still not convinced by their conclusion on the model of “multiple-origin/single domestication event”, in particular, on their conclusion of “single domestication event” reached by their reanalyzed results and by their ways in data analyses based on the following comments:

1. In the data analyses, the authors used “only the progenitors (Or-IIIa and Or-Ib) of other cultivated rice groups as the ancestral groups of japonica and indica for searching “domestication” segments across the genomes of the two subspecies. This was a valid comparison. The result “By genome-wide comparison of genetic diversity ratios (πw/πc) and Fst values between ancestral groups and cultivated groups, we identified 50 loci with strong selection signals, covering a total region of 12.35 Mb (Methods). Additionally, beyond SNP-based analysis, we identified domestication-related Presence/Absence Variations (domPAVs) using Fisher's exact tests to compare PAV frequencies between both populations.” This method effectively captured most well-known domestication genes in the initial stages of rice domestication within identified selective sweep peaks or domPAVs hotspots, including ……”.
However, the reported results did not seem sufficient to support their conclusion of “single domestication event” unless they could provide the following pieces of evidence: (1) Most, if not all, of the “domestication” loci identified through the genome-wide comparison of genetic diversity and Fst values between Or-IIIa and japonica are the same ones as those identified from the comparison between Or-Ib and indica; (2) The favored alleles (such as alleles for non-shattering at sh4, PROG1, etc) at most, if not all, domestication loci in japonica are the same ones as those in indica and aus; (3) Gene PAVs pattern in japonica and indica to show evidence for the absence of domestication activities (selection for less shattering, lodging, greater grain sizes, etc.) in the ‘pre-domesticated but cultivated (indica/aus) rice’ in South Asia.
2. To the reviewer, the reported results suffered greatly from insufficient population sizes for all rice populations and the authors did not show all results in their two genome-wide comparisons between Or-IIIa and japonica, and between Or-Ia and indica, reflected in the following two aspects: (1) the sampled OS accessions did not cover the whole spectrum of OS population structure which included four subpopulations in Indica, three subpopulations within Japonica, plus Aus and Bas of clear and distinct geographic origins (Wang et al. 2018); (2) their choices of 11 ‘domestication’ loci showing selective sweeps in only a few common Or-IIIa segments for deep analyses were highly biased and results of the remaining 39 identified domestication loci were presented.

Apparently, the authors did not read a recently reported study in J. Advanced Research titled “Selective and comparative genome architecture of Asian cultivated rice (Oryza sativa L.) attributed to domestication and modern breeding” by Wang et al. (2022). In the same comparisons between 456 OR genomes and 504 OS landrace genomes using the same methods, Wang et al. (2022) reported the identification of 624 (126.4 Mb) and 318 (44.5 Mb) domestication blocks (segments) japonica landraces and indica landraces, respectively, with only 179 (24.8 Mb) shared between japonica and indica. In these domestication segments, they found a total of 14,202 and 5,163 domestication selected genes in japonica and indica, respectively, with 3,115 domestication selected genes shared between the two subspecies. They also found that natural selection during the domestication process of japonica was much stronger than that of indica. They found that the detected domestication gene in japonica and indica shared many common gene families, but different members of the same gene families were selected in japonica and indica during domestication. Thus, their results indicated that evolution resulted from convergent selection was actually acting on different genes/alleles in populations japonica and indica, suggesting the domestication events in japonica and indica were largely independent from each other.

Minor comments
Several minor comments are listed as follows:
1. When compared the wild rice pangenome with the 129 OS pangenome reported previously, the authors indeed found 7,592 wild rice specific genes. However, in the Abstract, the authors still claimed that they found 13,728 wild rice specific genes, which is wrong. The reviewer is also wondering how many genes in their discovered genes are OS specific (present in the cultivated rice only, and absent in the wild rice accessions)?
2. Response Fig. 12 was incorrect because the earliest cultivation of rice and its domestication should have occurred simultaneously in the same process that had taken hundreds or even thousands of years in different places of Asia. Thus, the reviewer is wondering how it could have happened in South Asia where “pre-domesticated rice” was cultivated for several thousands of years without going through any artificial/natural selection for domestication traits, as argued by the authors or Fuller et al., (2010). If so, the authors should provide solid evidence to show why those “pre-domesticated” indica rice accessions were cultivated for ~4,000 – 5,000 years in South Asia without going through any artificial/natural selection for those domestication traits. Also, the authors have to clarify how they determine when the domestication process in japonica or indica was completed, or simply by citing previous archeological reports. To the reviewer, there is no universal unambiguous morphological domestication traits in rice, given the fact that indica and japonica are distributed in very different ecological environments because even today, the modern japonica and indica varieties differ greatly in these “domestication” traits!!
In summary, the reviewer believes that inferences on the origin of Cultivated rice have to be built on large numbers of sampled accessions of representative geographic origins because of the well-known population differentiation within the species as well as within the two major subspecies of rice and a significant amount of genetic variation exists among populations of different geographic origins. This is particularly true for the self-pollinated plant species like rice. Apparently, the sample sizes and their geographic representativeness of this study were insufficient for them reach their conclusions. Genetically, virtually all domestication traits are complex ones controlled by large numbers of genes/loci. The reported domestication genes only account for a small portion of genes/loci involved. In fact, all genes showing selective sweeps in the new comparisons by the authors should be deeply analyzed in order to support or reject their conclusion on the model of multiple-origin/single domestication event. Taking together, the reviewer is still not convinced by the authors’ responses and arguments to the questions.

Point-by-point responses to Referees' comments

Authors' Response to Referee #1

Referee #1 (Remarks to the Author):

Remarks: The manuscript by Guo et al. reported: 1) the generation of massive high quality de novo genome assemblies, which included 129 *Oryza. rufipogon* (wild rice) and 16 *O. sativa* (cultivated rice) accessions; 2) characterization of genome evolutionary patterns with these assemblies, particular on transposon elements; 3) insights and a new model of domestication origins and process for Asian cultivated rice.

Due to the domestication bottleneck, only a fraction of the genetic diversity present in wild progenitors was transferred to cultivated varieties. Wild rice thus represents a critical genetic reservoir to further improve cultivated rice. These unprecedent wild rice genome assemblies serve as invaluable resources to assist the community in identifying such missed beneficial genes/alleles. Just from this standpoint, this manuscript represents a significant milestone for crop genomics.

Response: Thank you very much for the comments.

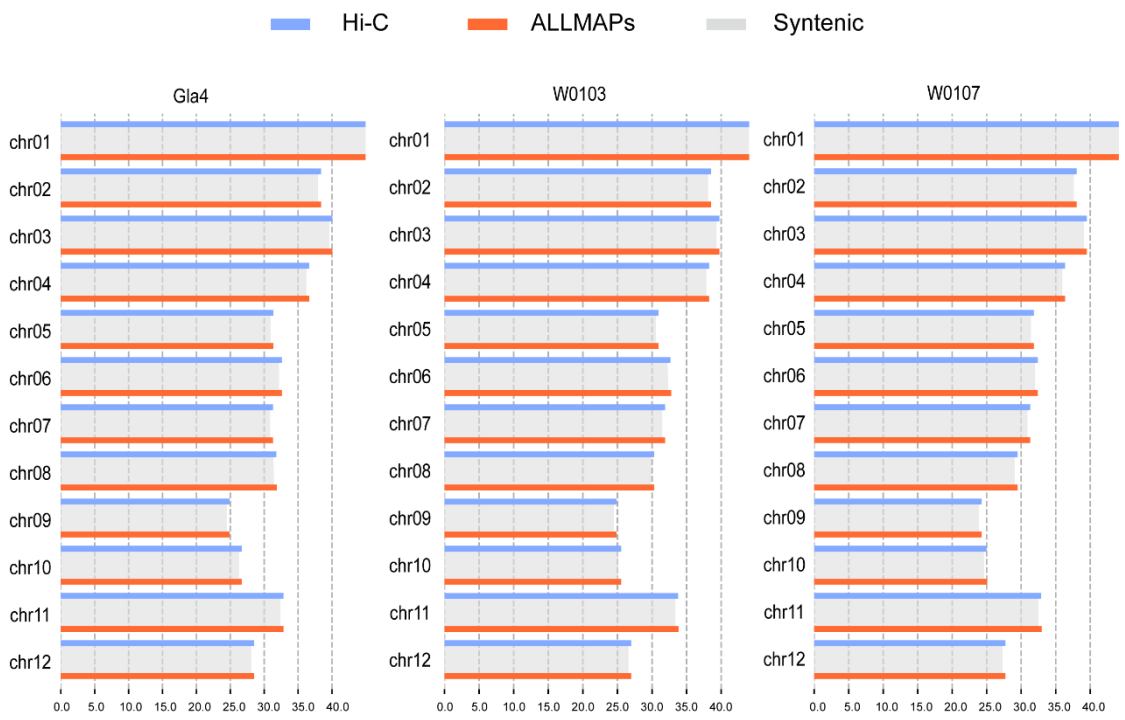
Comment (1): Meanwhile, it may be prudent to add "draft" in the title, given that only two accessions (Gla4 and W1943) were assembled at the chromosome level through a combination of PacBio HiFi and Hi-C techniques (Line 141-143), while other accessions were assembled to the contig level and subsequently anchored to pseudo-chromosomes (Line 601-602). Echoes the stages of "Draft" to "Finished" then "Complete" in the single genome era, adding "draft" follows the precedent set by "A draft human pangenome reference" (2023) and further reinforces this standard for future pangenome studies. Importantly, this simple term opens the possibility for a "Finished" and "Complete" pangenome reference in future.

Response: Thanks for the comments. To evaluate the quality of the pangenome sequences, we have further done Hi-C genome sequencing for 30 additional accessions, all of which have been successfully assembled to the chromosomal level (**listed in the table below**). Comparative analyses between these Hi-C assemblies and those generated by the ALLMAPs method mentioned in the manuscript show very few discrepancies (**Response Fig. 1**). These results suggest that our rice pseudo-chromosomes assemblies achieve a high level of accuracy and completeness comparable to established chromosome-level assemblies. These results have been included in **line 144-157** and **Supplementary Table 5** and **Supplementary Fig. 3** of the revised manuscript.

Response Table | Summary of chromosome assemble from 30 Hi-C sequencing samples

	Chromosome length (Hi-C)	anchor rate(Hi-C)	Chromosome length (ALLMAPs)	anchor rate (ALLMAPs)	syntenic length (Hi-C/ALLMAPs)	non-syntenic rate
W0179	404427253	98.38%	404427253	98.38%	404427253	0.00000%
W1677	393639671	98.75%	393848965	98.80%	393639671	0.00000%
W1919	402134563	98.59%	402345563	98.64%	402134563	0.00000%
W2275	404774783	97.49%	404774783	97.49%	404774783	0.00000%
Gla4	395345399	99.00%	395421328	99.01%	395345399	0.00000%
W1787	395506583	98.94%	395506483	98.94%	395506483	0.00003%
W1679	394917793	97.97%	395232693	98.05%	394917693	0.00003%
W1943	406029523	99.29%	406029323	99.29%	406029323	0.00005%
W2318	398841121	98.46%	399110911	98.53%	398840921	0.00005%
W0107	388305525	98.58%	388454169	98.61%	388305325	0.00005%
W1750	408343962	86.80%	409393723	87.02%	408343662	0.00007%
W1783	394557463	97.54%	398781860	98.58%	394557163	0.00008%
W1084	391848379	97.92%	392530386	98.09%	391847779	0.00015%
W2197	398948094	96.72%	405755141	98.37%	398905328	0.01072%
W3037	417602567	95.75%	420216500	96.35%	417553584	0.01173%
W0103	392965094	98.36%	393082868	98.39%	392896315	0.01750%
W0573	385158329	98.29%	385249953	98.31%	385062829	0.02479%
W3052	409358321	97.87%	412126937	98.53%	409181290	0.04325%
W1214	404465068	95.99%	408094745	96.85%	404284736	0.04459%
W0639	398004161	97.75%	398274781	97.82%	397683145	0.08066%
W0164	407859560	96.29%	412831144	97.46%	407508316	0.08612%
W1726	393129352	98.06%	392888760	98.00%	392741351	0.09870%
W3086	411959050	98.51%	413937249	98.99%	411529384	0.10430%
W0135	379581140	97.81%	379463092	97.78%	379152669	0.11288%
W3082	404572697	96.53%	407589310	97.25%	404114361	0.11329%
W1236	404932173	97.46%	407732984	98.14%	404424871	0.12528%
W0634	406347740	97.67%	406096108	97.61%	405726681	0.15284%
W3066	417468196	93.63%	425108918	95.34%	416276857	0.28537%
W2099	404684010	95.90%	404360424	95.82%	403251829	0.35390%
W1777	401065020	97.14%	402472566	97.48%	399252113	0.45202%

Note: the data in this table was collected from Supplementary Table 5.



Response Fig. 1 Several examples of whole genome synteny plot between chromosomes constructed using Hi-C technology and pseudo-chromosomes obtained via the ALLMAPS method.

We agree that future updates may further enhance the resolution and annotation quality for the dynamic and evolving nature of pangenome references, potentially warranting a future "finished" or "complete" status as more data become available.

Comment (2): The analyses and results of genome assemblies, evaluation, and genome evolution sections are sound and exciting. One potential issue is the differentially expressed alleles (DEA) part. Two out of three standards used for identifying DEA (Line 672-675), I believe, were wrong in the current written forms. For Equation 1 (Line 672) a division between two positive numbers should not yield a negative result, therefore, 'TPMp-gene / TPMa-gene < -2' is always false and won't do anything for DEA identification. The correct one should be 'TPMp-gene / TPMa-gene < ½'. Equation 3 (Line 675) also is unsuitable for identifying DEA (My guess '|TPMp-gene - TPMa-gene| >=10' would be the intended equation). However, based on a lot of unexpected typos in Figures (see below), it is hard to determine if both equations were also just typed wrong or if these they were indeed used in analyses. If the latter is the case, it would be necessary to redo the analyses using the corrected equations and update the results accordingly.

Response: Thank you for pointing out this miswriting. We apologize for a typo error in Equation 1, which should indeed be 'TPMp-gene / TPMa-gene > 2 or TPMp-gene / TPMa-gene < ½' as you pointed out. This correction had been made

in the revised manuscript **(Line 719)**. As for Equation 3, our intention was to employ a filter for heterozygous alleles with very low expression by using the criterion of total read counts ($NumReads_{p\text{-}gene}$ + $NumReads_{a\text{-}gene}$) $\geqslant$ 10 in an RNA-seq library. This ensured that only alleles with sufficient expression levels were considered, thereby enhancing the reliability of our analysis. We have synthesized the methods from multiple published papers (Tian et al. 2022; Hu et al. 2022; Hoopes et al. 2022) to define the criteria for identifying differential expression allelic genes (DEA). We believe that these methods provide a robust framework for our study, ensuring that our results are both accurate and reliable.

Comment (3): Another potential issue is related to Fig. 2a, which showed japonica is so different than other groups. Since all these variants were identified by mapping to Nipponbare (a japonica variety), a question is how much the reference bias contributed to this phenomenon. Using an outgroup accession, such as O. meridionalis, as the reference probably is a potential solution.

Response: Thank you for the comments on potential reference bias in the variant analysis. In response, we have conducted additional analyses to quantify the impact of reference bias on variant numbers across different rice groups. Here are the details of our analysis and findings:

First, we utilized the other three genomes of different groups as reference, *Oryza. sativa* ssp. *indica* (Gla4), *O. meridionalis* (OM1952) and *O. barthii* (public data, GenBank: GCA_000182155), for the comparative analysis. The variations of different types in each group were called by mapping reads to reference genomes (**Response Fig. 2a**). As expected, when mapping to the Gla4 reference, *indica* varieties were displayed to have the fewest variations, because of reference bias. However, using outgroup accessions such as *O. meridionalis* or *O. barthii*, we observed minimal variation discrepancies in the number of variations among the groups, supporting the hypothesis that reference bias influenced the initial findings.

We have therefore conducted a comparison of the number of SNPs per sample using Nipponbare, Gla4, and the outgroup OM1952 as reference genomes, respectively. For ease of understanding, the calculation formulas of reference biases (RefBias), the contribution of RefBias (BiasCon) were described in detail as follows:

$$\mathrm{RefBias} = \frac{\text{SNPs Number with OM1952} - \text{SNPs Number with Nipponbare or Gla4}}{\text{Total SNPs Number}}$$

(**Equation 1**)

$$\begin{aligned}&\text{SNPs Number with a fixed RefBias}\\ &= \text{SNPs Number with OM1952} - \text{Total SNPs Number} * \text{RefBias of Or}\\ &- \text{IIIb}\end{aligned}$$

(**Equation 2**)

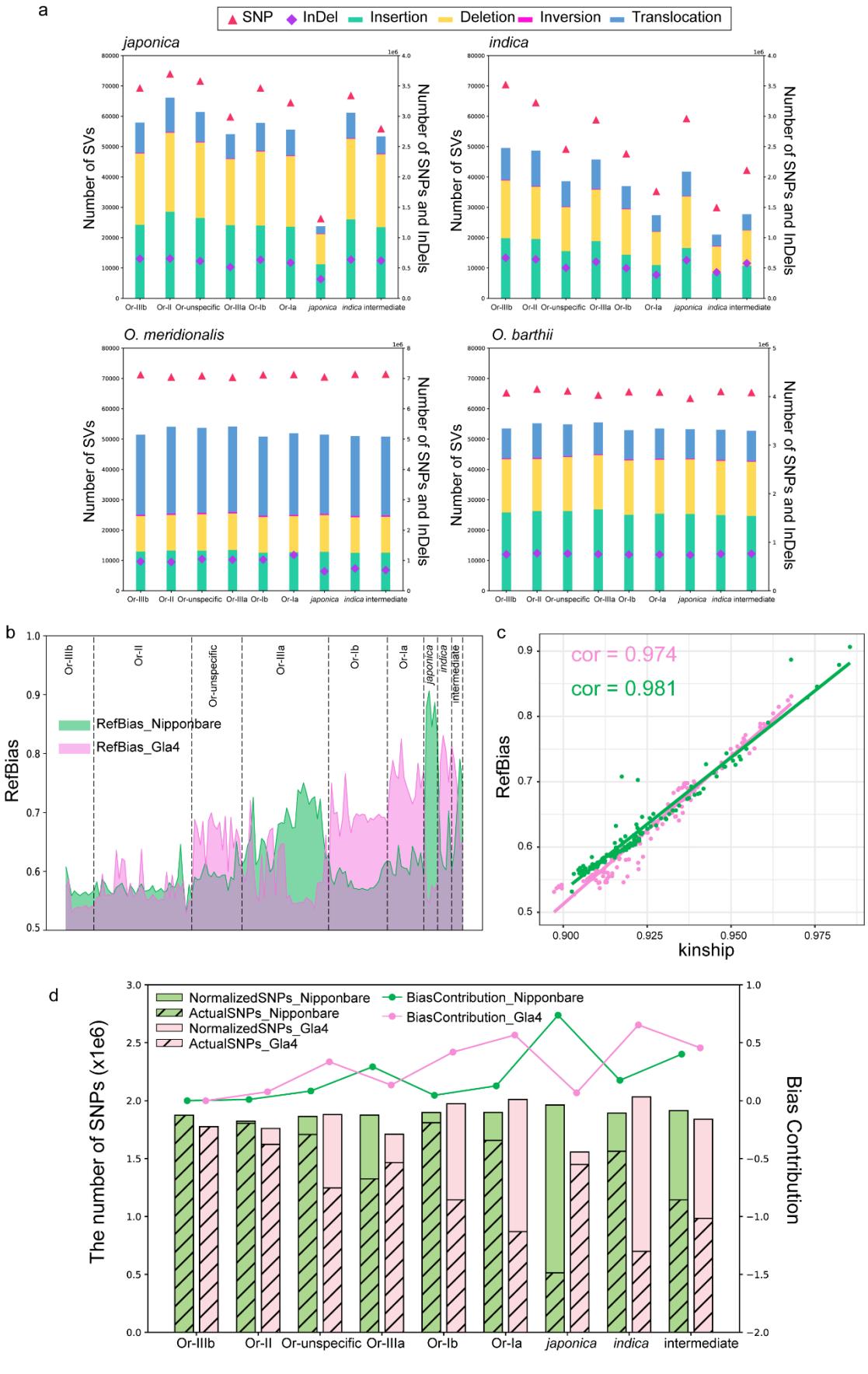
$$\mathrm{BiasCon} = \frac{\text{SNPs number with a fixed RefBias} - \text{SNPs Number with Nipponbare or Gla4}}{\text{SNPs number with a fixed RefBias}}$$

(**Equation 3**)

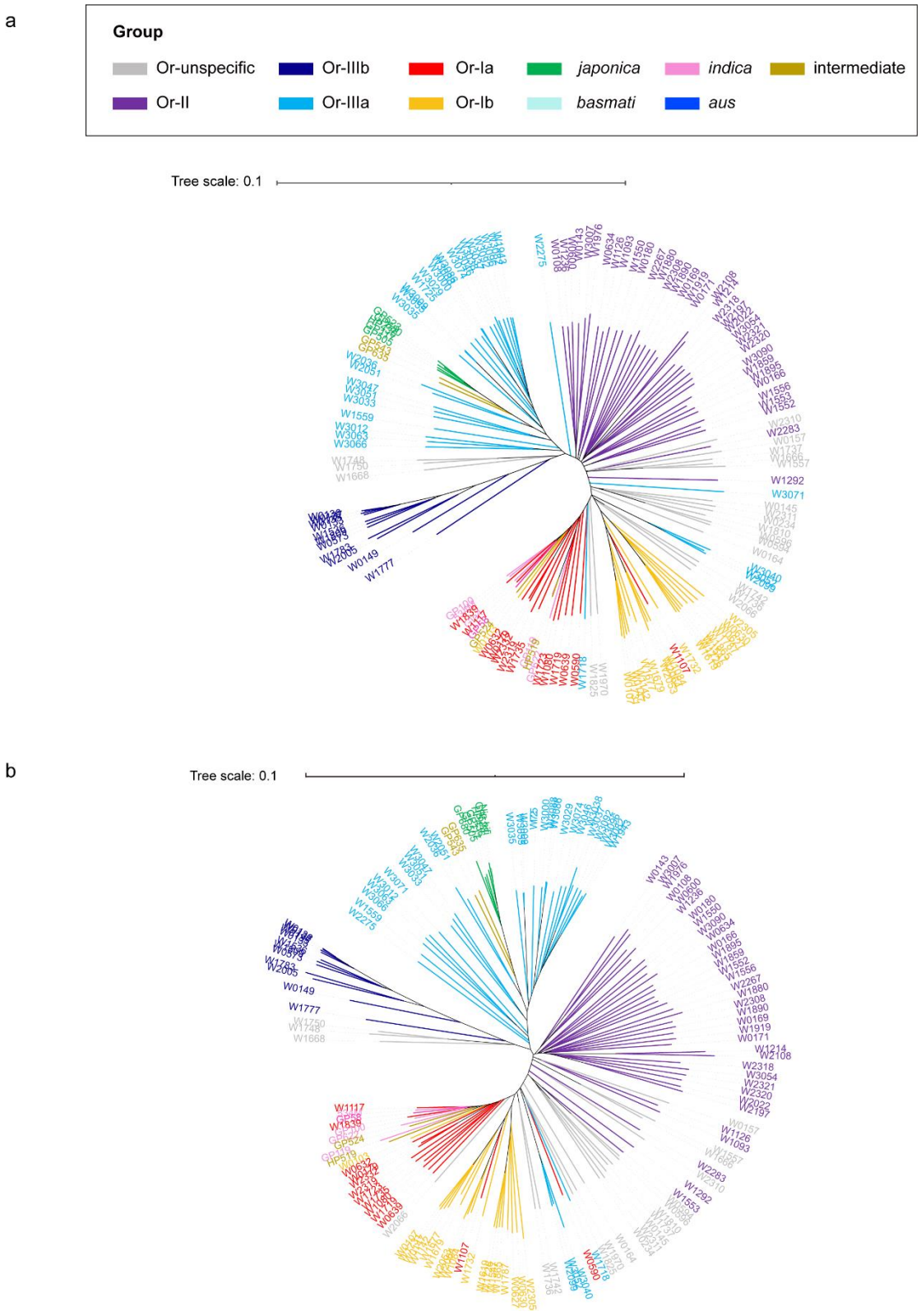
The largest RefBias for Nipponbare were observed in *japonica* and its close ancestor Or-IIIa. Similarly, the largest RefBias for Gla4 were in *indica* and its close ancestor Or-Ia (**Response Fig. 2b**). And the RefBias exhibited a positive correlation with the kinship (**Response Fig. 2c**). To better understand how RefBias affects the observed SNP variation among groups, we unified the standardization for the analysis by using the average RefBias values of Or-IIIb for Nipponbare (0.5688) and Gla4 (0.5460), which were the lowest of all groups of *O. rufipogon* and *O. sativa*. This standardization allowed us to calculate the number of SNPs with a fixed bias for each sample, thereby isolating the impact of reference bias and providing a clearer comparison of SNP counts without its influence. Our findings revealed that reference bias accounts for 73.82% of the observed variation in *japonica* with Nipponbare and 65.38% in *indica* with Gla4 (**Response Fig. 2d**). By incorporating these additional analyses, we confirmed your concerns that reference bias has contributed to the difference in the number of variations.

In addition, we further assessed the influence of different references on subsequent analyses such as evolution. Through constructing phylogenetic trees of Nipponbare and Gla4 as reference genomes, respectively, we found that they exhibited the similar structure and aligned with our initial classification using gene-based ML phylogenetic tree (**Response Fig. 3**). These consistent results are due to the use of a merged variation matrix in all subsequent analyses, which is free from RefBias. It's also worth noting that we attempted to construct an SNP matrix using *O. meridionalis* as a reference. However, this matrix showed extensive missing sites, totaling 81.44%. In contrast, when using the Nipponbare and Gla4 references, the percentages of missing sites were significantly lower, at 11.24% and 20.39% respectively. This discrepancy likely stems from poor alignment due to the distant phylogenetic relationship between *O. meridionalis* and the species under study. Therefore, we continue to use Nipponbare as the reference genome for evolutionary analysis in the revised manuscript.

We believed that these approaches provided an in-depth analysis of the impact of reference bias and enhanced the robustness of our findings. Thanks for your valuable feedback again, which has led to a more thorough and nuanced analysis.


a
SNP
InDel
Insertion
Deletion
Inversion
Translocation
japonica
indica
O. meridionalis
O. barthii
Number of SVs
Number of SNPs and InDels
Or-IIIb
Or-II
Or-unspecific
Or-IIIa
Or-Ib
Or-Ia
japonica
indica
intermediate
b
RefBias
RefBias_Nipponbare
RefBias_Gla4
c
cor = 0.974
cor = 0.981
RefBias
kinship
d
NormalizedSNPs_Nipponbare
ActualSNPs_Nipponbare
NormalizedSNPs_Gla4
ActualSNPs_Gla4
BiasContribution_Nipponbare
BiasContribution_Gla4
The number of SNPs (x1e6)
Bias Contribution

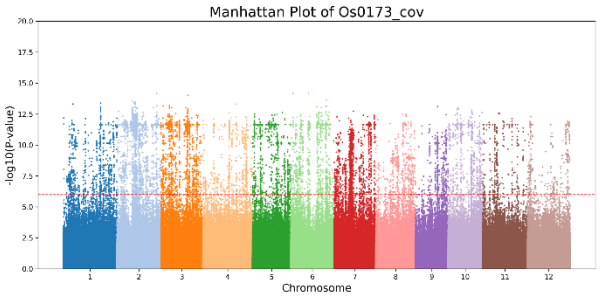
Response Fig. 2 The reference bias of Nipponbare and Gla4 as reference genomes. **a**, The number of various sequence variations under different reference genomes (*japonica*, *indica*, *O. meridionalis*, and *O. barthii*). The primary y-axis displays the number of structure variations (SVs), including insertions (INS), deletions (DEL), inversions (INV), and translocations (TRANS), each of which is indicated by a different color. The secondary y-axis displays the number of SNPs (triangle) and small Indels (diamond). **b**, The reference bias in each sample when using Nipponbare and Gla4 as reference genome. **c**, The correlation between reference bias (RefBias) and kinship is shown when using Nipponbare and Gla4 as reference genome. The Pearson's correlation coefficients (PCC) are depicted, and the lines are colored corresponding to the plots in (**b**). **d**, Comparation the actual number of SNPs to the number standardization for a fixed reference bias of Or-IIIa in each group of *O. sativa* and *O. rufipogon*. The comparisons are shown for scenarios where Nipponbare and Gla4 are used as reference genomes.

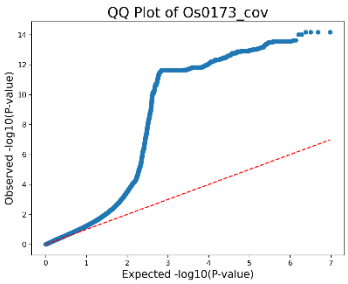


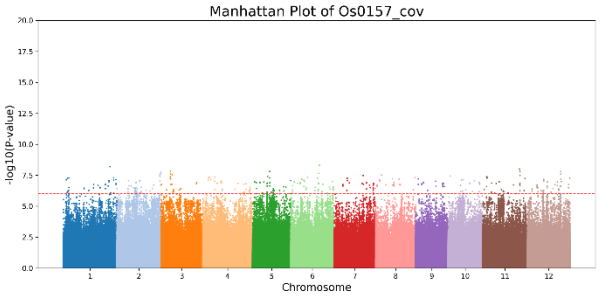
Response Fig. 3 The phylogenetic trees of 144 *Oryza* accessions when using Nipponbare (a) and Gla4 (b) as reference genome, respectively. The tree branches and labels are color-coded to represent different groups.

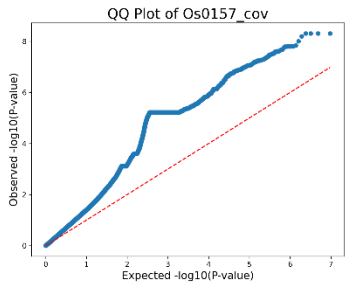
Comment (4): The assembled pangenome enables extensive analyses on the genome evolution. The authors revealed that Or-IIIa has a greater mean and variance in the genome size than japonica, which was mainly attributed to the Gypsy families. Or-IIIa and japonica have different outcome directions from illegitimate recombination among Gypsy retrotransposons: japonica tends to reduce Gypsy copies while Or-IIIa tends to increase. These initial characterizations likely stimulate further investigations of mechanisms on this fascinating genome phenomenon. For instance, an early study (10.1038/s41467-018-07974-5) conducted GWAS for inferred copy number variation from short reads with cultivated rice. Given the large variation between wild and cultivated rice (such as RIRE3, Os0016, and RIRE2 shown in sFig 13a) estimated from assemblies, conducting similar GWAS may lead interesting discoveries.

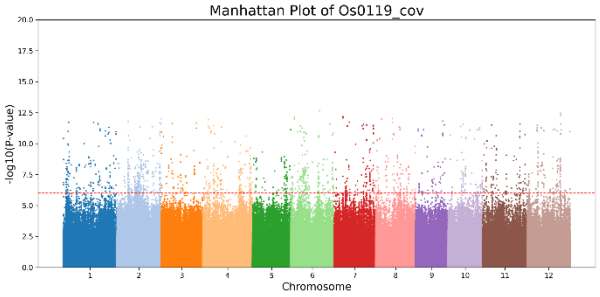
Response: Thank you for this constructive comment. Following the suggestions, we conducted a similar GWAS analysis focusing on 17 Gypsy families that are more prevalent in Or-IIIa compared to *japonica* (**Response Fig. 4**). Unfortunately, no significant association peaks and no satisfactory Q-Q (Quartile-Quartile) plots were found (**Response Fig. 4**). This may be due to strong population structure occurred in our wild population, making it less suitable for GWAS. Another potential reason could be the abundance of these 17 Gypsy families in the genomes (**Supplementary Fig. 19a** in the revised manuscript), as the clearest results were found in families with fewer repeats in the previous study (Carpentier et al. 2019). Nevertheless, your suggestions are highly valuable, and we are committed to further exploring the mechanisms behind this intriguing genomic phenomenon in future studies.

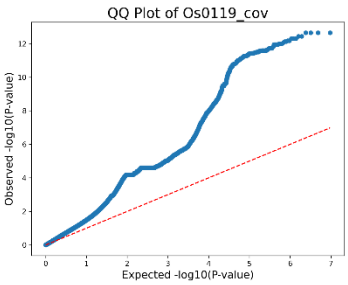

Manhattan Plot of Os0173_cov
-log10(P-value)
Chromosome

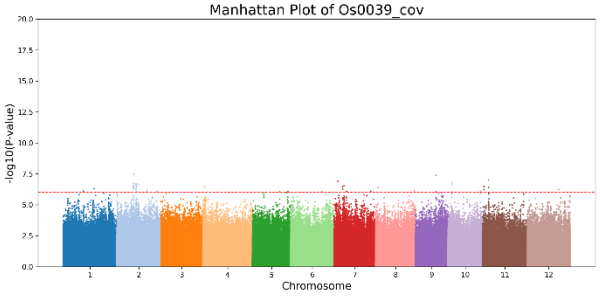

QQ Plot of Os0173_cov
Observed -log10(P-value)
Expected -log10(P-value)

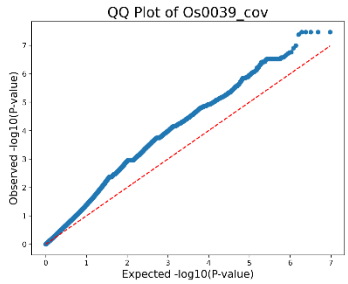

Manhattan Plot of Os0157_cov
-log10(P-value)
Chromosome

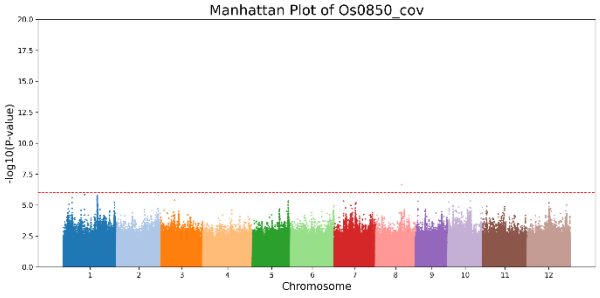

QQ Plot of Os0157_cov
Observed -log10(P-value)
Expected -log10(P-value)

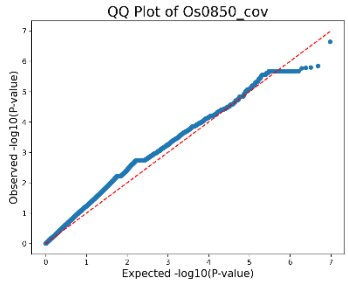

Manhattan Plot of Os0119_cov
-log10(P-value)
Chromosome

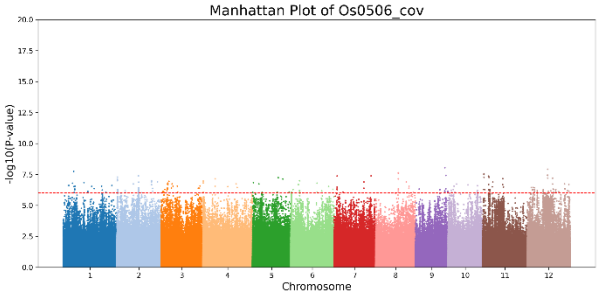

QQ Plot of Os0119_cov
Observed -log10(P-value)
Expected -log10(P-value)

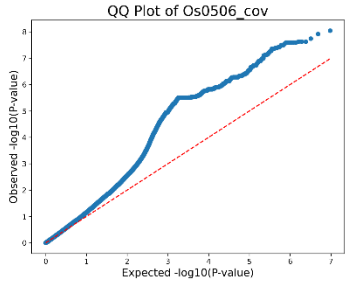

Manhattan Plot of Os0039_cov
-log10(P-value)
Chromosome

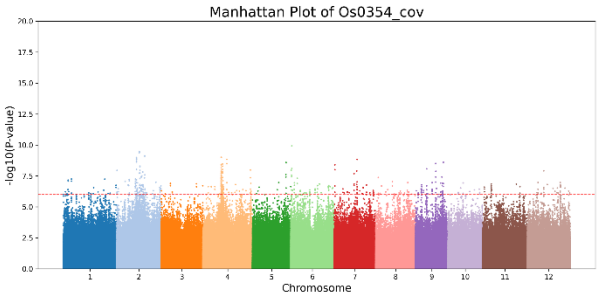

QQ Plot of Os0039_cov
Observed -log10(P-value)
Expected -log10(P-value)

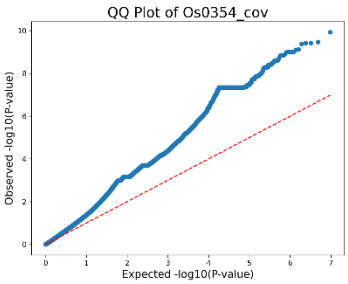

Manhattan Plot of Os0850_cov
-log10(P-value)
Chromosome

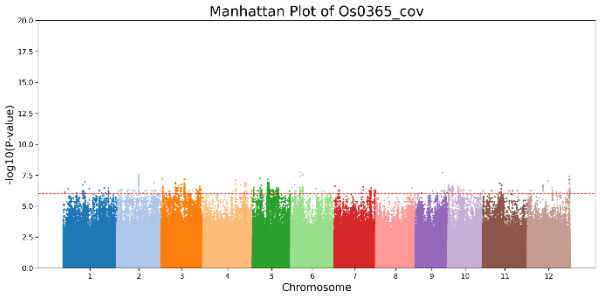

QQ Plot of Os0850_cov
Observed -log10(P-value)
Expected -log10(P-value)

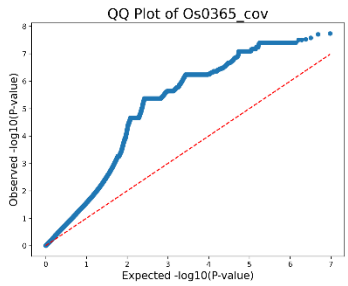

Manhattan Plot of Os0506_cov
-log10(P-value)
Chromosome

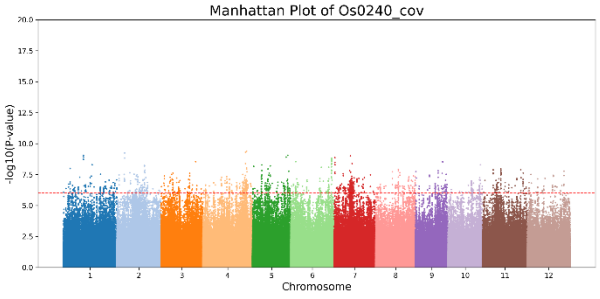

QQ Plot of Os0506_cov
Observed -log10(P-value)
Expected -log10(P-value)

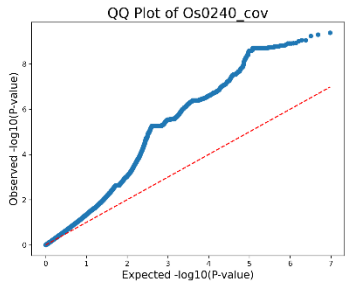

Manhattan Plot of Os0354_cov
-log10(P-value)
Chromosome

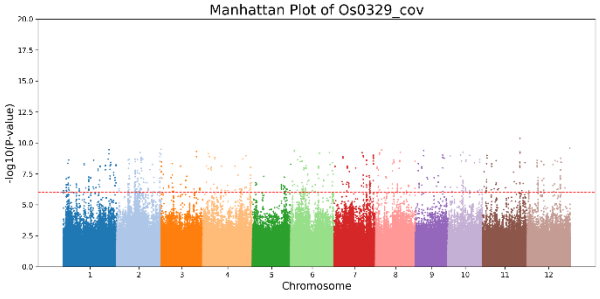

QQ Plot of Os0354_cov
Observed -log10(P-value)
Expected -log10(P-value)

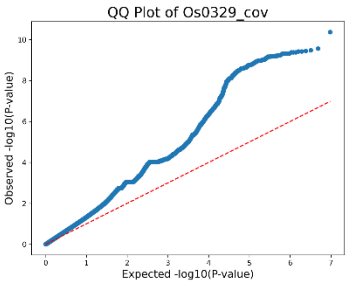

Manhattan Plot of Os0365_cov
-log10(P-value)
Chromosome

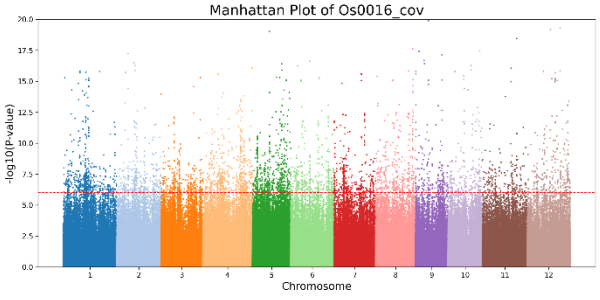

QQ Plot of Os0365_cov
Observed -log10(P-value)
Expected -log10(P-value)

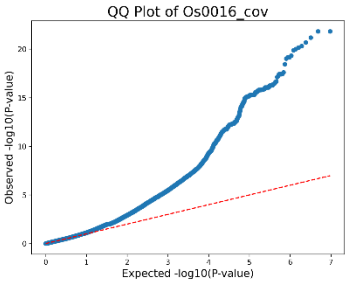

Manhattan Plot of Os0240_cov
-log10(P-value)
Chromosome

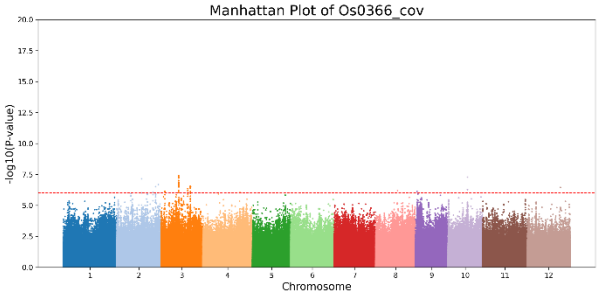

QQ Plot of Os0240_cov
Observed -log10(P-value)
Expected -log10(P-value)

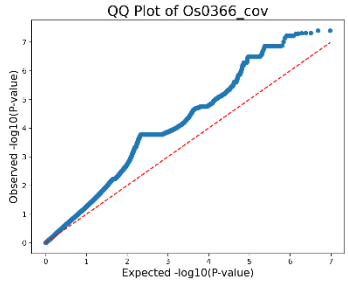

Manhattan Plot of Os0329_cov
-log10(P-value)
Chromosome

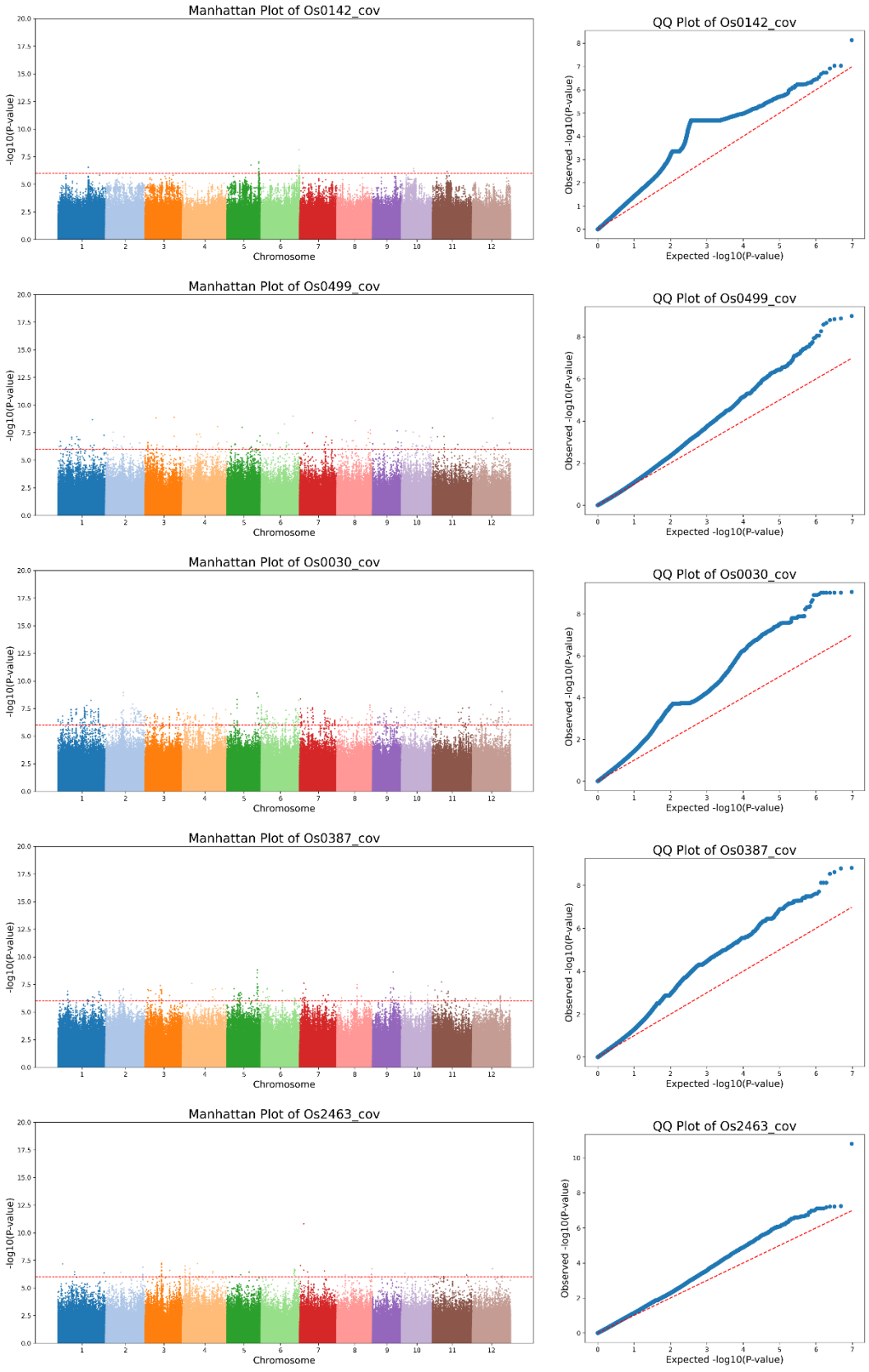

QQ Plot of Os0329_cov
Observed -log10(P-value)
Expected -log10(P-value)


Manhattan Plot of Os0016_cov
-log10(P-value)
Chromosome


QQ Plot of Os0016_cov
Observed -log10(P-value)
Expected -log10(P-value)


Manhattan Plot of Os0366_cov
-log10(P-value)
Chromosome


QQ Plot of Os0366_cov
Observed -log10(P-value)
Expected -log10(P-value)



Response Fig. 4 GWAS results for 17 Gypsy families that are larger in Or-IIIa genomes.

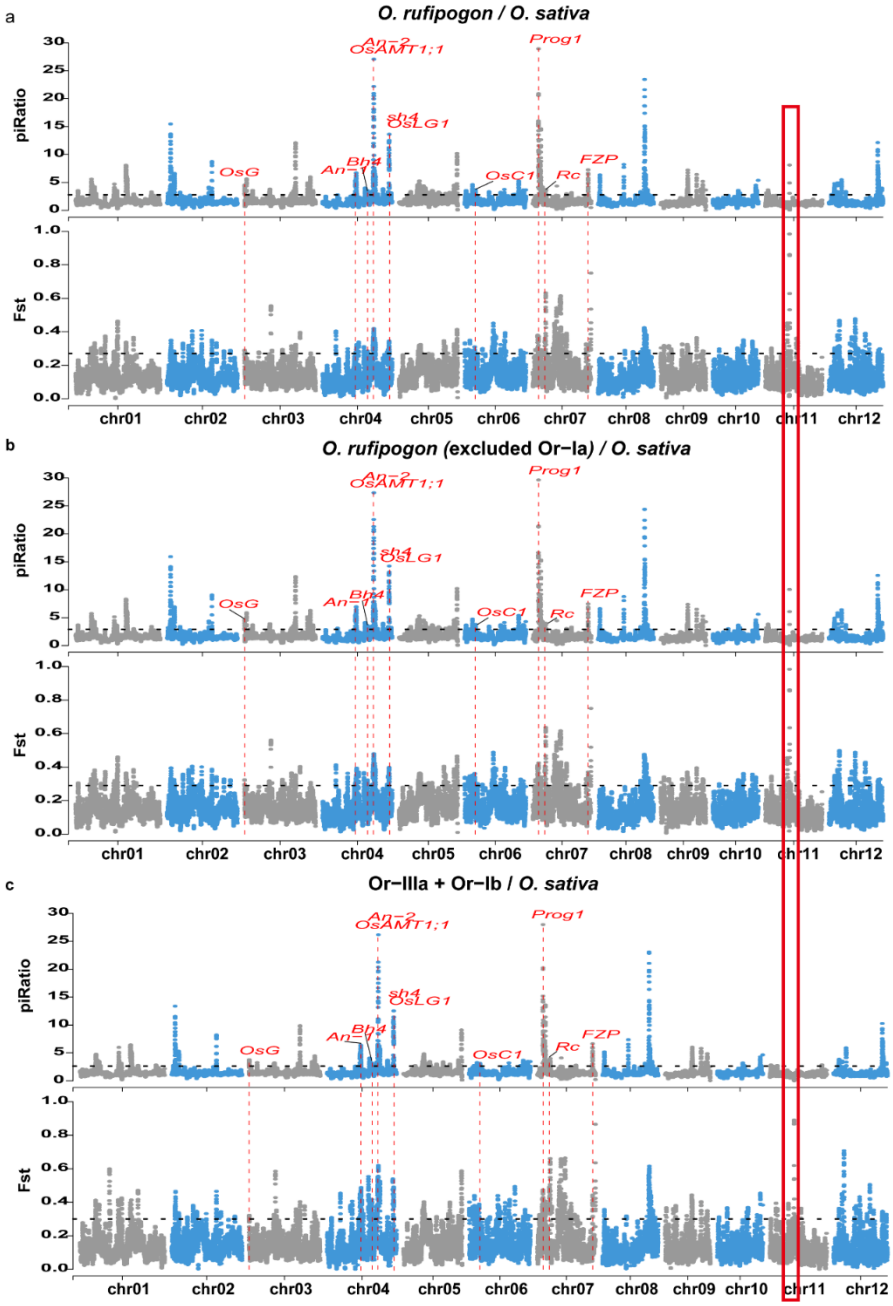
Comment (5): The question of how O. sativa was domesticated from wild progenitor is being actively debated. Early findings utilizing low-density SNP (Ref19) from the lab eventually led to the hypothesis of "a single domestication event with multiple origins". However, an alternative hypothesis was formulated as "multiple domestications" based on either the same genomics dataset in the Ref19 (Ref23) or new dataset (Ref24). This study generated novel insights with a higher resolution of phylogenetic relationships among wild rice, such as previously characterized Or-I can be further clustered into two subtype Or-Ia and Or-Ib; proposed a new cluster of cultivated rice, intro-indica, derived from crossing between indica and aus; then intro-indica crossing with japonica derived basmati rice. A refined "single domestication event with multiple origins" model was then proposed.

My major comment pertains to the identification of selective sweep regions. These selective regions established the foundation of further characterization between wild and cultivated rice. In this analysis, only Or-IIIa and Or-1b were used as the ancestral group (Line 375 & 964). While it is acceptable that Or-Ia was excluded because of extensive gene flow between Or-Ia and indica (Line 372-373), the exclusion of other groups, Or-IIIb, Or-II, and Or-Unspecific, seems unjustified. Including these groups should increase the power of detecting the same selective sweep regions. In fact, this strategy (all wild rice groups) was used in the Ref19 from the lab. Unless the authors had other reasons, which were unspecified in this version, it would be better to use the same strategy of including all other groups (Or-Ia can be excluded) in the analysis.

Response: We appreciate your comments regarding the identification of selective sweep regions. We have reconducted the analysis using three different strategies for defining the ancestral groups: including 1) all groups of *O. rufipogon*, 2) all groups of *O. rufipogon* except Or-Ia, and 3) only Or-IIIa and Or-Ib, as originally described in our manuscript (**Response Fig. 5**).

Our findings showed that the 11 known domestication genes mentioned in our manuscript were consistently located in the regions with the peaks identified by all three strategies, indicating a high detection ability. But the third strategy (including only Or-IIIa and Or-Ib) demonstrated a higher effect, particularly in terms of *F*st values (**Response Fig. 5**). This improved performance may be attributed to reduced false positives achieved by excluding wild rice groups (Or-IIIb, Or-II, and Or-unspecific) that are less relevant to *O. sativa*, and by removing Or-Ia, which has extensive gene flow from cultivated rice. For example, within a region on chromosome 11 (12.36M-12.50M), which is marked by a red rectangle, the first two strategies both indicated a domestication peak (**Response Fig. 5a-b**). However, annotation results from MSU7 reveal that this area consists solely of transposable elements, with no significant genes present. Therefore, this peak can be considered a false positive.

In summary, our original strategy appears to provide a more robust identification of selective sweep regions by minimizing the inclusion of unrelated wild rice groups and reducing the impact of gene flow from cultivated rice. Therefore, we have retained the third strategy in the revised manuscript. To enhance clarity for our readers, we have added a detailed description of this strategy (**Line 397-399 and Supplementary Fig. 25** in the revised manuscript).



Response Fig. 5 Analysis of selective sweeps using three different strategies. Nucleotide diversity ratio (piRatio) and *F*st tests are used for three ancestral groups: all groups of *O. rufipogon* (**a**), all groups except Or-Ia (**b**), and only Or-IIIa and Or-Ib groups (**c**). Horizontal dashed lines indicate the genome-wide threshold (top

5%) of selection signals. The 11 known domestication genes are marked in red text.

Comment (6): It may be good to combine all the sequenced accessions in this study and Ref24, which supported another hypothesis, to check the relationships among these different subtypes of wild rice (A sFigure with a phylogenetic tree would serve the purpose). The main reason is that Ref24 separated perennial and annual wild rice into two subgroups (O. rufipogon and O. nivara, respectively), while based on Ref19, the authors probably still treated these two as one O. rufipogon (If so, it is better to reiterate this point in this manuscript). Even though the geographic distributions of these sampled wild rice accessions in both studies are highly overlapped, the varied nomenclatures challenge readers attempting to make connections and comparisons. For instance, if the Niv2 group in Ref24, progenitor of indica, is clustered together with Or-Ia in this study, it would clearly address a potential concern that the author's 'single domestication event' model might be due to the fact that Niv2 was not included in the sample. This extra simple analysis might provide valuable insights into the ongoing debate.

Response: Thank you for your suggestions.

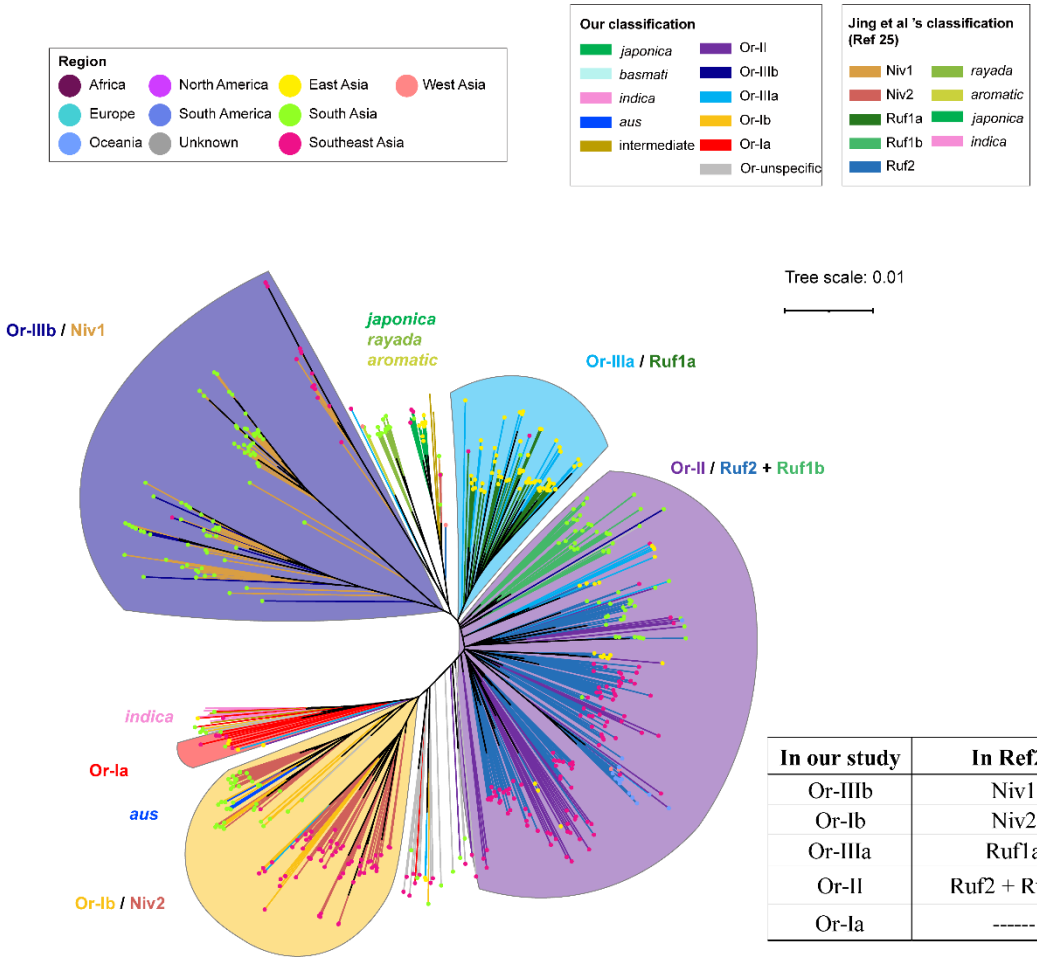
We have supplemented our original dataset (including 15 *O. sativa* and 127 *O. rufipogon*, the remaining 6 samples lacked next-generation data) with Illumina short paired-end reads. These were mapped against our study's graph-based pangenome using Giraffe (Sirén et al. 2021), along with the newly sequenced and non-redundant samples (33 *O. sativa* and 374 *O. rufipogon* or *O. nivara*) from Ref24 (**Supplementary Table 14** in the revised manuscript). SNPs/Indels were called using DeepVariants (Poplin et al. 2018) with the NGS model, and all individual variants were merged using GLnexus (Yun et al. 2021). We then constructed a phylogenetic tree for all 549 samples using the method described in the manuscript (**Supplementary Fig. 20a** in the revised manuscript and **Response Fig. 6**).

Our analysis revealed that *O. nivara* does not form an independent cluster but instead clusters with certain *O. rufipogon* groups, this supports our classification of *O. nivara* to be as ecotypes of *O. rufipogon*. Specifically, we found that Niv1 and Niv2 from Ref24 correspond to our Or-IIIb and Or-Ib groups, respectively. Additionally, the evolutionary relationships between cultivated and wild rice groups in this expanded analysis were consistent with our original findings. Or-IIIa (Ruf1a group in Ref24) were identified as the progenitor of *japonica*; Or-Ib (Niv2 group in Ref24) from South Asia were the progenitor of *aus*; and Or-Ia (missing in Ref24) were the progenitor of *indica*. An additional archetypal analysis also confirmed these relationships (**Supplementary Fig. 20b** in the revised manuscript).

Furthermore, we reidentified selective sweep regions in this expanded population and reconstructed a phylogenetic tree based on SNPs within these

regions (**Supplementary Fig. 28** in the revised manuscript). This analysis reconfirmed that the domestication-related genomic regions of all cultivated rice originated from Or-IIIa, supporting a single domestication model. Interestingly, we observed that a cluster of NIV1 samples from Southeast Asia (in Ref24) was grouped with cultivated rice, showing a domesticated state. This may be due to past gene flow from cultivated rice, which was supported by both whole genome-wide evolutionary tree and archetypal analyses (**Supplementary Fig. 20a-b** in the revised manuscript).

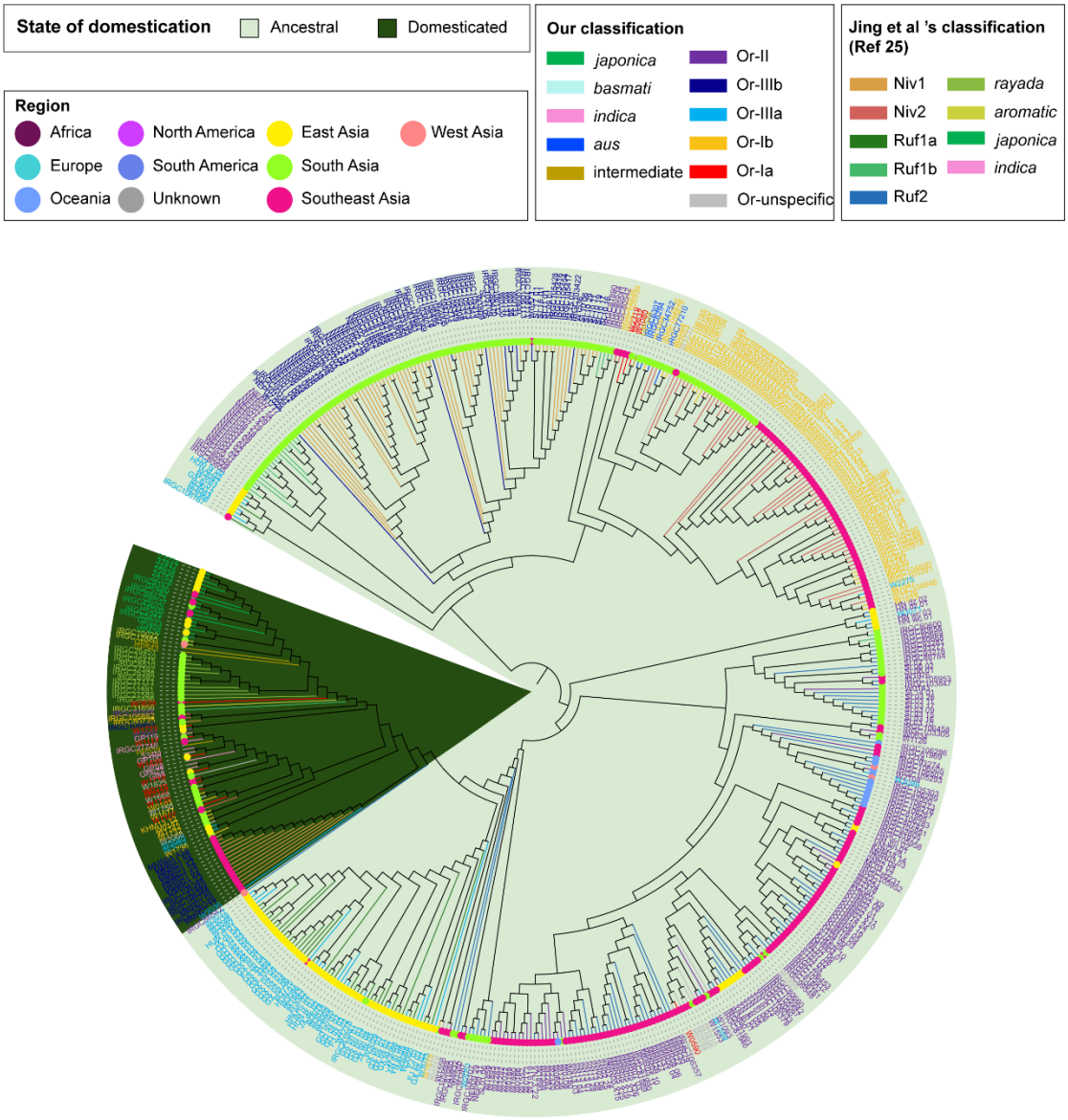
We are grateful for your valuable comment, which has led to a more robust and informative analysis. These results not only clarify the classification ambiguities between the different subtypes of wild rice into the ongoing debate but also significantly enhance our understanding of the complex patterns of rice evolution and domestication process. We have added this analysis into the revision (**Line 530-542**). In addition, I would like to mention that this study fully supports the previous finding regarding to Asian cultivated rice domestication that the *O. sativa* has a single origin but has undergone multiple domestication events.



In our study	In Ref24
Or-IIIb	Niv1
Or-Ib	Niv2
Or-IIIa	Ruf1a
Or-II	Ruf2 + Ruf1b
Or-Ia	------

Response Fig. 6 Phylogenetic analysis of 549 wild and cultivated rice accessions. The external circles denote the geographic distribution of each accession. Tree branches are color-coded to represent original groupings from two

studies. Background colors denote new groups based on our classification method derived from this evolutionary tree analysis. The text labels on the outside of the tree show the corresponding relationship between our grouping (before the slash) and the classification from another study (after the slash).



Supplementary Fig. 28 Phylogenetic tree of 549 *Oryza* accessions based on the SNPs within identified selective sweep regions. The external circles denote the geographic distribution of each accession. Tree branches are color-coded to represent original groupings from two studies (Jing et al. 2023). The tree features color-coded labels denote new groups based on our classification method derived from **Supplementary Fig.19a**. The color ranges indicate the state of domestication.

Some minor points:

Comment (7): Fig 2b was cited to support the "a propensity for enrichment in intergenic repetitive regions". However, reading this message from Fig 2b is not

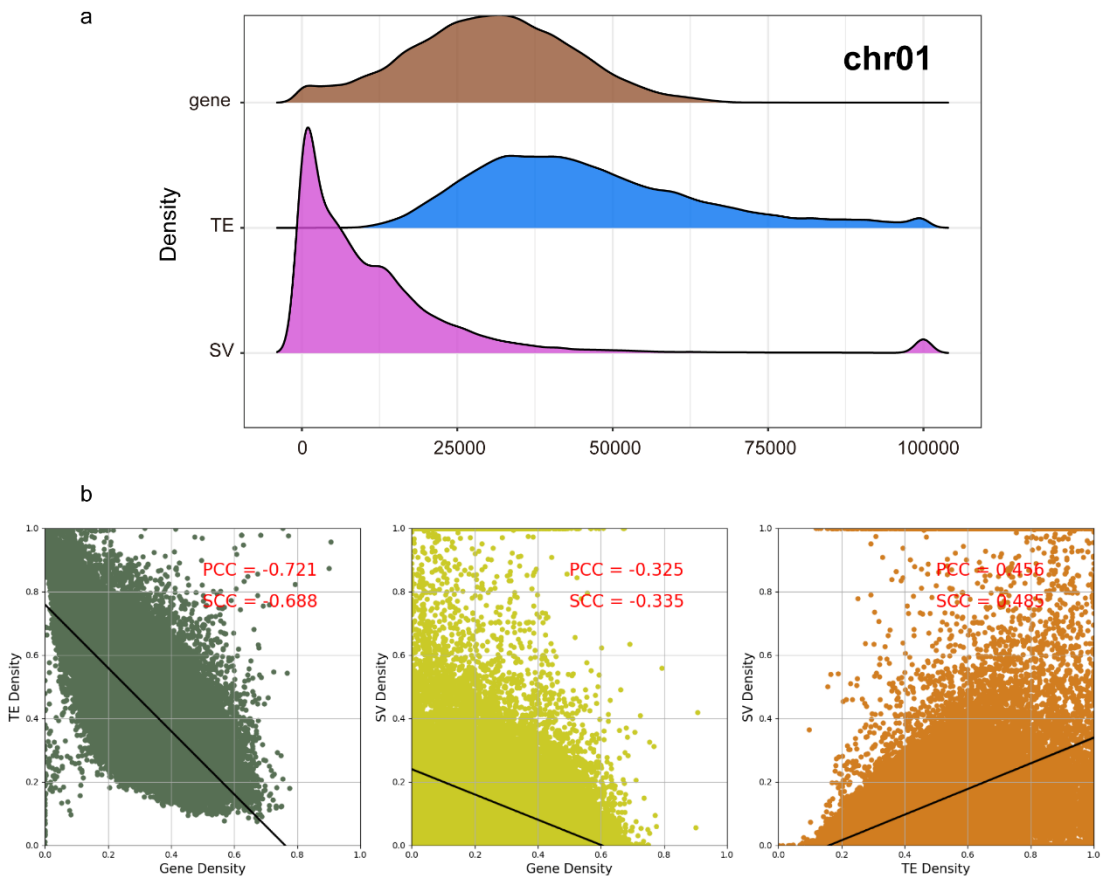
straightforward (the circular plot is too clouded). Why not use one chromosome as a representative, coupling with Pearson correlations between densities among these elements, to show the trend.

Response: Thank you for your valuable feedback regarding Fig. 2b. We acknowledge that the original circular plot may have been too cluttered to clearly convey the message regarding the propensity for enrichment in intergenic repetitive regions. We have moved Fig. 2b to the supplementary materials (**Supplementary Fig. 14**).

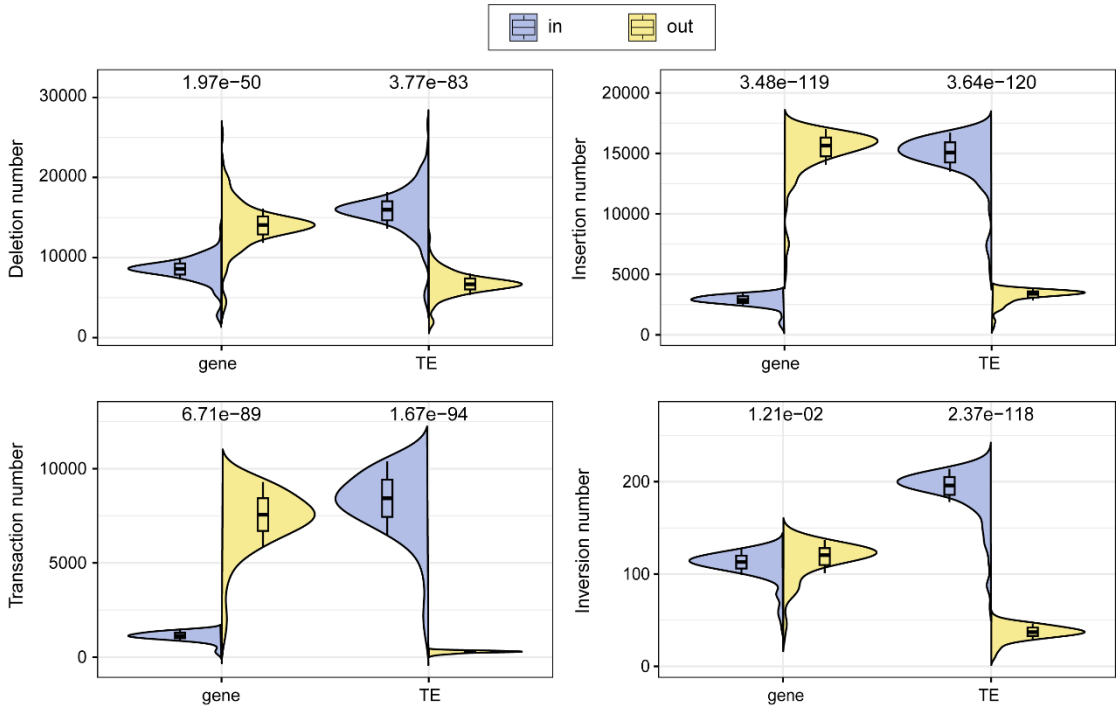
In response to the suggestions, we have calculated the Pearson correlations among the density of genes, transposable elements (TE) and structure variations (SVs) for each chromosome. Additionally, we calculated the Spearman correlations, which are more appropriate for data that do not conform to a normal distribution (**Response Fig. 7a**). Both correlation coefficients indicated a significant negative correlation between gene density and TE density, as well as between gene density and SV density (**Response Fig. 7b**). Conversely, there was a significant positive correlation between TE density and SV density.

It was worth mentioning that our SV density calculations included only insertions, translocations, and inversions, as deletion sequences are present in the reference genome. To account for this, we also used Student's *t*-test to analyze the distribution patterns of each SV type, including deletion sequences in the reference genome (**Supplementary Fig. 15** in the revised manuscript). Our findings consistently show that all four types of structural variations tend to be distributed in intergenic regions rich in repetitive elements.

This revised explanation provides a clearer and more concise overview of our methodology and results, addressing your concerns while maintaining the scientific integrity of our work.



Response Fig. 7 Correlation analysis of genes, transposable elements, and structural variants on chromosome 1. a, Density distribution of genes, transposable elements, and structural variants across chromosome 1. **b,** Correlation between gene density and TE density (left panel), correlation between gene density and SV density (middle panel), correlation between gene density and SV density (right panel). Linear regression lines are depicted, and both Pearson correlation coefficient (PCC) and Spearman's Correlation Coefficient (SCC) are displayed to quantify the linear and rank correlation, respectively.



Supplementary Fig. 15 Distribution of structural variations in gene and transposable elements (TEs) regions. Composition of four types of structural variations (deletion, insertion, translocation and inversion) between genes (left) and transposable elements (TEs, right) across genomes. Violin plots show the distribution of variation counts, with blue representing variations within the genetic element (in) and yellow representing variations outside the element (out). Embedded box plots display the median and quartiles. The numbers above each plot indicate *p* values.

Comment (8): Fig 2g. Both x and y are related to -log(P), but the significance threshold was only plotted for y, which may generate unnecessary confusions. I would recommend extending x-axis to 0 and add the threshold for x too, which follows the practice in the cited Ref40. Furthermore, modify Line "we designated these" (Line 733) to avoid the confusion of a new category being defined in this manuscript (Ref40 actually defined this category).

Response: Thank you for your concerns on Fig 2g (**Fig 2f** in the revised manuscript) and the related text. The confusion might be caused by only plotting the significance threshold for the y-axis. We have extended the x-axis to 0 and added the significance threshold for the x-axis. Regarding Line 733 (**Line 780-782** in the revised manuscript), we have revised the text to clarify that we are not defining a new category but rather referring to the classification established in Ref 40 (**Ref 50** in the revised manuscript) (Ou et al. 2022). We appreciate you for highlighting these issues, which have helped us enhance the accuracy and clarity of our manuscript, and hope these revisions address your concerns satisfactorily.

Comment (9): Fig 4a & sFig 17. Clearly define the last panel of blue and orange whether each group is based on one representative accession or the whole group.
Response: We apologize for the ambiguity in the figure legend. To address this, we have revised the legends to clearly describe the meaning of blue and orange in **Fig 4a** and Supplementary Fig 17 **(Supplementary Fig 26** in the revised manuscript).

Comment (10): For sFig7, it might be more informative to reorganize the accessions based on the presence/absence of the RLK gene and the size of the syntenic segment.
Response: Thank you for your suggestion. We have updated this figure (**Supplementary Fig. 9**) in the revised manuscript. This reorganization allows for a clearer visualization of the RLK gene, which is unique to wild rice, across different accessions. We believe this adjustment enhances the figure's informativeness and aids in better interpreting the data in relation to our study's objectives.

Comment (11): Line 201 "Acknowledging the propensity of RGAs to form clusters in the genome" can be removed as the analysis results didn't align well with this assumption.
Response: Thank you for your comments. We apologize for the misunderstanding caused by missing the corresponding literature, which has been added in the revised manuscript (**Line 208-209**). The characteristic that RGAs tend to cluster has been well-documented in the previous studies (Shang et al. 2022; Stein et al. 2018; Van de Weyer et al. 2019; Wang et al. 2019), which influenced our decision to categorize the identified RGAs based on their physical locations on chromosomes as singletons, pairs, and clusters. This categorization helps in refining the identification of RGA loci and facilitates subsequent analyses, such as synteny analysis. Therefore, we believe that retaining this statement, with appropriate references, is crucial for the accuracy and completeness of our discussion.

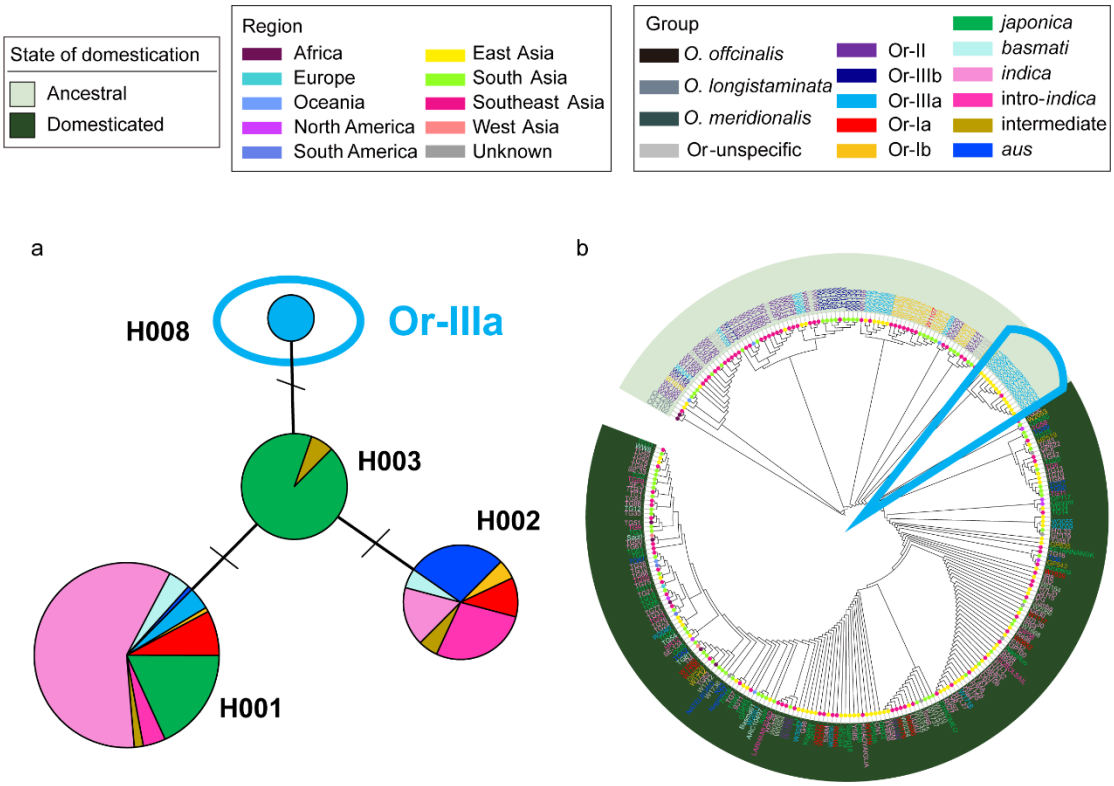
Comment (12): Line 457-467 and the related Fig. 5 introduced SNP categories as japonica/indica genetic bottleneck. I would suggest finding alternative terms for such category as "bottleneck" has special meaning in domestication related literature.
Response: Thank you for your suggestion. Yes, we have changed it to "*japonica* preference" or "*indica* preference".

Comment (13): Duplicated statements in Line 500-503.
Response: Thank you for pointing out this typo error, it has been corrected.

Comment (14): Line 512 to 518. The statement about LABA1 may require clarification for a better comprehension.

Response: We appreciate your comment regarding the need for a more comprehensive description of *LABA1*. We have supplemented more information in our revised manuscript (**Line 556-560**).

LABA1, as known as *An-2*, is a key gene in rice domestication that regulates awns length and grain yield. Previous researches (original researches) have suggested that *laba1* allele originated in the *japonica*-type wild rice and was subsequently introduced into the *indica* gene pool through introgression (Gu et al. 2015; Hua et al. 2015). Our haplotype analysis supports this view, indicating that *An-2* originated from Or-IIIa (the ancestor of *japonica*) (**Supplementary Fig. 31-32 and Response Fig. 8**). However, a recent study (Jing et al. 2023) has proposed an alternative origin for *LABA1*, suggesting it may be originated from *O. nivara* (Or-Ib in our population, the ancestor of *aus*). While we acknowledge this new perspective, we maintain that our analysis is more robust due to our higher quality genome data and more comprehensive haplotype analysis.



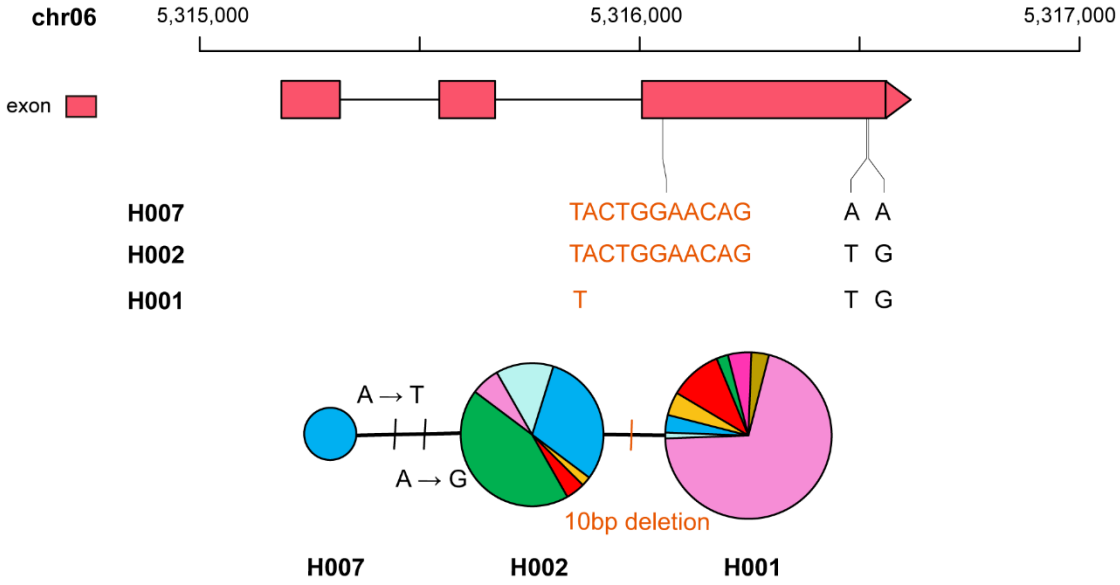
Response Fig. 8 Evidence supporting that *LABA1* (*An-2*) originated from *O. rufipogon*. a, Simplified haplotype networks for *LABA1* (*An-2*). Each circle represents a haplotype, with the circle size proportional to the number of accessions of that haplotype. Different colors indicate different groups. Variations of different haplotypes are indicated by short dashes. **b**, Phylogenetic trees for *LABA1* (*An-2*). The trees feature color-coded labels to distinguish various groups, and the surrounding external circles represent the geographic distribution of each

accession. The color ranges indicate the state of domestication.

Comment (15): Line 540 indicated some new functional polymorphisms were identified in known domestication genes. It would greatly enhance clarity to depict the locations and consequences for the newly identified polymorphisms for genes in sFig 21. For example, is the A:T mutation in OsC1 located in CDS and altered an amino acid?

Response: Thank you for your suggestion. We apologize for the oversight, which was not including detailed information on the newly identified polymorphisms in our initial manuscript. In the revised version, we have enhanced the analysis by focusing on the regions including known functional sites or quantitative trait nucleotides (QTNs) (**Supplementary Fig. 31-32** in the revised manuscript). This approach allows for a clearer haplotype network of domestication genes, without being disturbed by variations in the intron region. We have also improved the clarity of the corresponding statement (**Line 1046-1049**).

Regarding the specific mutation in *OsC1* that you highlighted, our analysis now confirms that A:T mutation, along with another identified SNP (A:G), results in an amino acid change (S:T) (**Response Fig. 9**).



Response Fig. 9 Gene structure and simplified haplotype of *OsC1*. Variations among different haplotypes are indicated by short dashes. Each dash represents a specific genetic variation. Short red line represents the known functional site.

Comment (16): Check Line 750-751 is the intended statement.

Response: We appreciate your attention to detail regarding Lines 750-751. Yes, it was that the statement in question lacked clarity. In response to your feedback, we have revised this section in the updated manuscript.

Comment (17): Although the main text is well written and clean, a lot of

unexpected typos occur in figure legends and labels: “Sourth” in most of Figures and sFigures when geographic regions were mentioned; “Hetrozygous allele” in Fig 1b&f, sFig 4, & sFig 24; “domsticattion gene” for Fig 4e; “State of demoestication” in Fig 4c, sFig 18 & sFig 22; “misiing allele” in Fig 5c; “phynotype” in Fig 5d; “maped” in sFig 10; “acitivity” and “oxidorreductase” in sFig 13; “America North” in sFig 15; “WastAsia” in sFig 18 & sFig 22; “Accesstions” in sFig 21; “MIT gene” in sFig 24; “Comparation” in sTable 3 & sTable 4; “out study” in sTable 8; “Grap-pangenome” in sTable 9. These typos did not impact the analyses and results but impair the manuscript quality.

Response: Thank you very much for pointing out these typo errors, which all have been corrected in the revised manuscript. Sorry for the any inconvenience.

Reference

Carpentier, Marie-Christine, Ernandes Manfroi, Fu-Jin Wei, Hshin-Ping Wu, Eric Lasserre, Christel Llauro, Emilie Debladis, Roland Akakpo, Yue-Ie Hsing, and Olivier Panaud. 2019. ‘Retrotranspositional Landscape of Asian Rice Revealed by 3000 Genomes’. *Nature Communications* 10 (1): 24. https://doi.org/10.1038/s41467-018-07974-5.

Gu, Benguo, Taoying Zhou, Jianghong Luo, Hui Liu, Yongchun Wang, Yingying Shangguan, Jingjie Zhu, et al. 2015. ‘An-2 Encodes a Cytokinin Synthesis Enzyme That Regulates Awn Length and Grain Production in Rice’. *Molecular Plant* 8 (11): 1635–50. https://doi.org/10.1016/j.molp.2015.08.001.

Hoopes, Genevieve, Xiaoxi Meng, John P. Hamilton, Sai Reddy Achakkagari, Fernanda de Alves Freitas Guesdes, Marie E. Bolger, Joseph J. Coombs, et al. 2022. ‘Phased, Chromosome-Scale Genome Assemblies of Tetraploid Potato Reveal a Complex Genome, Transcriptome, and Predicted Proteome Landscape Underpinning Genetic Diversity’. *Molecular Plant* 15 (3): 520–36. https://doi.org/10.1016/j.molp.2022.01.003.

Hu, Guibing, Junting Feng, Xu Xiang, Jiabao Wang, Jarkko Salojärvi, Chengming Liu, Zhenxian Wu, et al. 2022. ‘Two Divergent Haplotypes from a Highly Heterozygous Lychee Genome Suggest Independent Domestication Events for Early and Late-Maturing Cultivars’. *Nature Genetics* 54 (1): 73–83. https://doi.org/10.1038/s41588-021-00971-3.

Hua, Lei, Diane R. Wang, Lubin Tan, Yongcai Fu, Fengxia Liu, Langtao Xiao, Zuofeng Zhu, et al. 2015. ‘LABA1, a Domestication Gene Associated with Long, Barbed Awns in Wild Rice’. *The Plant Cell* 27 (7): 1875–88. https://doi.org/10.1105/tpc.15.00260.

Jing, Chun-Yan, Fu-Min Zhang, Xiu-Hua Wang, Mei-Xia Wang, Lian Zhou, Zhe Cai, Jing-Dan Han, et al. 2023. ‘Multiple Domestications of Asian Rice’. *Nature Plants* 9 (8): 1221–35. https://doi.org/10.1038/s41477-023-01476-z.

Ou, Shujun, Tyler Collins, Yinjie Qiu, Arun S. Seetharam, Claire C. Menard, Nancy Manchanda, Jonathan I. Gent, et al. 2022. ‘Differences in Activity and

Stability Drive Transposable Element Variation in Tropical and Temperate Maize'. Preprint. Genomics. https://doi.org/10.1101/2022.10.09.511471.

Poplin, Ryan, Pi-Chuan Chang, David Alexander, Scott Schwartz, Thomas Colthurst, Alexander Ku, Dan Newburger, et al. 2018. 'A Universal SNP and Small-Indel Variant Caller Using Deep Neural Networks'. *Nature Biotechnology* 36 (10): 983–87. https://doi.org/10.1038/nbt.4235.

Shang, Lianguang, Xiaoxia Li, Huiying He, Qiaoling Yuan, Yanni Song, Zhaoran Wei, Hai Lin, et al. 2022. 'A Super Pan-Genomic Landscape of Rice'. *Cell Research* 32 (10): 878–96. https://doi.org/10.1038/s41422-022-00685-z.

Sirén, Jouni, Jean Monlong, Xian Chang, Adam M. Novak, Jordan M. Eizenga, Charles Markello, Jonas A. Sibbesen, et al. 2021. 'Pangenomics Enables Genotyping of Known Structural Variants in 5202 Diverse Genomes'. *Science* 374 (6574): abg8871. https://doi.org/10.1126/science.abg8871.

Stein, Joshua C., Yeisoo Yu, Dario Copetti, Derrick J. Zwickl, Li Zhang, Chengjun Zhang, Kapeel Chougule, et al. 2018. 'Genomes of 13 Domesticated and Wild Rice Relatives Highlight Genetic Conservation, Turnover and Innovation across the Genus Oryza'. *Nature Genetics* 50 (2): 285–96. https://doi.org/10.1038/s41588-018-0040-0.

Tian, Yi, Amali Thrimawithana, Tiyu Ding, Jian Guo, Andrew Gleave, David Chagné, Charles Ampomah-Dwamena, et al. 2022. 'Transposon Insertions Regulate Genome-wide Allele-specific Expression and Underpin Flower Colour Variations in Apple (*Malus* Spp.)'. *Plant Biotechnology Journal* 20 (7): 1285–97. https://doi.org/10.1111/pbi.13806.

Van de Weyer, Anna-Lena, Freddy Monteiro, Oliver J. Furzer, Marc T. Nishimura, Volkan Cevik, Kamil Witek, Jonathan D. G. Jones, Jeffery L. Dangl, Detlef Weigel, and Felix Bemm. 2019. 'A Species-Wide Inventory of NLR Genes and Alleles in *Arabidopsis Thaliana*'. *Cell* 178 (5): 1260-1272.e14. https://doi.org/10.1016/j.cell.2019.07.038.

Wang, Long, Lina Zhao, Xiaohui Zhang, Qijun Zhang, Yanxiao Jia, Guan Wang, Simin Li, Dacheng Tian, Wen-Hsiung Li, and Sihai Yang. 2019. 'Large-Scale Identification and Functional Analysis of NLR Genes in Blast Resistance in the Tetep Rice Genome Sequence'. *Proceedings of the National Academy of Sciences* 116 (37): 18479–87. https://doi.org/10.1073/pnas.1910229116.

Yun, Taedong, Helen Li, Pi-Chuan Chang, Michael F Lin, Andrew Carroll, and Cory Y McLean. 2021. 'Accurate, Scalable Cohort Variant Calls Using DeepVariant and GLnexus'. *Bioinformatics* 36 (24): 5582–89. https://doi.org/10.1093/bioinformatics/btaa1081.

Authors' Response to Referee #2

Referee #2 (Remarks to the Author):

Remarks： The manuscript presents a comprehensive analysis of a pangenome reference of 149 wild and cultivated rice, providing valuable insights into rice genetics and evolution. The investigation into the origin and domestication of Asian cultivated rice, including the identification of introgression events and selection patterns, adds significant value to the field. The manuscript's findings are likely to interest a wide audience, ranging from geneticists and plant biologists to researchers studying crop domestication and evolution.

Response: Thank you very much for your comments.

Comment: However, I have the following major concerns:

Comment (1): The manuscript provides an important resource for the discovery of elite natural variations, and studies on the functional genomics, and evolutionary biology. However, the results in this version do not sufficiently highlight the importance of this resource. For example, in the identification of potential elite alleles present in wild species, the authors briefly examined the resistance gene analogs using this remarkable genetic resource but did not draw substantive conclusions or provide detailed examples, leaving this section feeling empty.

Response: Thanks for the comments and suggestions regarding the need to emphasize the significance of the genetic resources described in our manuscript.

We performed further analysis of RGA to be associated with functional variations. We have thoroughly revised and expanded the section on disease and pest resistance genes from *O. rufipogon* in our manuscript. Our research has identified 1,346 loci where the average number of (RGAs in *O. rufipogon* exceeds that in cultivated rice (**Fig 1g** in the revised manuscript). Notably, at one of these loci, we identified the receptor kinase gene *LOC_Os07g35680* from *O. rufipogon* accession Y476, which has been validated for its resistance to blast disease (Huang et al. 2024). Furthermore, upon detailed examination, we found some accessions in our *O. rufipogon* population that exhibit the exact synonymous mutation in the coding region as found in Y476 (**Supplementary Fig. 10** in the revised manuscript).

We also conducted an in-depth analysis of the well-known blast resistance *Pi5* locus (Lee et al. 2009), which includes two NBS-LRR genes, *Pi5-1* and *Pi5-2*. It is noteworthy that Nipponbare also possesses two *Pi5* genes, *Pi5-1* and *Pi5-3*, the latter of which lacks disease resistance capabilities. Through homology comparisons and synteny analysis, we discovered numerous accessions within our *O. rufipogon* population that carry this resistance locus (**Fig 1h and**

Supplementary Fig. 11 in the revised manuscript). These valuable results have already been included in the main text (**Line 213-231**).

These enhancements to the manuscript clarify the significant contributions of wild rice genetic resources to crop improvement and underscore the potential for their utilization in developing resilient rice varieties. We believe these revisions address the concerns raised and effectively communicate the importance of these findings. Thank you once again for your valuable suggestions.

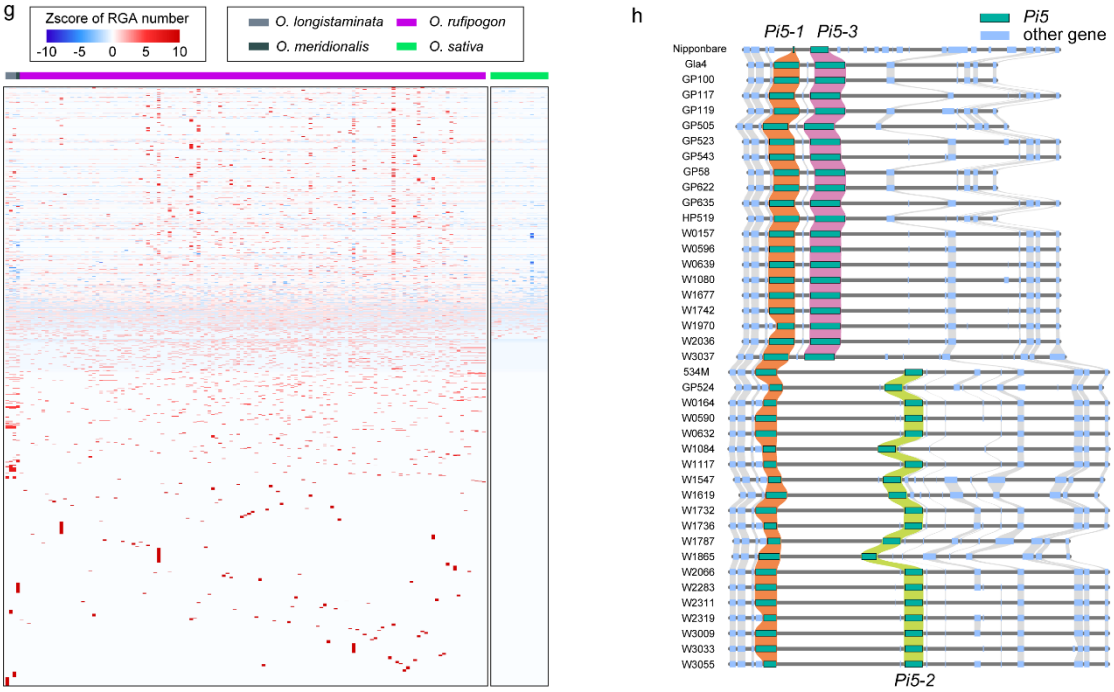
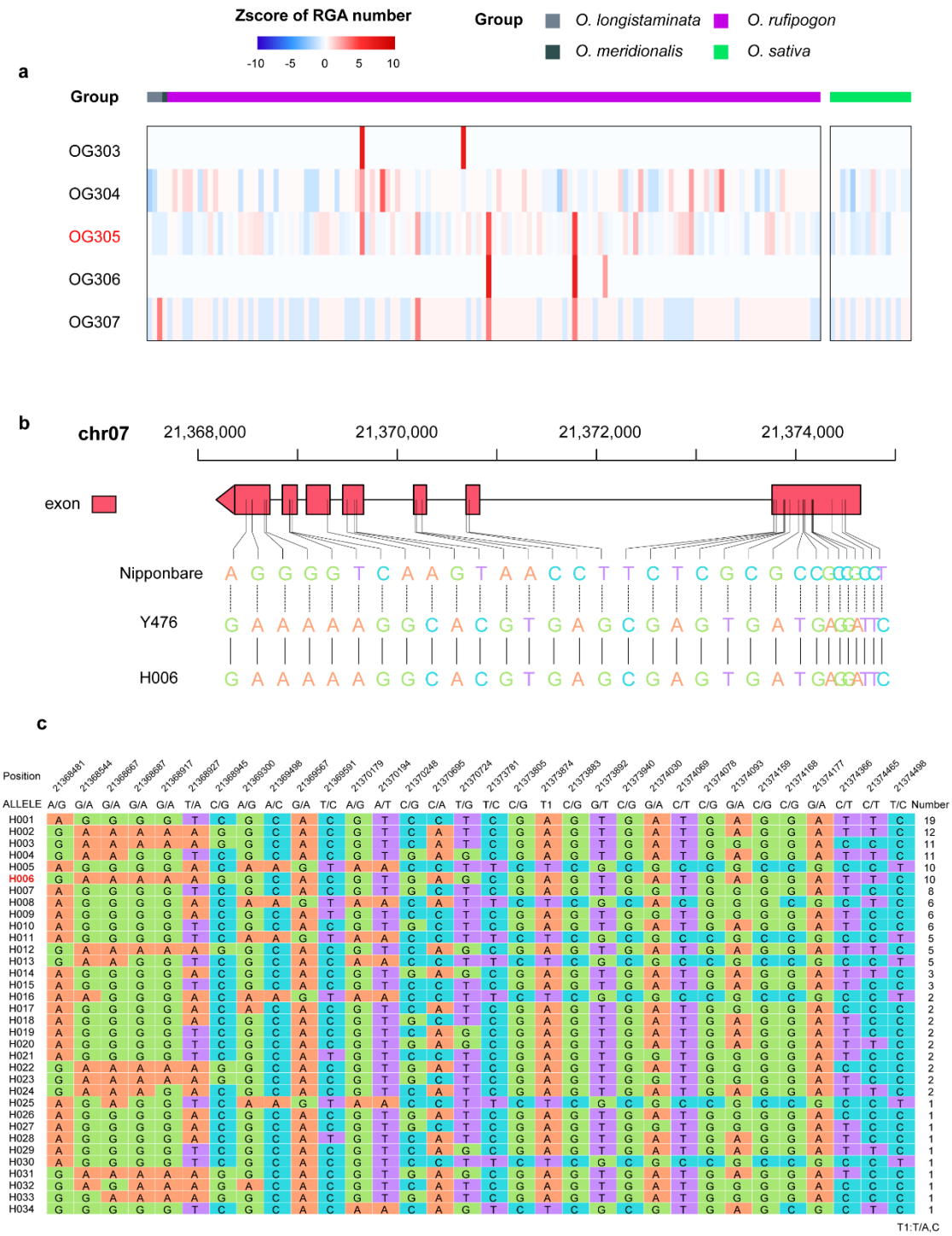


Fig. 1g-h g, The landscape of RGA loci that have a higher average copy number in wild rice than cultivated rice. Each row indicates a RGA locus, and each column indicates a variety. The data was normalized using a row-wise *Z*-score method. **h**, Local synteny plot of *Pi5* loci (chr09: 9.65-9.79 Mb in Nipponbare reference genome) from 41 accessions. Green boxes indicate *Pi5* genes (including *Pi5-1*, *Pi5-2* and *Pi5-3*) in each accession and different lines show their collinearity. Blue boxes indicate other adjacent genes and grey lines indicate their respective collinearity.



Supplementary Fig. 10 Genetic Diversity Analysis of *LOC_Os07g35680* in Wild and Cultivated Rice. a, The landscape of several RGAs loci that exhibit a higher average copy number in wild rice compared to cultivated rice. The locus containing the disease resistance gene *LOC_Os07g35680*, designated as OG305, is marked in red. **b,** The distribution of non-synonymous mutations between disease-resistant (the haplotype of Y476) and non-resistant (the haplotype of Nipponbare) variants within *LOC_Os07g35680*. **c,** A haplotype analysis of non-

synonymous mutations in *LOC_Os07g35680* among our wild rice population, with the disease-resistant haplotype H006 marked in red font.

Comment (2): Considering that many functional genomics research teams worldwide are not familiar with big data analysis, it would significantly enhance the paper's citations and impact if the manuscript could provide a convenient entry point for accessing related data for the resource, such as genome sequence variation queries, alignment, distribution ranges, or differentiation sites between indica and japonica rice.

Response: Thank you very much for the valuable comment and suggestions. Following the suggestion, we have made the data freely available through a comprehensive web-based pangenome database of wild and cultivated rice (**http://ricepandb.ncgr.ac.cn, http://119.78.67.204:8080**), which is now online available to access. The details of the dataset are as described below (**Response Fig. 10**):

The Asian cultivated rice (*Oryza sativa* L.), domesticated from its wild progenitor, *Oryza rufipogon*, is one of the most important food crops in the world. To fully exploit and utilize wild rice resources, we developed **RicePandb**, a pangenome database of wild and cultivated rice, to provide an intuitive user experience for functional genomics and evolution studies. RicePandb offers a range of resources and powerful tools, including detailed population information, an interactive genome browser, variation search, and syntelog analysis. These resources are designed to facilitate in-depth comparative, evolutionary, and functional analyses, supporting researchers in their quest to enhance rice improvements and enrich our understanding of its origin and domestication.

Redacted

Response Fig. 10 The homepage of RicePandb

Comment (3): Regarding heterozygous loci, the heterozygosity rate in wild rice species is high. However, I did not see relevant information on the heterozygosity

rates of wild species. Although the authors paid attention to heterozygous loci during genome assemblies, this consideration was not evident in the subsequent analyses, such as genetic variation, gene flow, population structure and introgression, selective sweep, and the genomic variation information used in the analysis of rice origin.

Response: We appreciate your observation regarding the details of the heterozygosity rates in wild rice species and apologize for the initial oversight in our initial submission. To address it, we have now updated detailed heterozygosity rate information in **Supplementary Table 6** of the revised manuscript.

In our initial analyses, to avoid errors due to differences in sequencing and analysis methods, we supplemented our dataset solely with assembled contig sequences of cultivated rice from published data. However, these data did not include heterozygous loci, an aspect we neglected in our subsequent analyses. We understand your concern about this limitation and its potential impact on our subsequent analyses. Recognizing the importance of this aspect as highlighted in your comment, we used unified short reads of our population and a recent study (Jing et al. 2023) to map against the graph-based pangenome we constructed, which now incorporated heterozygous loci information (**Line 295-302**). The additional analyses include the construction of the phylogeny tree, the archetypal analysis, and the identification of selective sweeps (**Supplementary Fig. 20 and Supplementary Fig. 28** in the revised manuscript), which confirmed the conclusions drawn in our initial findings.

The consistency can be attributed to several factors. Firstly, the heterozygosity rate of 0.37% in *O. rufipogon* suggests that heterozygosity may not have as significant an impact on population-level analyses as initially assumed. Secondly, the random distribution of heterozygous loci across the wild rice genome (**Supplementary Fig. 6c-d** in the revised manuscript), indicating that these variations are unlikely to systematically bias our evolutionary and domestication analyses. This randomness helps explain why our original methods, despite not explicitly accounting for heterozygosity, still captured the essential genetic relationships and evolutionary patterns.

Comment (4): I suggest that the author reconsider the importance of each section and remove or shorten some standard descriptions or contents without new findings. For example, the graph-based pan-genome did not provide any substantive content. Graph-based pan-genome analysis is only valuable if it yields new insights, so I suggest removing it. The content regarding gene-based pan-genomes and transposable elements mostly consists of conventional data descriptions and should be substantially shortened, as they did not yield new conclusions or exciting content.

Response: Thank you for your comments. We have carefully considered your suggestions and made the following revisions:

We have carefully streamlined our manuscript by reducing conventional data descriptions, as you recommended. This helps focus the paper on novel findings and key insights.

Regarding the graph-based pan-genome analysis, we have decided to retain this section. In our revised manuscript, we used the graph-based wild and cultivated pangenome to identify sequence variations in the next-generation sequencing data, which have demonstrated that graph-based methods are superior to linear reference genomes in some studies (Sirén et al. 2021; Zhou et al. 2022).

We believe these changes improve the overall clarity and impact of our paper while preserving essential content. We appreciate your thoughtful review, which has helped us refine our work.

Comment (5): Many descriptions in the results did not explicitly mention the datasets used. For example, lines 441-442: "Our analysis identified 442,855,122 SNPs and 13,853 PAVs differentiating indica from japonica." While I checked the methods section, I found that the authors used 90 indica and 36 japonica cultivars. This should be briefly mentioned in the results because different datasets may yield different results and conclusions. Therefore, the authors should also explain why they chose these datasets for each analysis when a new dataset was used.

Response: Thank you for your comment. Following your suggestion, we have now clarified the composition and rationale behind the choice of these datasets in the results section of our revised manuscript. For example, we have updated **Line 466-469** to include a brief description of the datasets used for identifying *indica-japonica* differentiated SNPs and PAVs. We used fewer samples to identify PAVs compared to SNPs because we combined only high-quality non-redundant genome data from 33 cultivated rice pan-genomic (Qin et al. 2021). This selection is crucial for the accurate identification of structural variants.

We appreciate your guidance on improving the manuscript and hope that these revisions meet your expectations for a clearer presentation of our research methods and results.

Comment (6): The readability of the paper needs improvement. The paper had a number of long sentences and paragraphs. For example, the first paragraph of "Extensive variation and graph-based pan-genome" was too long; readers may find it challenging to read such a lengthy paragraph.

Response: We appreciate your suggestions regarding the paper's readability. We have extensively revised the manuscript, particularly the sections on "Gene-based and graph-based pan-genomes construction" (**Line 232**) and " Extensive variations and transposable elements" (**Line 271**). These revisions include shortening long sentences and breaking up overly lengthy paragraphs to enhance clarity and reader engagement. Additionally, we have refined the language throughout to ensure smoother transitions and better coherence between paragraphs.

Reference

Huang, Jingfen, Yilin Zhang, Yapeng Li, Meng Xing, Cailin Lei, Shizhuang Wang, Yamin Nie, et al. 2024. 'Haplotype-Resolved Gapless Genome and Chromosome Segment Substitution Lines Facilitate Gene Identification in Wild Rice'. *Nature Communications* 15 (1): 4573. https://doi.org/10.1038/s41467-024-48845-6.

Jing, Chun-Yan, Fu-Min Zhang, Xiu-Hua Wang, Mei-Xia Wang, Lian Zhou, Zhe Cai, Jing-Dan Han, et al. 2023. 'Multiple Domestications of Asian Rice'.

Lee, Sang-Kyu, Min-Young Song, Young-Su Seo, Hye-Kyung Kim, Seho Ko, Pei-Jian Cao, Jung-Pil Suh, et al. 2009. 'Rice Pi5-Mediated Resistance to Magnaporthe Oryzae Requires the Presence of Two Coiled-Coil–Nucleotide-Binding–Leucine-Rich Repeat Genes'. *Genetics* 181 (4): 1627–38. https://doi.org/10.1534/genetics.108.099226.

Qin, Peng, Hongwei Lu, Huilong Du, Hao Wang, Weilan Chen, Zhuo Chen, Qiang He, et al. 2021. 'Pan-Genome Analysis of 33 Genetically Diverse Rice Accessions Reveals Hidden Genomic Variations'. *Cell* 184 (13): 3542-3558.e16. https://doi.org/10.1016/j.cell.2021.04.046.

Sirén, Jouni, Jean Monlong, Xian Chang, Adam M. Novak, Jordan M. Eizenga, Charles Markello, Jonas A. Sibbesen, et al. 2021. 'Pangenomics Enables Genotyping of Known Structural Variants in 5202 Diverse Genomes'. *Science* 374 (6574): abg8871. https://doi.org/10.1126/science.abg8871.

Zhou, Yao, Zhiyang Zhang, Zhigui Bao, Hongbo Li, Yaqing Lyu, Yanjun Zan, Yaoyao Wu, et al. 2022. 'Graph Pangenome Captures Missing Heritability and Empowers Tomato Breeding'. *Nature* 606 (7914): 527–34. https://doi.org/10.1038/s41586-022-04808-9.

Authors' Response to Referee #3

Referee #3 (Remarks to the Author):

Remarks： Comments on the MS "A pangenome reference of wild and cultivated rice" by Gu et al. This MS reported an effort to construct a pan-genome of Oryza rufipogon (OR), the presumed progenitor of Asian cultivated rice (Oryza sativa, or OS). This study had two key results. The first was the construction of the OR pan-genome comprising a total of 69,531 genes with 13,728 wild rice specific genes based on chromosome-level assemblies of 129 genetically diverse OR accessions and 16 diverse OS varieties. The second was to provide "strong" evidence in supporting their early hypothesis of single origin of OS, i.e. all Asian cultivated rice varieties today was originated from a single domestication event, based on simple population genetic analyses using their OR data combining with 128 OS genome data in the public domain. Although this study generated useful genomic information of the OR population interesting to the public rice scientific community, the total amount, quality and novelty of the reported work were far from the high standard required for publication in Nature. In particular, their conclusions on the OR pan genome and the single origin (domestication) hypothesis of OS were apparently based on incorrect comparisons and misinterpretation of the results. Thus, the MS is not acceptable for publication in Nature. Specific comments are given as follows:

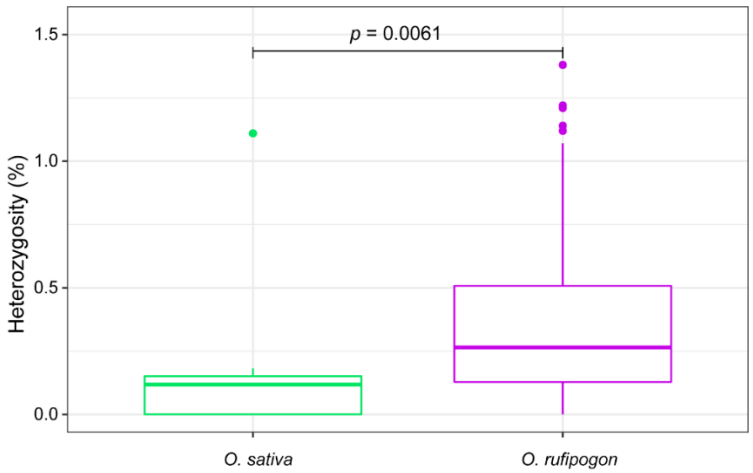
Response: Thanks for the comments.

Comment (1): The average 24.2x genome sequencing coverage for the 133 OR genomes was insufficient to ensure the claimed high-quality assemblies of most OR genomes, given the fact that most OR genomes contain highly heterozygous loci (58.9%). The relatively low quality of the genome assemblies was also evidenced by the unreasonably high heterozygous loci (34.3%) of the 9 OS genomes (Supplementary table 5), which are expected to be low given the self-pollinated nature of the OS accessions, including Nipponbare.

Response: I think these are misunderstanding points. Firstly, I would like to clarify that the percentage of heterozygous alleles mentioned in Supplemental Table 5 (**Supplemental Table 6** in the revised manuscript) refers specifically to the proportions of the heterozygous alleles found in genes on alternate assemblies (a-contigs) of PacBio HiFi genomes, not the overall heterozygous rate of the whole genome. To address it, we have added more comprehensive information on heterozygous rates in **Supplemental Table 6** in the revised manuscript. Our analysis shows that the average heterozygous rates are 0.37% for *O. rufipogon* and 0.08% for *O. sativa* (**Response Fig. 11**).

Secondly, as to the genome sequencing coverage for the 149 rice genomes, 16 of them were sequenced by ONT technology (103.0X ONT sequence depth, with

82.2X short reads for genome polish). And the other 133 rice genomes were sequenced by PacBio circular consensus sequencing (CCS), known for its ability to generate highly accurate (99.8% for single long-read) long high-fidelity (HiFi) reads (Wenger et al. 2019). Therefore, such high accuracy has led to its increasing adoption across various pangenome studies. For instance, in a human pangenome (much bigger and heterozygous genome compared to rice genome) study involving species with higher heterozygosity rates than wild rice, an average HiFi read depth of 28.82X (14.29-60.67X) has been used to construct a pangenome reference (Gao et al. 2023). In our study, the average HiFi read depth for the rice genomes was 24.2X (17.0-43.6X). This depth has proven to be adequate, as multiple indicators, particularly quality values (Q50, or >99.999% accuracy), have confirmed that we achieved the high-quality of genome assemblies of the 133 rice varieties (**Supplementary Table 3**), which were sufficient to the subsequent analysis. I hope this clarification will address the concerns raised.



Response Fig. 11 Box plot representing the distribution of heterozygosity in *O. sativa* and *O. rufipogon* populations. *O. sativa* and *O. rufipogon* are marked in green and purple, respectively.

Supplementary Table 3. Comparison of average assembly evaluation indexes for PacBio HiFi and ONT genomes.

	PacBio HiFi	ONT
Depth	24.20	103.00
Contig N50 (Mb)	14.69	17.04
LAI	24.42	21.73
BUSCO (%)	98.55	98.55
QV	57.92	24.55

Reference

Wenger, Aaron M., Paul Peluso, William J. Rowell, Pi-Chuan Chang, Richard J. Hall, Gregory T. Concepcion, Jana Ebler, et al. 2019. 'Accurate Circular Consensus Long-Read Sequencing Improves Variant Detection and Assembly of a Human Genome'. Nature Biotechnology 37 (10): 1155–62. https://doi.org/10.1038/s41587-019-0217-9.

Gao, Yang, Xiaofei Yang, Hao Chen, Xinjiang Tan, Zhaoqing Yang, Lian Deng, Baonan Wang, et al. 2023. 'A Pangenome Reference of 36 Chinese Populations'. Nature 619 (7968): 112–21. https://doi.org/10.1038/s41586-023-06173-7.

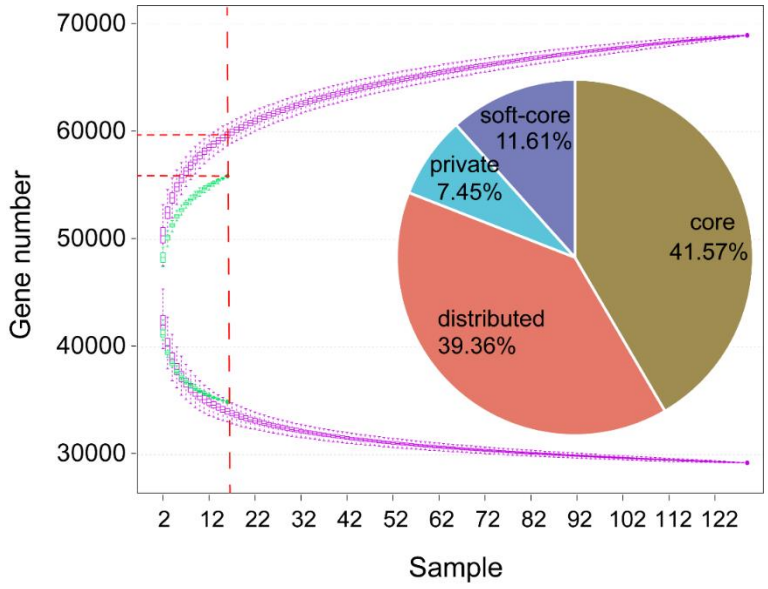
Comment (2): Because pan genome is a property of population or species, the comparison between the OR and OS genomes in pan-genome parameters of this study was invalid given the extremely small population size (16) of the OS samples. For example, the authors constructed the OR pan genome consisting of 69,531 genes, which was much bigger than the OS pan genome from the 16 OS accessions. However, Zhang et al. (2022) reported that the OS pan genome obtained from 105 OS accessions consists of 70,624 protein-coding genes, which is equivalent to the OR pan-genome reported in this study. Regarding the 13,728 wild rice specific genes, the reviewer did not see any evidence how the authors made the comparison between the OR pan-genome with the previously reported OS pan-genome to determine the 13,728 wild rice specific genes. According to Zhang et al. (2022), there are, on average, ~0.2% private genes in each OS accession. If so, most of the 13,728 wild rice specific genes could be well attributable to the small sample size of the OS accessions used in this study. This was obvious in their statement "We observed a greater abundance and diversity of RGAs in O. rufipogon, averaging 1,710, compared to around 1,664 RGAs in the 16 OS accessions". This difference by 46 (~2.7%) in the number of RGAs, is insignificant considering the possible large variation in RGA number among different accessions within OS or among different accessions within an OS population such as indica or japonica.

Response: Thanks for the concerns. I try to sort out the related questions from your concerns as following: 1) sample sizes for OR and OS; 2) annotated pan-genes and comparison with previous publications; 3) the number of RGAs in OR and OS.

I would like to clarify our approach and address the concerns:

Firstly, for sample size. As we have mentioned in the manuscript that most of current researches on the rice pan-genome has mainly focused on cultivated rice populations. So, this study focused on construction of a high-quality pangenome of wild rice *O. rufipogon* populations with a comprehensive collection of genetically and geographically diverse 129 *O. rufipogon* accessions and 16 *O. sativa* varieties. Our results showed that the *O. sativa* sample size does not affect

identification of pan-genes in wild rice. **Response Fig. 12** showed clearly that the number of pan-genes has reached ~60,000 for 16 *O. rufipogon* genomes, while that of the pan-genes only reached ~55,000 for 16 *O. sativa* genomes (**Fig.1i** in the revised manuscript).



Response Fig. 12 The detrended growth curve of the pan- (up) and core- (down) gene number of 16 *O. sativa* (green curve) and 129 *O. rufipogon* (purple curve) varieties. The red dashed lines represent the number of pan-genes reached by 16 *O. rufipogon* and 16 *O. sativa* genomes, respectively.

Secondly, for genome annotation and comparison. I would like to clarify that we performed *de novo* assembly for each individual genome and annotated genes especially using transcriptome data from 3-4 tissues per genome (encompassing a total of 554 tissue samples for this study). This approach allows us to identify high-confidence genes for each genome separately and construct a comprehensive pan-genome. This strategy differs from the "map-to-pan" method used in the 111 rice pan-genome study (105 *O. sativa* and **6 *O. rufipogon***), where pan-genes consist of a reference genome (55,986 genes) and additional genes from non-redundant novel sequences (19,319 genes). Given these methodological differences, direct comparisons of gene numbers between our pan-genome and those from previous studies may not be appropriate.

Following your suggestion, we have compared our identified wild rice-specific genes with three published rice pan-genomes (Qin et al. 2021; Zhang et al. 2022; Shang et al. 2022) (**Supplementary Table 10** in the revised manuscript). As expected, the number of genes exclusive to wild rice is different in this broader comparison. However, the results demonstrate that most of the wild rice-specific genes we identified (63%-82%) remain unique, even when compared against a larger set of cultivated rice genomes. These are truly important and novel findings.

Supplementary Table 10 The proportion of wild rice specific genes we identified in recent cultivated rice pangenome studies (Qin et al. 2021; Zhang et al. 2022; Shang et al. 2022).

	Homologous genes (13,728)	Total genes (90,805)
33 O. sativa	17.47%	9.46%
105 *O. sativa* + 6 *O. rufipogon*	18.44%	27.15%
202 *O. sativa*	36.87%	34.69%

Thirdly, for the relatively small difference in the average number of RGAs between *O. rufipogon* and cultivated rice. As you mentioned, we have already made the necessary updates to the manuscript to reflect this point (**Line 206-208**). More importantly, despite the slight numerical difference, the RGAs in *O. rufipogon* exhibit a broader diversity, both in terms of the total number of genome-wide resistance genes and the diversity of genes at specific loci (**Fig.1g and Supplementary Fig. 7b** in the revised manuscript). This extensive diversity provides a robust gene pool and donor resources for enhancing disease resistance in cultivated rice. This aspect underscores the significant potential of wild rice in contributing to the genetic improvement of cultivated varieties.

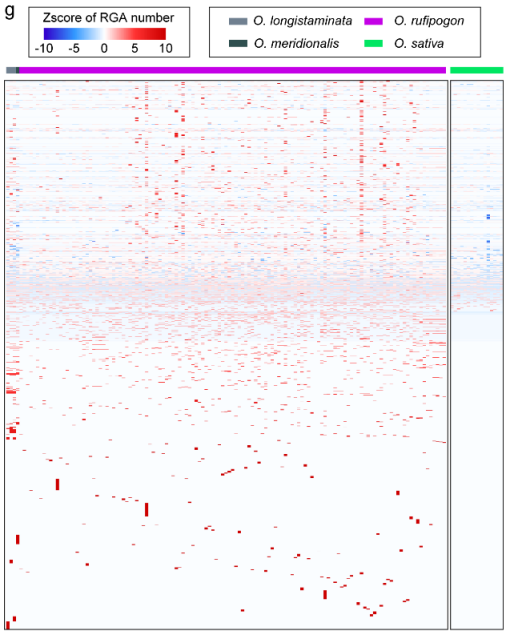
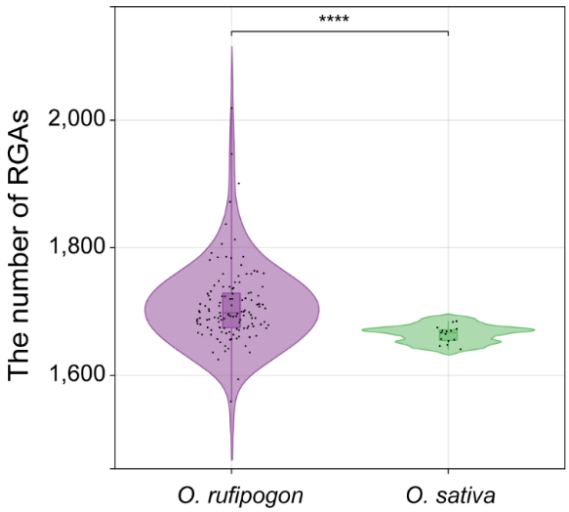


Fig.1g The landscape of RGA loci that have a higher average copy number in wild rice than cultivated rice. Each row indicates a RGA locus, and each column indicates a variety. The data was normalized using a row-wise Z-score method.



Supplementary Fig. 7b Differences in RGAs number between 16 *O. sativa* and 129 *O. rufipogon*. The 25% and 75% quartiles are marked by the lower and upper box edges in boxplots, with the median represented by the central line. Whiskers extend up to 1.5 times the interquartile range. Statistical significance was assessed using two-sided Wilcoxon tests, labeled above the plots (****p value $<$ 0.0001).

I think this additional analysis strengthens our findings while addressing the concerns about sample size and population representation of cultivated rice.

Reference

Qin, Peng, Hongwei Lu, Huilong Du, Hao Wang, Weilan Chen, Zhuo Chen, Qiang He, et al. 2021. 'Pan-Genome Analysis of 33 Genetically Diverse Rice Accessions Reveals Hidden Genomic Variations'. Cell 184 (13): 3542-3558.e16. https://doi.org/10.1016/j.cell.2021.04.046.

Shang, Lianguang, Xiaoxia Li, Huiying He, Qiaoling Yuan, Yanni Song, Zhaoran Wei, Hai Lin, et al. 2022. 'A Super Pan-Genomic Landscape of Rice'. Cell Research 32 (10): 878–96. https://doi.org/10.1038/s41422-022-00685-z.

Zhang, Fan, Hongzhang Xue, Xiaorui Dong, Min Li, Xiaoming Zheng, Zhikang Li, Jianlong Xu, Wensheng Wang, and Chaochun Wei. 2022. 'Long-Read Sequencing of 111 Rice Genomes Reveals Significantly Larger Pan-Genomes'. Genome Research 32 (5): 853–63. https://doi.org/10.1101/gr.276015.121.

Comment (3): In addition to its economic importance as the staple food feeding most Asian people, OS is well-known for its wide geographical distribution, subspecific and population differentiation, and extremely rich genetic/genomic

diversity within populations. Thus, the origin (domestication) of OS has been an issue of long debate, with two major hypotheses: the multiple origins of OS and single origin of OS. The fundamental difference between the two primary hypotheses is that OS was domesticated in different geographic locations at different times in the former hypothesis, while domestication of OS was presumed to have occurred only once in one location in the latter. In the reviewer's opinion, any conclusion on this issue has to reconcile with all evidence from genetic/genomic studies and archeological discoveries. The authors were apparently trying to obtain evidence in supporting the hypothesis of single origin, i.e. "OS was first domesticated as proto-japonica cultivars from OR population III (Or-III) in China and indica cultivars were subsequently domesticated as proto-japonica cultivars, which were then spread southward and eastwards in Asia and crossed with local OR population group I (Or-I)". Unfortunately, the authors did not perform sufficient population genetic analyses to support or reject any of the hypotheses, and their conclusions were based on the misinterpretation of their results with the following three major problems.

Comment (3.1): First, one most obvious expectation from their hypothesis of single origin was a severe genetic bottleneck and dramatic reduced genetic diversity genomewide in OS populations. With the inclusion of additional 128 OS accessions from the public domain in their data analyses, both OR and OS showed clear population differentiation of geographic correspondence (Fig. 3a). The phylogenetic analyses resolved the OR accessions into three well-differentiated populations, OrIII classified with OS japonica accessions, OrI with OS indica accessions and OrII in between, consistent with three previous reports based on SNPs from a low-coverage (~1x) sequencing of the much large OR and OS populations containing all OR and OS accessions used in this study. Compared to the OR populations, the OS populations (indica, japonica, aus and basmati) showed greater population differentiation but slightly reduced within population diversity (Fig. 3c). These results were consistent with virtually all previous reports of similar studies. The observed reduction of genetic diversity in japonica, when compared with its presumed ancestor, OrIIIa, was ~57%, indicating presence of a significant level of genetic diversity within population japonica, consistent with that reported by Wang et al. (2022). More severe genetic bottlenecks would have been expected for populations indica, aus and basmati, if they were derived from a single domestication event of japonica. Thus, there is no evidence for severe genetic bottlenecks (greatly reduced diversity genomewide) observed in all OS populations when compared with their most closely related OR populations of the same geographic origin.

Response: I think these comments are much involved in the previous studies. I am not intended to refer too much or comment on the previous studies on rice domestications, which have demonstrated clear and solid conclusion on the

evolutionary origin and domestication process of Asian cultivated rice. Most of the related papers were cited in this manuscript, and no evidence proved that those previous studies were based on misinterpretation of the results. In response the particular question regarding to this study, I would like to aggress it below:

The **Fig. 3c** showed accurate nucleotide diversity (π) in each group and mean population differentiation (F_{st}) between groups. In response to the concerns on the genetic diversity and genetic bottlenecks in cultivars, I would like to clarify that the hypothesis of a shared domestication event does not necessarily imply severe genetic bottlenecks and dramatically reduced genetic diversity across the entire genome of all *O. sativa* populations. It means that the reduced genetic diversity on some domestication trait loci in cultivars due to the genetic bottlenecks is not directly associated with the reduced whole-genome genetic diversity under long-term of artificial and natural selections. As the reviewer mentioned that "OS is well-known for its wide geographical distribution, subspecific and population differentiation, and extremely rich genetic/genomic diversity within populations". The modern *O. sativa* cultivars including *indica* and *japonica* subspecies, which have undergone long-term and many cycles of artificial and natural selections. These cultivars have been identified to be domesticated through crossing the proto-*japonica* type cultivar (with multi-domestication traits, such as non-shattering seed habit, from prostrate to erect growth, transition from black to straw-white seed hull, etc) with diverse wild rice *O. rufipogon* populations for adapting various environments (Gutaker et al. 2020). That is why the *O. sativa* cultivars still maintain genome-wide genetic diversity in spite of possessing reduced genetic diversity of several domestication trait loci. That is also why we are trying very hard to construct pan-genomes of diverse *O. rufipogon* populations for the rice genetic study and future rice breeding.

Furthermore, I may have to briefly explain to the study of the previous model that the proto-*japonica* rice was initially, domesticated from an Or-IIIa population. Subsequently, the ancestors of other cultivated rice varieties interbred with proto-*japonica* to acquire those key domestication traits (or loci). These ancestors diverged and adapted to diverse environmental conditions before domestication. Domestication primarily affects a limited number of genes controlling key traits (Stetter et al. 2017; Stetter, n.d.; Andersson and Purugganan 2022). While these specific loci may show reduced diversity, the rest of the genome can retain substantial diversity. This is reflected in our findings where specific loci show reduced diversity, but overall genetic variability remains high.

Indica has a high genetic diversity and does not experience severe genetic bottlenecks, indicating that it has not experienced intense artificial selection similar to *japonica*. This further illustrates the impossibility of *indica* independent domestication. Moreover, our model supports the idea that *indica* acquired domestication genes through gene flow, allowing it to retain more of its ancestral genetic diversity for adapting to the various environmental conditions.

In summary, our analysis indicates that while domestication has influenced the genetic diversity of specific genetic regions, the broader genetic landscape of *O. sativa* showcases a complex history of multiple ancestors and a shared domestication event, rather than a single, severe bottleneck.

Reference

Andersson, Leif, and Michael Purugganan. 2022. 'Molecular Genetic Variation of Animals and Plants under Domestication'. *Proceedings of the National Academy of Sciences of the United States of America* 119 (30): e2122150119. https://doi.org/10.1073/pnas.2122150119.

Gutaker, Rafal M., Simon C. Groen, Emily S. Bellis, Jae Y. Choi, Inês S. Pires, R. Kyle Bocinsky, Emma R. Slayton, et al. 2020. 'Genomic History and Ecology of the Geographic Spread of Rice'. *Nature Plants* 6 (5): 492–502. https://doi.org/10.1038/s41477-020-0659-6.

Stetter, Markus G. 'Limits and Constraints to Crop Domestication'. *American Journal of Botany* 107(12): 1617-1621. https://doi.org/10.1002/ajb2.1585.

Stetter, Markus G., Daniel J. Gates, Wenbin Mei, and Jeffrey Ross-Ibarra. 2017. 'How to Make a Domesticate'. *Current Biology* 27 (17): R896–900. https://doi.org/10.1016/j.cub.2017.06.048.

Comment (3.2): Secondly, the authors' inferences on the separation history of different OR and OS populations (Fig. 4) were largely reached based on their analytic results of the OR and OS genomic sequencing data using the Multiple Markovian Coalescent (MSMC2) method. However, according to Mather et al. (2020), the MSMC2 method developed to infer past population sizes and divergence time of related lineages has two assumptions, the neutrality of genome sequence variation and random mating of the studied population. When the assumptions are violated, the parameters of the demographic history inference obtained by MSMC2 could be incorrect and misleading. In particular, when MSMC2 is used for highly structured populations such as OR and OS, the results are likely to be extremely misleading. Moreover, the impact of population structure on the results of MSMC2 cannot be assessed directly from the observed coalescence times. "Concerningly, this problem appears to grow worse as larger numbers of genomes are used for MSMC2, so the problem cannot necessarily be overcome by adding more data" (Mather et al. 2020). Now, both the OR and OS accessions used in this study are highly structured. OS is known to be a highly self-pollinated species and the OR accessions used in this study are not random-mating populations (though OR accessions have higher outcrossing rates than OS accessions). This was clearly seen from their well-differentiated population structure. Most importantly, the OR and OS accessions from different geographic origins are known to adapt to a diverse range of environments, each carrying large numbers of genes/alleles that had gone through various kinds of selection during

evolution. Thus, both underlying assumptions of the MSMC2 method are severely violated in this study. Thus, the authors are required to explain how this had affected their results and conclusions.

Response: Thank you for the concern. As noted in the reviewing literature you referenced, MSMC2 has been widely used in population genetics research (Jing et al. 2023; Liu et al. 2023). We believe that fully considering the high population structure of *O. sativa* and *O. rufipogon* will be helpful to accurately estimate or infer the population separation history in the future. However, we also believe that the relative timing of differentiation between different groups, as inferred by this method of MSMC2, still provides valuable insights. To ensure rigor and clarity, we have moved this section to the supplementary materials (**Supplementary Fig. 13b**).

Reference

Jing, Chun-Yan, Fu-Min Zhang, Xiu-Hua Wang, Mei-Xia Wang, Lian Zhou, Zhe Cai, Jing-Dan Han, et al. 2023. 'Multiple Domestications of Asian Rice'. Nature Plants 9 (8): 1221–35. https://doi.org/10.1038/s41477-023-01476-z.

Liu, Chunhui, Tianyi Wang, Huicha Chen, Xiaoding Ma, Chengzhi Jiao, Di Cui, Bing Han, et al. 2023. 'Genomic Footprints of Kam Sweet Rice Domestication Indicate Possible Migration Routes of the Dong People in China and Provide Resources for Future Rice Breeding'. Molecular Plant 16 (2): 415–31. https://doi.org/10.1016/j.molp.2022.12.020.

Comment (3.3): Thirdly, it is well-known that results from population genetic analyses are extremely sensitive to sampling, and the sampling problem in this study were reflected in at least two aspects: (1) the sequenced OR and OS accessions do not cover the whole spectrum of OR and OS geographic distributions, because the sampled OR accessions were collected in relatively recent time and most and real 'ancestor' OR populations got lost as their habitats were largely occupied by OS varieties over time worldwide; and (2) the few selected "domestication genes" for detailed analyses by the authors only account for a very small portion of rice genes that had gone through varied types of selection during the process of domestication. In the former case, for example, the OR and OS accessions used in this study failed to include those (japonica landraces and OrIII accessions) from the foothills of the Himalayas (North India, Nepal and North Myanmar). In the latter cases, the sampled 11 "domestication-genes" used in gene PAV analysis (Fig. 4d) were too few to reveal the whole history of domestication. According to a recent report using much larger population sizes of OR and OS populations, Wang et al. (2022) reported a total of 14,202 and 5,163 domestication selected genes in OS subspecies japonica and indica, respectively, with 3,115 shared between the two subspecies. Of these, only a very small proportion of (3.4 %

in japonica and 1.9 % in indica) of the domestication selected genes were attributable to genetic bottlenecks that either showed uniform low polymorphism or no variation in gene CDS regions. They found that selection in japonica and indica during domestication tended to act on different genes in the same pathways, on different functional alleles of the same loci, or on different members of the same gene families, indicating that japonicas and indicas may have compensatory regulation and functional genetic networks affecting the same phenotypes for their adaptations to different environments and their responses to artificial selection during the process of domestication, indicating that genetic bottleneck played little role during domestication of both japonica and indica. These results led them to the proposition of the multiple origins (domestication) of OS. Unfortunately, the authors of this study failed to perform appropriate analyses which would allow them to test the major hypotheses regarding the OS origin.

Response: Thank you for your concern regarding the sampling issues in our population genetic analyses.

Geographic Coverage of Accessions: As noted in our manuscript, we observed significant gene flow from *indica* into some wild rice clades, such as Or-Ia and Or-unspecific, which we excluded from subsequent evolutionary analyses to avoid confounding effects. However, other *O. rufipogon* groups still exhibit wild rice characteristics both genomically and phenotypically, which are crucial to our analysis.

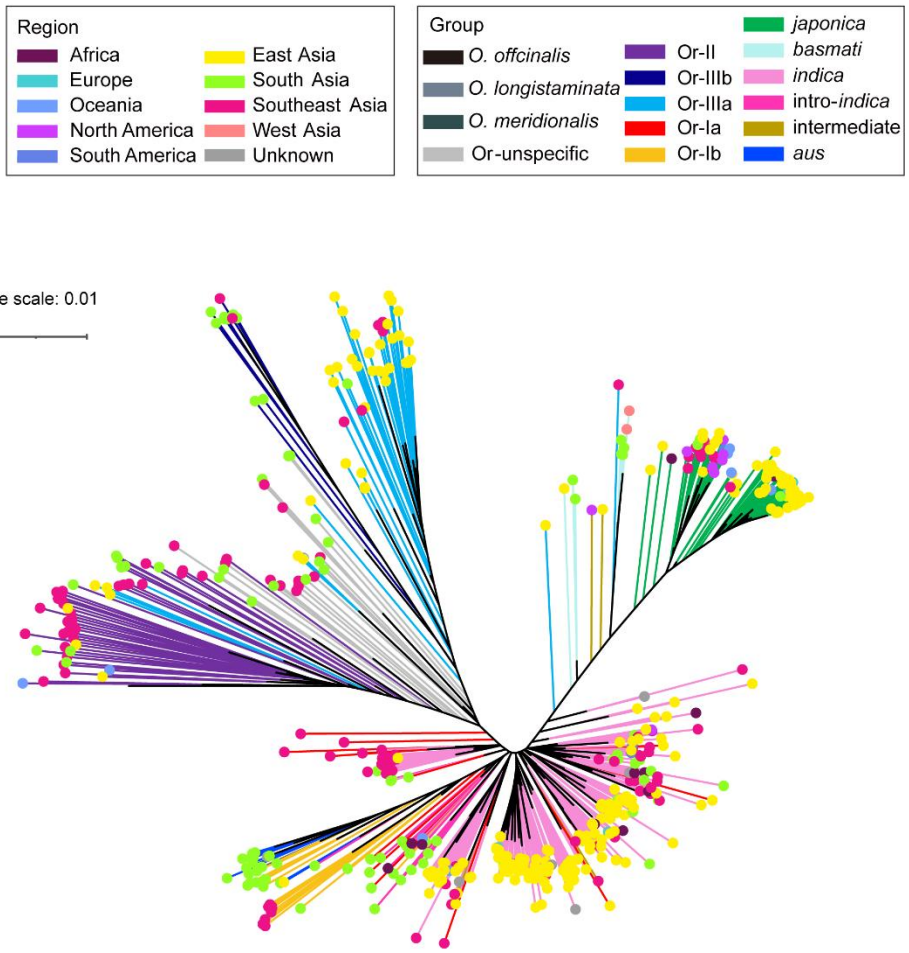
To address the limitations in geographic distribution of our sampled accessions, we supplemented our dataset with published data on cultivated (Shang et al. 2022) and wild rice (Jing et al. 2023). This has nearly integrated all currently available high-quality genome data of *O. sativa* and *O. rufipogon*. This expanded dataset improved the representation of geographic diversity (**Supplementary Table 14** and **Supplementary Table 17** in the revised manuscript). Our supplementary analyses, including phylogenetic trees and archetypal analysis (**Supplementary Fig. 20**, **Supplementary Fig. 28** and **Response Fig. 13**), corroborate our initial findings.

Selection of Domestication Genes: The 11 domestication genes analyzed represent only a subset of those involved in rice domestication. However, these genes were selected based on strong prior evidence from literature indicating their crucial roles in the domestication process. It is important to note that in crop plants, such as rice and maize, a relatively small number of genes were involved in key domestication syndrome (Stetter et al. 2017; Stetter, n.d.; Andersson and Purugganan 2022). In theory, the most comprehensive and accurate set of domestication genes provides a truer picture of the domestication process. We should focus on those genes that have been identified as playing a key role in early domestication, allowing us to draw meaningful conclusions about the process.

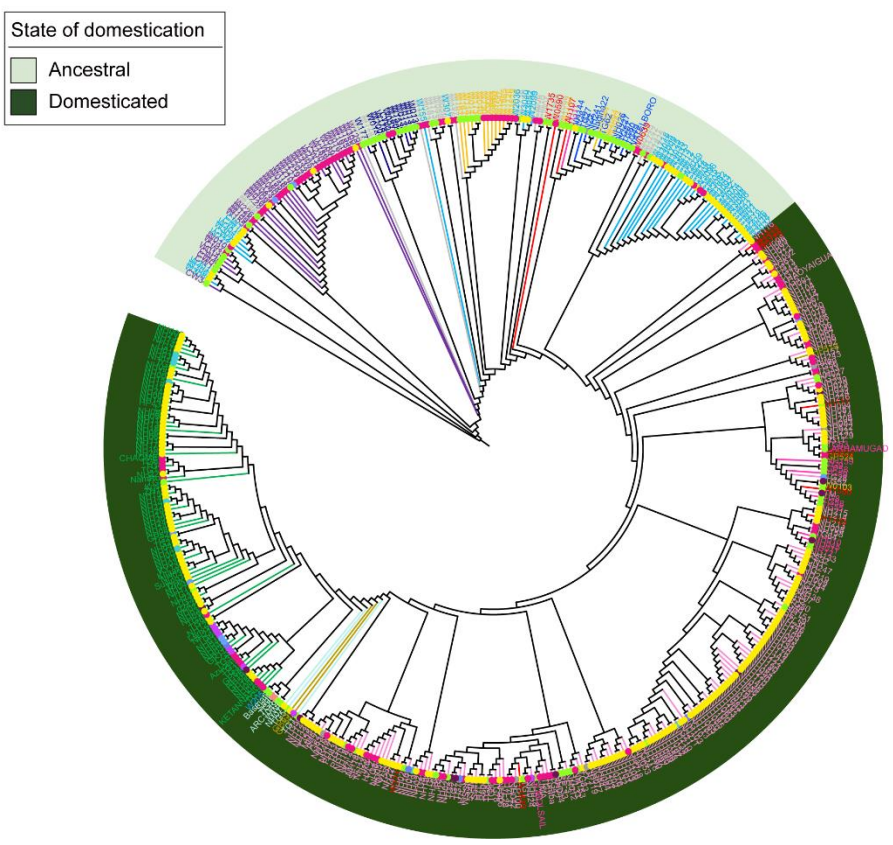
Simply grouping all genes selected by humans into one category may obscure genuine domestication clues.

We believe that our approach, while not without limitations, provides valuable insights into rice domestication history. We remain committed to refining our methods and expanding our datasets in future research to address these complex questions more comprehensively.

a


Region
Africa
Europe
Oceania
North America
South America
East Asia
South Asia
Southeast Asia
West Asia
Unknown
Group
O. offcinalis
O. longistaminata
O. meridionalis
Or-unspecific
Or-II
Or-IIIb
Or-IIIa
Or-Ia
Or-Ib
japonica
basmati
indica
intro-indica
intermediate
aus
Tree scale: 0.01

b


State of domestication
Ancestral
Domesticated

Response Fig. 13 Phylogenetic analysis of 510 wild and cultivated rice accessions. **a**, The external circles denote the geographic distribution of each accession. Tree branches are color-coded to represent group information. **b**, Phylogenetic tree of 510 Oryza accessions based on the SNPs within identified selective sweep regions. The tree features color-coded labels and branches to distinguish various groups, and the surrounding external circles represent the geographic distribution of each accession. The color ranges indicate the state of domestication.

Reference

Andersson, Leif, and Michael Purugganan. 2022. 'Molecular Genetic Variation of Animals and Plants under Domestication'. *Proceedings of the National Academy of Sciences of the United States of America* 119 (30): e2122150119. https://doi.org/10.1073/pnas.2122150119.

Jing, Chun-Yan, Fu-Min Zhang, Xiu-Hua Wang, Mei-Xia Wang, Lian Zhou, Zhe Cai, Jing-Dan Han, et al. 2023. 'Multiple Domestications of Asian Rice'. *Nature Plants* 9 (8): 1221–35. https://doi.org/10.1038/s41477-023-01476-z.

Shang, Lianguang, Xiaoxia Li, Huiying He, Qiaoling Yuan, Yanni Song, Zhaoran Wei, Hai Lin, et al. 2022. 'A Super Pan-Genomic Landscape of Rice'. *Cell Research* 32 (10): 878–96. https://doi.org/10.1038/s41422-022-00685-z.

Stetter, Markus G. 'Limits and Constraints to Crop Domestication'. *American Journal of Botany* 107(12): 1617-1621. https://doi.org/10.1002/ajb2.1585.

Stetter, Markus G., Daniel J. Gates, Wenbin Mei, and Jeffrey Ross-Ibarra. 2017. 'How to Make a Domesticate'. *Current Biology* 27 (17): R896–900. https://doi.org/10.1016/j.cub.2017.06.048.

Comment (4): In addition to above comments, the authors have to provide answers to the following two questions to support their hypothesis of single origin of OS:

Comment (4.1). First, it is well-known that the subspecific (japonica-indica) differentiation of OS involves a large portion of their genome constitutions, as indicated in the MS "We identified a total of 855,121 indica–japonica differentiated SNPs (Supplementary Table 16) based on a very stringent standard of 0.9 frequency difference. We also applied the same standard to identify 13,853 indica-japonica differentiated PAVs between 26 indica cultivars and 13 japonica cultivars". These results were consistent with the results of the 3,000 rice genome project (Wang et al. 2018; Zhang et al. 2021) that the indica–japonica subspecific differentiation represents the maximum direction of within OS diversity, and indica and japonica each contains many (hundreds of) subspecific genes/gene families, and thousands of subspecific alleles (functional haplotypes) at >13,000

gene loci (Zhang et al. 2021). It is also known for the presence of high levels of reproductive barriers (hybrid sterility and hybrid breakdown) between japonica and indica. The mountainous areas along the Himalayas including Southwest China are known as the center of diversity for OS, where various ecotypes of OS, ranging from typical indica to temperate japonica and upland (japonica) ecotype, are distributed across different longitudes and show tremendous phenotypic and genetic diversity. However, the recorded history of OS cultivation in Southwest China was <3,000 years. Then, the authors have to explain what genetic mechanism(s) could have generated such high levels of genetic diversity, population differentiation and reproductive barriers in >3,000 years if all OS accessions were originated from single ancestor not adapted to the diverse environments along the Himalayas. The authors proposed that the hybridization and introgression from population OrI into "ancient japonica" to explain how indica and aus were derived (Fig. 4e). However, this explanation is implausible because of the severe hybrid sterility and hybrid breakdown between japonica and OrIa would have blocked the massive gene flow between them. The alternative explanation would be that indica and aus accessions of South Asia were more likely to have been directly domesticated from the local OrIa and Or1b at ancient times independent from the domestication events in China, while the observed introgression between indica and OsIa and between aus and OsIb interpretation was more likely to have occurred after domestication. Thus, the authors are required to provide evidence to reject the alternative explanation.

Response: I would like to clarify that our model does not propose a single ancestor for all types of Asian cultivated rice. Instead, we suggest that different groups of *O. sativa* just shared a common domestication event, with their respective ancestors diverging and adapting to distinct ecological environment before this event. This resulted in significant genetic differentiation and diversity among these groups of *O. rufipogon*. Subsequent artificial selection and genetic improvements further enhanced this diversity, tailoring rice varieties to meet diverse farming and culinary needs (Jiang et al. 2022). Therefore, *indica*, through shares key domestication loci with *japonica*, shows larger diversity and higher genetic differentiation from *japonica*. The domestication process of wild to cultivated rice was gradual; *japonica* experienced intense artificial selection and significant genetic bottlenecks, whereas *indica* could have rapidly incorporated domestication genes through gene flow, preserving its diversity. This suggests that *indica* did not experience a separate, *de novo* domestication, challenging the multiple domestication hypothesis.

Regarding your concern about hybrid sterility between *japonica* and Or-Ia, it is that reproductive barriers exist. However, these barriers are not complete or absolute. The outcrossing nature and exposed stigmas of *O. rufipogon* allow for pollen acceptance from various sources, including cultivated rice. Even limited successful hybridization can significantly impact evolution and domestication.

Previous research by Wang et al. has shown evidence of gene flow between domesticated and wild rice (Wang et al. 2017). Additionally, the successful creation of chromosome segment substitution lines (CSSLs) from wild and cultivated rice crosses (Huang et al. 2012; Huang et al. 2024), and the existence of weedy rice (Li et al. 2017; Wu et al. 2023), further support the feasibility of such hybridization.

In conclusion, while hybrid sterility is a consideration, it does not preclude the possibility of *indica* acquiring domestication genes through gene flow. Our model aims to reconcile the observed genetic patterns with the complex history of rice domestication.

Reference

Huang, Jingfen, Yilin Zhang, Yapeng Li, Meng Xing, Cailin Lei, Shizhuang Wang, Yamin Nie, et al. 2024. 'Haplotype-Resolved Gapless Genome and Chromosome Segment Substitution Lines Facilitate Gene Identification in Wild Rice'. *Nature Communications* 15 (1): 4573. https://doi.org/10.1038/s41467-024-48845-6.

Huang, Xuehui, Nori Kurata, Xinghua Wei, Zi-Xuan Wang, Ahong Wang, Qiang Zhao, Yan Zhao, et al. 2012. 'A Map of Rice Genome Variation Reveals the Origin of Cultivated Rice'. *Nature* 490 (7421): 497–501. https://doi.org/10.1038/nature11532.

Jiang, Linlin, Lian Wu, Yun Wang, Quan Xu, Zhengjin Xu, and Wenfu Chen. 2022. 'Research Progress on the Divergence and Genetic Basis of Agronomic Traits in Xian and Geng Rice'. *The Crop Journal* 10 (4): 924–31. https://doi.org/10.1016/j.cj.2022.02.006.

Li, Lin-Feng, Ya-Ling Li, Yulin Jia, Ana L. Caicedo, and Kenneth M. Olsen. 2017. 'Signatures of Adaptation in the Weedy Rice Genome'. *Nature Genetics* 49 (5): 811–14. https://doi.org/10.1038/ng.3825.

Wang, Hongru, Filipe G. Vieira, Jacob E. Crawford, Chengcai Chu, and Rasmus Nielsen. 2017. 'Asian Wild Rice Is a Hybrid Swarm with Extensive Gene Flow and Feralization from Domesticated Rice'. *Genome Research* 27 (6): 1029–38. https://doi.org/10.1101/gr.204800.116.

Wu, Dongya, Lingjuan Xie, Yanqing Sun, Yujie Huang, Lei Jia, Chenfeng Dong, Enhui Shen, Chu-Yu Ye, Qian Qian, and Longjiang Fan. 2023. 'A Syntelog-Based Pan-Genome Provides Insights into Rice Domestication and de-Domestication'. *Genome Biology* 24 (1): 179. https://doi.org/10.1186/s13059-023-03017-5.

Comment (4.2) Secondly, the earliest archeological evidence in China (lower Yangtze River) and India (lower Ganges River) has traced the cultivation of OS back to ~12,000 years and ~9,000 years ago, respectively (Fuller et al. 2010; Gross and Zhao 2014; Ma et al. 2016; Zuo et al. 2017). Thus, the authors have to provide

evidence for the ancient human migration between the two countries that occurred at least 9,000 years ago if their hypothesis of single origin for OS holds true.

Response:

The archaeological studies about timing and spread route of rice domestication continues to be a vibrant area. A recent publication in *Science* shed new light on this topic. By sampling rice phytoliths from two archaeological sites near the lower reaches of the Yangtze River, researchers demonstrated that selection and/or predomestication cultivation of Asian rice began around 13,000 years ago, and domestication began around 11,000 years ago (Zhang et al. 2024). However, archaeological data from the Ganges basin are only starting to be accumulated, and as a result, the domestication status of proto-*indica* rice in South Asia before the introduction of domesticated japonica rice remains inconclusive (Fuller et al. 2010; Zhou et al. 2022). Despite these gaps, there is growing archaeological evidence that domesticated *japonica* rice spread from China along the Silk Rode, accompanying other crops (such as millet) and farm tools (such as the distinctive stone harvesting knives). The trade route started in the Yellow River valley, extending westward through the Hexi Corridor, then just south of the Inner Asian Mountains and thence to India (Fuller et al. 2010; Fuller 2011). This hypothesis had been reinforced by spatial modeling and simulations (Silva et al. 2018). Furthermore, an alternative southern migration route, proposed by other archaeologists, stretched from China through Myanmar, Assam, and Bangladesh to the Indian Plains (Kovach, Sweeney, and McCouch 2007; Vaughan, Lu, and Tomooka 2008; Ikehashi 2014). Regardless of the path, once domesticated *japonica* encountered local unimproved proto-*indica* or wild populations, hybridization and subsequent selection by local farmers led to the rapid development of the *indica* subgroup.

However, it is important to note that determining the origins of cultivated rice through archaeology alone is challenging (Jing et al. 2023). This is due to difficulties in distinguishing between cultivar subgroups based on the archaeobotanical remains and limited archaeological explorations in South and Southeast Asia, poor preservation conditions in tropical climates. We all believe that the advancements in genomics, functional studies offer an unprecedented opportunity to explore longstanding questions about rice domestication. As research progresses, we expect to gain a clearer understanding of the complex history of rice origin and its domestication across Asia.

Reference

Fuller, Dorian Q. 2011. ‘Pathways to Asian Civilizations: Tracing the Origins and Spread of Rice and Rice Cultures’. *Rice* 4 (3): 78–92. https://doi.org/10.1007/s12284-011-9078-7.

Fuller, Dorian Q., Yo-Ichiro Sato, Cristina Castillo, Ling Qin, Alison R. Weisskopf, Eleanor J. Kingwell-Banham, Jixiang Song, Sung-Mo Ahn,

and Jacob van Etten. 2010. ‘Consilience of Genetics and Archaeobotany in the Entangled History of Rice’. *Archaeological and Anthropological Sciences* 2 (2): 115–31. https://doi.org/10.1007/s12520-010-0035-y.

Ikehashi, Hiroshi. 2014. ‘Domestication and Long-Distance Dissemination of Rice: A Revised Version’. *Rice Research: Open Access* 03 (01). https://doi.org/10.4172/2375-4338.1000128.

Jing, Chun-Yan, Fu-Min Zhang, Xiu-Hua Wang, Mei-Xia Wang, Lian Zhou, Zhe Cai, Jing-Dan Han, et al. 2023. ‘Multiple Domestications of Asian Rice’. *Nature Plants* 9 (8): 1221–35. https://doi.org/10.1038/s41477-023-01476-z.

Kovach, Michael J., Megan T. Sweeney, and Susan R. McCouch. 2007. ‘New Insights into the History of Rice Domestication’. *Trends in Genetics* 23 (11): 578–87. https://doi.org/10.1016/j.tig.2007.08.012.

Silva, Fabio, Alison Weisskopf, Cristina Castillo, Charlene Murphy, Eleanor Kingwell-Banham, Ling Qin, and Dorian Q Fuller. 2018. ‘A Tale of Two Rice Varieties: Modelling the Prehistoric Dispersals of *Japonica* and Proto- *Indica* Rices’. *The Holocene* 28 (11): 1745–58. https://doi.org/10.1177/0959683618788634.

Vaughan, Duncan A., Bao-Rong Lu, and Norihiko Tomooka. 2008. ‘Was Asian Rice (Oryza Sativa) Domesticated More Than Once?’ *Rice* 1 (1): 16–24. https://doi.org/10.1007/s12284-008-9000-0.

Zhang, Jianping, Leping Jiang, Lupeng Yu, Xiujia Huan, Liping Zhou, Changsheng Wang, Jianhui Jin, et al. 2024. ‘Rice’s Trajectory from Wild to Domesticated in East Asia’. *Science*. https://doi.org/DOI: 10.1126/science.ade4487.

Zhou, Jiawu, Ying Yang, Yonggang Lv, Qiuhong Pu, Jing Li, Yu Zhang, Xianneng Deng, Min Wang, Jie Wang, and Dayun Tao. 2022. ‘Interspecific Hybridization Is an Important Driving Force for Origin and Diversification of Asian Cultivated Rice Oryza Sativa L.’ *Frontiers in Plant Science* 13 (June):932737. https://doi.org/10.3389/fpls.2022.932737.

Authors' Response to Referee #4

Referee #4 (Remarks to the Author):

Remarks： In this manuscript, Professor Han and colleagues provide an updated pangenome of wild and cultivated rice. The authors added 145 chromosome-level assemblies from a 129 genetically diverse Oryza rufipogon and 16 diverse O. sativa varieties. They integrated this data with previously published data to provide an important new pan-genome resource, for example, they identified resistance genes in wild rice not found in cultivated rice. Overall, the study is well-executed, and I have no technical objections to the methods used or the results presented.

Comment (1): Previous rice pan-genomes or super pan-genomes focused on cultivated varieties (papers cited include Zhao et al. 2018, Qin et al. 2021, Shang et al. 2022, Zhang et al. 2022, Zhou et al. 2023, Huang et al. 2021, Wu et al. 2023; however missing citation for Wang J et al. 2023 Genome Biology 24: article 19). The authors claim that an important scholarly gap is that previous rice pan-genomes focused on cultivated rice. However, looking at the literature cited were some O. rufipogon and other rice species included in previous pan-genomes (e.g., Zhao et al. 2018, Zhou et al. 2023, Wu et al. 2023). As a reviewer, the first question that I have is: What significant new insights were gained by adding additional O. rufipogon genomes?

Response: Thank you very much for your comments. The Asian cultivated rice was domesticated from its close ancestor wild rice *O. rufipogon*. It is that *O. rufipogon* made modern rice varieties or genetic diversity of *O. rufipogon* funded or shaped the genetic diversity of Asian cultivated rice. Although previous studies of rice pan-genomes or super pan-genomes included some *O. rufipogon* accessions, these studies were used insufficient wild rice *O. rufipogon* accessions, and the focus predominantly remained on cultivated varieties (**Supplementary Table 9** in the revise manuscript). Moreover, an analysis of our saturation curve of *O. rufipogon* pan-genome revealed that the number of pan genes for 28 *O. rufipogon* accessions (the highest number of *O. rufipogon* accessions in the super pan-genome of rice) was only 61,769, which is much lower than the number of pan genes (68,901) in our constructed pangenome of 129 *O. rufipogon* accessions. These observations indicate that constructing a nearly saturated *O. rufipogon* pangenome is essential to fully understand the diversity and complexity of the *O. rufipogon* genomes. We have added the reference of Wang J et al. 2023 to the revised manuscript (**Ref19**).

Here is a detailed response addressing the significant new insights gained by adding additional *O. rufipogon* genomes:

1) **High-Quality, Comprehensive Pan-genome Construction**: We sequenced 129 *O. rufipogon* genomes primarily using high-fidelity (HiFi) sequencing

technology, renowned for its long-reads and accuracy. Additionally, we retrieved alternate contigs (a-contigs) that included genetic information missing from the primary assemblies—an aspect previously overlooked in rice pangenome research. Building on this, we constructed the first, the most comprehensive, and highest-quality near-saturated wild rice pangenome to date. It encompasses all ecotypes of *O. rufipogon* and spans their major geographical habitats. This extensive dataset has allowed us to delineate the boundary of wild rice diversity comprehensively.

2) **Discovery of Important Genetic Resources of *O. rufipogon***: It is worth mentioning that for each accession, we conducted transcriptome sequencing across four different tissues, encompassing a total of 554 tissue samples. This extensive dataset helps us achieve the most reliable gene annotation. The high-resolution genomic data facilitated the identification of 13,728 *O. rufipogon*-specific genes and 1,346 RGA loci with higher copy numbers compared to cultivated rice. These discoveries significantly enhance the gene pool available for improving cultivated rice, offering new alleles for breeding resistance to various stresses.
3) **Enhanced Evolutionary Studies**: The improved quality of variant information in the *O. rufipogon* pan-genome, coupled with previous genomic data from cultivated rice, has solidified our 'single domestication event' model. The precise variations dataset have enabled detailed classifications, revealed the genetic evolutionary path of new subpopulations, and facilitated the analysis of the origins of *indica-japonica* differentiation.

In summary, the inclusion of these additional *O. rufipogon* genomes not only fills a crucial gap in wild rice genetics but also provides valuable resources for advancing the genetic improvement of cultivated rice. These insights are pivotal for both basic biological research and practical applications in agriculture.

Comment (2): There is an ongoing debate amongst scholars about the multiple or single origins of domesticated rice (e.g., Civan et al. 2015, Huang and Han 2015, Choi and Purugganan 2018, Gutaker et al. 2020, Xu and Sun 2021, Jing et al. 2023). Not being a specialist in rice domestication, I read Jing et al. 2023 (Multiple domestications of Asian rice. Nature Plants 9, 1221-1235). I found that paper to be an excellent overview of the issues surrounding how complicated the issues are in this debate. For example, there is no consensus as to whether the Asian rice progenitors (O. rufipogon and O. nivara) were two distinct species and whether these species are geographically and genetically subdivided. Or how inferring the number of origins or domestication “events” is deeply complicated by introgression among rice subpopulations. So, my second question that I have is: What are the significant new insights gained on the origin(s) or domestication of rice not previously reported?

Response: Thank you very much for your comments and suggestions. Previous studies have well addressed the origin or domestication of the Asian cultivated rice. We appreciate the opportunity to clarify our findings on the complex topic of Asian rice origins and domestication (**Response Fig. 14**).

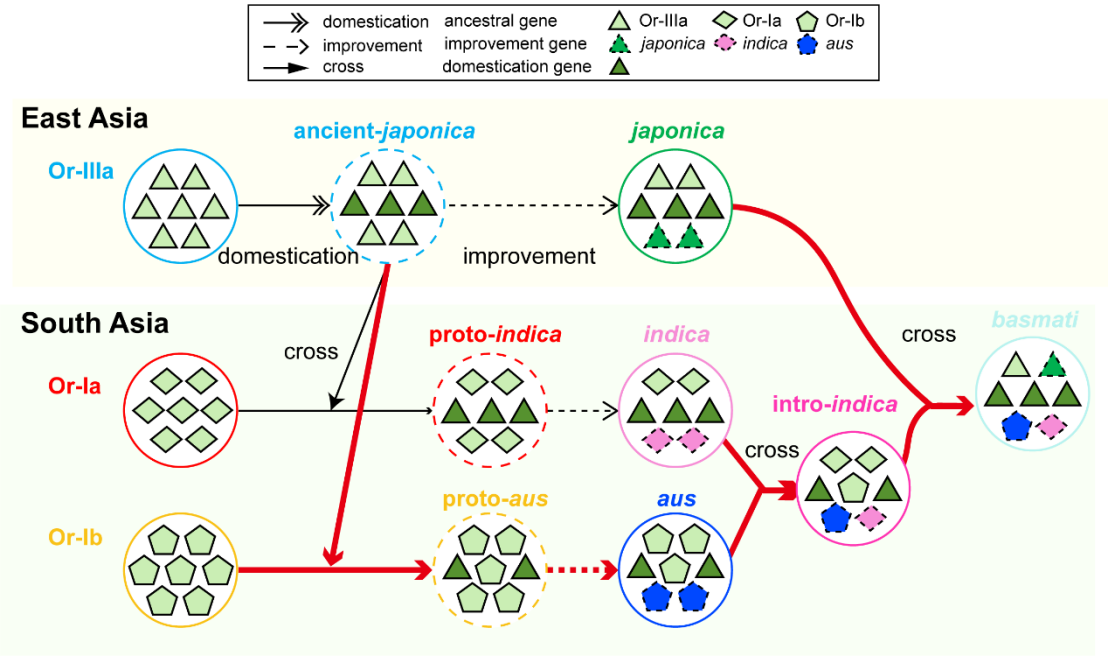
In this study, we conducted detailed evolutionary analyses and provided new evidence and findings on rice domestication:

1) We provided a new understanding of the phylogenetic relationships of wild rice. For example, the previously characterized Or-I can be further clustered into two subtypes, Or-Ia and Or-Ib.

2) We identified the ancestors of each cultivated rice group, including minor cultivated rice groups. For example, *aus* rice variety has a distinct ancestor from *indica*, originating from the Or-Ib group in South Asia.

3) We refined the evolutionary paths of two *O. sativa* subgroups. Intro-*indica*, a new classified subgroup, emerged from gene flow between *indica* and *aus*; then intro-*indica* crossing with *japonica* derived *basmati* rice.

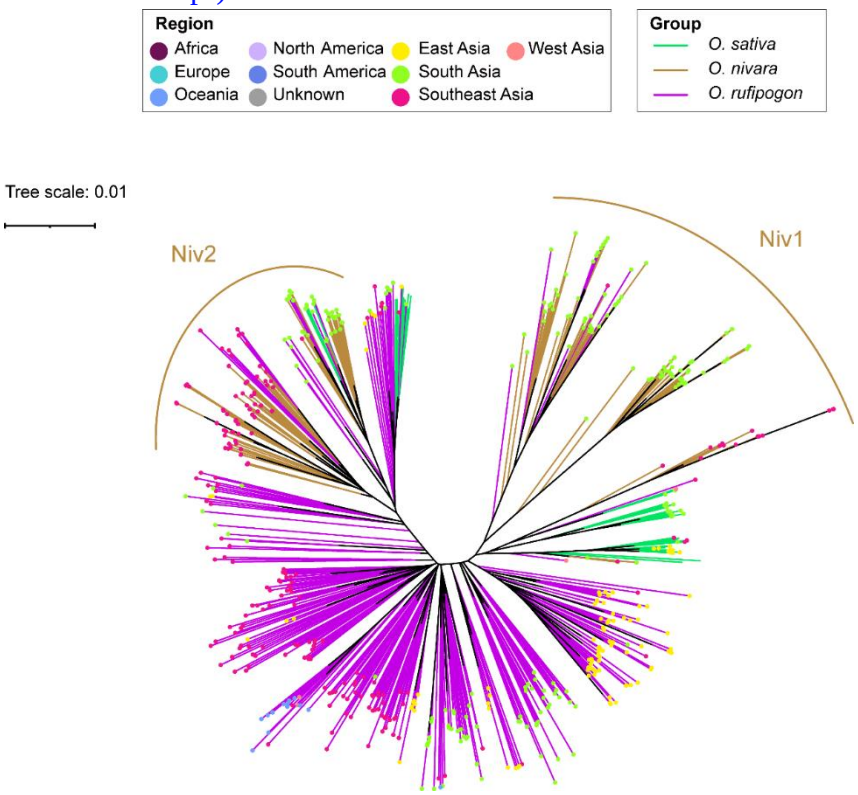
4) Our taxa-specific fragment method and haplotype analysis of domestication genes reinforced the hypothesis of a single domestication event for Asian cultivated rice.



Response Fig. 14 Overview of Asian cultivated rice evolutionary history. In the figure, the bold red arrows represent the new insights of this study.

We agree that this research field has long been debated with controversial ideas. We have tried to further address these issues in our revised manuscript (**Line 530-542**). Regarding the classification of Asian rice ancestor species (*O. rufipogon* and *O. nivara*), we expanded our dataset to include 549 wild and cultivated accessions by supplementing newly resequencing data from Jing et al. 2023

(**Supplementary Table 14**). Our evolutionary analyses revealed that *O. rufipogon* and *O. nivara* are not genetically and geographically distinct enough to warrant classification as separate species (**Response Fig. 15**). We have also clarified the connections between different nomenclatures, showing that Niv1 and Niv2 in correspond to our Or-IIIb and Or-Ib groups, respectively. The progenitor of *japonica* was identified as Or-IIIa (referred to as Ruf1a); the progenitor of *aus* rice was identified as Or-Ib (referred to as Niv2); and the progenitor of *indica* rice was identified as Or-Ia (not included in Jing et al. 2023) (**Supplementary Fig. 20a-b** in the revised manuscript).



Response Fig. 15 Phylogenetic analysis of 549 wild and cultivated rice accessions. The external circles denote the geographic distribution of each accession. Tree branches are color-coded to represent groups from our study and Jing et al. 2023. Clusters of Niv1 and Niv2 mixed with *O. rufipogon* are marked with arcs.

Intersubspecific crosses significantly complicate the study of domestication, as they can mask the genetic signals indicative of selection and domestication processes. Therefore, our study thoroughly considered the impact of gene introgression on genetic research. In addition to utilizing several statistical methods such as *D*-statistics and f3 tests, we conducted a meticulous analysis by identifying the origins of introgression segments. This confirmed that in Asian cultivated rice, particularly in *indica* and *basmati*, genomic fragments originating from the *japonica* ancestor Or-IIIa are predominantly concentrated in domestication regions (**Fig. 4a and b**).

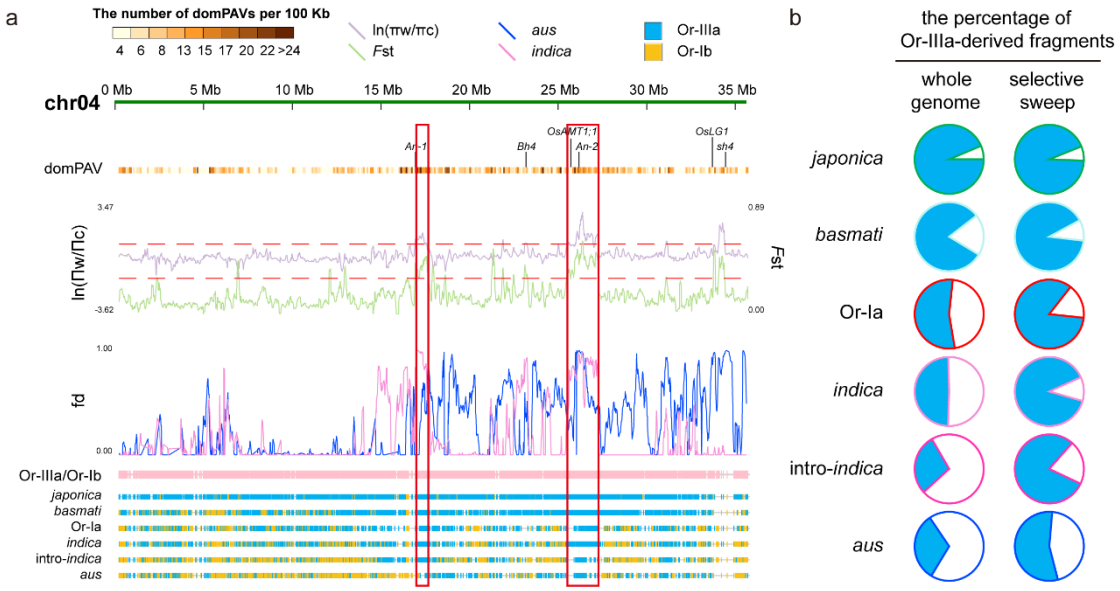


Fig. 4a-b Selective sweep of Asian cultivated rice. a, The graph presents the nucleotide diversity ratio (π_w/π_c) and fixation index (F_{ST}) values between wild rice (Or-IIIa and Or-Ib) and cultivated rice (*japonica*, *basmati*, *indica*, intro-*indica* and *aus*) along chromosome 4. Horizontal red dashed lines indicate genome-wide threshold of selection signals ($\ln(\pi_w/\pi_c) > 1.09$ and $F_{st} > 0.3$). In the line chart of fd values distribution, the population (Or-Ib, *aus*, *japonica*, *O. longistaminata*) is represented by the blue line, and the population (Or-Ia, *indica*, *japonica*, *O. longistaminata*) is represented by the pink line. 566,513 highly differentiated SNP sites between Or-IIIa (31 accessions) and Or-Ib (19 accessions) are indicated in pink. Within six populations (*japonica*, *basmati*, Or-Ia, *indica*, intro-*indica* and *aus*), the major alleles matching Or-IIIa-specific alleles are marked in blue, while those matching Or-Ib-specific alleles are marked in orange. In the figure, the red boxes represent segments including known domestication genes derived from Or-IIIa. **b,** The chart displays the percentage of Or-IIIa-derived fragments across the whole genome and in selective sweep regions for major groups.

Furthermore, we address the controversy surrounding the selection of domestication genes. The early rice domestication of rice primarily focused on key agronomic traits such as plant architecture, hull color, seed shattering, awn length, pericarp color, and grain size (Chen et al. 2019; Xu and Sun 2021). We note that Jing's study included genes related to environmental adaptation, disease, and stress resistance (e.g., *OsLCBK2*, *GLO4*, *UCL30*, and *OsFbox550*), which are traits associated with post-domestication or genetic improvement. We argue that such expansive selection might lead to potentially misleading conclusions about the initial domestication process. It is worth mentioning that *LABA1* (also known as *An-2*), a key domestication gene regulating awn length and grain yield, was considered to originate from Niv2 (Or-Ib, the ancestor of *aus* in our population) in

Jing's study. However, our haplotype analysis indicates that this gene originated from Or-IIIa (the *japonica* ancestor) (**Supplementary Fig. 27-28**), aligning with previous original research (Gu et al. 2015; Hua et al. 2015). This underscores the importance of our high-quality genomes and accurate gene annotations.

It is also important to emphasize that the number of origins or domestication events can be inferred from the count of quantitative trait nucleotides (QTNs) associated with key domestication genes. If domestication occurred as a singular event, these key domestication genes should be characterized by shared QTNs. Conversely, if domestication emerged from multiple origins, the corresponding functional sites would manifest as distinct alleles. The haplotype analysis of domestication genes conducted in our study supports the hypothesis that a single domestication event for Asian cultivated rice (**Supplementary Fig. 32**).

We believe these findings contribute significantly to the ongoing debate on rice origins and domestication. This was previously unachievable due to the lack of high-resolution, near-saturation pan-genomes of wild rice. Our research has filled this critical gap, offering insights that enhance our understanding of the genetic complexities involved in rice domestication.

Comment (3): In summary, the manuscript is a well-executed assembly of an updated rice pan-genome that should be an important resource for rice biology. However, as written, the insights provided seem incremental relative to previously published work (e.g., Jing et al. 2023). Perhaps what could be emphasized in a revision is how adding the new data (e.g., “outgroups”) dramatically changes how we understand rice origins and domestication.

Response: Thank you very much for your suggestions. Thanks for recognizing the efforts putting into assembling the updated rice pan-genome and acknowledging its potential as a valuable resource for rice biology. Following your suggestions, we have carefully rewritten the Discussion section with emphasizing the new insights in the revised manuscript.

In response to your comments, we revised our manuscript to better highlight how the integration of these outgroups significantly enhances our understanding of rice origins and domestication. These new datasets not only allow us to more accurately trace the origins of complex genomic variations, such as large inversions (**Supplementary Fig. 16**), but also provide important guidance for our evolutionary analysis of key domestication genes. Through these in-depth analyses, we have not only deepened our understanding of rice genetic diversity and evolutionary history but also challenged and refined the previous models of rice domestication.

Methodologically, we have made significant advancements, including more comprehensive sample collection, more advanced sequencing technologies, and more complex computational analyses. Based on this data, we employed a method not used in previous studies to conduct a detailed analysis of the origins of

introgression segments within the domestication regions of cultivated rice. These improvements have enabled us to produce more reliable and solid results, thereby advancing our understanding of rice origin and domestication.

Reference

Chen, Erwang, Xuehui Huang, Zhixi Tian, Rod A. Wing, and Bin Han. 2019. 'The Genomics of *Oryza* Species Provides Insights into Rice Domestication and Heterosis'. *Annual Review of Plant Biology* 70 (1): 639–65. https://doi.org/10.1146/annurev-arplant-050718-100320.

Gu, Benguo, Taoying Zhou, Jianghong Luo, Hui Liu, Yongchun Wang, Yingying Shangguan, Jingjie Zhu, et al. 2015. 'An-2 Encodes a Cytokinin Synthesis Enzyme That Regulates Awn Length and Grain Production in Rice'. *Molecular Plant* 8 (11): 1635–50. https://doi.org/10.1016/j.molp.2015.08.001.

Hua, Lei, Diane R. Wang, Lubin Tan, Yongcai Fu, Fengxia Liu, Langtao Xiao, Zuofeng Zhu, et al. 2015. 'LABA1, a Domestication Gene Associated with Long, Barbed Awns in Wild Rice'. *The Plant Cell* 27 (7): 1875–88. https://doi.org/10.1105/tpc.15.00260.

Jing, Chun-Yan, Fu-Min Zhang, Xiu-Hua Wang, Mei-Xia Wang, Lian Zhou, Zhe Cai, Jing-Dan Han, et al. 2023. 'Multiple Domestications of Asian Rice'. *Nature Plants* 9 (8): 1221–35. https://doi.org/10.1038/s41477-023-01476-z.

Xu, Ran, and Chuanqing Sun. 2021. 'What Happened during Domestication of Wild to Cultivated Rice'. *The Crop Journal* 9 (3): 564–76. https://doi.org/10.1016/j.cj.2021.02.005.

Point-by-point responses to Referees' comments

Referees' comments:

Authors' Response to Referee #1

Referee #1 (Remarks to the Author):

Remarks: I am impressed by the extra data, analyses, and reasoning of this revision. Overall, the manuscript is significantly improved, but a few minor issues still remain.

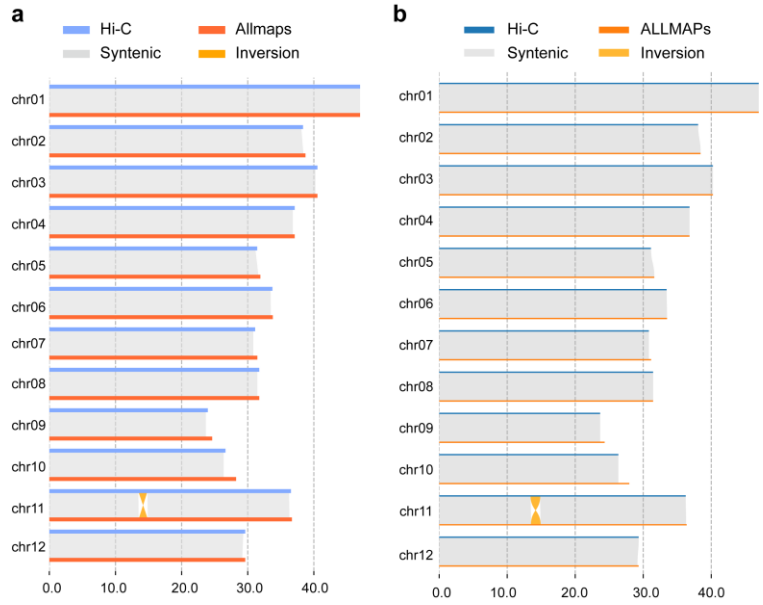
Response: Thank you very much for the comments.

Comment (1): Hi-C reads were added for 30 accessions to verify the quality of PacBio assemblies. sTable 5 and sFig. 3 showed the congruence between assemblies by PacBio only and PacBio + Hi-C, demonstrating the high quality of these assemblies. The current title is reasonable. Meanwhile, sFig. 3 showed that most chromosome pairs had an overhang at the right end, which means the discrepancies consistently occurred in the end of a chromosome. Please confirm this is not the artifact of plotting.

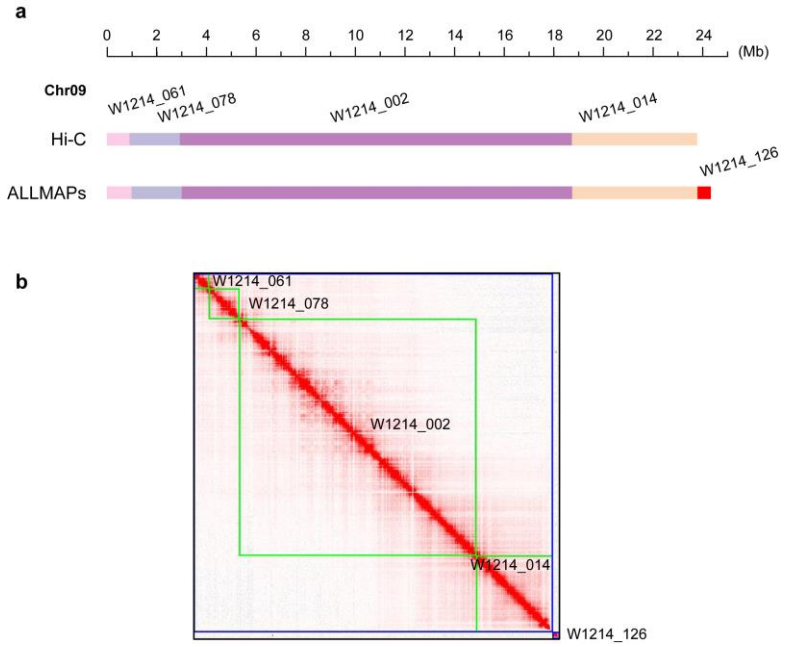
Response: Thank you for your careful review, and sorry for any misunderstandings caused by our plotting. Upon further examination, we found that the overhangs observed on the right ends of most chromosomes were caused by a bug in the drawing program (**Response Fig. 1**). This issue has been fixed in the latest version of plotsr (v1.1.1, update on May 8, 2023), and we have replaced the original graph (**Supplementary Fig. 3**) in the revised manuscript.

We observed that even after fixing the bug in the plotting program, there are still length discrepancies between the few chromosomes assembled using the two different methods. In some cases, the chromosomes assembled with Hi-C are slightly shorter than those assembled with the ALLMAPs method. This difference is likely due to weaker Hi-C signals, causing some smaller contigs to remain unanchored to the chromosomes. A typical example of this is W1214_126 contig, a short contig (~0.65 Mb) located at the end of chromosome 9 in W1214 (**Response Fig. 2**). However, this

phenomenon is infrequent across the 30 genomes studied. Therefore, we conclude that our pseudo-chromosomes assemblies (ALLMAPs) are in good agreement with the established chromosome-level assemblies (Hi-C).



Response Fig. 1 Whole genome synteny plot comparing chromosomes constructed using Hi-C technology and pseudo-chromosomes obtained via the ALLMAPs method. a, Plot generated using the previous version of plotsr (v0.1). **b**, Plot generated using the new version of plotsr (v1.1.1).



Response Fig. 2 The contig composition and Hi-C analysis of Chr09 in W1214. a, The schematic of contig composition of Chr09 in W1214 using Hi-C technology and

pseudo-chromosomes obtained via the ALLMAPs method. **b,** Hi-C contact matrices of Chr09 (**b**) in W1214.

Remarks: The authors further compiled a 549-accession new dataset by including resequencing data from Jing et al. (Ref 25). The results from this new dataset not only addressed the key concern of sampling issues but also connected both studies: the subgroups defined by two studies were aligned (such as Or-IIIb ~ Niv1 and Or-II ~ Ruf2 + Ruf1b). The discrepancy between this study and Jing et al. centered on the definitions of domestication and adaptation/ improvement. As stated in the manuscript (Line 549-553), this study represented the viewpoint of domestication as a process only involving selecting several obvious morphological traits and improvement/adaptation as a subsequence process, while another viewpoint is improvement/adaptation being a part of domestication. From the authors' point of view, the proposed model was supported by the results.

Response: We sincerely appreciate your thorough summary and the comments.

Comment (2): Meanwhile, I have a related question based on Line 556-559, which mentions "accurate haplotype analysis of LABA1 in this study". Was this accuracy due to the graph-based pangenome (indicated in Response), instead of a single reference genome, was used for calling variants?

Response: We apologize for any ambiguity in the description regarding the accuracy of our haplotype analysis. The accuracy can be attributed to several key factors:

1) **Higher Sequence Quality**: We utilized high-quality third-generation sequencing technologies (PacBio HiFi or Nanopore), which provided average quality values of 57.92 or 24.55, respectively (**Supplementary Table 3**). This enabled us to obtain more precise sequences directly, without relying on imputation methods for low-coverage next-generation sequencing data, which may overlook species- or group-specific variants.
2) **More Complete Gene Sequences**: By leveraging reference genome-level sequences, we annotated complete gene sequences (including promoter regions)

for haplotype analysis. This approach provides a more comprehensive view of the genetic variants compared to using only VCF files.

3) **More Effective Analysis Method**: We focused our analysis on regions containing quantitative trait nucleotides (QTNs) in key domestication genes, specifically their coding and promoter regions. By targeting these critical regions, we improved our ability to accurately identify sites selected or linked to functional sites of initial domestication genes more, compared to analyzing all variants within a 1 kb region around the gene.

Taken together, these factors allowed us to construct more accurate haplotype networks compared to the previous study (Ref25). We have clarified this in the revised manuscript (**Line 538-540**).

Comment (3): I appreciate the effort of conducting GWAS to identify potential loci/genes associated with the copy number variation of Gypsy families, which are main contributors to the genome evolution between wild and domesticated rice. Although some families showed nice GWAS signals, I agreed that exploring the mechanisms should be followed up in future studies.

Response: We are grateful for your suggestions, which will be instrumental in shaping our future research endeavors.

Comment (4): In addressing the comment for Fig. 2a about the reference bias, the author conducted comprehensive analyses and confirmed the result in Fig. 2a was biased but other evolutionary analyses were unaffected regardless which reference genome was used. However, despite confirming the bias, the original Fig. 2a and the corresponding statements were retained in this revision. In my opinion, these parts should be removed from manuscripts.

Response: We appreciate your suggestions and understand your concerns regarding Fig. 2a. This figure presents the number of variations relative to the Nipponbare (a *japonica* cultivar) across all groups, providing valuable insights for researchers to compare different types of variations and assess group differences. Furthermore, these

data provide a crucial foundation for the subsequent analyses in our study.

After careful consideration, we prefer to retain Fig. 2a in the revised manuscript. To address your concerns about reference bias, we have now included relevant analysis in the revised manuscript (**Supplementary Fig. 14**, **Line 276-278, Line 846-858**). However, we also defer to the judgment of the reviewer and editor and are open to their final decision on this matter.

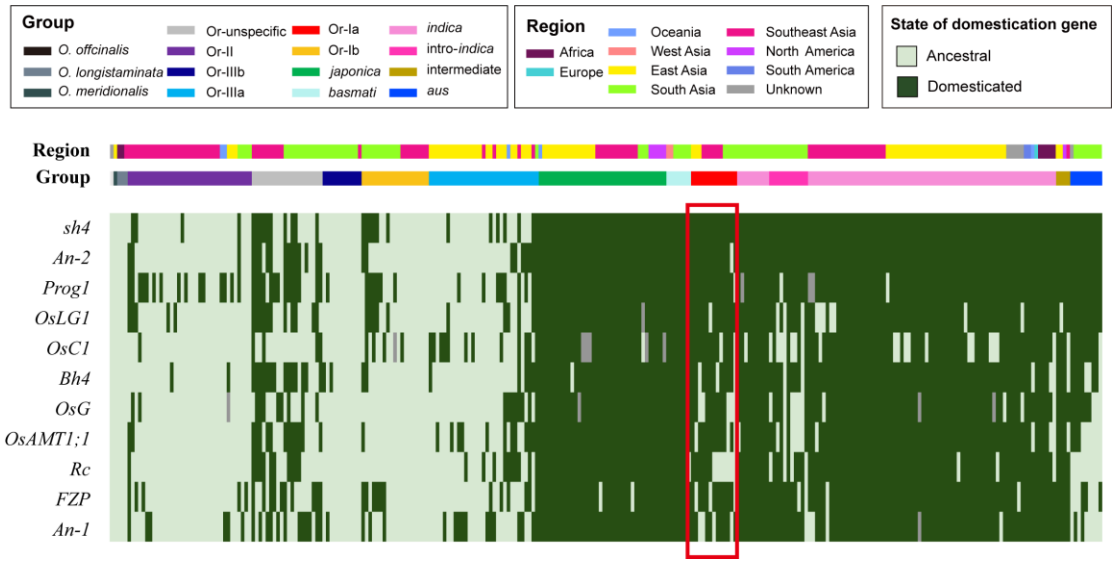
Comment (5): About the third DEA equation, the manuscript indicated three standards (Line 718), implying that any gene that meets one of these equations is DEA, but the Response indicated Equation 3 is a filter, or an extra condition for Equation 1 and/or 2. The equation 3 is good as a filter, but not a good standard for classifying DEA (for instance, p-gene reads = 6, a-gene reads = 5). Please confirm which is the case and revise the manuscript accordingly.

Response: We apologize for any confusion caused by the previous wording. To clarify, an allele is classified as differentially expressed allelic (DEA) only when all three equations are met simultaneously. The first two equations identify differentially expressed allelic genes, while the third equation acts as a filter to exclude alleles with low expression levels. We have revised the manuscript to clarify this definition (**Line 698-699**).

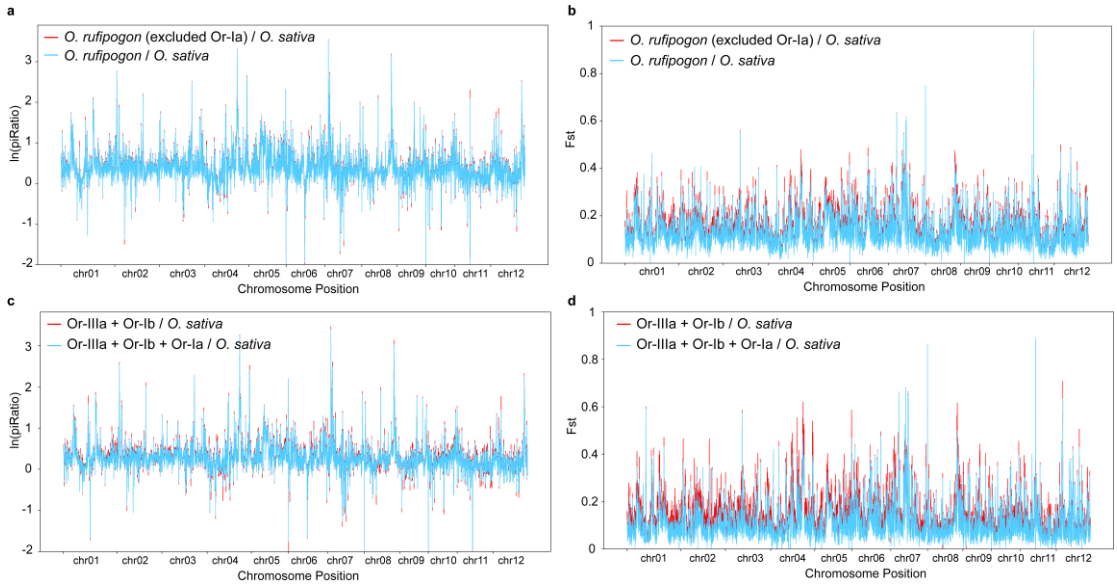
Comment (6): The new sFig. 25 compared Pi and Fst among different datasets to support that the applied standard of Or-IIIa + Or-Ib vs Os is reasonable. However, panel a and b are very similar. Please confirm that removing 13 Or-Ia accessions didn’t change the estimates that much.

Response: The results presented in the manuscript demonstrated that Or-Ia has undergone gene flow from cultivated rice, with some domestication genes exhibiting domesticated states (**Response Fig. 3**). Therefore, it should not be classified as part of the ancestral group. As previously demonstrated, removing the Or-Ia accessions did not significantly affect the piRatio and Fst estimates for the entire *O. rufipogon* dataset (129 accessions), likely because the Or-Ia (13 accessions) group constitutes a

relatively small proportion of the overall *O. rufipogon* population (**Response Fig. 4a-b**). However, when focusing on the Or-IIIa and Or-Ib groups (50 accessions), the inclusion of 13 Or-Ia accessions slightly reduced these estimates (**Response Fig. 4c-d**). This may be attributed to the higher proportion of Or-Ia in the ancestral groups. While the overall results remained largely unchanged, the removal of Or-Ia was intended to refine the dataset for more accurate comparisons. Therefore, we continue to classify only Or-IIIa and Or-Ib as the ancestral groups when identifying the selective sweep regions.



Response Fig. 3 The domestication state of 11 known domestication genes in each accession. Each row indicates a domestication gene, and each column indicates an accession. The red box indicates the states of domestication genes in Or-Ia (from **Fig. 4d**).



Response Fig. 4 Effect of Or-Ia inclusion on the piRatio and *F*st estimates in all *O. rufipogon* groups, and only ancestors' groups of *O. sativa*. (**a-b**). Genome-wide distribution of ln(piRatio) (**a**) and *F*st (**b**) values comparing the entire *O. rufipogon* populations with and without Or-Ia against *O. sativa*. (**c-d**). Genome-wide distribution of ln(piRatio) (**c**) and *F*st (**d**) values comparing ancestors' groups of *O. sativa* with and without Or-Ia against *O. sativa*.

Comment (7): Fig. 3b and Fig. 4c showed the results from Admixture with different K groups. A common practice is separating each group by, for example, a vertical line. However, this was not followed for most K groups. Furthermore, Line 383-385 cited Fig. 3b about "Intro-indica", but the "Intro-indica" cluster was not labeled in Fig. 3b.

Response: Thank you for your valuable suggestions. In the revised manuscript, we have updated Fig. 3b, Fig. 4c and Supplementary Fig. 20 by adding vertical lines to clearly separate each group. Additionally, we have ensured proper labeling to stress the "intro-*indica*" cluster in Fig. 3b.

Comment (8): A typo of "Transaction" in sFig. 15.

Response: Thank you for pointing out this typo, it has been corrected.

Authors' Response to Referee #2

Referee #2 (Remarks to the Author):

Remarks: I am satisfied with the responses and revisions to my concerns. However, I still have several issues that need addressing:

Response: Thank you very much for the comments.

Comment 1: I would like clarification on the criteria used to define a novel sequence, which resulted in a 3.87G sequence being absent from the Nipponbare reference genome. This criterion was not detailed in the methods section.

Response: We apologize for the lack of clarity regarding the criteria used to define novel sequences in the method section. We identified nodes that were not present in the Nipponbare reference path (Non-ref) as novel sequences and measured their total length separately in three constructed graph-pangenomes (**Supplementary Table 11**). In the graph pan-genome composed of 15 *O. sativa* and 129 *O. rufipogon* accessions, the length of Non-ref is 3.87 Gb. These detailed descriptions have now been included in the revised methods section (**Line 883-885**).

Comment 2: Regarding the domestication debate, the author suggested that the previous study utilized a gene set that included traits related to post-domestication (or genetic improvement), potentially leading to the conclusion of multiple domestications. The authors also emphasized that the initial stages of rice domestication primarily focused on key agronomic traits, such as plant architecture, hull color, seed shattering, awn length, pericarp color, and grain size.

This raises questions about the criteria for determining domestication traits. For instance, why are hull color and pericarp color considered key domestication syndromes at the initial stage? It's difficult to imagine why people would prioritize hull color and pericarp color during the initial stages of rice domestication. While

grain size, which is often associated with genetic improvement, is not?

Response: Thank you for your question, which highlights the complexities of domestication syndrome.

Plant domestication is a continuous process of increasing interdependence between plants and humans, making it difficult to pinpoint the exact timing of domestication or classify traits and genes directly associated with it (Gross and Olsen, 2010). However, a set of common traits, known as the "domestication syndrome", can distinguish most domesticated crops from their wild ancestors (Hammer, 1984). These traits include larger fruits or grains, more robust plant structures, more determinate growth and/or apical dominance, and other characteristics such as grain quality (changes in starch and other compounds), fruit or grain color, shape, flowering time, and plant height (Doebley et al., 2006; Gross and Olsen, 2010).

Notable phenotypic differences exist between Asian cultivated rice and its wild relatives. For instance, pericarp and hull color are among the most apparent domestication traits linked to grain quality, even although they may not significantly impact cultivation practices (Gross and Olsen, 2010). Wild grains have a red pericarp and black hulls at maturity, which are absent in modern Asian cultivars, resulting in straw-white pericarp and seed hulls in domesticated rice (Sweeney and McCouch, 2007). Although several domestication genes that regulate these phenotypes have been identified (Sweeney et al., 2006; Zhu et al., 2011), the reasons behind their selection and the pressures involved remain unclear. This may be linked to early preferences for white seeds among Asian cultures or their association with traits like seed shattering and dormancy (Sweeney et al., 2006; Purugganan and Fuller, 2009; Zhu et al., 2011; Vigueira et al., 2013).

Grain size is also a crucial domestication trait in rice, as farmers likely selected larger seeds to enhance yield and obtain more food during domestication (Liu et al., 2018). Wild grains are generally small, while domesticated grains exhibit a wider range of sizes (Sweeney and McCouch, 2007). Larger seeds may have been favored in open environments, where bigger seedlings had an advantage due to deeper burial in soils disturbed by human tillage (Purugganan and Fuller, 2009). Some researchers

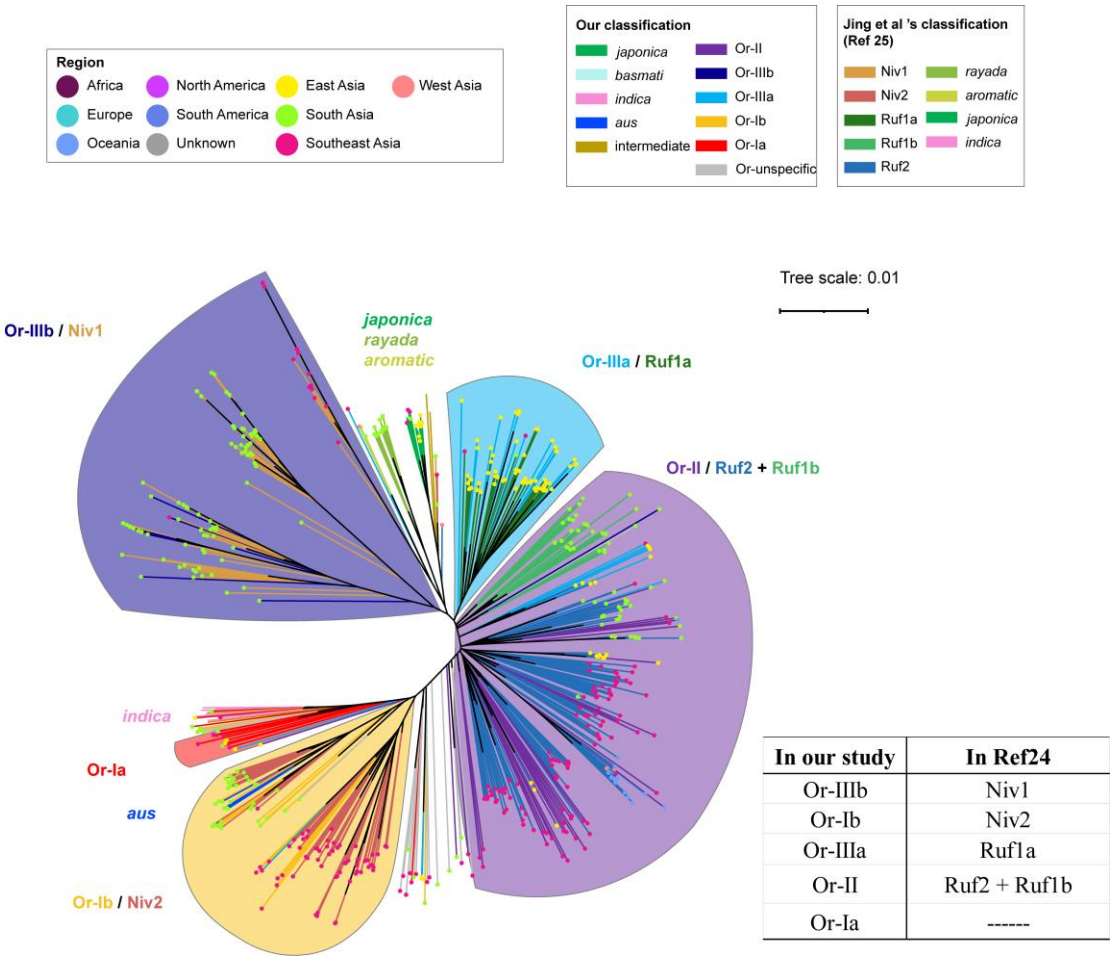
proposed that the non-shattering phenotype, a classic domestication trait, may have appeared only after an increase in grain size (Fuller, 2007; Gross and Olsen, 2010). Therefore, grain size may have been among the first traits subjected to selective pressure from human cultivation (Purugganan and Fuller, 2009).

However, it is important to emphasize that not all genes regulating these traits qualify as domestication genes; only those located in selective sweep regions and subject to selection pressure should be considered as such. This distinction is particularly important for complex traits like grain size, which is regulated by multiple genes and continues to evolve during genetic improvement and modern breeding. Including genes that were selected during later improvement processes could mask the original domestication signatures and potentially lead to misleading conclusions about the genomics and geographical origins of *O. sativa* domestication. Therefore, a more rigorous approach focusing specifically on genes with evidence of initial domestication is essential for accurately reconstructing the domestication history of *O. sativa*. For clarify, we have incorporated relevant descriptions into the revised manuscript (**Line 531-534**).

Additionally, the authors concluded that O. rufipogon and O. nivara are not genetically or geographically distinct enough to warrant classification as separate species. However, I could not find information on the genetic and geographical differences between O. nivara and O. rufipogon in the results section. Could you please clarify or add this information?

Response: We apologize for the lack of clarity on this point. In the last manuscript, we expanded our dataset to include 549 wild and cultivated accessions by incorporating newly resequencing data from Jing et al. 2023 (Ref25) (**Supplementary Table 14**). After calling variants using the graph-pangenome, we conducted phylogenetic and archetypal analyses (**Supplementary Fig. 20**). Our evolutionary analyses show that *O. nivara* (Niv1 and Niv2 in Ref25) does not form an independent cluster but instead groups with certain *O. rufipogon* populations (Niv1→Or-IIIb; Niv2

→Or-Ib), supporting our conclusion that *O. nivara* should be considered an ecotype of *O. rufipogon* (**Response Fig. 5**). Furthermore, there is no clear geographical separation among these groups. We have added these results to the manuscript for better clarity (**Line 347-351,514-522**).



In our study	In Ref24
Or-IIIb	Niv1
Or-Ib	Niv2
Or-IIIa	Ruf1a
Or-II	Ruf2 + Ruf1b
Or-Ia	------

Response Fig. 5 Phylogenetic analysis of 549 wild and cultivated rice accessions. The external circles denote the geographic distribution of each accession. Tree branches are color-coded to represent original groupings from two studies. Background colors denote new groups based on our classification method derived from this evolutionary tree analysis. The text labels on the outside of the tree show the corresponding relationship between our grouping (before the slash) and the classification from another study (after the slash).

Comment 3: Most of the first paragraph in the discussion repeats content from the results or introduction. It would be more effective to shorten this paragraph and merge

it with the second paragraph to create a more cohesive discussion.

Response: Thank you for your suggestion. We have removed redundant content and revised the discussion section, resulting in a more concise and focused narrative (**Line 492-509**).

Authors' Response to Referee #3

Referee #3 (Remarks to the Author):

Comments in attachment

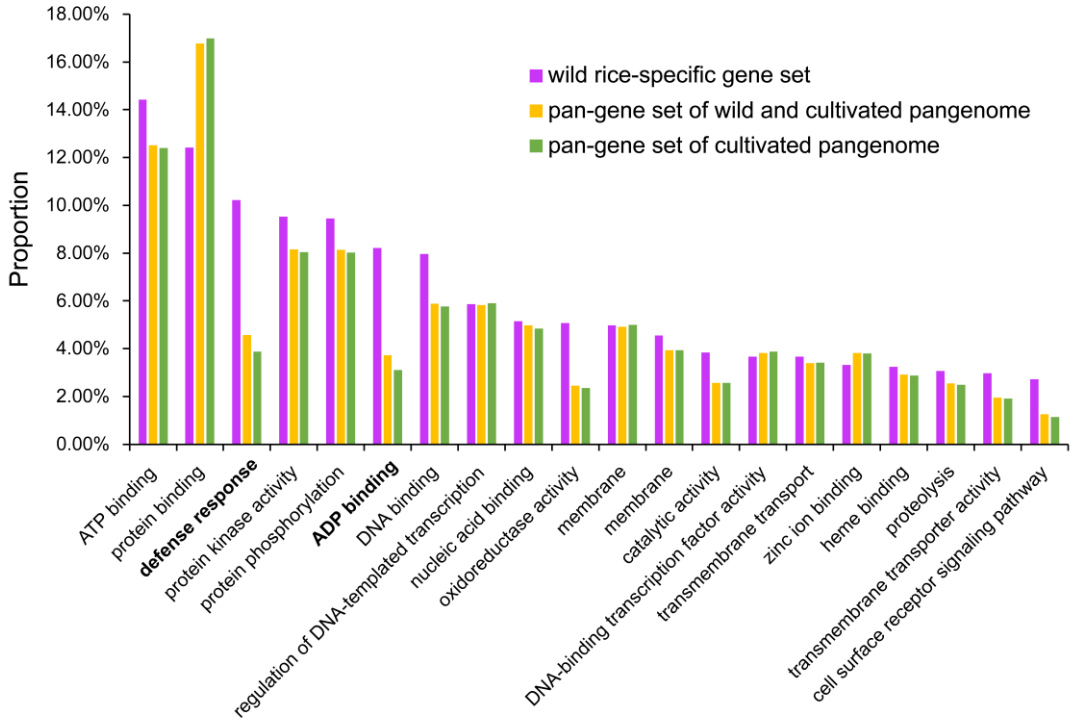
Comments on the revised MS of 2024-03-04666B

Comment 1. The authors did not understand my question on the invalid comparison between the pan genomes of OS and OR. Although the sampled 16 OS a.ccessions did not affect the pan genome estimation of the sampled OR accessions, the authors should use the pan genome of OS accessions of equivalent sample size and similar geographic distribution for their comparison and claim. In other words, the authors should compare all OR pan genes they discovered with all previously published OS pan genes in order to be sure those 13,728 wild rice "novel" genes are indeed OR specific, even though they are absent in the 16 sampled OS accessions. In the authors' response "Following your suggestion, we have compared our identified wild rice-specific genes with three published rice pan-genomes (Qin et al. 2021; Zhang et al. 2022; Shang et al. 2022) (Supplementary Table 10 in the revised manuscript). As expected, the number of genes exclusive to wild rice is different in this broader comparison. However, the results demonstrate that most of the wild rice-specific genes we identified (63%-82%) remain unique, even when compared against a larger set of cultivated rice genomes. These are truly important and novel findings." If so, the authors' claim for the discovering "13,728 wild rice-specific genes in their abstract does not hold true because they should include the results into the revised MS, stating only at most ~63% of the 13,728 genes are likely OR specific. Again, when the authors did not obtain an accurate estimate on the number and diversity of RGAs in OS (given the 16 OS accessions they used), their claim that OR population has more PGAs and higher diversity at PGAs than OS remains invalid, until an OS population of equivalent size and geographic distribution is used in their analyses.

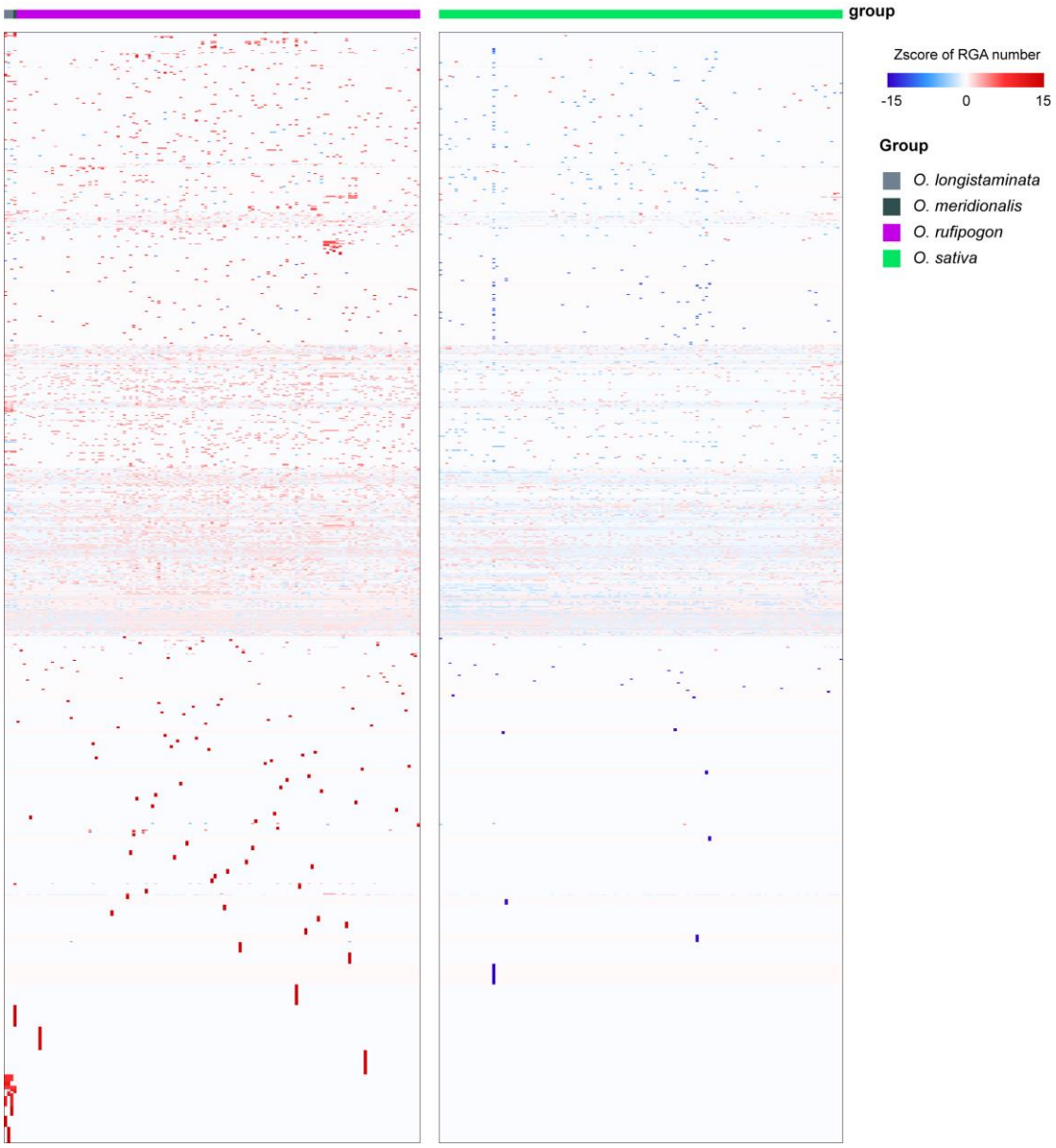
Response: Thank you for your suggestions and we regret that our previous response might not fully address your concerns. In our previous study, we considered that the direct comparison of genes across different datasets might introduce bias due to

methodological differences in sequencing and gene annotation across those datasets. Following your suggestion, we have now expanded our study by incorporating the 113 non-redundant cultivated rice genome data from the three previously published pan-genomic data (Qin et al., *Cell*, 2021; Zhang et al., *Genome Res.*, 2022; Shang et al., *Cell Res.*, 2022) and performing gene annotation across all samples using a consistent approach in the revised manuscript (**Supplementary Table 10**) In this dataset of 129 *O. rufipogon* and 129 *O. sativa*, we identified a total of 7,592 wild rice-specific genes. These genes were still more related to defense response and ADP binding processes (**Response Fig. 6**). The results were well matched with our original data analysis, which was indicated by the trended growth curve of the pan- and core-gene number of 16 *O. sativa* and 129 *O. rufipogon* genomes (Figure 1i). In comparison with an equal number of cultivated rice samples (n=129), we identified 1,184 (51.12%) collinear loci in wild rice that exhibited higher mean copy numbers. Among these, 638 (27.55%) loci were exclusively present in wild rice (**Response Fig. 7**).

These corresponding descriptions have now been included in the revised results and methods sections (**Line 200-211, Line 239-241, Line 706-708, Line 869-875, Fig. 1g**, **Supplementary Fig. 7-8, 10a and 11-12, Supplementary Table 8, 10**).



Response Fig. 6 Proportion of genes with different functional categories in Gene Ontology (GO) in specific gene set of *O. rufipogon*, pan-gene set of wild-cultivated pangenome and pan-gene set of *O. sativa* pangenome. Items with names in bold indicate their proportions in specific gene set of *O. rufipogon* are significantly greater than that of the other two gene sets (from **Supplementary Fig. 12**).



Response Fig. 7 The landscape of RGA loci that have a higher average copy number in wild rice than cultivated rice. Each row indicates an RGA locus, and each column indicates a variety. The data was normalized using a row-wise Z-score method (from **Fig. 1g**).

Comment 2. The reviewer does not understand how the authors reached the speculation "These cultivars have been identified to be domesticated through crossing the proto-japonica type cultivar (with multi-domestication traits, such as non-shattering seed habit, from prostrate to erect growth, transition from black to straw-white seed hull, etc) with diverse wild rice O. rufipogon populations for

adapting various environments (Gutaker et al. 2020)”, given the fact that japonica cultivars or japonica type of ORs are known to have very low outcrossing rates, and “the modern breeding process” involving hybridization and artificial selection of rice has a short history < 130 years.

Response: Thank you for raising your concerns about the possibility of proto-*japonica* hybridization during domestication.

O. rufipogon exhibited a mixed mating system of both outcrossing and selfing, with an average outcrossing rate of 0.478, which is not particularly low (Morishima and Barbier, *Plant Species Biol.*, 1990; Yang et al., *Genet. Resour. Crop Evol.*, 2024). Additionally, one report has shown that bidirectional gene flow between *O. rufipogon* and *O. nivara* (an annual ecotype of *O. rufipogon*), particularly with gene flow from *O. rufipogon* to *O. nivara*, which occurred repeatedly during their divergence (Zheng and Ge, *Mol. Ecol.*, 2010). These findings suggested that hybridization among the ancestors of Asian cultivated rice was indeed possible. Moreover, it is worth noting that domestication did not require widespread and extensive gene flows–even limited hybridization could have affected the evolutionary trajectory of a species during domestication, unlike the modern breeding process.

Archaeological and genetic evidence shows that human societies began transitioning from hunting and gathering to agriculture around 10,000 years ago. with major crops including rice, wheat and maize being domesticated by 4,000 years ago (Doebley et al., *Cell*, 2006). This was a complex co-evolutionary process that enhances the interdependence between humans and plants (Purugganan, *Trends Ecol. Evol.*, 2022) and is driven by human selection and crops recurrent adaptations (Purugganan and Fuller, *Nature*, 2009; Purugganan, *Curr. Biol.*,2019).

Taking together, we hope this explanation clarifies that, although ancient farming techniques were not as advanced as modern breeding, early farmers likely employed selection and multi-generational hybridization, whether consciously and unconsciously, to contribute to the domestication of Asian cultivated rice by ten thousand years ago.

Comment 3. Unfortunately, the authors did not answer to my question regarding how the serious violation of the assumptions "the neutrality and random mating" for software MSMC2 influences their conclusions. In fact, the authors should use computer simulation to address how the breeding systems and different types of selection affect their results. They did not even try to test the alternative hypothesis of the multiple origins using their data.

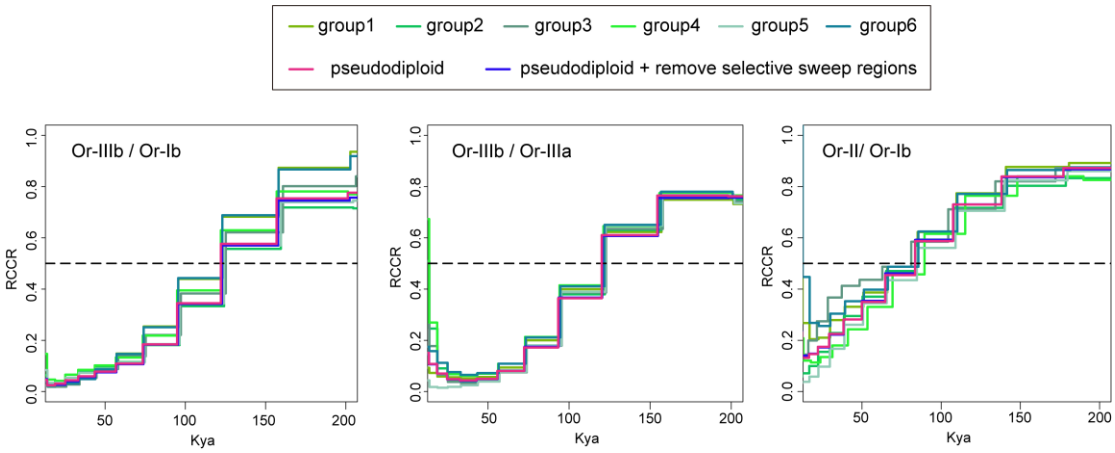
Response: Thank you for your concerns regarding the use of software MSMC2. Following your suggestion, we have now conducted additional analyses to access the impact of the breeding systems and different types of selection on the results.

The *O. rufipogon* species, used in our study, naturally employs both cross-pollination (~48%) and self-pollination (~52%) (Morishima and Barbier, *Plant Species Biol.*, 1990; Yang et al., *Genet. Resour. Crop Evol.*, 2024). To access the impact of breeding systems on the results, we followed an established approach of constructing pseudodiploids, which was widely used in similar studies of inbreeding species such as Caenorhabditis, *Arabidopsis thaliana*, soybean, African rice and Asian rice (Thomas et al., *Genome Res.*, 2015; Beissinger et al., *Nat. Plants*, 2016; Meyer et al., *Nat. Genet.*, 2016; Durvasula et al., *Proc. Natl. Acad. Sci.*, 2017; Cubry et al., *Curr. Biol.*, 2018; Kim et al., *Nat. Commun.*, 2021; Jing et al., *Nat. Plants*, 2023). We randomly selected 4 samples from each population and treated each sample as a single haplotype. We then paired chromosomes from haplotypes within the same population to construct pseudodiploids. When we compared to previous strategy that preserves two haplotypes per sample, the differences in the estimated split times were minimal (**Response Fig 8**). Therefore, the breeding system has little influence on the results when using MSMC2 to split time of *O. rufipogon*.

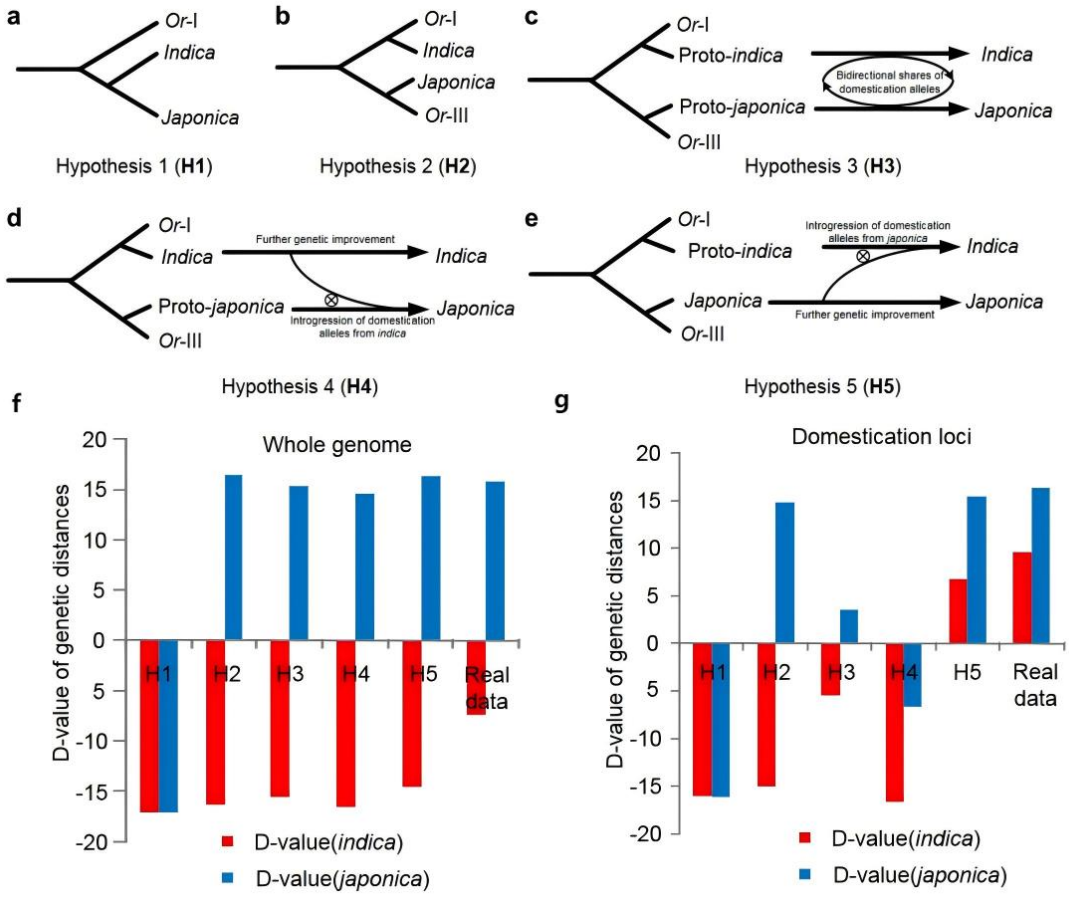
Regarding the effect of selection pressure, we have to mention that we focused exclusively on non-coding sequences to infer demographic history (**Line 952-953**). This allows us to preserve neutral sites and minimize the influence of selection on the results. In addition, we evaluated the effect of selective sweep regions and found minimal differences when compared to results that did not exclude these regions (**Response Fig 8**).

We would like to clarify that we used MSMC2 specifically to infer the population history of *O. rufipogon* and estimate the split time among various groups in our study, not to investigate the origin and domestication of cultivated rice. Regarding the multiple origin hypothesis of Asian cultivated rice, we previously tested this using computer simulations in our paper (Huang et al., *Nature*, 2012) (**Response Fig 9**). Therefore, we did not repeat this analysis in the current study.

In conclusion, we have updated our approach to infer population history, including constructing pseudohaploids and excluding coding regions (**Line 960-966**). These new results were presented in the revised manuscript (**Supplementary Fig 13**). We hope this explanation clarifies our methods and results.



Response Fig. 8 Effect of breeding systems and selective sweep regions on the split time of *O. rufipogon* estimated by MSMC2. Cross-population split time estimates shown for (**a**) Or-IIIb versus Or-Ib, (**b**) Or-IIIb versus Or-IIIa, and (**c**) Or-II versus Or-Ib. In each panel, green lines represent combinations of different four samples (two samples of each group), using a strategy of retaining two haplotypes per sample. Pink lines represent results using pseudodiploids constructed from 4 samples per group, with each genome acting as a single haplotype. Blue lines represent the results after removing selective sweep regions on the basis of the construction of the pseudodiploid.



Response Fig. 9 Five models for the domestication of *indica* and *japonica* rice. (**a**) The hypothesis of a simple single origin of cultivated rice. (**b**) The hypothesis of independent domestication origins of cultivated rice. (**c**) This hypothesis considers two origins of cultivated rice followed with bidirectional shares of domestication alleles. (**d**) This hypothesis considers that indica rice was first domesticated and *japonica* rice was subsequently developed from crosses between *indica* rice and another clade of *O. rufipogon*. (**e**) This hypothesis considers that *japonica* rice was first domesticated and *indica* rice was subsequently developed from crosses between *japonica* rice and another clade of *O. rufipogon*. (**f-g**) Computational simulations of in silico data sets under five demographic scenarios using the forwardsimulation program SFS_CODE. The *in silico* domestication loci of *indica* and *japonica* were created by either the domestication model or the admixture model in SFS_CODE. The D-values of the genetic distances between the subspecies of *O. sativa* and each clade of *O. rufipogon* were calculated for both the simulated datasets and the real data. D-values were computed as followed: D-value(*indica*) = d(*indica*,Or-I) -

d(*indica*,Or-IIIa) and D-value(*japonica*) = d(*japonica*,Or-I) - d(*japonica*,Or-IIIa). The genetic distance between any two groups A and B, which was expressed as d(A,B), was computed based on the simple matching coefficient matrix of the SNP data (from **Supplementary Figure 24-25** of Huang et al., *Nature*, 2012).

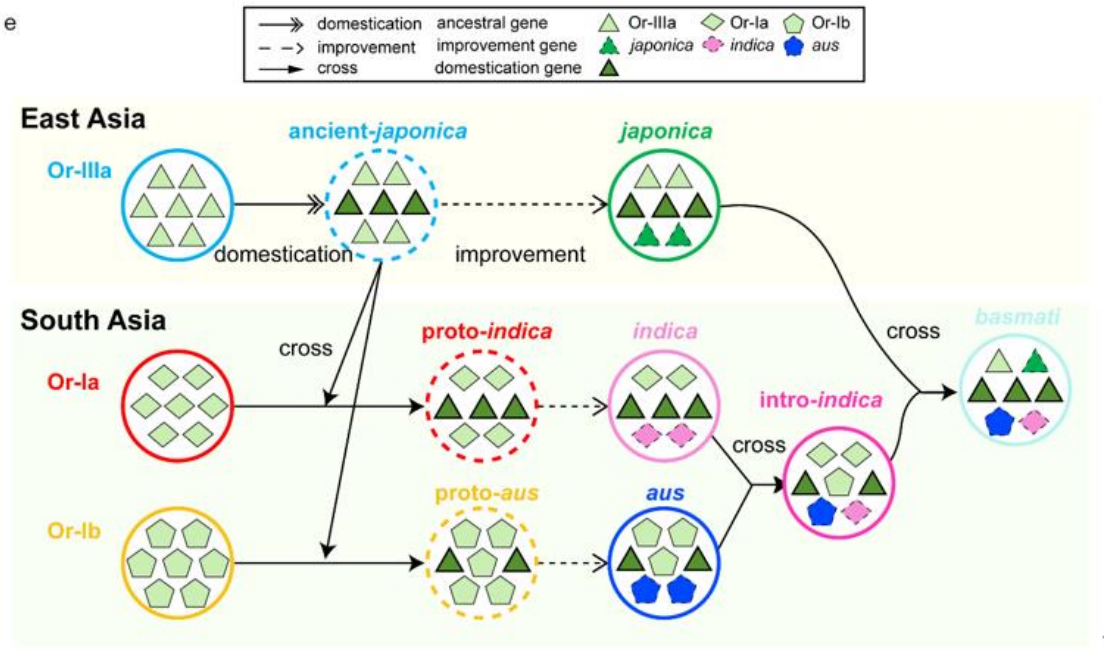
Comment 4. In the authors' response, "I would like to clarify that our model does not propose a single ancestor for all types of Asian cultivated rice. "**Instead, we suggest that different groups of O. sativa just shared a common domestication event, with their respective ancestors diverging and adapting to distinct ecological environment before this event.**" This resulted in significant genetic differentiation and diversity among these groups of O. rufipogon. Subsequent artificial selection and genetic improvements further enhanced this diversity, tailoring rice varieties to meet diverse farming and culinary needs (Jiang et al. 2022). Therefore, indica, through shares key domestication loci with japonica, shows larger diversity and higher genetic differentiation from japonica. The domestication process of wild to cultivated rice was gradual; japonica experienced intense artificial selection and significant genetic bottlenecks, whereas indica could have rapidly incorporated domestication genes through gene flow, preserving its diversity. This suggests that indica did not experience a separate, de novo domestication, challenging the multiple domestication hypothesis." To the reviewer, the authors' arguments are in conflict to their conclusion "We also provided strong evidence to support the initial domestication of all Asian cultivated rice occurred only once" because this conclusion would imply all Asian cultivated rice has a single japonica type of OR ancestor. Unfortunately, the authors did not seem to understand the most recent multi-origin hypothesis of OS which has two key elements: 1) the Asian cultivated rice has multiple ancestors, ancient japonica types in different places of China, ancient japonica types in South and southeast Asia, ancient indica types in China, India and Southeast Asia; 2) the domestication events may have occurred multiple times in different places of China, South and Southeast Asia at different times, while the OR populations had become well differentiated into different populations in different

geographic regions (clearly indicated from all previous and the present studies).

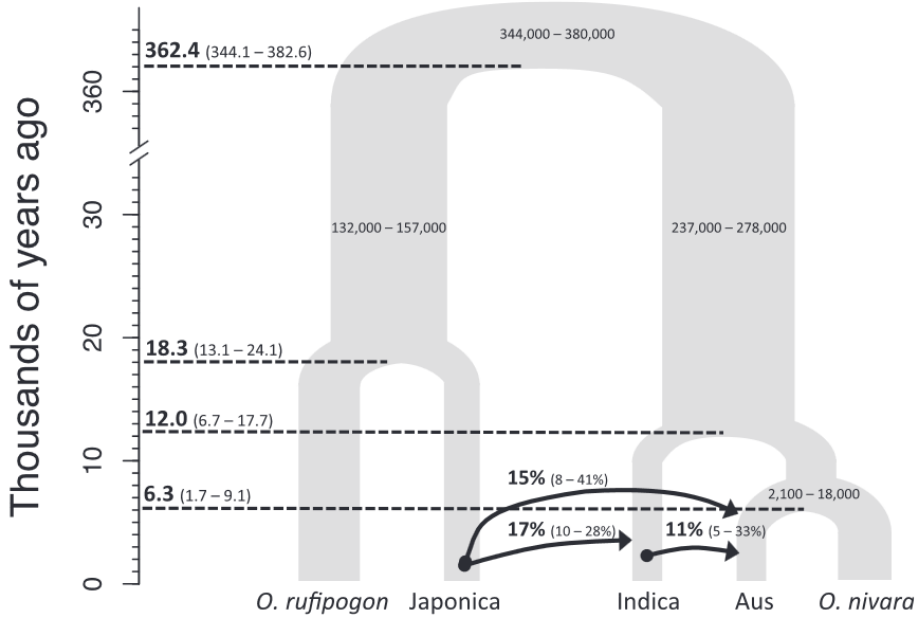
Response: Thank you for raising the concerns.

We would like to review some of the previous reports, although these are not the subject in this study. We would also like to clarify an important distinction in our statement about "the initial domestication of all Asian cultivated rice occurred only once". This does not mean that "all Asian cultivated rice has a single *japonica* type of OR ancestor". The origin of the ancestor and the domestication process are distinct concepts that occurred at different stages. This is visually represented in our model, where *japonica*, *indica* and *aus* have their respective ancestors: Or-IIIa, Or-Ia, Or-Ib, respectively (**Response Fig 10**). This has been supported by our whole genome-wide phylogenetic tree and archetypal analysis (**Fig. 3a-b**). Additionally, our data showed that domestication regions in the different groups of cultivated rice are predominantly derived from Or-IIIa, much more than the proportion seen across the whole genome (**Fig. 4a-c and Supplementary Fig 26-27**). This suggested that while the rice groups had different ancestors, **the crucial domestication traits that made domesticated rice distinct from their wild progenitors** were likely introduced through gene flow from early *japonica* rice.

Our hypothesis is also consistent with the "Multiple Origin but Single Domestication Led to *O. sativa*" model proposed by Choi and D. Purugganan (Choi et al., *Mol. Biol. Evol.*, 2017; Choi and Purugganan, *GenesGenomesGenetics*, 2018). Their results supported that "different rice subspecies had separate origins, but that ***de novo* domestication occurred only once**, **in *O. sativa* ssp. *japonica*, and introgressive hybridization from early *japonica* to proto-*indica* and proto-*aus* led to domesticated *indica* and *aus* rice**" (**Response Fig 11**). Therefore, we all pay more attention to the initial domestication (de novo domestication) event of *O. sativa*, which can be inferred from our analysis of key domestication genes (*sh4*, *PROG1*, *Bh4* and so on). They all shared the same domesticated-state haplotypes (**Supplementary Fig. 32**), a pattern that would be unlikely under multiple independent domestication events in different places.



Response Fig 10 Overview of Asian cultivated rice evolutionary history. This schematic graph outlines the key milestones in the domestication and improvement of various groups of Asian cultivated rice, highlighting notable crossbreeding events (from **Fig. 4e**).



Response Fig. 11 G-PhoCS estimated demographic model of the Asian rice complex. Each internal node has a median mutation rate calibrated divergence time (T) estimate (ka) with its 95% Highest Posterior Density (HPD) in parenthesis. Only the 95% HPD is shown for each ancestral effective population size (Ne). Arrows

indicate the migration band and direction of gene flow. Arrows are labeled with median and 95% HPD for the total migration rate estimates (from **Fig 2** of Choi et al., *Mol. Biol. Evol.*, 2017).

Comment 5. The authors' response to the reviewer's question on the consistency between the genetic/genomic evidence and the archaeological evidence was extremely weak. The earliest histories of rice cultivation in China and India were completely independent according to the archaeological evidence, and both could be traced back to the time that far preceded to the earliest recorded interaction event(s) between the two ancient civilizations. Thus, to the reviewer, all these arguments provided no evidence regarding direct relationships between domesticated rice types (indica, Aus, Basmati and japonica) in South Asia and those domesticated japonica in China, nor any direct relationship between the japonicas and indicas in China.

Response: Thank you for your comments. We would like to address your concerns and clarify our points.

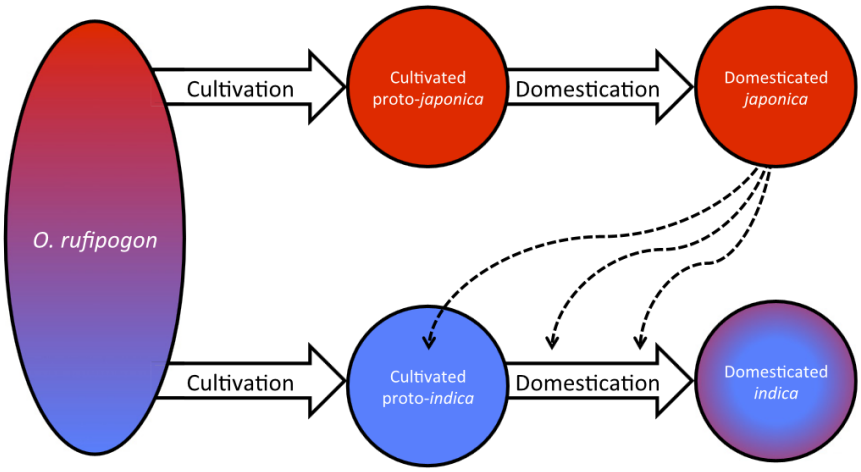
Firstly, we may have to distinguish between the "cultivation" and "domestication" when studying the origins of rice agriculture (Zhao, *Curr. Anthropol.*, 2011; Gross and Zhao, *Proc. Natl. Acad. Sci.*, 2014) (**Response Fig. 12**). Domestication is a complex co-evolutionary process that enhances the interdependence between humans and plants (Purugganan, *Trends Ecol. Evol.*, 2022). Cultivation, on the other hand, refers to a set of intentional human activities aimed at increasing plant yield, which may involve either wild or domesticated plants, both morphologically and genetically. Research has shown that "predomestication cultivation sometimes lasting thousands of years is being increasingly documented in the Old World" (Gross and Zhao, *Proc. Natl. Acad. Sci.*, 2014) and "it is possible for early rice cultivation to be practiced without domesticated rice" (Zhao, *Curr. Anthropol.*, 2011). Therefore, "the earliest histories of rice cultivation" should not be interpreted as evidence of completed domestication, in either China or India.

In last comment, you mentioned that "the earliest archeological evidence in China (lower Yangtze River) and India (lower Ganges River) has traced the cultivation of

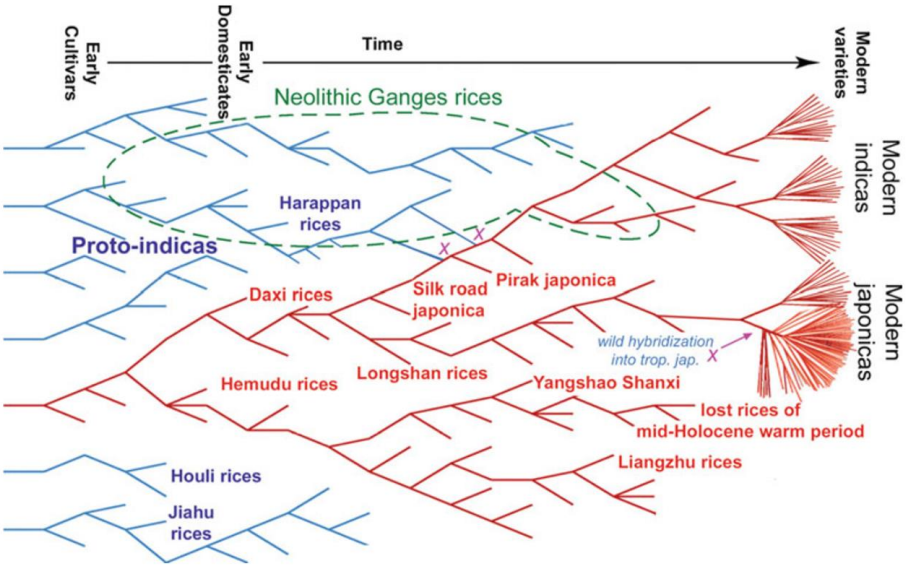
OS back to ~12,000 years and ~9,000 years ago, respectively (Fuller et al. 2010; Gross and Zhao 2014; Ma et al. 2016; Zuo et al. 2017)". However, based on unambiguous morphological domestication traits, the study suggested the presence of domesticated rice can be traced "from 6000 cal. BC in the Lower Yangtze and from ca. 2000 cal. BC in the Ganges" (Fuller et al., *Archaeol. Anthropol. Sci.*, 2010). Based on this, the hybridization event between ancient-*japonica* and *indica*-type ancestors in our hypothesis should occur before the appearance of domesticated rice in India (~4000 years ago), rather than 9000 years ago as you suggested.

Additionally, the archaeological studies you cited (**Fuller et al. 2010; Gross and Zhao 2014; Ma et al. 2016; Zuo et al. 2017**) do not prove that domesticated rice in China and India appeared long before the earliest recorded interactions between these ancient civilizations. On the contrary, these studies support the possibility that hybridization between early *japonica* and *indica* lineages may have taken place near the beginning of domestication (**Response Fig. 12-13**) *(*Fuller et al., *Archaeol. Anthropol. Sci.*, 2010; Fuller, *Rice*, 2011; Gross and Zhao, *Proc. Natl. Acad. Sci*, 2014). Furthermore, Fuller et al suggested that an intensive and expansive agricultural economy focused on rice in the Upper Ganges and Upper Indus regions developed after *japonica* introduced key domestication genes (Fuller et al., *Archaeol. Anthropol. Sci.*, 2010).

Therefore, we think that our hypothesis is consistent with above archaeological findings and hope this clarification will addresses your concerns.



Response Fig. 12. Possible scenarios for the origin of japonica and indica from a genetically diverse *O. rufipogon* population, with genetic differentiation represented by different colors. Possible timing for the movement of domestication alleles from *japonica* into *indica* via hybridization is shown with dashed lines. Colors in the final domesticated *indica* circle represent the contributions from the original *O. rufipogon* populations (blue) and introgression from *japonica* (red) (from Fig 2 of Gross and Zhao, *Proc. Natl. Acad. Sci*, 2014).



Response Fig. 13 Hypothetical representation of the phylogenetic history within rice over time, with hybridizations, and lost lineages, and diversity not sampled in modern germplasm represented (from Fig 2 of Fuller, *Archaeol. Anthropol. Sci.*, 2011)**.**

Remarks: In summary, the reviewer is not convinced by the authors' responses and arguments to the questions. Given the presence of large numbers of previous studies on rice population structure and genomic variation. The population size of 129 OR accessions does not warrant the high quality of the OR pan genome and their comparison between the pan genomes of OS and OR remains invalid. In particular, their selection of "domestication genes" and their interpretations on the analyzed results were inappropriate and highly biased. Their selection of references is highly biased in favor of their conclusions. They failed to test the alternative hypothesis of the multiple origins (domestications) of rice using their data. Taking together, the reviewer feel strongly that the MS does not meet the high standard for publication in Nature.

Response: We would like to say that our study is fully based on the analysis of original genomic data, which were generated from diverse accessions carefully collected by Japanese and Chinese scientists. We hope that we have fully addressed these concerns.

Authors' Response to Referee #4

Referee #4 (Remarks to the Author):

Remarks: This is a revised manuscript by Professor Han and colleagues. I (Reviewer 4) had two simple questions about the original manuscript:

(1) What significant new insights were gained by adding additional O. rufipogon genomes?

(2) What are the significant new insights gained on the origin(s) or domestication of rice not previously reported?

In their rebuttal to question 1, the authors replied that they constructed high-quality pan-genomes, created important genetic resources, and enhanced evolutionary studies. Specifically, they "solidified our 'single domestication event' model. The precise variations dataset have enabled detailed classifications, revealed the genetic evolutionary path of new subpopulations, and facilitated the analysis of the origins of indica-japonica differentiation."

In their rebuttal to question 2, the authors replied that they provided a new understanding of the phylogenetic relationships of wild rice, identified the ancestors of each cultivated rice group, refined the evolutionary paths of two O. sativa subgroups, and reinforced the hypothesis of a single domestication event for Asian cultivated rice.

I'm still not sure where on the spectrum these two replies are between an incremental improvement and an advance worthy of publication in Nature, so I bow to the wisdom of other reviewers who had methodological issues with the manuscript or who are experts in rice domestication.

Response: We sincerely appreciate your thoughtful questions about the significance of our findings. Briefly, this study is fully based on the analysis of original genomic data, which were generated from diverse and representative wild rice *O. rufipogon* accessions carefully collected by the Japanese (Prof. Kurata) and Chinese scientists. The combined data from 129 *O. rufipogon* accessions has created valuable reference

resource that reflects rice diversity and will enhance future rice breeding and genetics study. The several fundamental contributions of this study to both wild rice conservation and cultivated rice breeding can be summarized below:

1) The most comprehensive wild rice pan-genome reference. Wild rice face critical threats from habitat destruction and limited global dispersal, which hinder research and conservation efforts (Zheng et al., 2024). To preserve and integrate these valuable resources, we have constructed the highest quality wild rice pan-genome to date using primarily PacBio HiFi sequencing technology. This comprehensive pan-genome reference encompasses the largest number of *O. rufipogon* accessions, capturing the highest levels of genetic diversity across the widest geographic distribution.

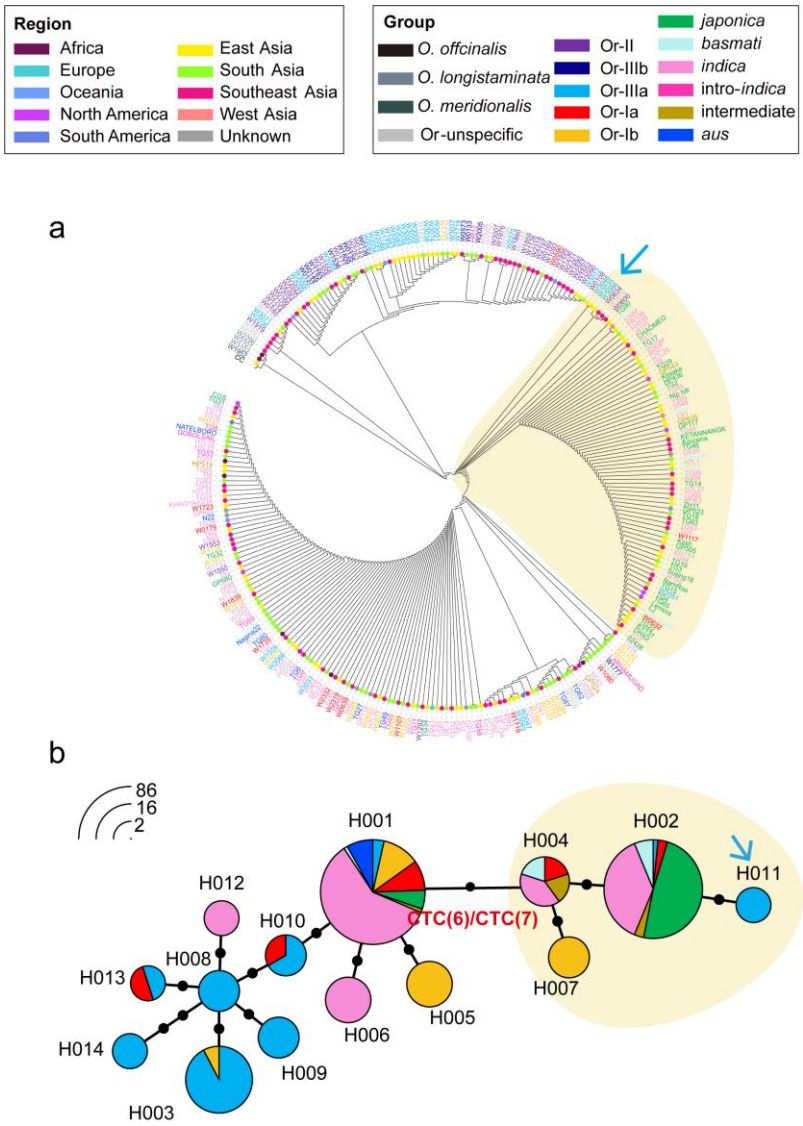
This unprecedented dataset will enable researchers to trace the functional alleles of key genes, advancing our understanding of rice environmental adaptation, phenotypic plasticity, and regeneration potential (Wang and Han, 2022). One notable example involves *COLD6*, a recently characterized gene that negatively regulates cold tolerance through a 3bp InDel in CTC repeat region (Luo et al., 2024). Our analysis revealed the cold-tolerant *COLD6*^{jap} haplotype in 16 *O. rufipogon* accessions, particularly in Or-Ia from India and Or-IIIa from China (**Response Fig. 14**). These findings significantly expand upon previous studies, which, due to limited sampling and sequencing quality constraints, only detected this haplotype in one *O. rufipogon* accession from Nepal and two from India (**Response Fig. 15**). This led to the incomplete conclusion that "the *COLD6*^{jap} allele likely underwent introgression from *O. rufipogon* in India or Nepal during *japonica* rice domestication". Our results suggested that this haplotype in *japonica* is more likely to origin directly from its ancestor Or-IIIa.

2) The most extensive wild rice gene pool. The decrease in the diversity of cultivars is an important constraint on breeding at present, so it is more important to find the genetic variation pool using the wild relatives of crops. In our study, this extensive dataset enabled us to identify 68,901 genes of wild rice, of which ~20% were specific compared to cultivated rice. These genes were more related to defense response and

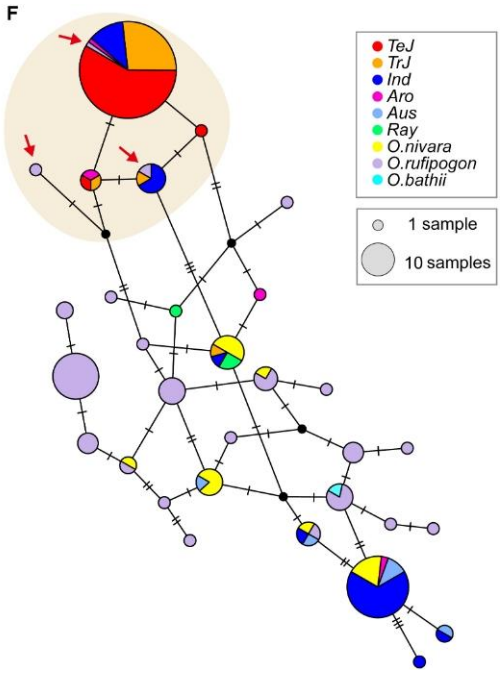
ADP binding processes (**Supplementary Fig. 12**), which can provide direct wild rice sources for breeding resistant varieties. This unprecedented wild rice gene pool offers invaluable resources for future further improvement of crops or de novo domestication of wild plants in a fast-forward way (Huang et al., 2022).

3) The most detailed origin and domestication route of Asian cultivated rice. By precisely classifying *O. rufipogon* populations, we've mapped the evolutionary pathway leading to major cultivated rice groups. This not only helps us to understand the complete evolutionary history of rice, but also helps us to further explore the origin of important functional genes of rice, especially those that are both involved in *indica–japonica* differentiation. It is desirable to further study the relationships among *indica* and *japonica* differentiation genes and to open new possibilities for crop improvement by combining beneficial genes from different rice subspecies (Jiang et al., 2022).

Thank you for your valuable feedback. In response to your suggestion, we have expanded the discussion section to include additional insights (**Line 492-509, Supplementary Fig. 35**), which have significantly contributed to improving the manuscript.



Response Fig. 14 Phylogenetic analysis of *COLD6* gene in our population. a, Phylogenetic tree of *COLD6* gene. The trees feature color-coded labels to distinguish various groups, and the surrounding external circles represent the geographic distribution of each accession. **b,** Haplotype network of *COLD6* gene. In the plots, each circle represents a haplotype, with the circle size proportional to the number of accessions of that haplotype. Different colors indicate different groups. Variations of different haplotypes are indicated by dots. Only haplotypes with at least two samples are included **(**from **Supplementary Fig. 35)**.



Response Fig. 15 Haplotype network of *COLD6* (from Fig 2 of Luo *et al*., 2024).

Reference

Beissinger, T. M., Wang, L., Crosby, K., Durvasula, A., Hufford, M. B., and Ross-Ibarra, J. (2016). Recent demography drives changes in linked selection across the maize genome. *Nat. Plants* **2**:1–7.

Choi, J. Y., and Purugganan, M. D. (2018). Multiple Origin but Single Domestication Led to Oryza sativa. *G3 GenesGenomesGenetics* **8**:797–803.

Choi, J. Y., Platts, A. E., Fuller, D. Q., Hsing, Y.-I., Wing, R. A., and Purugganan, M. D. (2017). The rice paradox: Multiple origins but single domestication in Asian rice. *Mol. Biol. Evol.* Advance Access published 2017, doi:10.1093/molbev/msx049.

Cubry, P., Tranchant-Dubreuil, C., Thuillet, A.-C., Monat, C., Ndjiondjop, M.-N., Labadie, K., Cruaud, C., Engelen, S., Scarcelli, N., Rhoné, B., et al. (2018). The Rise and Fall of African Rice Cultivation Revealed by Analysis of 246 New Genomes. *Curr. Biol.* **28**:2274-2282.e6.

Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The Molecular Genetics of Crop Domestication. *Cell* **127**:1309–1321.

Durvasula, A., Fulgione, A., Gutaker, R. M., Alacakaptan, S. I., Flood, P. J., Neto, C., Tsuchimatsu, T., Burbano, H. A., Picó, F. X., Alonso-Blanco, C., et al. (2017). African genomes illuminate the early history and transition to selfing in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* **114**:5213–5218.

Fuller, D. Q. (2007). Contrasting Patterns in Crop Domestication and Domestication Rates: Recent Archaeobotanical Insights from the Old World. *Ann. Bot.* **100**:903–924.

Fuller, D. Q. (2011). Pathways to Asian Civilizations: Tracing the Origins and Spread of Rice and Rice Cultures. *Rice* **4**:78–92.

Fuller, D. Q., Sato, Y.-I., Castillo, C., Qin, L., Weisskopf, A. R., Kingwell-Banham, E. J., Song, J., Ahn, S.-M., and van Etten, J. (2010). Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeol. Anthropol. Sci.* **2**:115–131.

Gross, B. L., and Olsen, K. M. (2010). Genetic perspectives on crop domestication. *Trends Plant Sci.* **15**:529–537.

Gross, B. L., and Zhao, Z. (2014). Archaeological and genetic insights into the origins of domesticated rice. *Proc. Natl. Acad. Sci.* **111**:6190–6197.

Hammer, K. (1984). Das Domestikations syndrom. *Kult.* **32**:11–34.

Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**:497–501.

Huang, X., Huang, S., Han, B., and Li, J. (2022). The integrated genomics of crop domestication and breeding. *Cell* **185**:2828–2839.

Jiang, L., Wu, L., Wang, Y., Xu, Q., Xu, Z., and Chen, W. (2022). Research progress on the divergence and genetic basis of agronomic traits in xian

and geng rice. *Crop J.* **10**:924–931.

Jing, C.-Y., Zhang, F.-M., Wang, X.-H., Wang, M.-X., Zhou, L., Cai, Z., Han, J.-D., Geng, M.-F., Yu, W.-H., Jiao, Z.-H., et al. (2023). Multiple domestications of Asian rice. *Nat. Plants* **9**:1221–1235.

Kim, M.-S., Lozano, R., Kim, J. H., Bae, D. N., Kim, S.-T., Park, J.-H., Choi, M. S., Kim, J., Ok, H.-C., Park, S.-K., et al. (2021). The patterns of deleterious mutations during the domestication of soybean. *Nat. Commun.* **12**:97.

Liu, H., Li, Q., and Xing, Y. (2018). Genes Contributing to Domestication of Rice Seed Traits and Its Global Expansion. *Genes* **9**.

Luo, W., Xu, Y., Cao, J., Guo, X., Han, J., Zhang, Y., Niu, Y., Zhang, M., Wang, Y., Liang, G., et al. (2024). COLD6-OSM1 module senses chilling for cold tolerance via 2′,3′-cAMP signaling in rice. *Mol. Cell* **84**: 4224–4238.

Meyer, R. S., Choi, J. Y., Sanches, M., Plessis, A., Flowers, J. M., Amas, J., Dorph, K., Barretto, A., Gross, B., Fuller, D. Q., et al. (2016). Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat. Genet.* **48**:1083–1088.

Morishima, H., and Barbier, P. (1990). Mating System and Genetic Structure of Natural Populations in Wild Rice Oryza rufipogon. *Plant Species Biol.* **5**:31–39.

Purugganan, M. D. (2019). Evolutionary Insights into the Nature of Plant Domestication. *Curr. Biol.* **29**:R705–R714.

Purugganan, M. D. (2022). What is domestication? *Trends Ecol. Evol.* **37**:663–671.

Purugganan, M. D., and Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature* **457**:843–848.

Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., He, Q., Ou, S., Zhang, H., Li, X., et al. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**:3542-3558.e16.

Shang, L., Li, X., He, H., Yuan, Q., Song, Y., Wei, Z., Lin, H., Hu, M., Zhao, F., Zhang, C., et al. (2022). A super pan-genomic landscape of rice. *Cell Res.* **32**:878–896.

Sweeney, M., and McCouch, S. (2007). The Complex History of the Domestication of Rice. *Ann. Bot.* **100**:951–957.

Sweeney, M. T., Thomson, M. J., Pfeil, B. E., and McCouch, S. (2006). Caught Red-Handed: Rc Encodes a Basic Helix-Loop-Helix Protein Conditioning Red Pericarp in Rice. *Plant Cell* **18**:283–294.

Thomas, C. G., Wang, W., Jovelin, R., Ghosh, R., Lomasko, T., Trinh, Q., Kruglyak, L., Stein, L. D., and Cutter, A. D. (2015). Full-genome evolutionary histories of selfing, splitting, and selection in Caenorhabditis. *Genome Res.* **25**:667–678.

Vigueira, C. C., Li, W., and Olsen, K. M. (2013). The role of *Bh4* in parallel

evolution of hull colour in domesticated and weedy rice. *J. Evol. Biol.* **26**:1738–1749.

Wang, C., and Han, B. (2022). Twenty years of rice genomics research: From sequencing and functional genomics to quantitative genomics. *Mol. Plant* **15**:593–619.

Yang, J., Liang, X., Wang, Y., Duan, J., Tang, L., and Zheng, X. (2024). Exploring the mating systems of wild rice Oryza rufipogon and O. nivara: implications for population genetic variation. *Genet. Resour. Crop Evol.* Advance Access published June 12, 2024, doi:10.1007/s10722-024-02043-x.

Zhang, F., Xue, H., Dong, X., Li, M., Zheng, X., Li, Z., Xu, J., Wang, W., and Wei, C. (2022). Long-read sequencing of 111 rice genomes reveals significantly larger pan-genomes. *Genome Res.* **32**:853–863.

Zhao, Z. (2011). New Archaeobotanic Data for the Study of the Origins of Agriculture in China. *Curr. Anthropol.* Advance Access published October 1, 2011, doi:10.1086/659308.

Zheng, X.-M., and Ge, S. (2010). Ecological divergence in the presence of gene flow in two closely related Oryza species (Oryza rufipogon and O. nivara). *Mol. Ecol.* **19**:2439–2454.

Zheng, X., Ratnasekera, D., Fan, J., Henry, R. J., Song, B.-K., Olsen, K. M., Joshi, B. K., Banaticla-Hilario, M. C. N., Pusadee, T., Melaku, A. G., et al. (2024). Global Wild Rice Germplasm Resources Conservation Alliance: WORLD WILD-RICE WIRING. *Mol. Plant* Advance Access published March 2024, doi:10.1016/j.molp.2024.03.001.

Zhu, B.-F., Si, L., Wang, Z., Jingjie Zhu, Y. Z., Shangguan, Y., Lu, D., Fan, D., Li, C., Lin, H., Qian, Q., et al. (2011). Genetic Control of a Transition from Black to Straw-White Seed Hull in Rice Domestication. *Plant Physiol.* **155**:1301–1311.

Point-by-point responses to Referee's comments

Referees' comments:

Authors' Response to Referee #1

Referee #1 (Remarks to the Author):

My only comment is about keeping Fig. 2a (and the new sFig. 14a). Variation of polymorphism counts across different subgroups doesn't provide significant insights. And it suffers strong reference bias. Unless I missed something, removing Fig. 2a, sFig. 14a, and Line 273-278 won't impact the subsequent analyses.

Response: Thank you for your comment. We agree and have removed Figure 2a, Supplementary Figure 14, and the corresponding text from the manuscript as suggested.

In terms of statistic test, the error bars in Fig. 2d and sFig. 34 were not defined.

Response: Thank you for pointing out these missing definitions. They have been added to the revised manuscript (**Fig. 2c and Supplementary Fig. 29**).

sFig. 15, the p-value of inversions in genes was 1.21e-02, which didn't pass the threshold of 0.01. Maybe it is better to annotate this was not significant while others were significant.

Response: Thank you for this observation. We have marked non-significant result with '-' symbol in the revised manuscript (**Supplementary Fig. 11**).

Authors' Response to Referee #2

Referee #2 (Remarks to the Author):

I am satisfied with the responses and revisions to my concerns, and don't have any more concerns.

Response: Thank you for your time and positive feedback. We appreciate your thorough review that helped improve the clarity and accuracy of our manuscript.

Authors' Response to Referee #3

Reviewer #3 (Remarks to the Author)

Remarks: In summary, the reviewer is not convinced by the authors'responses and arguments to the questions. Given the presence of large numbers of previous studies on rice population structure and genomic variation. The population size of 129 OR accessions does not warrant the high quality of the OR pan genome and their comparison between the pan genomes of OS and OR remains invalid. In particular, their selection of "domestication genes" and their interpretations on the analyzed results were inappropriate and highly biased. Their selection of references is highly biased in favor of their conclusions. They failed to test the alternative hypothesis of the multiple origins (domestications) of rice using their data. Taking together, the reviewer feel strongly that the MS does not meet the high standard for publication in Nature.

Response: We would like to say that our study is fully based on the analysis of original genomic data, which were generated from diverse accessions carefully collected by Japanese and Chinese scientists. We hope that we have fully addressed these concerns.

The reviewer sees the considerable efforts from the authors in addressing my comments. However, the reviewer is still not convinced by their conclusion on the model of "multiple-origin/single domestication event", in particular, on their conclusion of "single domestication event" reached by their reanalyzed results and by their ways in

data analyses based on the following comments:

Response: Many thanks for the comments regarding to our previous replies---mostly related to the core contents of the manuscript for a high-quality original genomic resource of wild and cultivated rice.

We also appreciate the assistance provided in enhancing the quality of our work. We performed domestication analyses as suggested. The result indicated that the majority of key domestication genes, especially those responsible for critical domestication traits such as reduced shattering and erect growth, underwent a single domestication event.

1. In the data analyses, the authors used "only the progenitors (Or-IIIa and Or-Ib) of other cultivated rice groups as the ancestral groups of *japonica* and *indica* subspecies for searching "domestication" segments across the genomes of the two subspecies. This was a valid comparison. The result "By genome-wide comparison of genetic diversity ratios (πw/πc) and Fst values between ancestral groups and cultivated groups, we identified 50 loci with strong selection signals, covering a total region of 12.35 Mb (Methods). Additionally, beyond SNP-based analysis, we identified domestication-related Presence/Absence Variations (domPAVs) using Fisher's exact tests to compare PAV frequencies between both populations." This method effectively captured most well-known domestication genes in the initial stages of rice domestication within identified selective sweep peaks or domPAVs hotspots, including ……".

However, the reported results did not seem sufficient to support their conclusion of "single domestication event" unless they could provide the following pieces of evidence: (1) Most, if not all, of the "domestication" loci identified through the genome-wide comparison of genetic diversity and *F*st values between Or-IIIa and japonica are the same ones as those identified from the comparison between Or-Ib and indica; (2) The favored alleles (such as alleles for non-shattering at sh4, PROG1, etc) at most, if not all, domestication loci in japonica are the same ones as those in indica and aus; (3) Gene PAVs pattern in japonica and indica to show evidence for the absence of domestication activities (selection for less shattering, lodging, greater grain sizes, etc.) in the 'pre-

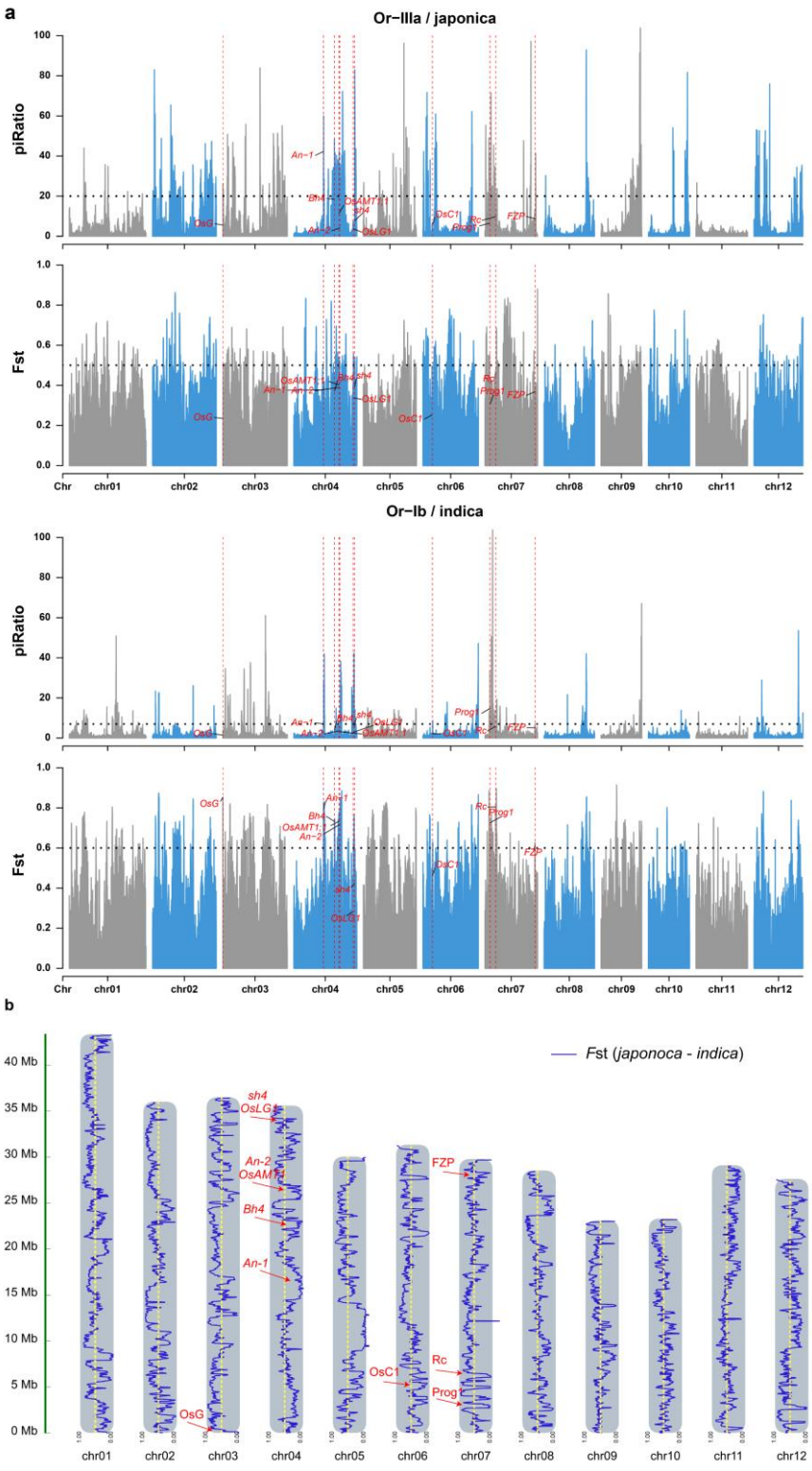
domesticated but cultivated (indica/aus) rice' in South Asia.

Response: We understand the concerns raised regarding our conclusion of "single domestication event" and appreciate the specific solutions provided. Following the suggestions, we have added the analyses below:

(1) Comparison of the "domestication" loci between Or-IIIa and *japonica*, and Or-Ib and *indica*: Following your suggestion, we compared the genetic diversity ratios (π_w/π_c) and *F*st values for both Or-IIIa vs. *japonica* and Or-Ib vs. *indica* (**Response Fig. 1a**). This analysis identified 70 loci in *japonica* and 82 loci in *indica*, with only 7 loci shared between the two groups. This limited overlap is primarily attributed to the severe genetic bottleneck experienced by *japonica* during early domestication and later improvement (Gutaker et al., 2020; Wu et al., 2023). A previous study found that 78.4% of the *japonica* genome exhibited selection signals exceeding the permutation noise threshold, compared to only 41% in *aus* and 16.4% in *indica*, highlighting the significant impact of the genetic bottleneck on *jap*onica (Civan et al., 2015). This strong bottleneck effect reduces the effectiveness of identifying domesticated regions in japonica (Civan et al., 2015). To address this limitation, we believe it is more reasonable to treat the entire cultivar groups as a single domesticated population for detecting selective sweep regions. This approach, as adopted in many studies (Huang et al., 2012; Jing et al., 2023), can help mitigate the influence of domestication bottlenecks and provide a more comprehensive analysis of selective sweeps.

Nevertheless, we demonstrated through an alternative approach that some domestication loci are shared in *indica* and *japonica*. Specifically, we observed extremely low *F*st values between *indica* and *japonica* in many regions including known domestication genes (**Response Fig. 1b**), despite their distinct ancestors and significant genetic differentiation (the average *F*st value across the whole genome between *indica* and *japonica* was 0.55). This suggested that domestication genes were likely shared between these groups through introgression.

These corresponding discussion have been included in **Supplementary Note 3.3, 3.4** and **Supplementary Fig. 19** in the revised manuscript.

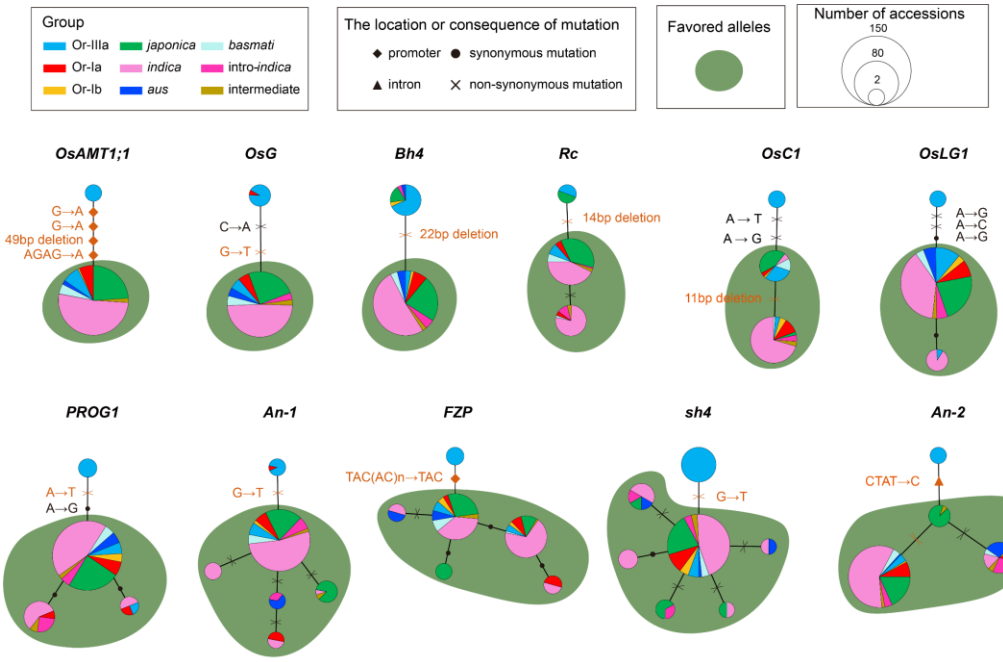


Response Fig. 1 Analysis of selective sweeps in two populations. a, Genome-wide analysis of nucleotide diversity ratio (π Ratio) and *F*st tests are used for two comparisons: Or-IIIa vs. *japonica* (up) and Or-Ib vs. *indica* (down). The horizontal dashed lines indicate the genome-wide threshold (top 5%) of selection signals. The 11 known domestication genes are marked in red text. **b**, Genome-wide distribution of *F*st values between *japonica* and *indica*. The yellow dashed line indicates the genome-wide

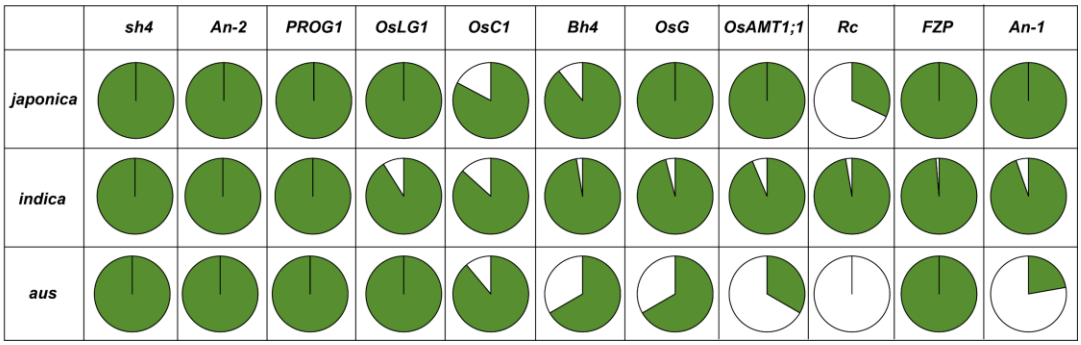
average *F*st value (0.55). The 11 known domestication genes are marked in red text.

(2) Haplotype analysis of domestication genes in *japonica*, *indica*, and *aus*: if all Asian cultivated rice had a single domestication origin, the key domestication genes should share the same favored alleles across all cultivated groups. Therefore, we conducted a haplotype analysis of known domestication genes as suggested (**Response Fig. 2a** and **Supplementary Fig. 26-27**). The results revealed that favored domesticated alleles of most domestication genes, particularly *sh4*, *An-2* and *PROG1*, are the same across *japonica*, *indica* and *aus* (**Response Fig. 2b**). Furthermore, the shared major haplotypes are either directly linked to the haplotype of Or-IIIa (such as *sh4* and *PROG1*) or connected through a haplotype predominantly composed of *japonica* (such as *An-2*) (**Response Fig. 2a**). These findings provided strong evidence that *japonica*, *indica*, and *aus* shared a domestication origin from Or-IIIa.

a



b



Response Fig. 2 Haplotype analysis of 11 known domestication genes. a, Simplified haplotype network of 11 known domestication genes. Only major haplotypes (the largest circle of each gene network) and its linked haplotypes are included, and distantly related haplotypes are omitted for clarity. Each haplotype circle is a pie plot showing the group composition. Symbols between two haplotypes indicate variations, with orange representing known causal functional variants according to existing literature, and black representing variants not previously reported. The green ranges indicate the favored domesticated haplotypes of each gene. **b**, The distribution of favored alleles of 11 domestication genes in *japonica*, *indica* and *aus*. Each pie chart represents the proportion of the favored domesticated alleles (green) and other alleles (white) within each subspecies.

(3) Gene PAVs pattern in the "pre-domesticated but cultivated" rice: We think that the reviewer might want us to validate that the original ancestors (pre-domesticated but cultivated) of *indica* and *aus* in South Asia did not undergo artificial domestication activities before gene flow from *japonica*. However, collecting such 'pre-domesticated but cultivated rice' poses significant challenges, as they have either evolved into cultivated rice through continuous selection or have been lost over time and no longer exist. As an alternative, we think that the results of haplotype analysis of the 11 domestication genes mentioned above can also address the confusion (**Response Fig. 2**). If *aus* and *indica* had experienced independent domestication events, particularly for critical early domestication traits such as reduced shattering and erect growth, it would be highly unlikely for these groups to have the same alleles of key genes regulating these traits as *japonica*. The fact that these shared haplotypes are present across all three groups (*japonica*, *indica*, and *aus*) supports the conclusion that the key domestication genes were fixed first in *japonica*, with gene flow spreading the favored alleles to other populations.

In conclusion, we appreciate the comments, which have helped us to refine our analysis and interpretation of the domestication process. We believe that the additionally analyses and results presented here effectively address the concerns and provide robust support for our hypothesis of a single domestication event.

2. To the reviewer, the reported results suffered greatly from insufficient population sizes for all rice populations and the authors did not show all results in their two

genome-wide comparisons between Or-IIIa and japonica, and between Or-Ia and indica, reflected in the following two aspects: (1) the sampled OS accessions did not cover the whole spectrum of OS population structure which included four subpopulations in Indica, three subpopulations within Japonica, plus Aus and Bas of clear and distinct geographic origins (Wang et al. 2018); (2) their choices of 11 'domestication' loci showing selective sweeps in only a few common Or-IIIa segments for deep analyses were highly biased and results of the remaining 39 identified domestication loci were presented.

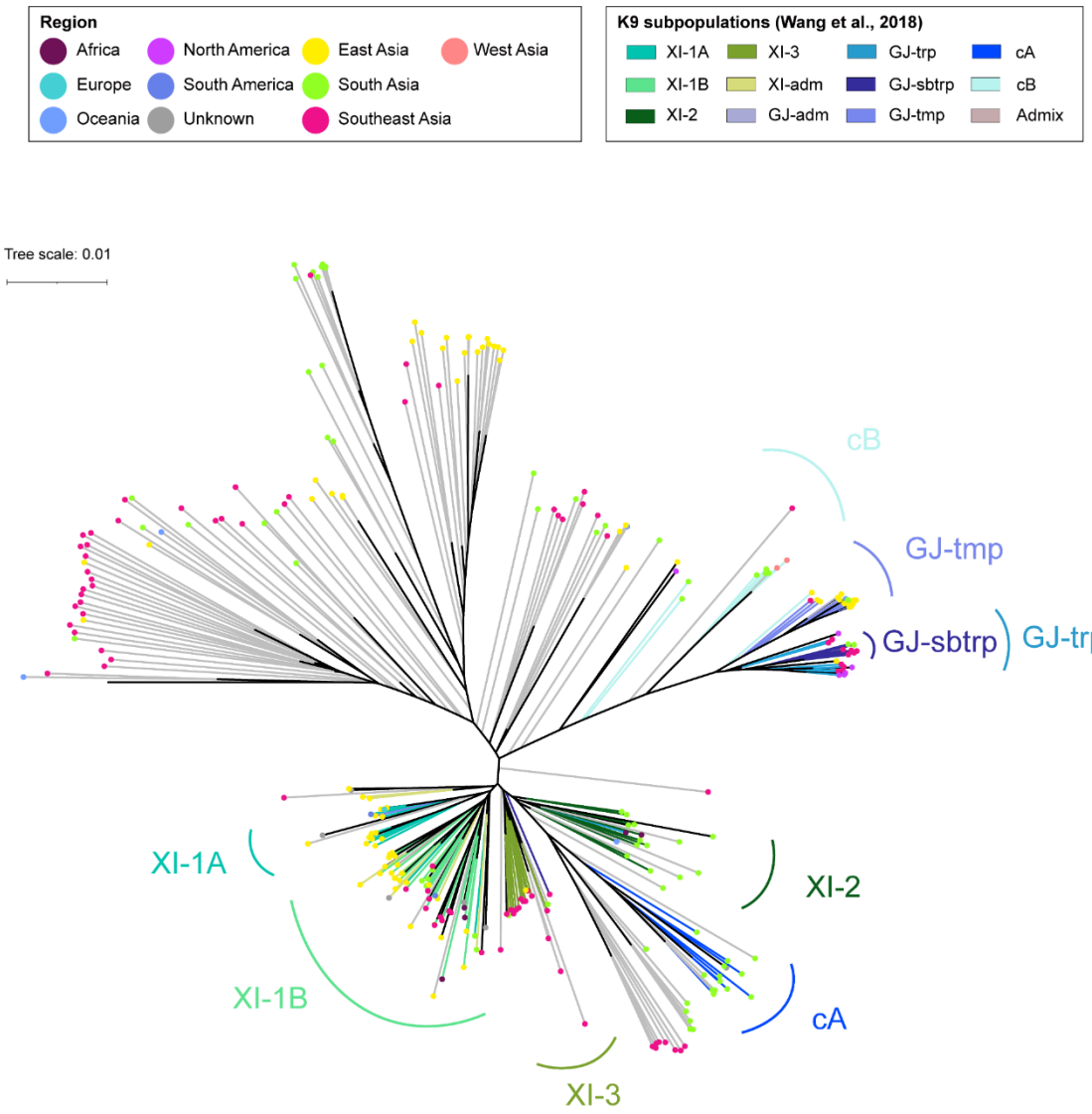
Response: we have further supplemented the following analysis to fully address the concerns.

(1) Subpopulation Representation and Population Size: We acknowledge the importance of representing all rice subpopulations in evolutionary analyses. The classification of cultivated rice into 9 subpopulations ($k = 9$) by the 3K Rice Genomes Project (Wang et al., 2018a) provides a comprehensive and detailed framework for evolutionary research. However, for simplicity, our study grouped cultivated rice into 5 major subpopulations: *japonica*, *indica*, intro-*indica*, *basmati*, *aus*. Despite this simplification, we ensured that our sampled 280 accessions **do** represent all key subgroups, including the four *indica* subpopulations, three *japonica* subpopulations, and the geographically distinct *aus* and *basmati* subpopulations (**Response Table 1**). These subpopulations are explicitly marked in the phylogenetic tree constructed for the 280 accessions (**Response Fig. 3**).

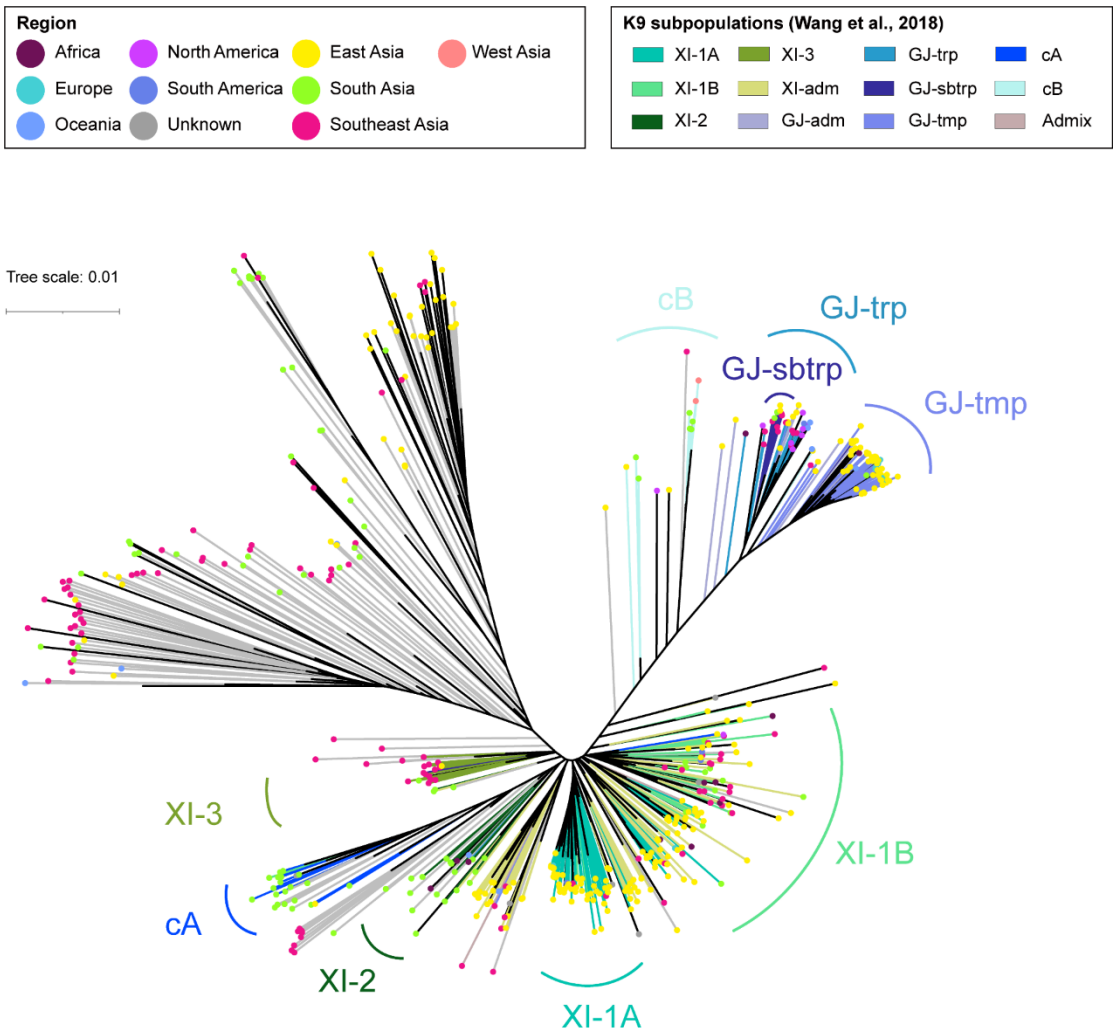
To address potential limitations due to sample size, we expanded our analysis by incorporating additional cultivated rice samples from the super pan-genome dataset (Shang et al., 2022) (**Response Fig. 4** and **Response Table 1**). This enlarged dataset enabled us to validate our findings and confirm that our original conclusions from the original 280 accessions are reliability and robust (**Response Fig. 4-5**).

Response Table 1 Summary of *O. sativa* composition (k = 9) in two populations.

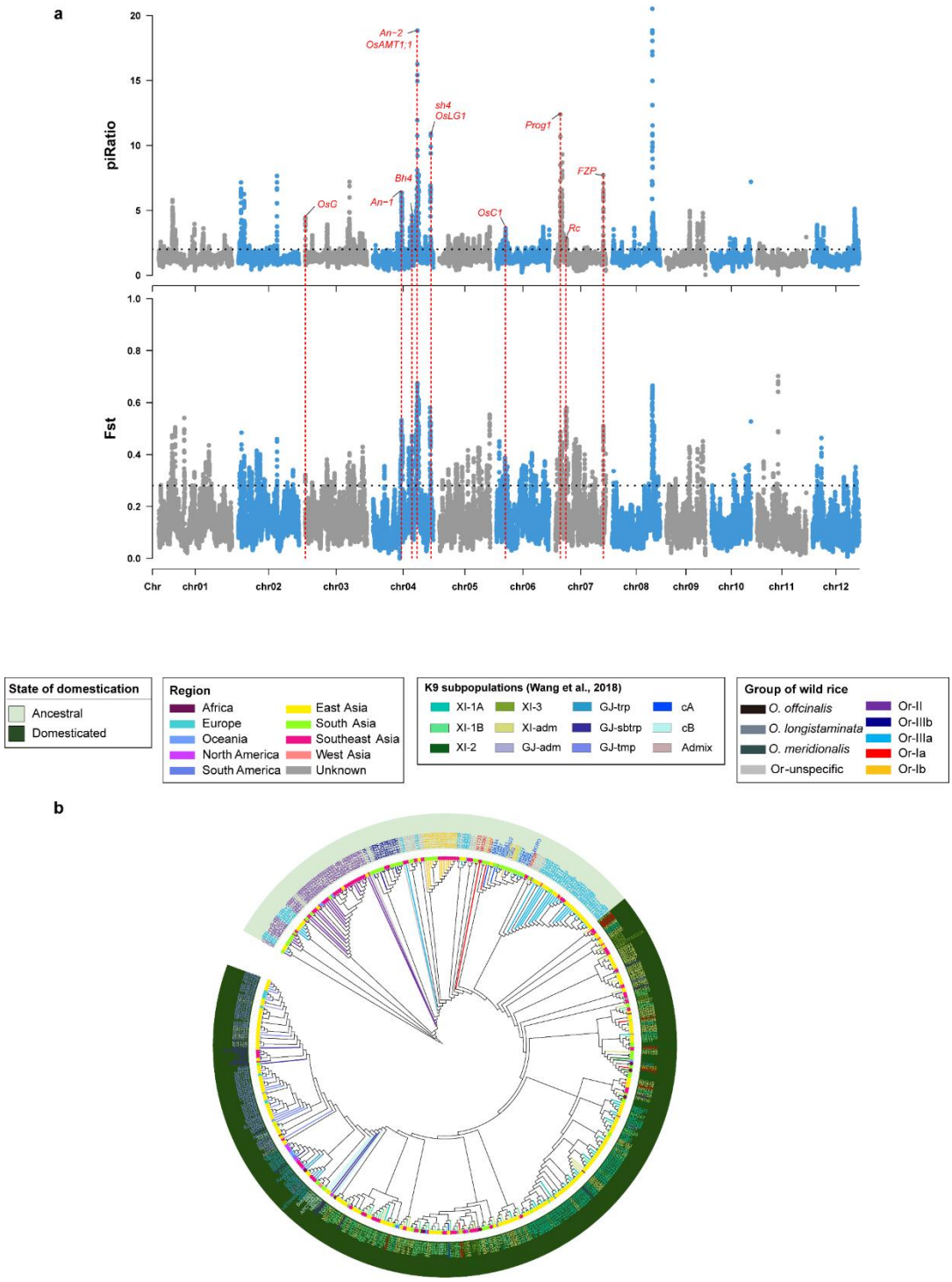
	indica					*japonoca*				*aus*	*basmati*	Admix	Total number of *O. sativa*
	XI-1A	XI-1B	XI-2	XI-3	XI-adm	GJ-tmp	GJ-trp	GJ-sbtrp	GJ-adm	cA	cB		
Population of 280 accessions	16	36	12	13	9	16	12	9	1	10	9	3	146
Population of 510 accessions	66	58	14	14	67	57	20	12	10	14	10	6	348



Response Fig. 3 Phylogenetic analysis of 280 wild and cultivated rice accessions. This phylogenetic tree was constructed based on whole-genome SNPs of 280 accessions. Tree branches are color-coded to represent subpopulation information, with all wild rice branches depicted in gray. The text labels on the outside of the tree show the corresponding k = 9 subpopulation from 3K-RG study (Wang et al., 2018a). The external circles denote the geographic distribution of each accession.



Response Fig. 4 Phylogenetic analysis of 510 wild and cultivated rice accessions. This phylogenetic tree based on whole-genome SNPs of 510 accessions. Tree branches are color-coded to represent subpopulation information, with all wild rice branches depicted in gray. The text labels on the outside of the tree show the corresponding k = 9 subpopulation from 3K-RG study (Wang et al., 2018a). The external circles denote the geographic distribution of each accession.

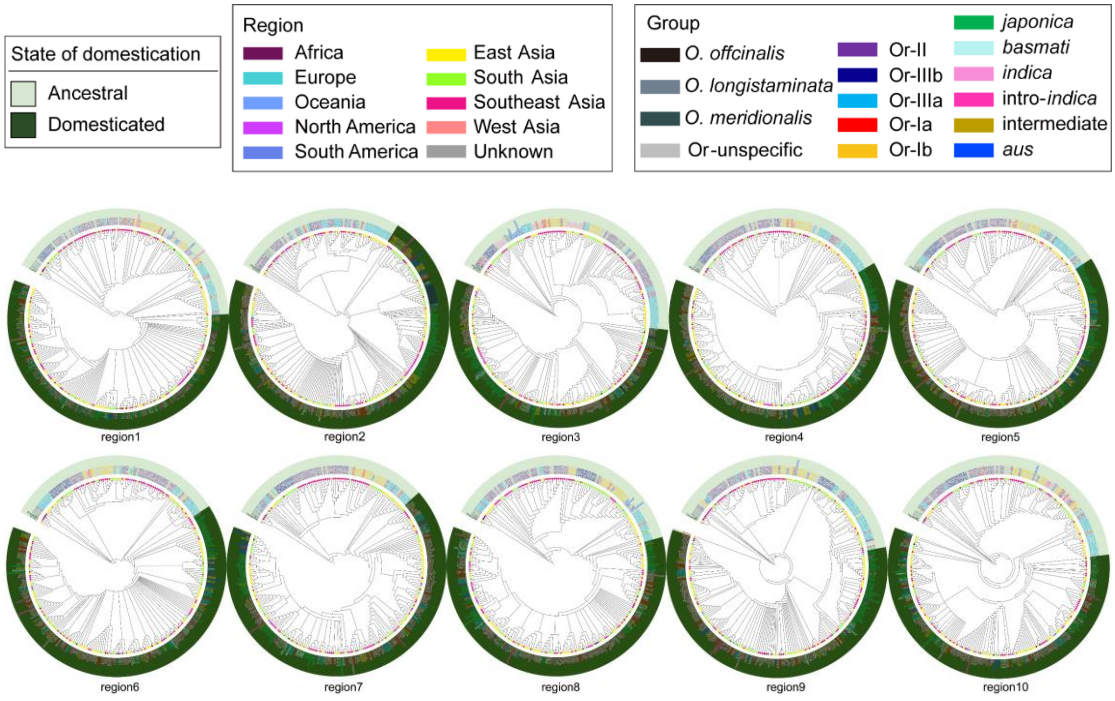


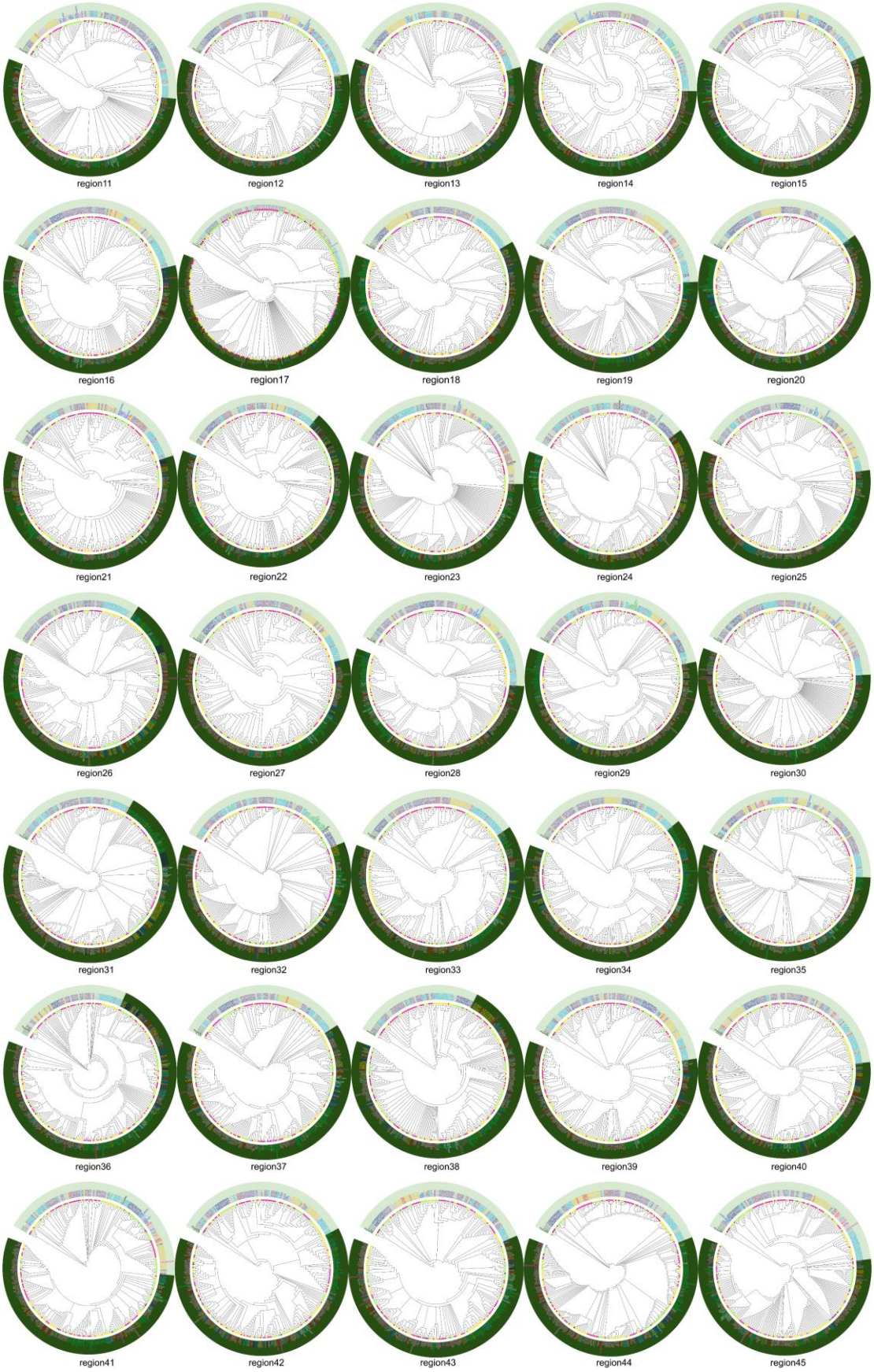
Response Fig. 5 Selective sweep analysis of 510 wild and cultivated rice accessions. **a,** Genome-wide analysis of nucleotide diversity ratio (piRatio) and *F*st values are showed for Or-IIIa and Or-Ib vs *O. sativa*. Horizontal dashed lines indicate the genome-wide threshold (top 5%) of selection signals. The 11 known domestication genes are marked in red text. **b,** Phylogenetic tree of 510 *Oryza* accessions based on the SNPs within identified selective sweep regions. The tree features color-coded labels and branches to distinguish the $k = 9$ subpopulation from 3K-RG study (Wang et al., 2018a), and the surrounding external circles represent the geographic distribution of each accession. The color ranges indicate the state of domestication.

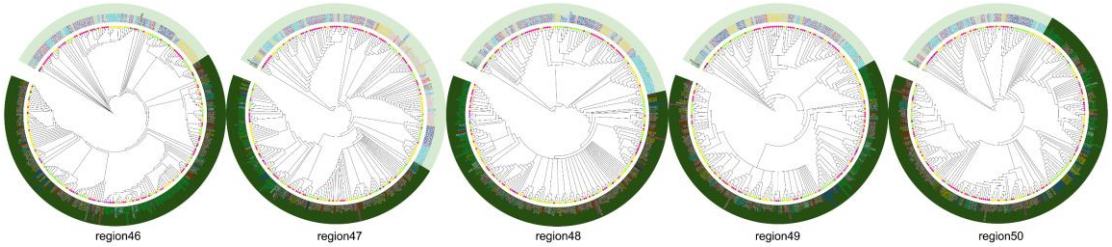
(2) Selective Sweep Analysis and Selection of Domestication Genes: We focus on 11 well-known domestication genes because these genes are widely recognized for their roles in regulating major traits during domestication. All 11 genes are located within the selective sweep peaks, including *OsG* (region10), *An-1* (region14), *Bh4* (region16), *OsAMT1;1* (region17), *An-2* (region18), *sh4* and *OsLG1* (region20), *OsC1* (region28), *PROG1* (region33), *Rc* (region35).

To clarify, we expanded our analysis by constructing phylogenetic trees for each of the 50 identified selective sweep regions (**Response Fig. 6**). Our results show that, in 43 of the 50 regions (except region 23,24,29,32,38,41,46), cultivated rice clustered together, with Or-IIIa tended to be the direct domestication origin of all cultivars. Additionally, we have also conducted phylogenetic trees specifically for the 11 domestication genes, which revealed clearer patterns and further supported our conclusions that all cultivated rice shared a domestication event originated from Or-IIIa (**Response Fig. 7**).

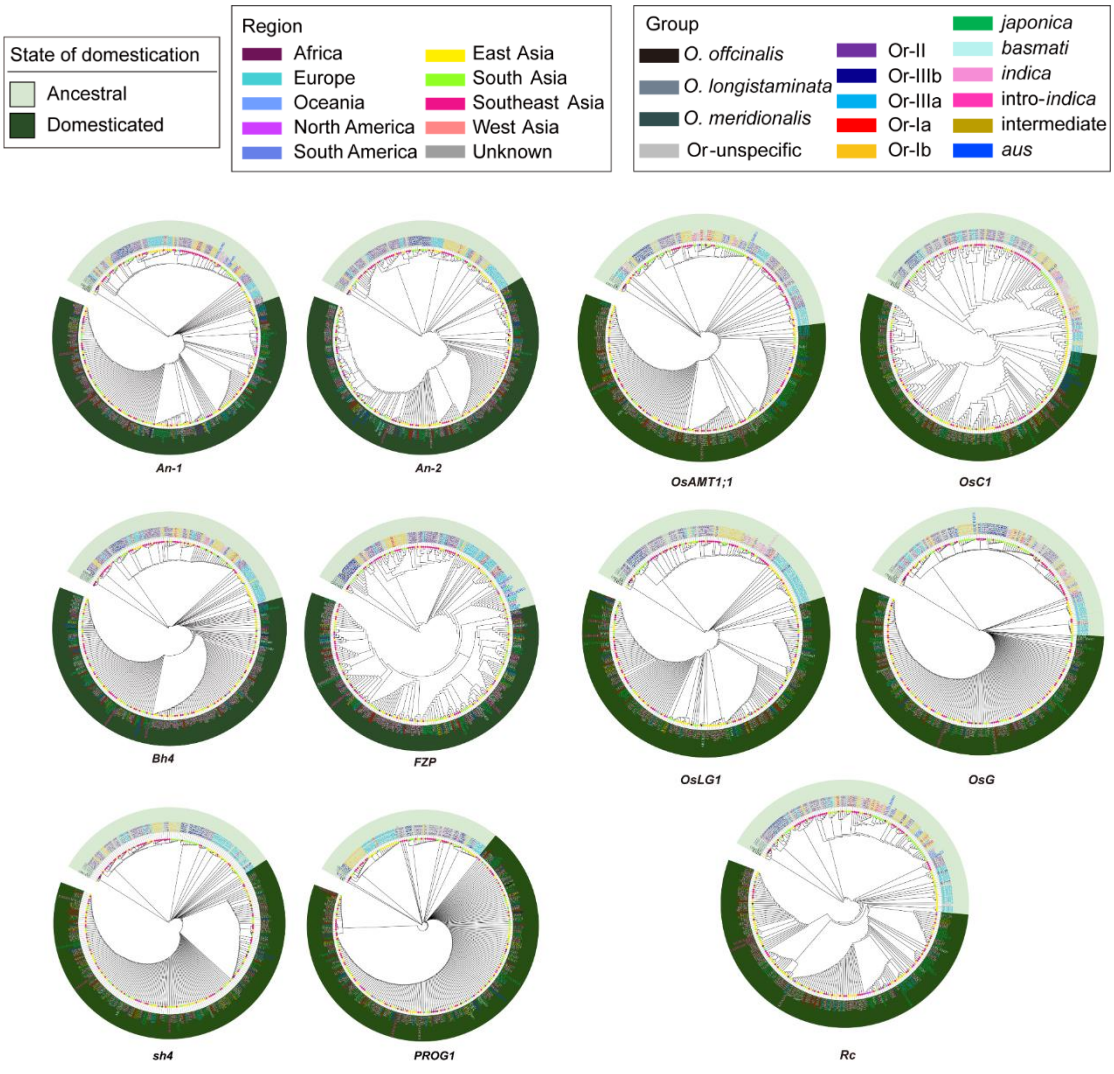
These corresponding discussions have been included in **Supplementary Note 3.2** and **Supplementary Fig. 22** in the revised manuscript. We believe these additional results addressed your concerns about potential bias.




region11
region12
region13
region14
region15
region16
region17
region18
region19
region20
region21
region22
region23
region24
region25
region26
region27
region28
region29
region30
region31
region32
region33
region34
region35
region36
region37
region38
region39
region40
region41
region42
region43
region44
region45



Response Fig. 6 Phylogenetic tree of 280 *Oryza* accessions based on the SNPs within each identified selective sweep region. The tree features color-coded labels to distinguish various groups, and the surrounding external circles represent the geographic distribution of each accession. The color ranges indicate the state of domestication.



Response Fig. 7 Phylogenetic trees of 11 known domestication genes. The trees feature color-coded labels to distinguish various groups, and the surrounding external circles represent the geographic distribution of each accession. The color ranges indicate the state of domestication.

Apparently, the authors did not read a recently reported study in J. Advanced Research titled

"Selective and comparative genome architecture of Asian cultivated rice (Oryza sativa L.) attributed to domestication and modern breeding" by Wang et al. (2022). In the same comparisons between 456 OR genomes and 504 OS landrace genomes using the same methods, Wang et al. (2022) reported the identification of 624 (126.4 Mb) and 318 (44.5 Mb) domestication blocks (segments) japonica landraces and indica landraces, respectively, with only 179 (24.8 Mb) shared between japonica and indica. In these domestication segments, they found a total of 14,202 and 5,163 domestication selected genes in japonica and indica, respectively, with 3,115 domestication selected genes shared between the two subspecies. They also found that natural selection during the domestication process of japonica was much stronger than that of indica. They found that the detected domestication gene in japonica and indica shared many common gene families, but different members of the same gene families were selected in japonica and indica during domestication. Thus, their results indicated that evolution resulted from convergent selection was actually acting on different genes/alleles in populations japonica and indica, suggesting the domestication events in japonica and indica were largely independent from each other.

Response: Thank you for the information. This study in *J. Advanced Research* by Wang et al. (2022) provided us with a valuable reference. We have carefully read the paper and cited it as Ref #10 in our manuscript. However, this study reported 14,202 and 5,163 domestication selected genes in *japonica* and *indica*, respectively, meaning that a significant portion, up to 35%, of the rice genes were associated with domestication. We think that the number of selected domestication genes may be overestimated in that study. Domestication loci are limited in number within the genome, as strong selection on too many loci can lead to excessive loss of competing lineages and population collapse, with estimates suggesting an upper limit of 50 to 100 loci under selection during domestication (Allaby et al., 2016; Alam and Purugganan, 2024). Many detected selective sweep regions with strong artificial selection signals do not overlap with any cloned domestication genes and may represent false positives or driven by more recent development selective pressures (Huang et al., 2022; Alam and Purugganan, 2024). To ensure the consistency and robustness of our study, we used the top 5% of selective sweep signals as the threshold for identifying domestication-related regions,

considering both the reduction of false positives and the strong genetic bottleneck experienced by *japonica* (Gutaker et al., 2020; Wu et al., 2023).

The major genes controlling the critical changes during rice domestication have been genetically identified and functionally characterized by the great efforts of the rice community (Li et al., 2006; Li et al., 2006; Sweeney et al., 2007; Jin et al., 2008; Tan et al., 2008; Ding et al., 2011; Ishii et al., 2013; Luo et al., 2013; Hua et al., 2015; Gu et al., 2015; Wang et al., 2018b; Huang et al., 2018; Zheng et al., 2019). As proofed in a recent work, *de novo* domesticated rice from allotetraploid wild rice was rapidly and successfully developed by editing these major genes (Yu et al., 2021). Therefore, the known domestication genes likely contribute significantly to the genetic basis for crucial domestication alterations, and we also focus on the known genes and their domestication mutations in these analyses.

Minor comments

Several minor comments are listed as follows:

1. When compared the wild rice pangenome with the 129 OS pangenome reported previously, the authors indeed found 7,592 wild rice specific genes. However, in the Abstract, the authors still claimed that they found 13,728 wild rice specific genes, which is wrong. The reviewer is also wondering how many genes in their discovered genes are OS specific (present in the cultivated rice only, and absent in the wild rice accessions)?

Response: Thank you for highlighting the discrepancy in the number of wild rice-specific genes between the Abstract and Results section. To clarify, the 13,728 wild rice-specific genes reported in the Abstract were derived from our original analysis using our own wild and cultivated rice pan-genome data (129 *O. rufipogon* and 16 *O. sativa*). This is valid at the specified dataset and is not wrong. We chose to retain this number in the Abstract to ensure consistency with the other findings from our pan-genome population we reported. For additional validation, we further identified 7,592 wild rice-specific genes as an equal number of cultivated rice samples and wild rice (n=129).

Regarding the question on OS-specific genes, we also performed an in-depth analysis. We identified 630 OS-specific genes in our own dataset (129 *O. rufipogon* and 16 *O. sativa*), and 3,348 OS-specific genes in the extended dataset (129 *O. rufipogon* and 129 *O. sativa*). Notably, this number is smaller than the count of wild rice-specific genes, which further underscores the genetic diversity of wild rice.

2. Response Fig. 12 was incorrect because the earliest cultivation of rice and its domestication should have occurred simultaneously in the same process that had taken hundreds or even thousands of years in different places of Asia. Thus, the reviewer is wondering how it could have happened in South Asia where "pre-domesticated rice" was cultivated for several thousands of years without going through any artificial/natural selection for domestication traits, as argued by the authors or Fuller et al., (2010). If so, the authors should provide solid evidence to show why those "pre-domesticated" indica rice accessions were cultivated for ~4,000 – 5,000 years in South Asia without going through any artificial/natural selection for those domestication traits. Also, the authors have to clarify how they determine when the domestication process in japonica or indica was completed, or simply by citing previous archeological reports. To the reviewer, there is no universal unambiguous morphological domestication traits in rice, given the fact that indica and japonica are distributed in very different ecological environments because even today, the modern japonica and indica varieties differ greatly in these "domestication" traits!!

Response: Yes, it is conceivable that early rice cultivation could occur without domesticated rice (Zhao et al, 2011). However, it is difficult to ascertain whether these "pre-domesticated rice" in South Asia underwent natural or successful artificial selection for domestication traits. These "pre-domesticated but cultivated rice" are no longer available, as they have either evolved into modern cultivated rice through continuous selection or have been lost over time. As a result, proving this through genomic or genetic analysis is not feasible. Nevertheless, our haplotype analysis of key early domestication genes revealed that the major domesticated haplotypes in cultivated rice all originate from Or-IIIa group (**Response Fig. 2**).

Additionally, we also want to clarify that our study did not attempt to pinpoint when the domestication process of *japonica* or *indica* rice was completed. Establishing such a timeline would require substantial archaeological and genetic evidence, which is beyond the scope of this research.

While we acknowledge the significant physiological and phenotypic differences between modern *indica* and *japonica* rice—for example, in grain shape, heading date, tiller angle and chilling tolerance—these differences largely reflect the ecological adaptations of cultivated rice to diverse environments (Jiang et al., 2022), rather than domestication traits. Our results showed that their genetic differentiation was mainly due to the divergence of their respective ancestors (~60%) and a larger genetic bottleneck in *japonica* (~30%). However, these findings are not contradictory with the hypothesis that they shared the same domestication origin. **Traits such as the reduction in prostrate growth and decreased shattering are widely recognized as critical early domestication traits, not only in rice but also in other crop species** (Doebley et al., 2006). These traits are universal and unambiguous in both *japonica* and *indica*, and the key domestication genes regulating these traits (e.g., *sh4* and *Prog1*) shared the same domestication mutations across cultivated rice.

We hope these explanations have fully addressed your concerns.

In summary, the reviewer believes that inferences on the origin of Cultivated rice have to be built on large numbers of sampled accessions of representative geographic origins because of the well-known population differentiation within the species as well as within the two major subspecies of rice and a significant amount of genetic variation exists among populations of different geographic origins. This is particularly true for the self-pollinated plant species like rice. Apparently, the sample sizes and their geographic representativeness of this study were insufficient for them reach their conclusions. Genetically, virtually all domestication traits are complex ones controlled by large numbers of genes/loci. The reported domestication genes only account for a small portion of genes/loci involved. In fact, all genes showing selective sweeps in the new comparisons by the authors should be deeply analyzed in order to support or reject

their conclusion on the model of multiple-origin/single domestication event. Taking together, the reviewer is still not convinced by the authors' responses and arguments to the questions.

Response: The phylogenetic tree in Response Fig. 3 demonstrates our pan-genome dataset covering a broad spectrum of wild and cultivated rice accessions. Our genome analysis, based on high-quality original genomic resources, offers a sound basis for domestication analyses. In future, the resource will also facilitate identifications of functionally significant genes in wild rice and enhance future rice breeding efforts.

Although domestication traits are complex ones controlled by multiple genes/loci, the major genes controlling the critical changes during rice domestication have been genetically identified and functionally characterized by the great efforts of the rice community (Li et al., 2006; Sweeney et al., 2007; Jin et al., 2008; Tan et al., 2008; Zhu et al., 2011; Ding et al., 2011; Ishii et al., 2013; Luo et al., 2013; Hua et al., 2015; Gu et al., 2015; Wang et al., 2018b; Huang et al., 2018; Zheng et al., 2019). As proofed in a recent work, *de novo* domesticated rice from allotetraploid wild rice was rapidly and successfully developed by editing these major genes such as *An-1, Sh4* and *Bh4* etc (Yu et al., 2021). The reported domestication genes likely contribute significantly to the genetic basis for crucial domestication alterations.

References

Alam, O., and Purugganan, M. D. (2024). Domestication and the evolution of crops: variable syndromes, complex genetic architectures, and ecological entanglements. *Plant Cell* **36**:1227–1241.

Allaby, R. G., Kitchen, J. L., and Fuller, D. Q. (2016). Surprisingly Low Limits of Selection in Plant Domestication. *Evol. Bioinforma. Online* **11**:41–51.

Civan, P., Craig, H., Cox, C. J., and Brown, T. A. (2015). Three geographically separate domestications of Asian rice. *Nat Plants* **1**:15164.

Ding, Z., Wang, C., Chen, S., and Yu, S. (2011). Diversity and selective sweep in the OsAMT1;1genomic region of rice. *BMC Evol. Biol.* **11**:61.

Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The Molecular Genetics of Crop Domestication. *Cell* **127**:1309–1321.

Gu, B., Zhou, T., Luo, J., Liu, H., Wang, Y., Shangguan, Y., Zhu, J., Li, Y., Sang, T., Wang, Z., et al. (2015). An-2 Encodes a Cytokinin Synthesis Enzyme that Regulates Awn Length and Grain Production in Rice. *Mol. Plant* **8**:1635–1650.

Gutaker, R. M., Groen, S. C., Bellis, E. S., Choi, J. Y., Pires, I. S., Bocinsky, R. K., Slayton, E. R., Wilkins, O., Castillo, C. C., Negrão, S., et al. (2020). Genomic history and ecology of the geographic spread of rice. *Nat. Plants* **6**:492–502.

Hua, L., Wang, D. R., Tan, L., Fu, Y., Liu, F., Xiao, L., Zhu, Z., Fu, Q., Sun, X., Gu, P., et al. (2015). LABA1, a Domestication Gene Associated with Long, Barbed Awns in Wild Rice. *Plant Cell* **27**:1875–1888.

Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**:497–501.

Huang, Y., Zhao, S., Fu, Y., Sun, H., Ma, X., Tan, L., Liu, F., Sun, X., Sun, H., Gu, P., et al. (2018). Variation in the regulatory region of *FZP* causes increases in secondary inflorescence branching and grain yield in rice domestication. *Plant J.* **96**:716–733.

Huang, X., Huang, S., Han, B., and Li, J. (2022). The integrated genomics of crop domestication and breeding. *Cell* **185**:2828–2839.

Ishii, T., Numaguchi, K., Miura, K., Yoshida, K., Thanh, P. T., Htun, T. M., Yamasaki, M., Komeda, N., Matsumoto, T., Terauchi, R., et al. (2013). OsLG1 regulates a closed panicle trait in domesticated rice. *Nat. Genet.* **45**:462–465.

Jiang, L., Wu, L., Wang, Y., Xu, Q., Xu, Z., and Chen, W. (2022). Research progress on the divergence and genetic basis of agronomic traits in xian and geng rice. *Crop J.* **10**:924–931.

Jin, J., Huang, W., Gao, J.-P., Yang, J., Shi, M., Zhu, M.-Z., Luo, D., and Lin, H.-X. (2008). Genetic control of rice plant architecture under domestication. *Nat. Genet.* **40**:1365–1369.

Jing, C.-Y., Zhang, F.-M., Wang, X.-H., Wang, M.-X., Zhou, L., Cai, Z., Han, J.-D., Geng, M.-F., Yu, W.-H., Jiao, Z.-H., et al. (2023). Multiple domestications of Asian rice. *Nat. Plants* **9**:1221–1235.

Li, C., Zhou, A., and Sang, T. (2006). Rice Domestication by Reducing Shattering. *Science* **311**:1936–1939.

Luo, J., Liu, H., Zhou, T., Gu, B., Huang, X., Shangguan, Y., Zhu, J., Li, Y., Zhao, Y., Wang, Y., et al. (2013). An-1 Encodes a Basic Helix-Loop-Helix Protein That Regulates Awn Development, Grain Size, and Grain Number in Rice. *Plant*

Cell **25**:3360–3376.

Shang, L., Li, X., He, H., Yuan, Q., Song, Y., Wei, Z., Lin, H., Hu, M., Zhao, F., Zhang, C., et al. (2022). A super pan-genomic landscape of rice. *Cell Res.* **32**:878–896.

Sweeney, M. T., Thomson, M. J., Cho, Y. G., Park, Y. J., Williamson, S. H., Bustamante, C. D., and McCouch, S. R. (2007). Global Dissemination of a Single Mutation Conferring White Pericarp in Rice. *PLOS Genet.* **3**:e133.

Tan, L., Li, X., Liu, F., Sun, X., Li, C., Zhu, Z., Fu, Y., Cai, H., Wang, X., Xie, D., et al. (2008). Control of a key transition from prostrate to erect growth in rice domestication. *Nat. Genet.* **40**:1360–1364.

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R. R., Zhang, F., et al. (2018a). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**:43–49.

Wang, M., Li, W., Fang, C., Xu, F., Liu, Y., Wang, Z., Yang, R., Zhang, M., Liu, S., Lu, S., et al. (2018b). Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nat. Genet.* **50**:1435–1441.

Wu, D., Xie, L., Sun, Y., Huang, Y., Jia, L., Dong, C., Shen, E., Ye, C.-Y., Qian, Q., and Fan, L. (2023). A syntelog-based pan-genome provides insights into rice domestication and de-domestication. *Genome Biol.* **24**:179.

Yu, H., Lin, T., Meng, X., Du, H., Zhang, J., Liu, G., Chen, M., Jing, Y., Kou, L., Li, X., et al. (2021). A route to de novo domestication of wild allotetraploid rice. *Cell* **184**:1156-1170 e14.

Zhao, Z. (2011). New Archaeobotanic Data for the Study of the Origins of Agriculture in China. *Curr. Anthropol.* Advance Access published October 1, 2011, doi:10.1086/659308.

Zheng, J., Wu, H., Zhu, H., Huang, C., Liu, C., Chang, Y., Kong, Z., Zhou, Z., Wang, G., Lin, Y., et al. (2019). Determining factors, regulation system, and domestication of anthocyanin biosynthesis in rice leaves. *New Phytol.* **223**:705–721.

Zhu, B.-F., Si, L., Wang, Z., Jingjie Zhu, Y. Z., Shangguan, Y., Lu, D., Fan, D., Li, C., Lin, H., Qian, Q., et al. (2011). Genetic Control of a Transition from Black to Straw-White Seed Hull in Rice Domestication. *Plant Physiol.* **155**:1301–1311.

Comments on the revised MS of 2024-03-04666B

1. The authors did not understand my question on the invalid comparison between the pan genomes of OS and OR. Although the sampled 16 OS accessions did not affect the pan genome estimation of the sampled OR accessions, the authors should use the pan genome of OS accessions of equivalent sample size and similar geographic distribution for their comparison and claim. In other words, the authors should compare all OR pan genes they discovered with all previously published OS pan genes in order to be sure those 13,728 wild rice "novel" genes are indeed OR specific, even though they are absent in the 16 sampled OS accessions. In the authors' response "Following your suggestion, we have compared our identified wild rice-specific genes with three published rice pan-genomes (Qin et al. 2021; Zhang et al. 2022; Shang et al. 2022) (Supplementary Table 10 in the revised manuscript). As expected, the number of genes exclusive to wild rice is different in this broader comparison. However, the results demonstrate that most of the wild rice-specific genes we identified (63%-82%) remain unique, even when compared against a larger set of cultivated rice genomes. These are truly important and novel findings." If so, the authors' claim for the discovering "13,728 wild rice-specific genes in their abstract does not hold true because they should include the results into the revised MS, stating only at most ~63% of the 13,728 genes are likely OR specific. Again, when the authors did not obtain an accurate estimate on the number and diversity of RGAs in OS (given the 16 OS accessions they used), their claim that OR population has more PGAs and higher diversity at PGAs than OS remains invalid, until an OS population of equivalent size and geographic distribution is used in their analyses.
2. The reviewer does not understand how the authors reached the speculation "These cultivars have been identified to be domesticated through crossing the proto-japonica type cultivar (with multi-domestication traits, such as non-shattering seed habit, from prostrate to erect growth, transition from black to straw-white seed hull, etc) with diverse wild rice O. rufipogon populations for adapting various environments (Gutaker et al. 2020)", given the fact that japonica cultivars or japonica type of ORs are known to have very low outcrossing rates, and "the modern breeding process" involving hybridization and artificial selection of rice has a short history < 130 years.

3. Unfortunately, the authors did not answer to my question regarding how the serious violation of the assumptions "the neutrality and random mating" for software MSMC2 influences their conclusions. In fact, the authors should use computer simulation to address how the breeding systems and different types

of selection affect their results. They did not even try to test the alternative hypothesis of the multiple origins using their data.

4. In the authors' response, "I would like to clarify that our model does not propose a single ancestor for all types of Asian cultivated rice. **"Instead, we suggest that different groups of O. sativa just shared a common domestication event, with their respective ancestors diverging and adapting to distinct ecological environment before this event."** This resulted in significant genetic differentiation and diversity among these groups of O. rufipogon. Subsequent artificial selection and genetic improvements further enhanced this diversity, tailoring rice varieties to meet diverse farming and culinary needs (Jiang et al. 2022). Therefore, indica, through shares key domestication loci with japonica, shows larger diversity and higher genetic differentiation from japonica. The domestication process of wild to cultivated rice was gradual; japonica experienced intense artificial selection and significant genetic bottlenecks, whereas indica could have rapidly incorporated domestication genes through gene flow, preserving its diversity. This suggests that indica did not experience a separate, de novo domestication, challenging the multiple domestication hypothesis." To the reviewer, the authors' arguments are in conflict to their conclusion "We also provided strong evidence to support the initial domestication of all Asian cultivated rice occurred only once" because this conclusion would imply all Asian cultivated rice has a single japonica type of OR ancestor. Unfortunately, the authors did not seem to understand the most recent multi-origin hypothesis of OS which has two key elements: 1) the Asian cultivated rice has multiple ancestors, ancient japonica types in different places of China, ancient japonica types in South and southeast Asia, ancient indica types in China, India and Southeast Asia; 2) the domestication events may have occurred multiple times in different places of China, South and Southeast Asia at different times, while the OR populations had become well differentiated into different populations in different geographic regions (clearly indicated from all previous and the present studies).

5. The authors' response to the reviewer's question on the consistency between the genetic/genomic evidence and the archaeological evidence was extremely weak. The earliest histories of rice cultivation in China and India were completely independent according to the archaeological evidence, and both could be traced back to the time that far preceded to the earliest recorded interaction event(s) between the two ancient civilizations. Thus, to the reviewer, all these arguments provided no evidence regarding direct

relationships between domesticated rice types (indica, Aus, Basmati and japonica) in South Asia and those domesticated japonica in China, nor any direct relationship between the japonicas and indicas in China.

In summary, the reviewer is not convinced by the authors' responses and arguments to the questions. Given the presence of large numbers of previous studies on rice population structure and genomic variation. The population size of 129 OR accessions does not warrant the high quality of the OR pan genome and their comparison between the pan genomes of OS and OR remains invalid. In particular, their selection of "domestication genes" and their interpretations on the analyzed results were inappropriate and highly biased. Their selection of references is highly biased in favor of their conclusions. They failed to test the alternative hypothesis of the multiple origins (domestications) of rice using their data. Taking together, the reviewer feel strongly that the MS does not meet the high standard for publication in Nature.