Title page

Structural variations contribute to subspeciation and yield heterosis in rice

Zhiwu Dan[1,†], Yunping Chen[1,†], Wenchao Huang[1,*]

[1]State Key Laboratory of Hybrid Rice, Key Laboratory for Research and Utilization of Heterosis in Indica Rice, the Ministry of Agriculture, The Yangtze River Valley Hybrid Rice Collaboration & Innovation Center, College of Life Sciences, Wuhan University, Wuhan, China.

[*]Author for correspondence: wenchaoh@whu.edu.cn (W.H.)

[†]These authors contributed equally.

Abstract

Yield heterosis has been extensively exploited in hybrid breeding, with inter-subspecies hybrids often exhibiting the most pronounced effects. However, developing elite hybrids remains a laborious and time-consuming process. The genetic basis of heterosis has been debated for over a century, hindered largely by the lack of high-quality genomes. Here, we assembled genomes for 12 representative *indica* and *japonica* rice accessions. Using sequence variants of the *Phr1* gene, we functionally validated two deletions responsible for phenol reaction variation between the subspecies. Comparative genomic analyses revealed extensive sequence variation among these inbred lines and highlighted the pivotal role of structural variants (SVs) in rice subspeciation. Importantly, the number of SVs between parental inbred lines significantly correlated with heterosis across 17 agronomic traits, with distinct correlation patterns for intra- and inter-subspecific F_1 hybrids. We identified SVs associated with *S5-ORF5* and *OsBZR1* and validated their function to heterosis for seed setting rate and yield heterosis, respectively, underscoring the importance of SVs in breeding inter-subspecific hybrids. The genomic SVs altered gene expression and these transcriptional changes effectively explained the variances in heterosis. Furthermore, translocations outperformed other SVs and their heterozygous haplotypes exhibited heterosis over homozygous ones. Our findings establish SVs as pivotal drivers in subspeciation and highlight the overdominance model for harnessing rice heterosis.

Keywords: Structural variant, Yield heterosis, Overdominance model, Subspeciation, Rice

Introduction

Heterosis refers to the phenomenon where heterozygous F_1 hybrids exhibit superior performance compared to their parental lines (Shull, 1948), significantly enhancing crop yield and contributing to global food security (Hochholdinger and Yu, 2025). Though several well-known genetic models—including dominance (complementation of deleterious loci) (Bruce, 1910; Jones, 1917; Gu et al., 2023; Wang et al., 2023), overdominance (heterozygosity of loci) (Shull, 1908, 1911; East, 1936; Semel et al., 2006; Krieger et al., 2010), and epistasis (interaction of loci) (Powers, 1944; Jiang et al., 2017)—have been proposed to explain heterosis, the genetic mechanisms underlying crop heterosis remain debated. Over a century ago, Shull proposed the overdominance model, emphasizing heterozygosity of "elements", now always interpretated as genes or genomic sequences. He also suggested that the degree of heterosis correlates with the number of parental differential "characters", now may be understood as genomic variants (Shull, 1911). Consistent with the opinion of Shull, East postulated that heterosis arises from the action of non-defective alleles and is linked to parental genetic disparity (East, 1909; East, 1936).

The advent of DNA marker technology in the late 20th century enabled investigations into the relationship between parental genetic disparity and heterosis in F_1 hybrids using various molecular markers (Stuber et al., 1992; Xiao et al., 1995; Xiao et al., 1996; Zhang et al., 1996). Studies reported significant correlations between heterosis (or hybrid performance) and genetic distance or marker heterozygosity (Xiao et al., 1996; Zhang et al., 1996; Schrag et al., 2006; Schrag et al., 2009; Schrag et al., 2010; Dan et al., 2014; Liu et al., 2014). However, due to technology limitations, these studies relied on small number of marker sets, unable to reflect genome-wide sequence variation.

Advances in genome sequencing over the past two decades have dramatically improved the detection ability and completeness of genomic variants. High-density single-nucleotide polymorphism (SNP) and small insertion and deletion (InDel) markers (< 50 bp) have been widely used with various mathematical models to predict

heterosis or hybrid performance in maize, rice, and wheat (Xu et al., 2014; Zhao et al., 2014; Zhao et al., 2015). Sequencing thousands of rice hybrids revealed that the number of superior alleles correlated strongly with grain number per panicle (Huang et al., 2015). In-depth sequencing analysis of 17 representative hybrid populations demonstrated that heterozygosity at heterosis-associated loci underlay rice heterosis (Huang et al., 2016). The sequencing of 1,143 *indica* inbred lines further underscored the importance of genomic variants in breeding superior hybrid rice (Lv et al., 2020). Based on sequencing results of 2,839 hybrids and their 9,839 segregation individuals, the authors found that the breeding of elite *indica-indica* hybrids was a process of pyramiding favorable alleles from broadened genetic resources (Gu et al., 2023). And they revealed widespread genomic complementarity in *indica-japonica* heterosis (Gu et al., 2023). Moreover, genomic sequencing analysis of two test-cross populations elucidated that rice heterosis was due to homozygous disadvantage at insufficient genetic background, rather than heterozygous advantage (Xie et al., 2022).

The contribution of large size genomic variants (≥ 50 bp), e.g., genome assembly-dependent structural variants (SVs), to rice heterosis remains poorly understood. Recent progress in sequencing technologies, including Pacific Biosciences (PacBio), Oxford Nanopore Technologies, and Hi-C platforms (Li and Durbin, 2024), has enabled the assembly of high-quality genomes. Over the past five years, the completeness and accuracy of genome assemblies have been significantly improved, leading to the release of multiple high-quality rice genomes (Qin et al., 2021; Shang et al., 2022; Shang et al., 2023; Sun et al., 2023). Given the important roles of SVs in tomato improvement and maize heterosis (Alonge et al., 2020; Wang et al., 2023), SVs are likely also critical for rice heterosis.

In this study, we selected 12 representative *indica* and *japonica* inbred lines and assembled their genomes for comparative analysis. We identified genomic variants, particularly SVs, and assessed their functional implications in rice sub-speciation. Using these inbred lines as parents, we constructed a diallel-cross population and analyzed the relationship between parental genomic variants and heterosis of 17 agronomic traits. We identified SVs in *S5-ORF5* and *OsBZR1* and validated their

contributions to heterosis for seed setting rate and yield heterosis, respectively. In combination with gene expression data, we found that SVs drove transcriptional changes and that these effectively explained variance in rice heterosis. Furthermore, translocations outperformed other SVs and their heterozygous haplotypes conferred heterosis over the homozygous ones, supporting the overdominance model of heterosis in rice.

Results

Selection of 12 representative rice inbred lines

To investigate the genetic mechanisms underlying rice heterosis, we selected 12 representative *indica* and *japonica* accessions from 223 inbred lines (ILs) based on previous genotyping data (Dan et al., 2024) (Supplementary Fig. S1). These 12 ILs were equally clustered into *indica* and *japonica* groups (Fig. 1A), based on different types of genomic variants analyzed from the reported Illumina sequencing data (Dan et al., 2024). Among the selected ILs, seven—Yuetai B (YB), Mianhui725, Chenghui9348 (9311K), 610234, C418, Wuyunjing8, and Balilla—are well-established cultivars known for high grain yield or their widespread exploitation in breeding elite inbred or hybrid rice (https://www.ricedata.cn). Other ILs show highly recognizable phenotypes, e.g., large grain/primary branch/secondary branch number for Qianlijing, long panicle for R4115, and tall plant height for JR2 (Dan et al., 2024). These ILs represent phenotypically and geographically diverse varieties originating from China, India, and Italy (Fig. 1B).

Using a complete diallel-cross design, we generated F_1 hybrids from the 12 ILs and calculated better-parent heterosis (BPH) of 17 agronomic traits for the finally obtained 120 F_1 hybrids. We observed that F_1 hybrids derived from inter-subspecies exhibited significantly reduced seed setting rate and increased plant height compared to those of intra-subspecific hybrids (Fig. 1C and Supplementary Fig. S2), which is consistent with known breeding challenges. Furthermore, heterosis of the inter-subspecific hybrids showed significant differences from the pooled averages of *indica-indica* and *japonica-japonica* hybrids.

Genome assembly and annotation of 12 inbred lines

To generate genome assemblies, we obtained genomic data using single-molecule real-time (SMRT) technology on the PacBio Revio platform for the 12 ILs, achieving an average sequencing depth at 30.46× (Supplementary Tables S1-S2). We performed *de novo* assembly for two representative ILs, YB (*indica*) and Balilla (*japonica*), with the assistance of Hi-C data, yielding an average sequencing depth at 127.80× (Supplementary Fig. S3 and Supplementary Table S3). The genomes at the chromosome level of the remaining ten ILs were assembled using YB and Balilla as reference genomes for *indica* and *japonica* subgroups, respectively. The average contig N50, scaffold N50, and total assembled genome sizes for final genome assemblies of the 12 ILs were 28.74 Mb, 31.75 Mb, and 396.36 Mb, respectively, with an average anchor ratio of 98.18%. The high-quality and completeness of these genome assemblies were further confirmed by metrics such as average quality value of bases (63.06), re-mapping rate of HiFi reads (99.51%), Benchmarking Universal Single-Copy Orthologs (BUSCO) (98.61%), and long terminal repeats (LTRs) assembly index (22.13) (Supplementary Fig. S4).

We annotated repetitive sequences (tandem repeats and transposable elements), protein-coding genes, and non-coding RNAs for YB and Balilla (Supplementary Table S4). Repetitive sequences constituted 57.35% (YB) and 55.44% (Balilla) of their genomes, with transposable elements (TEs) comprising 55.09% and 53.10%, respectively. Among these TEs, DNA transposons and LTRs were the predominant types, accounting for 24.26-25.67% and 24.9-28.64% of the genomes, respectively. Three strategies, including homology-based (using five previously reported genomes), *de novo*, and transcriptome-based (full-length RNA sequencing) predictions, were integrated for the annotation of protein-coding genes. This resulted in YB containing 44,332 genes and Balilla containing 43,706 genes. We revealed that 94.4% of the annotated genes in YB and 95.35% in Balilla have assigned functions in one of the following six databases: InterPro, Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Swiss-Prot, TrEMBL, and Non-Redundant. Compared to the five established genomes (Nipponbare, 93-11, Minghui63, Zhenshan97, and R498),

these newly assemblies showed significant different in the lengths of genes and coding sequences, highlighting their value to current rice genomic resources (Supplementary Fig. S5). Moreover, miRNA and rRNA were the two predominant types of non-coding RNAs in the genome assemblies of both YB and Balilla, which accounted for 0.26-0.28% and 0.34-0.96% of the genomes, respectively.

Accuracy validation of genome assemblies using *Phr1*

To further validate assembly accuracy, we analyzed *Phr1*, a gene underlying phenol reaction phenotype between *indica* and *japonica* (Yu et al., 2008). Phylogenetic result of *Phr1* sequences from the 12 ILs, along with 9311 and Nipponbare, displayed slightly changed clustering patterns in subspecies, with Qianlijing grouped with *japonica* (Fig. 1D). Using the sequence of 9311 as reference, *Phr1* contained 24 SNPs and six InDels. Among these InDels, there existed three insertions (1-2, 6, and 7 or 14 bp) and three deletions (2, 18, and 29 bp), with only the 18 and 29 bp deletions located in exons. The result of phenol reaction experiment revealed that seed hulls of five ILs without the 18/29 bp deletions immediately turned black after one day's treatment, while the other nine remained yellow or brown even after 124 days (Fig. 1E). These results showed the high assembly accuracy of these new genomes and confirmed important roles of the 18/29 bp deletions in outcomes of phenol reaction.

Pan-genome assembly of 12 inbred lines

A pan-genome of the 12 ILs was constructed, comprising 47,490 gene families (Fig. 1F). Among the whole gene families, 26,556 (with 345,960 genes) were identified in all 12 ILs and defined as core-gene families (55.92% of gene families and 70.70% of genes), 2,037 (with 24,040 genes) identified in 11 ILs were defined as softcore gene families (4.29% of gene families and 4.91% of genes), 18,850 (with 118,436 genes) identified in 2-10 ILs were defined as dispensable families (39.69% of gene families and 24.20% of genes), and 47 (with 892 genes) identified in only one IL were defined as private gene families (0.10% of gene families and 0.18% of genes). The number of gene family increased with the addition of genomes, reaching a plateau at n = 10, indicating the representativeness of the selected ILs (Fig. 1F). Unlike the trend observed in pan genome, the number of core gene family decreased as the number increased,

with the lowest point reached at n = 12, underscoring the specificity of each IL (Fig. 1F). For the number of non-core gene families among these ILs, YB and Qianlijing had the highest and lowest amount, respectively, showing functional variations between ILs (Fig. 1G). Furthermore, we analyzed the proportion of genes with annotated functions in protein domain annotation (Pfam), GO, and KEGG. The annotation ratios decreased progressively from core to softcore to dispensable genes (Pfam: 63.23%, 43.05%, and 35.77%; GO: 69.55%, 56.17%, and 51.02%; and KEGG: 27.97%, 13.12%, and 10.26%), demonstrating the conservation of core genes amidst diversification (Fig. 1H).

Pan-structural variants in the 12 inbred lines

Genomic variants—especially SVs (presence/absence variants, copy number variants, translocations, and inversions)—were identified through comparing the reference genome YB to the other 11 ILs. Consistent with the patterns observed for SNPs and InDels (Fig. 2A), both the number and size (about 57-168 Mb) of SVs gradually increased from *indica* to *japonica* (Fig. 2B-C). All *japonica* ILs harbored more SVs than the *indica*. Notably, two large inversions were detected on chromosome 6 (inversion_6: 5.83 Mb) and 8 (inversion_8: 5.90 Mb) (Fig. 2D), which were between the six *indica* and six *japonica* or exclusive to JR2 (Supplementary Fig. S6-S10), respectively, indicating that inversions contribute to the evolutionary divergence of *indica* and *japonica*.

Unlike the changing trend of pan-genome, the number of pan-SVs did not plateau even at n = 11, underscoring the necessity of including more accessions for a comprehensive exploration of SVs (Fig. 2E). Core, softcore, dispensable, and private SVs were assigned using the same criteria as the pan-genome. Among the total SVs, 878 (0.73%) identified in 11 ILs were classified as core SVs, while 1,640 (1.36%) detected in 10 ILs were categorized as softcore SVs (Fig. 2F). Additionally, 66,655 (55.28%) SVs identified in 2-9 ILs were classified as dispensable SVs, and 51,413 (42.64%) identified in only one IL were considered private SVs. Different from the trends of core, softcore, and private SVs, the number of dispensable SVs increased from *indica* to *japonica* (Fig. 2G), suggesting the roles of these dispensable SVs played in rice subspeciation. The length of different SV types varied (Fig. 2H), with three large

absence variants (AVs) detected on chromosome 6 (AV_6, 0.73 Mb), 8 (AV_8, 1.16 Mb), and 11 (AV_11 and 0.94 Mb; Fig. 2I). Given that repetitive sequences accounted for more than 50% of the genomes, we analyzed the distributions of SVs and found that these SVs were predominantly detected in DNA transposons and LTRs (Fig. 2J), suggesting regulatory functions of SVs in non-coding regions of the genomes.

Associations between structural variants and heterosis of 17 agronomic traits

We performed paired genomic comparisons and identified both unique and common genomic variants for each pair of ILs, representing different and shared variants, respectively (Fig. 3A and Supplementary Table S5). On average, unique genomic variants between any two ILs comprised 2,018,539 SNPs, 391,298 InDels, and 40,082 SVs (17,690 AVs, 11,975 PVs, 6,327 CNVs, 3,941 translocations, and 149 inversions). The number of these unique variants was highly correlated (Fig. 3B), with the exception of translocation, inferring co-locations of these variants (Supplementary Fig. S11). Approximately half of these variants located in intergenic regions (57.45% of SNPs, 50.42% of InDels, and 50.56% of SVs), emphasizing the significance of these regions in differentiating the ILs. In contrast to other SV types, the majority (77.94%) of unique inversions occurred in exonic regions. Furthermore, genetic relatedness of the 12 ILs assessed using unique SVs, particularly translocations, showed variation compared to those based on unique SNPs or InDels, indicating the distinct functional contributions of different variant types to genomic divergence (Fig. 3C and Supplementary Fig. S12).

Correlation analysis between the number of parental genomic variants and heterosis of the 17 traits revealed divergent patterns for unique versus common variants (Fig. 3D). According to the number of significant correlations, the unique variants exhibited stronger correlations with heterosis than common variants (Fig. 3E). Meanwhile, no notable change in correlation patterns was observed between stop-gain and stop-loss variants in SNPs and InDels, synonymous and nonsynonymous SNPs, or frameshift and nonframeshift InDels. These findings emphasized the importance of genomic differences between the two parents in determining heterosis.

Though unique SVs accounted for only 1.99% and 10.24% of the total SNPs and InDels, respectively, their correlations with heterosis were comparable, probably due to

their larger size. Among the five types of SVs, inversions, with fewer number than translocations, showed stronger correlations with heterosis for plant height, inferring a size-dependent influence of genomic variants on heterosis of vegetative traits. Different from correlation patterns of heterosis for other yield-related traits, the number of unique variants showed negative correlations with heterosis for seed setting rate, showing reproductive isolation due to genomic disparity. Notably, we observed that an increase in unique SVs corresponded with a slight rise in heterosis for seed setting rate of intra-subspecific F_1 hybrids, but a dramatic decrease in *indica-japonica* F_1 hybrids (Fig. 3F). The inter-subspecific hybrids possessed significantly more unique SVs in their parents than their intra-subspecific counterparts (Fig. 3G and Supplementary Fig. S13), suggesting the roles of these SVs in heterosis of inter-subspecies. The unique SVs between the parental lines also exhibited stronger correlations with heterosis for tiller number in inter-subspecific F_1 hybrids, showing the potential for enhancing grain yield through utilizing heterosis for tiller number (Supplementary Fig. S14). However, the number of unique SVs displayed weak or no correlation with yield heterosis in inter-subspecific F_1 hybrids, likely due to reproductive isolation-driven reductions in seed setting rate.

Absence variant of *S5-ORF5* underlies heterosis for seed setting rate

Given the established role of the *S5* loci in embryo-sac fertility and seed setting rate of inter-subspecific F_1 hybrids (Chen et al., 2008; Yang et al., 2012), we performed phylogenetic analyses of its three genes (*S5-ORF3*, *S5-ORF4*, and *S5-ORF5*) in the 12 ILs (Fig. 4A). Unexpectedly, two ILs—YB and Qianlijing—clustered together and were identified as wide compatibility varieties (WCVs). With gene sequence of Balilla as the reference, we detected eight SNPs, three deletions (1, 10, and 42 bp), and one 94 bp AV in *S5-ORF5* (Fig. 4A), which acts as the "killer" gene. The 94 bp AV was exclusive to the two WCVs, potentially rescuing embryo-sac viability in their inter-subspecific F_1 hybrids.

To investigate contribution of the 94 bp AV to heterosis, we first compared seed setting rate between intra- and inter-subspecific F_1 hybrids lacking the WCV parentage. The seed setting rate of *indica-indica* and *japonica-japonica* F_1 hybrids were similar,

both significantly higher than those of *indica-japonica* F_1 hybrids (Fig. 4B). The heterosis for seed setting rate was also lower in inter-subspecific F_1 hybrids (Fig. 4C), compared to those of intra-subspecific F_1 hybrids and their pooled average values. However, when YB or Qianlijing was used as a parent, the *indica-japonica* F_1 hybrids achieved seed setting rate as high as the *indica-indica* F_1 hybrids (Fig. 4D-H). Importantly, heterosis for seed setting rate was also comparable between the inter- and intra-subspecific F_1 hybrids. We therefore identified YB and Qianlijing as two *indica*-type WCVs and confirmed important role of the 94 bp AV in heterosis for seed setting rate.

Presence and absence variants of *OsBZR1* contribute to yield heterosis

OsBZR1, a positive regulator of the brassinosteroid signaling pathway (He et al., 2002), regulates phenotypic differentiation between *indica* and *japonica* subspecies (Yang et al., 2025), including yield per plant, tiller number, grain number, thousand-grain weight, panicle length, and plant height. Clustering of the 12 ILs based on phenotypic data of these traits supported *OsBZR1*'s role in subspecific differentiation (Supplementary Fig. S15). Given that sequence variants in the *OsBZR1* promoter region—particularly a 4,808 bp retrotransposon—are critical for its function, we extracted and compared *OsBZR1* sequences in the 12 ILs. Phylogenetic analysis revealed that 11 of the 12 ILs clustered into distinct *indica* and *japonica* groups, with JR2 as an exception (Fig. 5A). We thus classified the ILs into three haplotypes: haplotype_i (*indica*), haplotype_j (*japonica*), and haplotype_JR2 (divergent).

Using the sequence of Balilla as reference, we identified 82 SNPs, four insertions (1, 4, 9, and 12 bp), four deletions (1, 2, 2, and 8 bp), one 266 bp PV, and six AVs (356, 523, 1,008, 1,486, 1,743, and 4,808 bp) (Fig. 5B). The 266 bp PV was exclusive to the six *japonica* accessions, while the 4,808 bp AV was present in the six *indica* lines, indicating contributions of these SVs to subspeciation. The remaining five AVs, with four of which mainly derived from the 4,808 bp AV, were unique to JR2.

Since the retrotransposon influences gene transcription, we analyzed the relationship between expression levels of *OsBZR1* and heterosis for seedling length in six pairs of reciprocal F_1 hybrids, which were divided into high-, low-, and no-heterosis

groups (Dan et al., 2024). We observed a significantly negative correlation, suggesting contributions of low expression levels of *OsBZR1* to heterosis for seedling length (Fig. 5C). When expression levels were compared between F_1 hybrids and corresponding parents, we detected overdominant effect, dominant effect, and unclassifiable effects, indicating diverse inheritance patterns of *OsBZR1* across different hybrid backgrounds (Supplementary Fig. S16).

To assess the impact of genomic variants in *OsBZR1* on heterosis of field traits, we compared heterosis of the seven traits between F_1 hybrids derived from ILs carrying different haplotypes (Fig. 5D-J). Except for seed setting rate and plant height, heterosis of haplotype_i/haplotype_j heterozygous F_1 hybrids did not significantly differ from the average of their homozygous counterparts (haplotype_i and haplotype_j). Strikingly, F_1 hybrids with the haplotype_i/haplotype_JR2 exhibited significantly stronger heterosis than the haplotype_i homozygotes for most traits, including yield, tiller number, and plant height. In contrast, hybrids combining haplotype_j and haplotype_JR2 showed no significant yield heterosis advantage over haplotype_j homozygotes, highlighting the importance of the haplotype_i/haplotype_JR2 heterozygous state for robust yield heterosis.

Changes in gene expression by structural variants explain variance in heterosis

To further explore the impacts of SVs on gene expression, we performed RNA-sequencing and differential expression analyses on seedling samples of the 12 ILs (Fig. 6A). The differentially expressed genes (DEGs) clustered the ILs into two major groups, with C418 and R4115 in *indica* group, inferring changes in genetic information from the genomic to transcriptomic levels (Supplementary Fig. S17). Drawing on previously reported inheritance patterns of metabolites in rice F_1 hybrids (Dan et al., 2024), we demonstrated that transcriptomic additive and partially dominant effects played more important roles than dominant effects in heterosis across the 17 traits, reinforcing contributions of the two main effects to heterosis across different genetic levels (Fig. 6B and Supplementary Fig. S17).

We next focused on expression levels of genes within the two large inversions: inversion_6 (304 genes) and inversion_8 (331 genes) (Fig. 6C). Consistent with the

presence/absence pattern of inversion_6, the 12 ILs were accurately classified into *indica* and *japonica* groups based on expression levels of these genes (Fig. 6D-E). Unexpectedly, genes within inversion_6 with additive or partially dominant effects were predictive of heterosis in partial least square (PLS) models (Supplementary Fig. S18). Based on genes in inversion_8, JR2 was found to be closer to *indica* than to *japonica* (Supplementary Fig. S19). Though only a limited number of genes differed in expression levels between the *indica* and *japonica*, the genes within inversion_8 effectively explained the variance in heterosis (Supplementary Fig. S19). Among the three large size AVs, genes in AV_8 (22 genes) and AV_11 (22 genes) were seldomly expressed in seedlings. For AV_6 (34 genes), most genes were expressed and effectively divided these ILs into two groups, with C418 in *indica* group (Fig. 6F-G). The genes in AV_6 also demonstrated predictability for heterosis comparable to the genome-wide DEGs of 12 ILs and the two inversions (Fig. 6H and Supplementary Fig. S20). These results confirmed the direct impact of SVs on gene expression, which in turn explained the variance in heterosis.

Translocation outperforms other SVs in heterosis

To comprehensively investigate associations between different types of parental genomic variants and expression changes in F_1 hybrids, we performed PLS analysis between additive expression levels of 1,759 DEGs from the 12 ILs and the number of genomic variants in each pair of ILs. Based on the R^2 (adjusted) values of PLS models with varying number of latent factors, the predictability of these DEGs was higher for SVs than those for SNPs and InDels (Fig. 6I). When SNPs and InDels were further subdivided by location or function, the predictability remained low, except for variants located in multiple regions with their number below ten (Supplementary Fig. S21-22). In contrast, the predictability varied considerably among different types of SVs, with translocation predicted at the highest accuracy (Fig. 6J). Among the top five largest translocations, genes associated with translocation_8.1, translocation_12.1, and translocation_8.2 were rarely expressed in seedling samples (Supplementary Fig. S23). Translocation_4 was an inverted translocation on chromosome 4, detected between six *indica* and four *japonica*, except for JR2 and Balilla (Fig. 6C). This translocation

364 contained 17 genes, eight of which had detectable expression levels (Supplementary
365 Fig. S24). Due to loss of this translocation, Balilla was clustered with *indica* while JR2
366 was remained in *japonica* group, suggesting divergent effect of the same translocation
367 on gene expression under different genetic background. The predictability of genes in
368 translocation_4 for heterosis was lower than those of DEGs, the two inversions, and the
369 AV_6 (Fig. 6H), which was mainly caused by the limited number of expressed genes.
370 Translocation_12.2, between chromosomes 12 and 11, was present fully in some ILs
371 and partially in others. Based on the normalized expression levels of its 17 genes,
372 110080 and Wuyunjing8 with partial translocations were clustered with *indica* ILs,
373 while three *indica* with translocations were remained in their group (Supplementary Fig.
374 S24), inferring differed translocation effect on gene expression between subspecies.
375 Despite its smaller size, the predictability of expressed genes in translocation_12.2 for
376 heterosis was as effectively as those within inversions and AVs (Fig. 6H).

377 When different types of SVs (with average variant number over ten) were detailly
378 analyzed by location, the predictability was higher for exonic region of AV
379 (Supplementary Fig. S25), UTR5 region of CNV (Supplementary Fig. S26), splicing
380 region of PV (Supplementary Fig. S27), and intergenic region of inversion
381 (Supplementary Fig. S28), suggesting diversified function of SVs on gene expression
382 through different locations. The predictability for translocation also varied across
383 locations, with splicing, upstream, and intronic regions having higher accuracy (Fig.
384 6K). To further investigate functions of different types of translocations, we selected
385 three top size ones in combination with the identified DEGs: splicing_6, upstream_7,
386 and intronic_4 (Fig. 6C), which had available gene expression levels. These three
387 representative translocations were invariant in *indica* but variable in *japonica*. Using a
388 haplotype analysis strategy similar to that for *OsBZR1*, we classified the 12 ILs into
389 two haplotypes based on the presence/absence of each translocation. Then, we
390 compared heterosis of the 17 agronomic traits between hybrids with heterozygous and
391 homozygous haplotypes to investigate their relationships with heterosis. The
392 associations varied across translocations, with the associated genes showing either
393 positive or negative correlations with heterosis (Fig. 6L). We also analyzed three other

394 representative translocations, including splicing_11, upstream_1, and intronic_11,
395 which were detected in at least five ILs but with nearly undetectable gene expression
396 (Supplementary Fig. S23). These translocations showed distinct associations with
397 heterosis (Fig. 6L). After the haplotype analysis was extended to all the above-
398 mentioned typical SVs, we found the largest number of associations for inversion_6
399 and AV_6 (n = 10; Fig. 6L). Then, the upstream_7, splicing_11, and upstream_1 were
400 associated with heterosis of nine traits. The upstream_7 was the only translocation for
401 yield heterosis, warranting further validation of this region or its associated gene by
402 sequence modulation technologies. Taken together, the heterozygous haplotypes of SVs
403 displayed significant heterosis to their homozygous ones, supporting the
404 overdominance model in rice heterosis.

405 **Discussion**

406 The genetic mechanism of heterosis has been a central debate in plant genetics for over
407 a century. The overdominance model emphasizes the contribution of genomic
408 heterozygosity and the co-existence of non-defective alleles (Shull, 1908; East, 1909;
409 Shull, 1911; East, 1936). In this study, genomic variants in parental lines are at
410 heterozygous states in F_1 hybrids. SNPs, InDels, PAVs, CNVs, inversions, and
411 translocations represent heterozygosity of single bases (1 bp), several bases (2-50 bp),
412 some bases (> 50 bp), sequence repeat number, sequence direction, and sequence
413 location, respectively (Supplementary Fig. S29). When genomic variants are combined
414 for haplotype analysis, F_1 hybrids with heterozygous haplotypes show heterosis to those
415 with homozygous haplotypes, underscoring the critical role of genomic heterozygosity
416 in rice heterosis. For the *S5-ORF5* locus, typical *indica-japonica* F_1 hybrids with
417 heterozygous *ORF5+* and *ORF5-*, which are functional alleles of aspartic protease
418 (Chen et al., 2008; Yang et al., 2012), exhibit negative heterosis for seed setting rate. In
419 contrast, hybrids carrying the *ORF5n*, which is assigned as a nonfunctional allele due
420 to a 94 bp AV spanning the translation start site, alleviate the negative heterosis. Thus,
421 the "nonfunctional" *ORF5n* allele shows significant function to heterosis for seed
422 setting rate. Since inbred lines carrying any of the three alleles have normal embryo-

423 sac fertility and seed setting rate, it is not appropriate to classify *ORF5n* as a dominant
424 superior allele and *ORF5*+/*ORF5*- as recessive detrimental alleles under the dominance
425 model. Similarly, heterozygous haplotypes of *OsBZR1* (haplotype_i/haplotype_JR2)
426 outperformed homozygous ones for yield and multiple yield-related traits. Most
427 importantly, our haplotype-based analysis of representative translocations and other
428 SVs consistently indicates that the heterozygous state confers heterosis over the
429 homozygous counterparts.

430 The overdominance model also posits the relationship between parental genetic
431 disparity and the degree of heterosis. Previous studies using low- or high-density DNA
432 markers have supported the close associations between genetic distance or genetic
433 variants and heterosis of F_1 hybrids in rice (Xiao et al., 1996; Zhang et al., 1996; Dan
434 et al., 2014; Huang et al., 2015). We here offer a more comprehensive genomic variant
435 picture based on high-quality genome assemblies and demonstrate that the number of
436 genomic variants in parents are powerful predictors of heterosis for multiple agronomic
437 traits. Crucially, the relationships differ between inter- and intra-subspecific hybrids,
438 especially for heterosis of seed setting rate, tiller number per plant, and yield per plant.
439 Based on the genome-wide heterozygosity of genomic variants, particularly SVs, and
440 their correlations with heterosis, we support for the overdominance model and propose
441 that heterozygosity of allelic/sequences underlies heterosis in rice.

442 Although we generated reference-level genome assemblies, telomere-to-telomere
443 (T2T) genomes for all investigated inbred lines are needed to display the complete
444 genomic landscape in subsequent heterosis studies. Concurrently, advanced variant
445 detection and description tools should be developed to accurately characterize genomic
446 variants in parental inbred lines. While large hybrid populations (often hundreds of F_1
447 hybrids) are standard for robust heterosis studies, and hybridization and phenotyping
448 are feasible in rice, generating hundreds of reference- or T2T-level genomes remains
449 challenging for most research teams currently. Complete or half diallel-cross designs,
450 which efficiently utilize parental genomic information, are practical strategies for
451 deeper genomic investigation on rice heterosis. Additionally, the representativeness of
452 parental inbred lines is important to outcomes. The 12 inbred lines used here capture

453 only a portion of rice genetic diversity. The inclusion of accessions from *aus*, *tropical*
454 *japonica*, *basmati*, and wild accessions (e.g. *rufipogon*) will provide updated genomic
455 insights into rice heterosis.

456 The SVs detected in our study ranged from about 57 Mb to168 Mb, constituting
457 approximately 14% to 42% of the genomes. These SVs, particularly subspecies-specific
458 inversions and AVs: inversion_6 and AV_6, provide a structural basis for *indica-*
459 *japonica* divergence and genetic variation among inbred lines. Importantly, specific
460 SVs are associated with heterosis of key agronomic traits, highlighting the application
461 potential of SVs in genomic prediction. As noted, the 94 bp AV in *S5-ORF5* and the
462 complex PAVs linked to *OsBZR1* drive heterosis for seed setting rate and yield,
463 respectively. We propose that SVs distinguishing *indica* from *japonica* or divergent
464 haplotypes are candidates for precise engineering to harness strong inter-subspecific
465 heterosis. The identified translocations, particularly intronic_4, splicing_6, and
466 upstream_7 related to changes in gene expression levels, are important targets.
467 Although precise modulation of genomic SVs remains challenging, we believe that
468 advances in genome editing technologies, such as DualPE (Zhao et al., 2025), will
469 elucidate function of SVs and facilitate the utilization of heterosis in crops. Prior to
470 establishing precise editing methods for SVs, e.g., translocations, the effects of specific
471 SVs on heterosis can be indirectly validated via a population genetic strategy. Firstly,
472 two inbred lines carrying targeted SVs are crossed to generate heterozygous F_1 hybrids
473 and subsequent recombinant inbred lines. These recombinant inbred lines—typically
474 requiring at least five generations of self-crossing to achieve phenotypic uniformity—
475 are then crossed with a male-sterile line to produce F_1 hybrids for heterosis
476 quantification of corresponding traits. To classify genotypes/haplotypes, the
477 recombinant inbred lines and the sterile line are analyzed using standard PCR assays
478 targeting the SVs. Finally, the functional impact of single SVs can be assessed by
479 comparing heterosis between heterozygous and homozygous genotypes/haplotypes.

480 Genomic SVs always induce gene expression changes (Alonge et al., 2020; Shang
481 et al., 2022). Moving beyond individual loci, our transcriptomic analyses reveal that
482 rice genomic SVs, especially large size inversions and AVs, not only correctly classify

483 inbred lines by subspecies but also effectively explain the variance in heterosis through
484 the cumulative, additive expression of multiple genes. These SVs are disproportionately
485 important drivers of transcriptional changes, with translocations (e.g., translocations
486 located in splicing, upstream, and intronic regions) outperforming others. We detect
487 associations among SVs, expression changes, and heterosis based on seedling
488 transcriptomes. However, many genes are expressed in specific tissues (e.g., young
489 panicles), stages, or environments. In addition to gene expression, genomic SVs may
490 be linked to DNA methylation levels, chromatin accessibility (Wang et al., 2024), and
491 histone modifications (Qi et al., 2024). Therefore, the collection of diversified genomic-
492 downstream data, including epigenomic, transcriptomic, and translational data, is
493 essential to design optimal SV heterozygosity in generating rice heterosis.

494 Materials and methods

495 Plant materials

496 Twelve rice inbred lines (ILs) were selected from accessions reported in previous
497 studies (Dan et al., 2014; Dan et al., 2016; Dan et al., 2020; Dan et al., 2024). Based on
498 clustering results from metabolite profiles generated using GC- or LC-MS platforms
499 (Dan et al., 2016; Dan et al., 2020), low-density SSR markers (Dan et al., 2014), and
500 high-density SNPs or InDels (Dan et al., 2024), these 12 inbred lines are representatives
501 of *indica* and *japonica* accessions. Plants of the selected inbred lines for genome
502 sequencing were grown under long-day conditions, maintaining a 16-hour light/8-hour
503 dark photoperiod at 30 °C in a plant culture room. The seedlings were sown in 96-well
504 PCR plates (without bottoms) and placed in a plastic basin filled with Yoshida nutrient
505 solution (pH = 5.5). After 20 days of growth, leaves from each genotype were harvested
506 for experimentation. For full-length RNA sequencing of YB and Balilla, the plants were
507 divided into four tissue parts: root, middle leaf (the youngest leaf), leaf without middle
508 leaf, and stem.

509 Additionally, seedlings of the 12 rice inbred lines were placed in an intelligent
510 breeding room for RNA sequencing. The environmental conditions were set to 28 °C
511 with a 16-hour light/8-hour dark photoperiod and controlled humidity. Each inbred line

occupied half of a PCR plate and was maintained in a single basin containing four liters of nutrient solution, which was changed every three days. Fifteen-day-old seedings (without root) were harvested and frozen in liquid nitrogen. The middle five plants from each row were pooled to create one biological replicate, with three biological replicates established for each genotype.

The 12 inbred lines were crossed using a complete diallel design. As previously detailed (Dan et al., 2024), a total of 17 agronomic traits were recorded for both the inbred lines and their corresponding F_1 hybrids. These traits included grain yield per plant (YPP), grain number per panicle (GNP), seed setting rate (SSR), thousand-grain weight (TGW), plant height/tiller number (at seedling, elongation, flowering and maturation stages) (SSPH, ESPH, FSPH, MSPH and SSTN, ESTN, FSTN, MSTN), heading date (HD), tiller number per plant (TPP), panicle length (PL), and the number of primary and secondary branches per panicle (PBN and SBN). Better-parent heterosis was calculated using the equation: $BPH = (F_1 - P_{high})/P_{high}$, where P_{high} represents the value of the high parent. For the heading date, parents with shorter heading dates were designated as P_{high}, and the calculated values were subsequently multiplied by -1.

Genomic sequencing

SMRTbell libraries (~17 kb) for each rice inbred line were constructed using the SMRTbell Express Template Prep Kit 3.0 and sequenced on the Pacbio Revio platform (Pacific Biosciences) at Frasergen Bioinformatics (Wuhan, China). The SMRTbell libraries were size-selected using the BluePippin system (Sage Science) and purified with a 1× concentration of SMRTbell cleanup beads. Library quality was assessed using Femto pulse and Qubit 3.0 (Life Technologies), resulting in an average of 12.31 Gb of HiFi reads (approximately 30.46× coverage of the genome) obtained for each inbred line.

Hi-C libraries for YB and Balilla were prepared and sequenced using the BGI PE150 platform, generating a total of 51.81 Gb and 49.27 Gb of reads (approximately 127.80× coverage of the genome) for the two samples, respectively. The raw data were filtered using Trimmomatic (v0.40) (Bolger et al., 2014) with parameters LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:15. The quality of the clean data

542 was evaluated with default parameters using FastQC (Wingett and Andrews, 2018).

543 **RNA-sequencing**

544 Full-length RNA sequencing technology was employed to aid in the annotation of genes
545 in YB and Balilla. Total RNA was extracted from four tissues of the seedling samples,
546 and equal amounts of RNA were pooled before constructing SMRTbell libraries. Full-
547 length cDNA was synthesized using the SMARTer PCR cDNA Synthesis Kit
548 (Clonetech) with the primer sequences
549 "AAGCAGTGGTATCAACGCAGAGTACATGGGG" and
550 "AAGCAGTGGTATCAACGCAGAGTAC". The products were purified and size-
551 filtered (>1 kb) prior to sequencing. The PacBio Sequel II platform at Frasergen
552 Bioinformatics was used for sequence analysis, with raw data (an average of 36.14 Gb)
553 filtered and preprocessed using SMRTlink (v10.1, --minLength=50). Iso-seq analysis
554 was performed to obtain full-length and non-concatemer (FLNC) sequences. These
555 FLNC sequences were mapped to reference genomes using GMAP (V19.06.10) (Wu
556 and Watanabe, 2005), retaining those with the highest PID to determine gene structure.

557 To investigate the impact of structural variants on gene expression changes, RNA
558 sequencing was also conducted to seedlings of the 12 rice inbred lines (with three
559 biological replicates per genotype). Total RNA was extracted for each sample to
560 construct a DNA library. The Illumina Hiseq PE150 platform at Frasergen
561 Bioinformatics was selected to generate 150 bp paired-end reads, with more than 6 Gb
562 of raw data per sample. The raw data were filtered using SOAPnuke (v2.1.0, -
563 lowQual=20, -nRate=0.005, -qualRate=0.5) (Chen et al., 2018). Clean reads were
564 aligned to YB genome using Hisat2 (v2.2.1, --no-unal) (Kim et al., 2019). Transcripts
565 were assembled using StringTie (Pertea et al., 2015), and clean reads were mapped to
566 reference transcripts using bowtie2 (v2.3.5, --sensitive -k 50 --no-mixed --no-
567 discordant) (Langmead, 2010). The read count of each transcript was calculated using
568 RSEM (v1.3.3) (Li and Dewey, 2011).

569 **Genome assembly**

570 Draft genomes of the 12 rice inbred lines were assembled using HiFi reads with
571 parameter -l3 via HiFiasm (v0.19.5) (Cheng et al., 2022). The clean Hi-C reads for YB

572 and Balilla were aligned to contigs using Juicer (v1.6) (Durand et al., 2016). Contact
573 matrices were generated with 3D-DNA (v180922, -r0) (Dudchenko et al., 2017), and
574 corrections were applied using JuiceBox (v1.11.08). The genomes of the remained ten
575 ILs were assembled based on the genomes of YB and Balilla using Ragtag (v2.1.0,
576 scaffold) (Alonge et al., 2022). We assessed the accuracy, identity, and completeness of
577 the 12 assemblies. The completeness of the assemblies was assessed using
578 Benchmarking Universal Single-Copy Orthologue (v3.0.2, embryophyta_odb10)
579 analysis (Simao et al., 2015). The long terminal repeats (LTR) assembly index (LAI)
580 was calculated for each chromosome using LTR tools (Ou et al., 2018), including
581 LTR_retriever (v2.9.0, -infinder ltrfinder.scn -inharvest ltrharvest.scn), LTR_FINDER
582 (v1.0.7, -D 15000 -L 7000 -C -M 0.7), and GenomeTools (v1.6.2, ltrharvest -similar 70
583 -vic 10 -seed 20 -seqids yes -maxlenltr 7000 -maxtsd 6). HiFi reads were re-mapped to
584 the assemblies with parameters "-ax map-hifi" using Minimap2 (v2.21) (Li, 2018).
585 Additionally, consensus quality values and completeness of the assemblies were scored
586 using a k-mer-based method with the Merqury pipeline (v1.3) (Rhie et al., 2020).

587 **Genome annotation**

588 Repetitive sequences (including tandem repeats and transposable elements), protein-
589 coding genes, and non-coding RNAs were annotated for the genome assemblies of YB
590 and Balilla. Tandem repeats were identified using the Tandem Repeats Finder (TRF,
591 v4.09.1, 2 7 7 80 10 50 2000 -d -h) (Benson, 1999). Transposable elements were
592 annotated through an integrated approach combining do novo and homology-based
593 (Repbase, www.girinst.org/repbase (Bao et al., 2015)) strategies. The de novo repeat
594 library was constructed using LTR_FINDER (v1.0.7, -w 2 -C) (Xu and Wang, 2007)
595 and RepeatModeler (v2.0.1) (Flynn et al., 2020). Transposable elements were detected
596 using RepeatMasker (v4.1.2) (Tarailo-Graovac and Chen, 2009) with parameters "-
597 nolow -no_is -norna -parallel 2", searching against both the Repbase (Bao et al., 2015)
598 and de novo repeat library. Additionally, RepeatProteinMask (1.36) (Tarailo-Graovac
599 and Chen, 2009) with parameters "-engine ncbi -noLowSimple -pvalue 0.0001" was
600 utilized to identify transposable elements in Repbase (Bao et al., 2015).

601 To annotate gene structure and function, we integrated three strategies: homology-

602 based (using five genomes), de novo, and transcriptome-based (full-length RNA
603 sequencing) predictions. In the homology-based approach, protein-coding sequences
604 from the genomes of 9311 (Yu et al., 2002), Minghui63 (Zhang et al., 2016),
605 Zhenshan97 (Zhang et al., 2016), R498 (Du et al., 2017), and Nipponbare (Shang et al.,
606 2023) were compared to the YB and Balilla genomes using tblastn (v2.11.0+, -evalue
607 1e-05) (Gertz et al., 2006). Gene structures of the aligned sequences were predicted
608 using Exonerate (v2.4.0, --model protein2genome) (Slater and Birney, 2005). De novo
609 annotation was conducted using Augustus (v3.4.0, --noInFrameStop=true --
610 strand=both) (Stanke et al., 2006), GlimmerHMM (v3.0.4) (Majoros et al., 2004), and
611 GeneMark (v4.65) (Borodovsky and McIninch, 1993) . The annotated gene sets were
612 integrated to non-redundant sets using MAKER (v3.01.03) (Holt and Yandell, 2011),
613 and further polished by incorporating full-length RNA sequencing results using PASA
614 (v2.4.1) (Haas et al., 2003). RNA-sequencing results were aligned using HiSat2 (v.2.2.1)
615 (Kim et al., 2019), and transcripts were assembled with StringTie (v.2.1.7) (Kovaka et
616 al., 2019). Additionally, transcripts were de novo assembled using Trinity (v2.8.5)
617 (Grabherr et al., 2011). Gene functional analysis was conducted with Diamond (v2.0.7,
618 -c 1 -f 5 -e 1e-5 --very-sensitive -k 10) (Buchfink et al., 2021) against five databases:
619 Non-Redundant (NR, ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz), Kyoto
620 Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2014), Gene Ontology
621 (GO) (Ashburner et al., 2000), TrEMBL (Boeckmann et al., 2003), and Swiss-Prot
622 (Boeckmann et al., 2003). Protein domains were annotated using InterProScan (v5.50-
623 84.0, --applications Pfam) (Jones et al., 2014) based on the InterPro database (Mitchell
624 et al., 2015).

625 Non-coding RNAs, including rRNA, tRNA, snRNA, and miRNA, were also
626 annotated for the two genomes. tRNA was identified using tRNAscan-SE (v2.0.9)
627 (Lowe and Eddy, 1997), while rRNA was detected using RNAmmer (v1.2, -S euk -
628 multi -m lsu,ssu,tsu) (Lagesen et al., 2007). miRNA and snRNA were annotated with
629 Infernal (v1.1.4) (Nawrocki and Eddy, 2013) using Rfam (v14.6) (Griffiths-Jones et al.,
630 2005). miRNA and rRNA were present in higher percentages within the genomes. The
631 completeness of the genomic annotations was further assessed using BUSCO (v5.2.2,

632 embryophyta_odb10) (Simao et al., 2015). Here, 98.5% and 98.6% of the gene sets in
633 the two genomes contained complete BUSCOs in embryophyta.
634 For the annotation of the remaining ten genomes, Liftoff (Shumate and Salzberg,
635 2021) was employed. The YB genome was chosen as the reference genome for five
636 *indica* accessions: R465, Mianhui725, 9311K, 610234, and Qianlijing. The Balilla
637 genome served as the reference for five *japonica* accessions: R4115, C418, 110080,
638 Wuyunjing8, and JR2.

639 **Pan-genome analysis**

640 Protein coding genes of the 12 inbred lines were initially aligned using BLASTp (-
641 evalue 1e-5) (Altschul et al., 1997). The BLAST results were processed with
642 OrthoFinder (v2.3.12) (Emms and Kelly, 2019) to classify gene families. Gene families
643 shared among all 12 inbred lines were defined as core gene families, while those
644 identified in more than 90% of inbred lines were termed softcore gene families. Gene
645 families specific to a single inbred line were classified as private gene families, and the
646 remaining gene families were considered dispensable. Genes in the four types of gene
647 families were further analyzed in the Pfam (Mistry et al., 2021), GO (Ashburner et al.,
648 2000), and KEGG (Kanehisa et al., 2014) databases to determine their conservation.

649 **Identification of genomic variants**

650 Genomic variants, including single nucleotide polymorphisms (SNPs), insertions and
651 deletions (InDels) (2-50 bp), and structural variants (SVs), were detected in the 12 rice
652 inbred lines, with the genome of YB selected as the reference genome and the remained
653 11 as query genomes. SNPs and InDels were detected using MUMmer4 (nucmer: -c 90
654 -l 40, delta-filter: -1, show-SNPs: -ClrTH) (Marcais et al., 2018). SVs, including
655 presence-absence variants (PAVs), inversions, translocations, and copy number variants
656 (CNVs), were detected with the assistance of SyRI (v1.4, nucmer: -c 90 -l 40 -t 16 ref.fa
657 query.fa, delta-filter: -1 -i 90 -l 100 ref_query.delta, show-coords: -THrd) (Goel et al.,
658 2019). The distribution of SNPs, InDels, and SVs was analyzed across various genomic
659 regions or functions, including: 1. Upstream (1 kb regions before the gene start codon);
660 2. Downstream (1 kb regions after the gene stop codon); 3. Intergenic regions; 4. UTR3
661 (3' untranslated regions); 5. UTR5 (5' untranslated regions); 6. Splicing junction

662 regions (< 10 bp); 7. Intronic regions; 8. Exonic regions; 9. Synonymous or
663 nonsynonymous mutants; 10. Frameshift or nonframeshift insertions/deletions; 11.
664 Stopgain (gain of stop codons); 12. Stoploss (loss of stop codons). Variants
665 simultaneously located in two regions, such as UTR5 and UTR3, were showed as
666 Region 1;Region 2. The functions of SNPs, InDels, and SVs was annotated with
667 ANNOVAR (Wang et al., 2010).

668 **Pan-SV analysis**

669 A pan-SV analysis was conducted for SVs of the 11 query genomes and the results were
670 further used for constructing graph-based genome using Vg (v1.43) (Garrison et al.,
671 2018). For SVs excluding PVs, an SV was assigned when the ratio of overlapping to
672 total locations in two inbred lines exceeded 50%. For PVs, a PV was designated when
673 the distance between PV locations in two inbred lines was less than 5 bp and the length
674 ratio between the shorter and longer samples was greater than 50%. Circles illustrating
675 genomic variants on chromosomes were generated using Circos (Krzywinski et al.,
676 2009). Genetic relatedness among the 12 inbred lines was visualized using Cytoscape
677 (Reimand et al., 2019).

678 **Construction of phylogenetic trees**

679 A phylogenetic tree for 223 rice inbred lines was constructed using genotypic data based
680 on 33 pairs of single sequence repeat markers from a previous study (Dan et al., 2024).
681 Phylogenetic trees based on the number of differential SNPs, InDels, and CNVs among
682 the 12 rice inbred lines were built using the neighbour-joining method with MEGA
683 (v10.2.6) (Stecher et al., 2020). The number of SNPs, InDels, and CNVs of the inbred
684 lines were retrieved from a previous report (Dan et al., 2024).

685 For phylogenetic analysis of *Phr1* and the *S5* loci, gene sequences from reference
686 inbred lines were obtained from the GenBank database. The *Phr1* sequences of
687 Nipponbare and 9311 corresponded to accession number DQ532376 and DQ532391,
688 respectively (Yu et al., 2008). For *S5-ORF3-5* (Chen et al., 2008; Yang et al., 2012), the
689 Balilla sequences were retrieved under accession number JX138498, JX138502,
690 EU889294. Gene sequences from the 12 newly assembled inbred lines were extracted
691 and aligned using Jalview (version 2.11.4.1) with default clustal parameter s

692 (Waterhouse et al., 2009). Phylogenetic trees were constructed with the neighbour-
693 joining method. Genotypes of inbred lines were assigned based on previous reports
694 (Chen et al., 2008; Yu et al., 2008; Yang et al., 2012). Gene structures were visualized
695 using the exon-intron graphic maker (http://wormweb.org/exonintron).

696 To investigate genomic variants of *OsBZR1*, especially SVs in promoter region,
697 sequences (including 6,345 bp before the ATG) of the 12 ILs were used for analysis. In
698 sequence alignment, the mafft combined with E-INS method was chosen. The
699 dendrogram tree of the 12 ILs based on phenotypic data of seven traits were
700 implemented using MetaboAnalyst 6.0 (Pang et al., 2022), with Spearman and ward as
701 the distance measure and clustering algorithm, respectively.

702 **Phenol reaction experiment**

703 Seeds from 14 inbred lines were harvested at maturity and stored at -20 ºC. For each
704 inbred line, three seeds were chosen and treated with a 2% (volume/volume) phenol
705 solution at 37 ºC. Changes in hull colour of the seeds were recorded after 1, 2, 3, 5, 7,
706 31, 62, 93, and 124 days of treatment. After 7 days of treatment, the seed were placed
707 at room temperature.

708 **Normalization and differential analysis of transcriptomics data**

709 Read counts of 53,564 genes in each experimental sample were used for data
710 normalization and differential analyses through MetaboAnalyst 5.0 (Pang et al., 2022).
711 The interquantile range was selected to filter genes, with a threshold set at 40%. Sum,
712 none, and auto-scaling were chosen for sample normalization, data transformation, and
713 data scaling, respectively. Differential analysis among the 12 inbred lines was
714 implemented using the ANOVA module in MetaboAnalyst 5.0 (Pang et al., 2022), with
715 a raw *P*-value cutoff set at 0.001 for post-hoc analysis. Principal component analysis
716 (PCA) and heatmap results for the 12 inbred lines were obtained through corresponding
717 modules on the website. PERMANOVA was conducted in PCA and the permutation
718 number was 999. Euclidean for distance measure and ward for clustering method were
719 set in heatmaps. Differential genes were selected from the original read count table and
720 averaged across the three biological replicates. The averaged data were further filtered
721 and normalized for each genotype with the above parameters. Due to the limited

722 number of genes in inversions and absence variants, no filtering was applied to these
723 genes. The genes in three large AVs, including AV_6, AV_8, and AV_11, were analyzed.
724 Despite AV_8 and AV_11 being larger in size, the genes in these AVs were rarely
725 expressed in the seedling samples. Only genes in AV_6 was further analyzed.

726 To investigate the relationship between expression level of *OsBZR1*
727 (*Os07g0580500*) and heterosis, gene expression data and seedling length (15-day old
728 seedlings) of six pairs of reciprocal hybrids and corresponding parental inbred lines
729 were obtained from a previous study (Dan et al., 2024). The transcriptomic data were
730 Log2-normalized using NetworkAnalyst 3.0 (Zhou et al., 2019). Since several
731 genotypes had six biological replicates, seeding length of the first three replicates was
732 used for the calculation of better-parent heterosis and Spearman correlation analysis.
733 To determine inheritance patterns of *OsBZR1*, the reciprocal hybrids from two inbred
734 lines were treated as single genotype. Average expression levels of *OsBZR1* were
735 calculated between each pair of inbred lines and were treated as additive expression
736 levels. Gene expression levels were compared for the parents, F_1 hybrids, and additive
737 expression levels. Inheritance patterns were assigned based on previous rules (Dan et
738 al., 2024).

739 **Partial least square (PLS) analysis**

740 PLS analysis was conducted to investigate the relationship between inheritance patterns
741 of expressed genes and better-parent heterosis of 17 agronomic traits using SPSS 20.0.
742 The maximum number of latent factors was set to 50. The different inheritance patterns
743 were manually calculated based on normalized gene expression levels using equations
744 outlines in a previously study (Dan et al., 2024). The genetic effects included: 1.
745 Additive effect: average levels of the female (P_f) and male (P_m) parents; 2. Partially
746 dominant effects (PD): partially dominant _0.3 (PD_0.3) = $P_m+0.3\times(P_f-P_m)$ and
747 partially dominant _0.7 (PD_0.7) = $P_m+0.7\times(P_f-P_m)$; 3. Dominant effects: expression
748 levels of the female (dominant_female) and male (dominant_ male) parents. The model
749 with the highest R^2 (adjusted) was selected as the final model for each trait.

750 PLS was also used to study the relationship between differentially expressed genes
751 in 12 inbred lines and the number of parental genomic variants. Additive effect of the

752 differential genes was used as predictive variables to predict the number of unique
753 variants between each two inbred lines.

754 **Statistical analysis**

755 Boxplots were created using BoxPlotR (Spitzer et al., 2014). Analysis of variance
756 (ANOVA) with the least significant difference for post hoc comparisons was performed
757 using SPSS 20.0. This analysis compared the following parameters: lengths of genes,
758 lengths of coding sequences, lengths of exons, lengths of introns, GC content of genes,
759 GC content of coding sequences, number of exons, and number of introns across seven
760 genomes (including two newly assembled genomes, YB and Balilla, along with five
761 representative rice genomes). Additionally, we assessed the number of genomic variants
762 and the degrees of better-parent heterosis for 17 traits in F_1 hybrids from both intra- and
763 inter-subspecies. The adjusted R^2 values of partial least square models for heterosis
764 across traits were also compared. An independent samples *t*-test (weighted) was
765 conducted to evaluate the number of correlations between unique and common genomic
766 variants and heterosis. SPSS was also employed to conduct Spearman correlation to
767 investigate the relationships between different types of unique genomic variants and
768 between genomic variants and heterosis.

769 ## Acknowledgements

770 We thank C. Zhu for funding of our rice research. We thank J. Bao, L. Rong, and S.
771 Zheng from the Frasergen Bioinformatics for their help in genomics and
772 transcriptomics analyses.

773 ## Author contributions

774 Z.D. conceived and designed the project. Z.D. and Y.C. prepared rice seedling samples,
775 conducted genomic, transcriptomic, phenotypic, and statistical analyses, and
776 interpretated the results. W.H. collected the rice accessions. Z.D. and W.H. constructed
777 the diallel cross population and recorded the phenotypes. Z.D., Y.C., and W.H. wrote
778 the manuscript.

779 ## Supplementary data

780 The following materials are available in the online version of this article.

Supplementary Figure S1. Neighbour-joining tree of 223 rice accessions.

Supplementary Figure S2. Heterosis of 16 agronomic traits in F_1 hybrids from intra- or inter-subspecies.

Supplementary Figure S3. Contact matrices and collinearity diagrams for genomes of YB and Balilla.

Supplementary Figure S4. Long terminal repeat assembly index (LAI) for 12 chromosomes of 12 rice accessions.

Supplementary Figure S5. Comparison of two newly assembled genomes with five established rice genomes.

Supplementary Figure S6. Genomic comparisons between Mianhui725, 610234, and YB.

Supplementary Figure S7. Genomic comparisons between 9311K, R465, and YB.

Supplementary Figure S8. Genomic comparisons between Qianlijing, C418, and YB.

Supplementary Figure S9. Genomic comparisons between R4115, Wuyunjing8, and YB.

Supplementary Figure S10. Genomic comparisons between 110080, JR2, and YB.

Supplementary Figure S11. Genomic variants displayed on the genome of YB.

Supplementary Figure S12. Genetic relatedness of 12 rice inbred lines based on different types of genomic variants.

Supplementary Figure S13. Number of genomic variants for F_1 hybrids from intra- and inter-subspecies.

Supplementary Figure S14. Correlations between the number of structural variants and heterosis of intra- and inter-subspecific F_1 hybrids.

Supplementary Figure S15. Dendrogram of 12 inbred lines based on phenotypic data of seven agronomic traits.

Supplementary Figure S16. Inheritance patterns of *OsBZR1* across F_1 hybrids.

Supplementary Figure S17. Transcriptomic additive and partially dominant effects contribute to heterosis of 17 traits.

Supplementary Figure S18. Associations between expression levels of genes in inversion_6 and heterosis.

Supplementary Figure S19. Associations between expression levels of genes in inversion_8 and heterosis.

Supplementary Figure S20. Genomic structure of two AVs and the prediction of heterosis with AV_6.

Supplementary Figure S21. Adjusted R^2 for different types of SNPs.

Supplementary Figure S22. Adjusted R^2 for different types of InDels.

Supplementary Figure S23. Genomic structure and expression patterns of five translocations.

Supplementary Figure S24. Associations between genes in translocation_4/translocation_12.2 and heterosis.

Supplementary Figure S25. Adjusted R^2 for different types of AVs.

Supplementary Figure S26. Adjusted R^2 for different types of CNVs.

Supplementary Figure S27. Adjusted R^2 for different types of PVs.

Supplementary Figure S28. Adjusted R^2 for different types of inversions.

Supplementary Figure S29. Heterozygosity of genomic variants in F_1 hybrids.

Supplementary Table S1. Genome assembly profile of 12 rice inbred lines.

Supplementary Table S2. Consensus quality values and completeness of the 12 assemblies.

Supplementary Table S3. Hi-C profile of two rice genomes.

Supplementary Table S4. Annotation of two rice genomes.

Supplementary Table S5. Genomic variants of each pair of inbred lines and heterosis of their F_1 hybrids.

Funding

This study was supported by the National Natural Science Foundation of China (31801439 to Z.D., 32472185 to W.H., and 32101667 to Y.C.), the China Postdoctoral Science Foundation (2019M660186 to Z.D., and 2022T150500 to Z.D., and 2023T160497 to Y.C.), the Key Research and Development Program of Hubei Province (2022BFE003 to W.H.), the National Key R&D Program of China (2017YFD0100400 to W.H.) and the Hubei Agriculture Science and Technology Innovation Center Program

to W.H.

Data availability

Raw reads of PacBio SMRT sequencing and RNA-sequencing of the 12 rice inbred lines have been deposited in the NCBI database under BioProject accession PRJNA1178436. Genome assemblies of the 12 rice inbred lines are available from Figshare repository with DOI: 10.6084/m9.figshare.28397960. Source data are available with this paper.

Competing interests

The authors declare no competing interests.

References

Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, and Soyk S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 2022:23(1):258. https://doi.org/10.1186/s13059-022-02823-7

Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, et al. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*. 2020:182(1):145-161. https://doi.org/10.1016/j.cell.2020.05.021

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, and J. LD. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997:25(17):3389-3402. https://doi.org/10.1093/nar/25.17.3389

Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, Davis A, Dolinski K, Dwight S, Eppig J, et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* 2000:25(1):25-29. https://doi.org/10.1038/75556

Bao W, Kojima KK, and Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 2015:6(11. https://doi.org/10.1186/s13100-015-0041-9

Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999:27(2):573-580. https://doi.org/10.1093/nar/27.2.573

Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 2003:31(1):365-370. https://doi.org/10.1093/nar/gkg095

Bolger AM, Lohse M, and Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014:30(15):2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Borodovsky M, and McIninch J. GENMARK: Parallel gene recognition for both DNA strands. *Comput. Chem.* 1993:17(2):123-133. https://doi.org/10.1016/0097-8485(93)85004-V

Bruce AB. The Mendelian theory of heredity and the augmentation of vigor. *Science*. 1910:32(827):627-628. https://doi.org/10.1126/science.32.827.627-a

Buchfink B, Reuter K, and Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods*. 2021:18(4):366-368. https://doi.org/10.1038/s41592-021-01101-x

Chen J, Ding J, Ouyang Y, Du H, Yang J, Cheng K, Zhao J, Qiu S, Zhang X, Yao J, et al. A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica-japonica hybrids in rice. *Proc Natl Acad Sci U S A*. 2008:105(32):11436-11441. https://doi.org/10.1073/pnas.0804761105

Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience*. 2018:7(1):1-6. https://doi.org/10.1093/gigascience/gix120

Cheng H, Jarvis ED, Fedrigo O, Koepfli KP, Urban L, Gemmell NJ, and Li H. Haplotype-resolved assembly of diploid genomes without parental data. *Nat. Biotechnol.* 2022:40(9):1332-1335. https://doi.org/10.1038/s41587-022-01261-x

Dan Z, Chen Y, Zhao W, Wang Q, and Huang W. Metabolome-based prediction of yield heterosis contributes to the breeding of elite rice. *Life Sci. Alliance*. 2020:3(1):e201900551. https://doi.org/10.26508/lsa.201900551

Dan Z, Hu J, Zhou W, Yao G, Zhu R, Zhu Y, and Huang W. Metabolic prediction of important agronomic traits in hybrid rice (*Oryza sativa* L.). *Sci. Rep.* 2016:6(21732. https://doi.org/10.1038/srep21732

Dan Z, Liu P, Huang W, Zhou W, Yao G, Hu J, Zhu R, Lu B, and Zhu Y. Balance between a higher degree of heterosis and increased reproductive isolation: a strategic design for breeding inter-subspecific hybrid rice. *PLoS ONE*. 2014:9(3):e93122. https://doi.org/10.1371/journal.pone.0093122

Dan Z, Chen Y, Zhou W, Xu Y, Huang J, Chen Y, Meng J, Yao G, and Huang W. Additive and partially dominant effects from genomic variation contribute to rice heterosis. *bioRxiv*. 2024:10.1101/2024.1107.1116.603817. https://doi.org/10.1101/2024.07.16.603817

Du H, Yu Y, Ma Y, Gao Q, Cao Y, Chen Z, Ma B, Qi M, Li Y, Zhao X, et al. Sequencing and de novo assembly of a near complete indica rice genome. *Nat. Commun.* 2017:8(15324. https://doi.org/10.1038/ncomms15324

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017:356(6333):92-95. https://doi.org/10.1126/science.aal3327

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, and Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 2016:3(1):95-98. https://doi.org/10.1016/j.cels.2016.07.002

East EM. The distinction between development and heredity in inbreeding. *Am. Nat.* 1909:43(507):173-181. https://doi.org/https://doi.org/10.1086/279042

East EM. Heterosis. *Genetics*. 1936:21(4):375-397. https://doi.org/10.1093/genetics/21.4.375

Emms DM, and Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019:20(1):238. https://doi.org/10.1186/s13059-019-1832-y

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, and Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 2020:117(17):9451-9457. https://doi.org/10.1073/pnas.1921046117

Garrison E, Siren J, Novak AM, Hickey G, Eizenga JM, Dawson ET, Jones W, Garg S, Markello C, Lin MF, et al. Variation graph toolkit improves read mapping by representing genetic variation in the

922 reference. *Nat. Biotechnol.* 2018:36(9):875-879. https://doi.org/10.1038/nbt.4227
923 Gertz EM, Yu YK, Agarwala R, Schaffer AA, and Altschul SF. Composition-based statistics and
924 translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol.*
925 2006:4(41. https://doi.org/10.1186/1741-7007-4-41
926 Goel M, Sun H, Jiao WB, and Schneeberger K. SyRI: finding genomic rearrangements and local
927 sequence differences from whole-genome assemblies. *Genome Biol.* 2019:20(1):277.
928 https://doi.org/10.1186/s13059-019-1911-0
929 Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,
930 Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a
931 reference genome. *Nat. Biotechnol.* 2011:29(7):644-652. https://doi.org/10.1038/nbt.1883
932 Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, and Bateman A. Rfam: annotating non-
933 coding RNAs in complete genomes. *Nucleic Acids Res.* 2005:33(Database issue):D121-124.
934 https://doi.org/10.1093/nar/gki081
935 Gu Z, Gong J, Zhu Z, Li Z, Feng Q, Wang C, Zhao Y, Zhan Q, Zhou C, Wang A, et al. Structure and
936 function of rice hybrid genomes reveal genetic basis and optimal performance of heterosis. *Nat.*
937 *Genet.* 2023:55(10):1745-1756. https://doi.org/10.1038/s41588-023-01495-8
938 Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Jr., Hannick LI, Maiti R, Ronning CM,
939 Rusch DB, Town CD, et al. Improving the Arabidopsis genome annotation using maximal
940 transcript alignment assemblies. *Nucleic Acids Res.* 2003:31(19):5654-5666.
941 https://doi.org/10.1093/nar/gkg770
942 He JX, Gendron JM, Yang Y, Li J, and Wang ZY. The GSK3-like kinase BIN2 phosphorylates and
943 destabilizes BZR1, a positive regulator of the brassinosteroid signaling pathway in Arabidopsis.
944 *Proc. Natl. Acad. Sci. USA*. 2002:99(15):10185-10190. https://doi.org/10.1073/pnas.152342599
945 Hochholdinger F, and Yu P. Molecular concepts to explain heterosis in crops. *Trends Plant Sci.*
946 2025:30(1):95-104. https://doi.org/10.1016/j.tplants.2024.07.018
947 Holt C, and Yandell M. MAKER2: an annotation pipeline and genome-database management tool for
948 second-generation genome projects. *BMC Bioinformatics*. 2011:12(491.
949 https://doi.org/10.1186/1471-2105-12-491
950 Huang X, Yang S, Gong J, Zhao Q, Feng Q, Zhan Q, Zhao Y, Li W, Cheng B, Xia J, et al. Genomic
951 architecture of heterosis for yield traits in rice. *Nature*. 2016:537(7622):629-633.
952 https://doi.org/10.1038/nature19760
953 Huang X, Yang S, Gong J, Zhao Y, Feng Q, Gong H, Li W, Zhan Q, Cheng B, Xia J, et al. Genomic
954 analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat.*
955 *Commun.* 2015:6(6258. https://doi.org/10.1038/ncomms7258
956 Jiang Y, Schmidt RH, Zhao Y, and Reif JC. A quantitative genetic framework highlights the role of
957 epistatic effects for grain-yield heterosis in bread wheat. *Nat. Genet.* 2017:49(12):1741-1746.
958 https://doi.org/10.1038/ng.3974
959 Jones DF. Dominance of linked factors as a means of accounting for heterosis. *Proc. Natl. Acad. Sci.*
960 *USA*. 1917:3(4):310-312. https://doi.org/https://doi.org/10.1073/pnas.3.4.310
961 Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka
962 G, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*.
963 2014:30(9):1236-1240. https://doi.org/10.1093/bioinformatics/btu031
964 Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, and Tanabe M. Data, information,
965 knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 2014:42(Database

issue):D199-205. https://doi.org/10.1093/nar/gkt1076
Kim D, Paggi JM, Park C, Bennett C, and Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 2019:37(8):907-915. https://doi.org/10.1038/s41587-019-0201-4
Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, and Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 2019:20(1):278. https://doi.org/10.1186/s13059-019-1910-1
Krieger U, Lippman ZB, and Zamir D. The flowering gene *SINGLE FLOWER TRUSS* drives heterosis for yield in tomato. *Nat. Genet.* 2010:42(5):459-463. https://doi.org/10.1038/ng.550
Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, and Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009:19(9):1639-1645. https://doi.org/10.1101/gr.092759.109
Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, and Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007:35(9):3100-3108. https://doi.org/10.1093/nar/gkm160
Langmead B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics*. 2010:Chapter 11(Unit 11 17. https://doi.org/10.1002/0471250953.bi1107s32
Li B, and Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011:12(323. https://doi.org/10.1186/1471-2105-12-323
Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018:34(18):3094-3100. https://doi.org/10.1093/bioinformatics/bty191
Li H, and Durbin R. Genome assembly in the telomere-to-telomere era. *Nat Rev Genet*. 2024:25(9):658-670. https://doi.org/10.1038/s41576-024-00718-w
Liu P, Dan Z, Wang Z, Li S, Li N, Yan H, Cai X, and Lu B. Predicting hybrid fertility from maker-based genetic divergence index of parental varieties. *Euphytica*. 2014:203(1):47-57. https://doi.org/10.1007/s10681-014-1249-3
Lowe TM, and Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997:25(5):955-964. https://doi.org/10.1093/nar/25.5.955
Lv Q, Li W, Sun Z, Ouyang N, Jing X, He Q, Wu J, Zheng J, Zheng J, Tang S, et al. Resequencing of 1,143 indica rice accessions reveals important genetic variations and different heterosis patterns. *Nat. Commun.* 2020:11(1):4778. https://doi.org/10.1038/s41467-020-18608-0
Majoros WH, Pertea M, and Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics*. 2004:20(16):2878-2879. https://doi.org/10.1093/bioinformatics/bth315
Marcais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, and Zimin A. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* 2018:14(1):e1005944. https://doi.org/10.1371/journal.pcbi.1005944
Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 2021:49(D1):D412-D419. https://doi.org/10.1093/nar/gkaa913
Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S, et al. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* 2015:43(Database issue):D213-221. https://doi.org/10.1093/nar/gku1243
Nawrocki EP, and Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*.

2013:29(22):2933-2935. https://doi.org/10.1093/bioinformatics/btt509
Ou S, Chen J, and Jiang N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* 2018:46(21):e126. https://doi.org/10.1093/nar/gky730
Pang Z, Zhou G, Ewald J, Chang L, Hacariz O, Basu N, and Xia J. Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat. Protoc.* 2022:https://doi.org/10.1038/s41596-022-00710-w
Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, and Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 2015:33(3):290-295. https://doi.org/10.1038/nbt.3122
Powers L. An expansion of Jones's theory for the explanation of heterosis. *Am. Nat.* 1944:78(776):275-280. https://doi.org/10.1086/281199
Qi T, Wang M, Wang P, Wang L, and Wang J. Insights into heterosis from histone modifications in the flag leaf of inter-subspecific hybrid rice. *BMC Plant Biol.* 2024:24(1):767. https://doi.org/10.1186/s12870-024-05487-6
Qin P, Lu H, Du H, Wang H, Chen W, Chen Z, He Q, Ou S, Zhang H, Li X, et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell*. 2021:184(13):3542-3558 e3516. https://doi.org/10.1016/j.cell.2021.04.046
Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* 2019:14(2):482-517. https://doi.org/10.1038/s41596-018-0103-9
Rhie A, Walenz BP, Koren S, and Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 2020:21(1):245. https://doi.org/10.1186/s13059-020-02134-9
Schrag TA, Melchinger AE, Sorensen AP, and Frisch M. Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL. *Theor Appl Genet*. 2006:113(6):1037-1047. https://doi.org/10.1007/s00122-006-0363-6
Schrag TA, Mohring J, Melchinger AE, Kusterer B, Dhillon BS, Piepho HP, and Frisch M. Prediction of hybrid performance in maize using molecular markers and joint analyses of hybrids and parental inbreds. *Theor. Appl. Genet.* 2010:120(2):451-461. https://doi.org/10.1007/s00122-009-1208-x
Schrag TA, Mohring J, Maurer HP, Dhillon BS, Melchinger AE, Piepho HP, Sorensen AP, and Frisch M. Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses. *Theor. Appl. Genet.* 2009:118(4):741-751. https://doi.org/10.1007/s00122-008-0934-9
Semel Y, Nissenbaum J, Menda N, Zinder M, Krieger U, Issman N, Pleban T, Lippman Z, Gur A, and Zamir D. Overdominant quantitative trait loci for yield and fitness in tomato. *Proc. Natl. Acad. Sci. USA*. 2006:103(35):12981-12986. https://doi.org/10.1073/pnas.0604635103
Shang L, He W, Wang T, Yang Y, Xu Q, Zhao X, Yang L, Zhang H, Li X, Lv Y, et al. A complete assembly of the rice Nipponbare reference genome. *Mol. Plant*. 2023:16(8):1232-1236. https://doi.org/10.1016/j.molp.2023.08.003
Shang L, Li X, He H, Yuan Q, Song Y, Wei Z, Lin H, Hu M, Zhao F, Zhang C, et al. A super pan-genomic landscape of rice. *Cell Res.* 2022:32(10):878-896. https://doi.org/10.1038/s41422-022-00685-z
Shull GH. The composition of a field of maize. *J. Hered.* 1908:4(1):296-301.

https://doi.org/10.1093/jhered/os-4.1.296
Shull GH. The genotypes of maize. *Am. Nat.* 1911:45(532):234-252. https://doi.org/10.1086/279207
Shull GH. What is "heterosis"? *Genetics*. 1948:33(5):439-446. https://doi.org/10.1093/genetics/33.5.439
Shumate A, and Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics*. 2021:37(12):1639-1643. https://doi.org/10.1093/bioinformatics/btaa1016
Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, and Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015:31(19):3210-3212. https://doi.org/10.1093/bioinformatics/btv351
Slater GS, and Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. 2005:6(31. https://doi.org/10.1186/1471-2105-6-31
Spitzer M, Wildenhain J, Rappsilber J, and Tyers M. BoxPlotR: a web tool for generation of box plots. *Nat. Methods*. 2014:11(2):121-122. https://doi.org/10.1038/nmeth.2811
Stanke M, Keller O, Gunduz I, Hayes A, Waack S, and Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 2006:34(Web Server issue):W435-439. https://doi.org/10.1093/nar/gkl200
Stecher G, Tamura K, and Kumar S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol Biol Evol*. 2020:37(4):1237-1239. https://doi.org/10.1093/molbev/msz312
Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, and Lander ES. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics*. 1992:132(3):823-839. https://doi.org/https://doi.org/10.1093/genetics/132.3.823
Sun Z, Peng J, Lv Q, Ding J, Chen S, Duan M, He Q, Wu J, Tian Y, Yu D, et al. Dissecting the genetic basis of heterosis in elite super-hybrid rice. *Plant Physiol.* 2023:192(1):307-325. https://doi.org/10.1093/plphys/kiad078
Tarailo-Graovac M, and Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 2009:Chapter 4(4 10 11-14 10 14. https://doi.org/10.1002/0471250953.bi0410s25
Wang B, Hou M, Shi J, Ku L, Song W, Li C, Ning Q, Li X, Li C, Zhao B, et al. De novo genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. *Nat. Genet.* 2023:55(2):312-323. https://doi.org/10.1038/s41588-022-01283-w
Wang F, Xi Z, Wang M, Wang L, and Wang J. Genome-wide chromatin accessibility reveals transcriptional regulation of heterosis in inter-subspecific hybrid rice. *Plant J.* 2024:119(5):2331-2348. https://doi.org/10.1111/tpj.16920
Wang K, Li M, and Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010:38(16):e164. https://doi.org/10.1093/nar/gkq603
Waterhouse AM, Procter JB, Martin DM, Clamp M, and Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009:25(9):1189-1191. https://doi.org/10.1093/bioinformatics/btp033
Wingett SW, and Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res*. 2018:7(1338. https://doi.org/10.12688/f1000research.15931.2
Wu TD, and Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*. 2005:21(9):1859-1875. https://doi.org/10.1093/bioinformatics/bti310
Xiao J, Li J, Yuan L, and Tanksley SD. Dominance is the major genetic basis of heterosis in rice as

revealed by QTL analysis using molecular markers. *Genetics*. 1995:140(2):745-754. https://doi.org/10.1093/genetics/140.2.745

Xiao J, Li J, Yuan L, McCouch S, and Tanksley SD. Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. *Theor. Appl. Genet.* 1996:92(637-643. https://doi.org/10.1007/BF00226083

Xie J, Wang W, Yang T, Zhang Q, Zhang Z, Zhu X, Li N, Zhi L, Ma X, Zhang S, et al. Large-scale genomic and transcriptomic profiles of rice hybrids reveal a core mechanism underlying heterosis. *Genome Biol.* 2022:23(1):264. https://doi.org/10.1186/s13059-022-02822-8

Xu S, Zhu D, and Zhang Q. Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proc. Natl. Acad. Sci. USA*. 2014:111(34):12456-12461. https://doi.org/10.1073/pnas.1413750111

Xu Z, and Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 2007:35(suppl_2):W265-W268. https://doi.org/10.1093/nar/gkm286

Yang J, Zhao X, Cheng K, Du H, Ouyang Y, Chen J, Qiu S, Huang J, Jiang Y, Jiang L, et al. A killer-protector system regulates both hybrid sterility and segregation distortion in rice. *Science*. 2012:337(6100):1336-1340. https://doi.org/10.1126/science.1223702

Yang X, Zhang J, Wang L, Zhang C, Xu P, Li Y, Yu S, and Li Y. BR signalling haplotypes contribute to indica-japonica differentiation for grain yield and quality in rice. *Plant Biotechnol. J.* 2025:https://doi.org/10.1111/pbi.14610

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*. 2002:296(5565):79-92. https://doi.org/10.1126/science.1068037

Yu Y, Tang T, Qian Q, Wang Y, Yan M, Zeng D, Han B, Wu CI, Shi S, and Li J. Independent losses of function in a polyphenol oxidase in rice: differentiation in grain discoloration between subspecies and the role of positive selection under domestication. *Plant Cell*. 2008:20(11):2946-2959. https://doi.org/10.1105/tpc.108.060426

Zhang J, Chen L-L, Xing F, Kudrna DA, Yao W, Copetti D, Mu T, Li W, Song J-M, Xie W, et al. Extensive sequence divergence between the reference genomes of two elite *indica* rice varieties Zhenshan 97 and Minghui 63. *Proc. Natl. Acad. Sci. USA*. 2016:113(35):E5163-E5171. https://doi.org/10.1073/pnas.1611012113

Zhang QF, Zhou ZQ, Yang GP, Xu CG, Liu KD, and Saghai Maroof MA. Molecular marker heterozygosity and hybrid performance in indica and japonica rice. *Theor Appl Genet*. 1996:93(1218-1224. https://doi.org/10.1007/BF00223453

Zhao Y, Mette MF, Gowda M, Longin CF, and Reif JC. Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity*. 2014:112(6):638-645. https://doi.org/10.1038/hdy.2014.1

Zhao Y, Li Z, Liu G, Jiang Y, Maurer HP, Würschum T, Mock H-P, Matros A, Ebmeyer E, Schachschneider R, et al. Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proc. Natl. Acad. Sci. USA*. 2015:112(51):15624-15629. https://doi.org/10.1073/pnas.1514547112

Zhao Y, Huang Z, Zhou X, Teng W, Liu Z, Wang W, Tang S, Liu Y, Liu J, Wang W, et al. Precise deletion, replacement and inversion of large DNA fragments in plants using dual prime editing. *Nat. Plants*. 2025:11(191-205. https://doi.org/10.1038/s41477-024-01898-3

1142 Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, and Xia J. NetworkAnalyst 3.0: a visual analytics
1143 platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.*
1144 2019:47(W1):W234-W241. https://doi.org/10.1093/nar/gkz240

1145

1146

Figure and figure legends

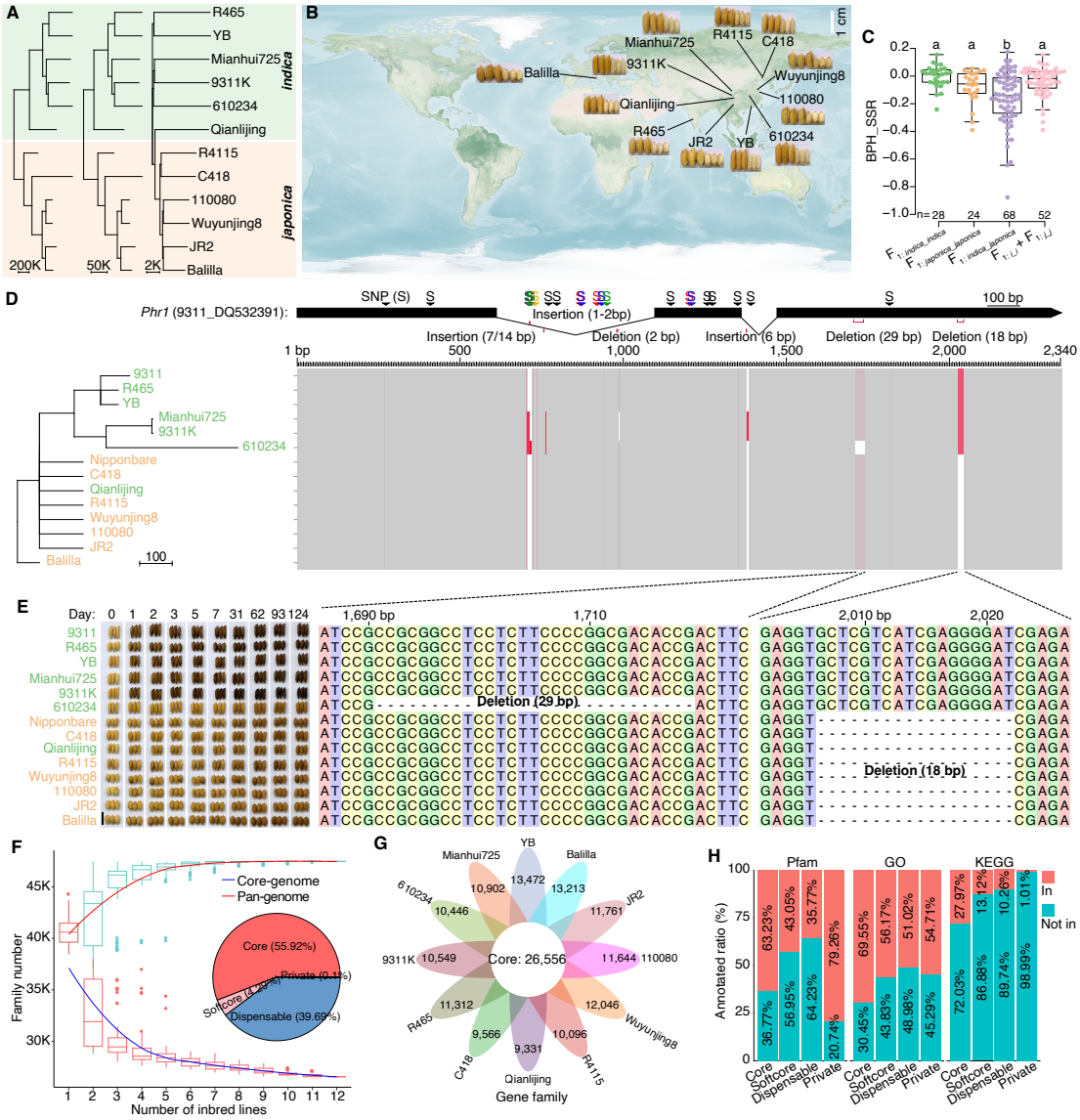

Figure 1. Genome assemblies and pan-genome of 12 rice inbred lines. **A)** Neighbour-joining trees of 12 representative rice inbred line (ILs) based on genomic SNPs (left side), InDels (middle), and CNVs (right side) analyzed from previously reported Illumina sequencing results. **B)** Geographic origins and seeds of the 12 ILs. **C)** Better-parent heterosis for seed setting rate (BPH_SSR) of intra- and inter-subspecific F_1 hybrids. The pooled average values of *indica-indica* (*i_i*) and *japonica-japonica* (*j_j*) F_1 hybrids were also included for comparison. Analysis of variance (ANOVA) with LSD in post-hoc test (P-value < 0.05) was performed. The number of F_1 hybrids are indicated. **D)** Gene structures and neighbour-joining tree of the 12 ILs based on gene sequences of *Phr1*. Nipponbare (*japonica*) and 9311 (*indica*) served as reference accessions. Red regions denote genomic variants, and dashed lines represent deletions. **E)** Phenol reaction phenotypes of 14 ILs. The bar represents 1 cm. **F)** Variation of gene family sizes in the pan- and core-genome. **G)** Core and non-core gene families across the 12 ILs. **H)** Gene annotation ratios cross three databases. Three databases, including Pfam, GO, and KEGG, were selected for annotation analysis.

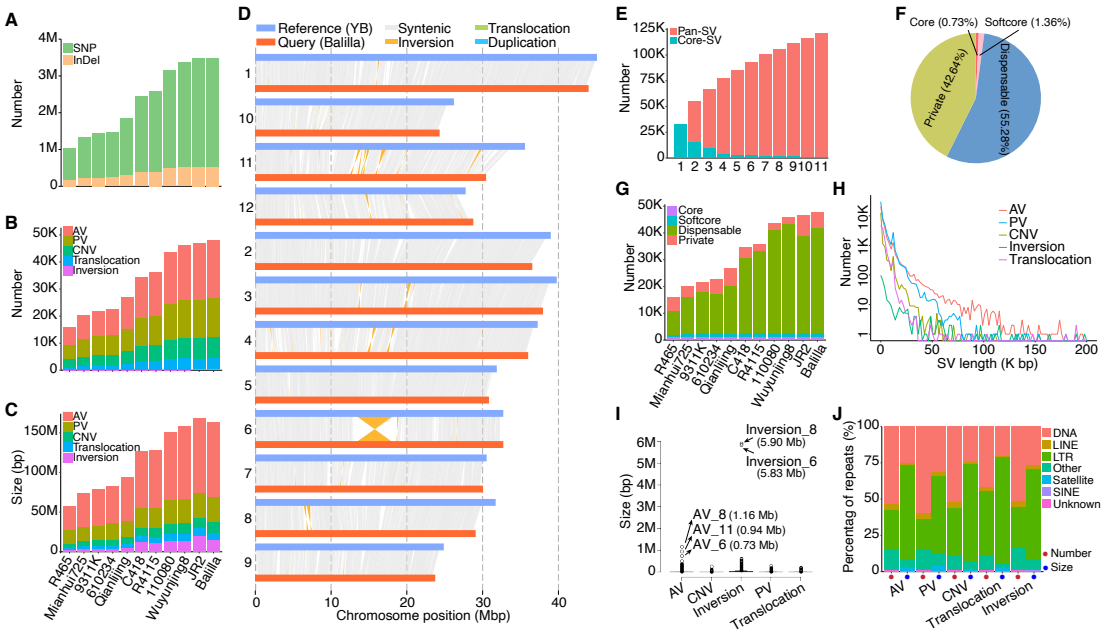

Figure 2. Genomic variants of 12 inbred lines. **A-C)** Number and size distribution of genomic variants in 11 inbred lines with genome assembly of YB as the reference. **D)** Comparison of genomic profiles between YB and Balilla. **E)** Variation of SVs in the pan- and core-SVs with varied number of inbred lines. **F)** Percentages of core, softcore, private, and dispensable SVs. **G)** Number of core, softcore, private, and dispensable SVs in 11 inbred lines. **H-I)** Number and size of SVs. **J)** Percentages of different types of repeats in SVs. The number and size are analyzed.

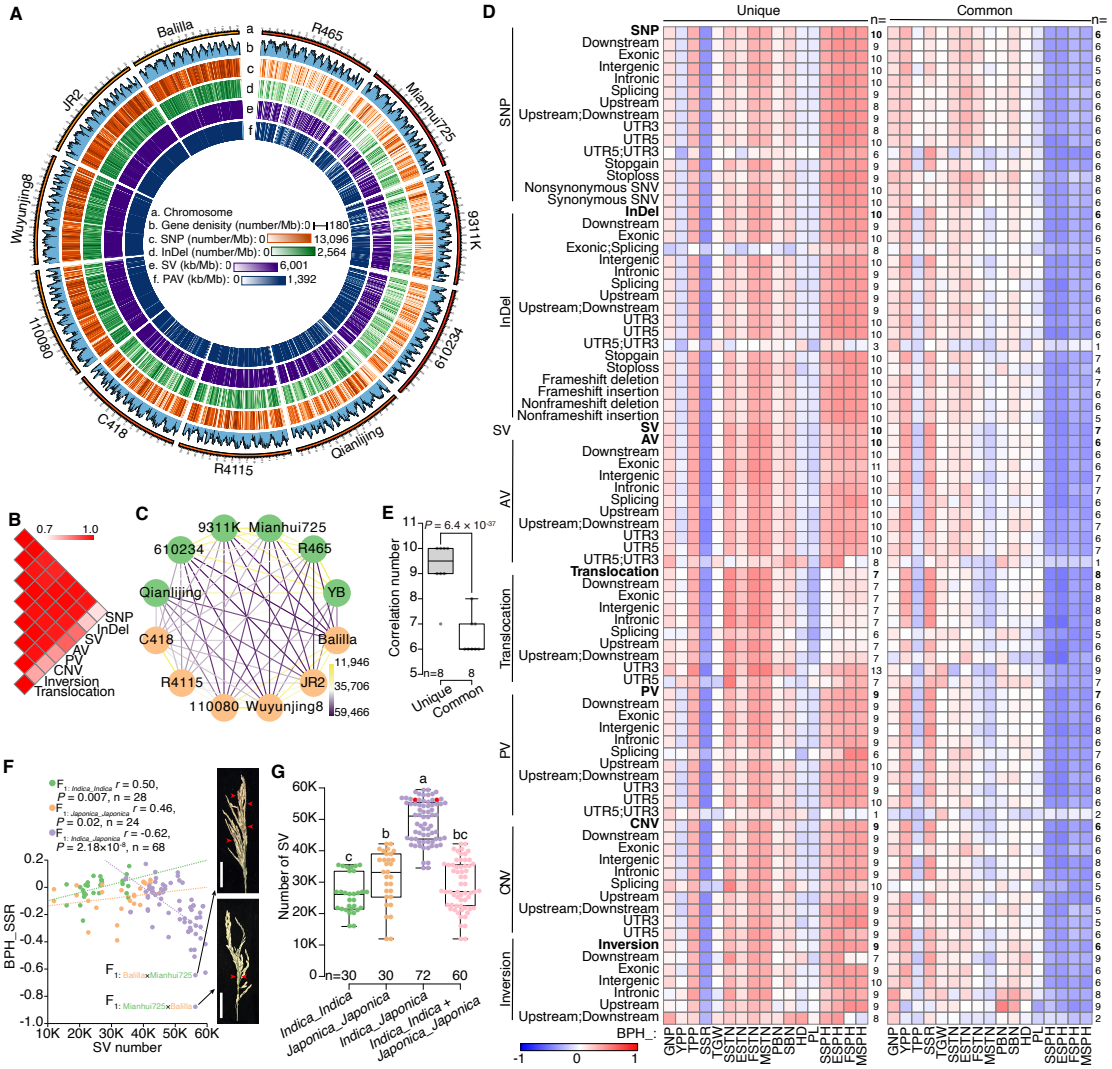

Figure 3. Correlations between parental genomic variants and heterosis of 17 agronomic traits. **A)** Genomic comparisons of YB against the other 11 inbred lines (ILs). **B)** Heatmap for the correlations between different types of unique genomic variants. **C)** Genetic relatedness of 12 ILs based on unique SVs. **D)** Correlation patterns between the number of parental genomic variants and heterosis of 17 traits. Parental unique and common variants were classified into different types according to their location or function. Spearman correlation analysis was performed with better-parent heterosis (BPH) of the following traits: grain number per panicle (GNP), tiller number per plant (TPP), seed setting rate (SSR), thousand-grain weight (TGW), seedling stage tiller number (SSTN), elongation stage tiller number (ESTN), flowering stage tiller number (FSTN), maturation stage tiller number (MSTN), yield per plant (YPP), seedling stage plant height (SSPH), elongation stage plant height (ESPH), flowering stage plant height (FSPH), maturation stage plant height (MSPH), panicle length (PL), primary branch number (PBN), secondary branch number (SBN), and heading date (HD). The number besides each row indicates the number of significant correlations (P-value < 0.05). **E)** Comparison of correlation frequencies between unique and common genomic variants. P-value indicates independent t-test (weighted). **F)** Correlations between the number of unique SVs and heterosis for seed setting rate in intra- and inter-subspecific F_1 hybrids. A pair of reciprocal hybrids ($F_{1:\ \mathrm{Mianhui725\times Balilla}}$ and $F_{1:\ \mathrm{Balilla\times Mianhui725}}$) with extremely low heterosis are shown. Mature seeds with black and white hulls are full and empty seeds, respectively. P-values indicate Spearman correlation. **G)** The number of unique SVs in intra- and inter-subspecific F_1 hybrids. The two reciprocal hybrids are marked in red. ANOVA with LSD in post-hoc test (P-value < 0.05) was performed.

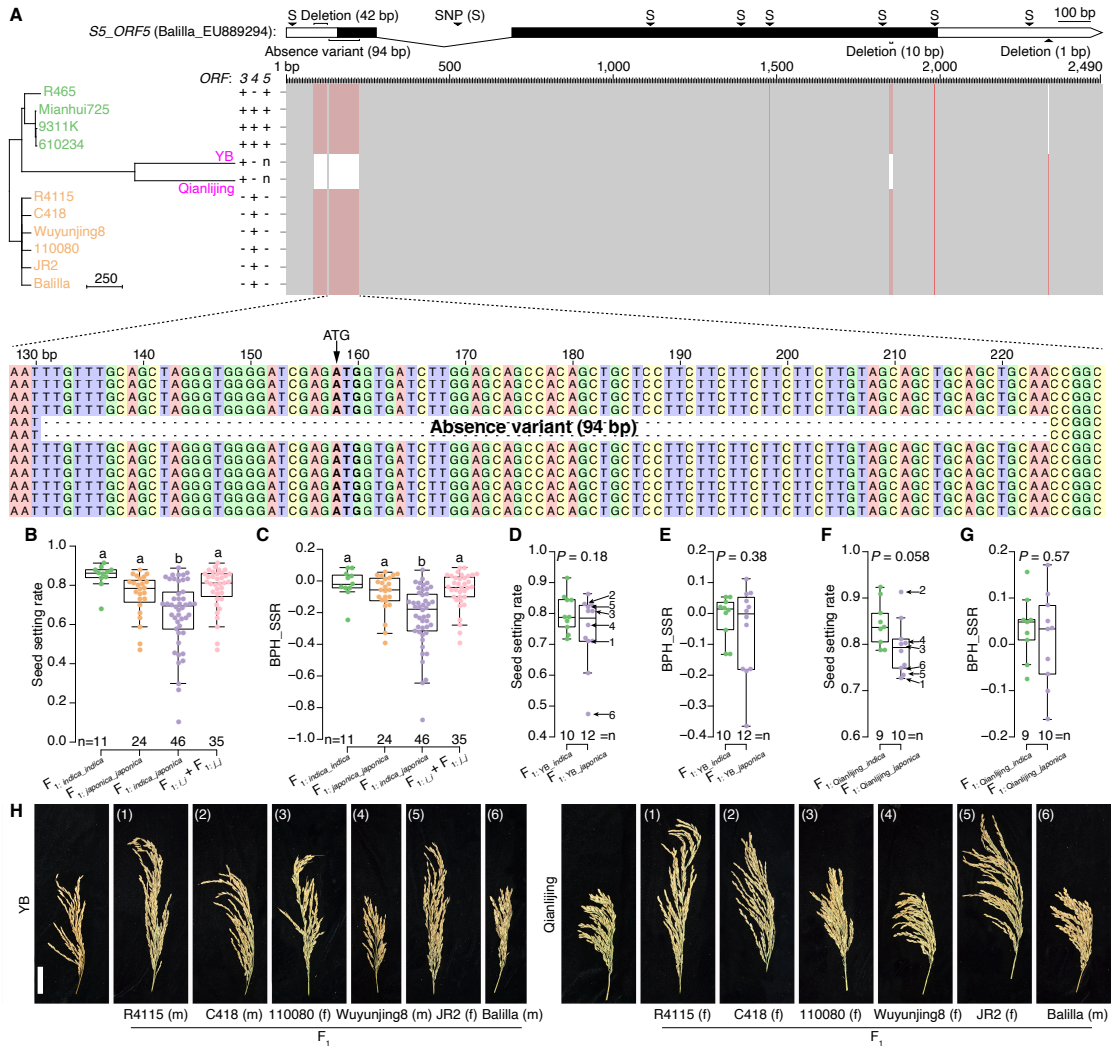

Figure 4. Contribution of AVs in *S5-ORF5* to heterosis for seed setting rate. **A)** Neighbour-joining tree of 12 inbred lines (ILs) based on *S5* loci (three genes) and gene structure of *S5-ORF5*. Genotypes of *S5-ORF3*, *S5-ORF4*, *S5-ORF5* for 12 ILs were assigned with previously reported rules. The "n" represents wide compatibility varieties. The sequence of Balilla serves as the reference and sequence variants are indicated in red. **B-C)** Comparisons of seed setting rate and better-parent heterosis for seed setting rate in intra- and inter-subspecific F_1 hybrids. Two wide compatibility varieties, namely YB and Qianlijing, were excluded for analysis. ANOVA with LSD in post-hoc test (*P*-value < 0.05) was performed. **D-E)** Comparisons of seed setting rate and better-parent heterosis for seed setting rate between intra- and inter-subspecific F_1 hybrids with YB as a parent. **F-G)** Comparisons of seed setting rate and heterosis for seed setting rate between intra- and inter-subspecific F_1 hybrids with Qianlijing as a parent. *P* values in **D-G** indicate independent sample *t*-test. **H)** Panicles of inter-subspecific F_1 hybrids with YB or Qianlijing as a parent. The letter f and m represent female and male parent, respectively. Seed setting rate of these F_1 hybrids are displayed in **D** and **F**.

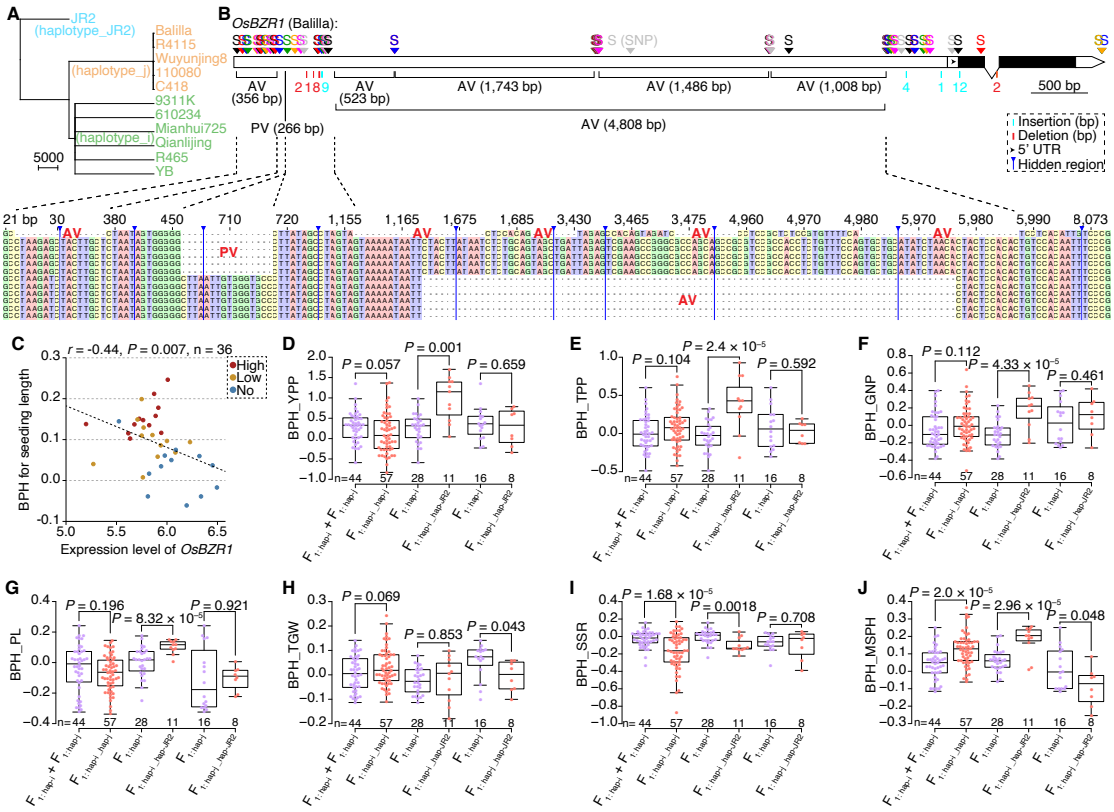

Figure 5. Contributions of SVs in *OsBZR1* to yield heterosis. **A)** Neighbour-joining tree of 12 inbred lines (ILs) based on sequences *OsBZR1*. The 12 ILs were divided into three haplotypes. **B)** Structure variations of *OsBZR1* across 12 ILs. The sequence of Balilla serves as the reference. Single nucleotide polymorphism (SNP), insertion and deletion (InDel), absence variant (AV), and presence variant (PV) were indicated. **C)** Correlation between expression level of *OsBZR1* and better-parent heterosis (BPH) for seedling length. Six pairs of reciprocal hybrids, which were classified as with high-, low-, and no-BPH for seedling length, were adopted for correlation analysis. *P* value indicates Spearman correlation. **D-J)** Comparisons of BPH for seven *OsBZR1*-associated traits between heterozygous and homozygous haplotypes. The seven traits are yield per plant (YPP), tiller number per plant (TPP), grain number per panicle (GNP), thousand-grain weight (TGW), panicle length (PL), maturation stage plant height (MSPH), and seed setting rate (SSR). Hap, i, and j represent haplotype, *indica*, and *japonica*, respectively. *P* values in **D-J** are for independent sample *t*-test.

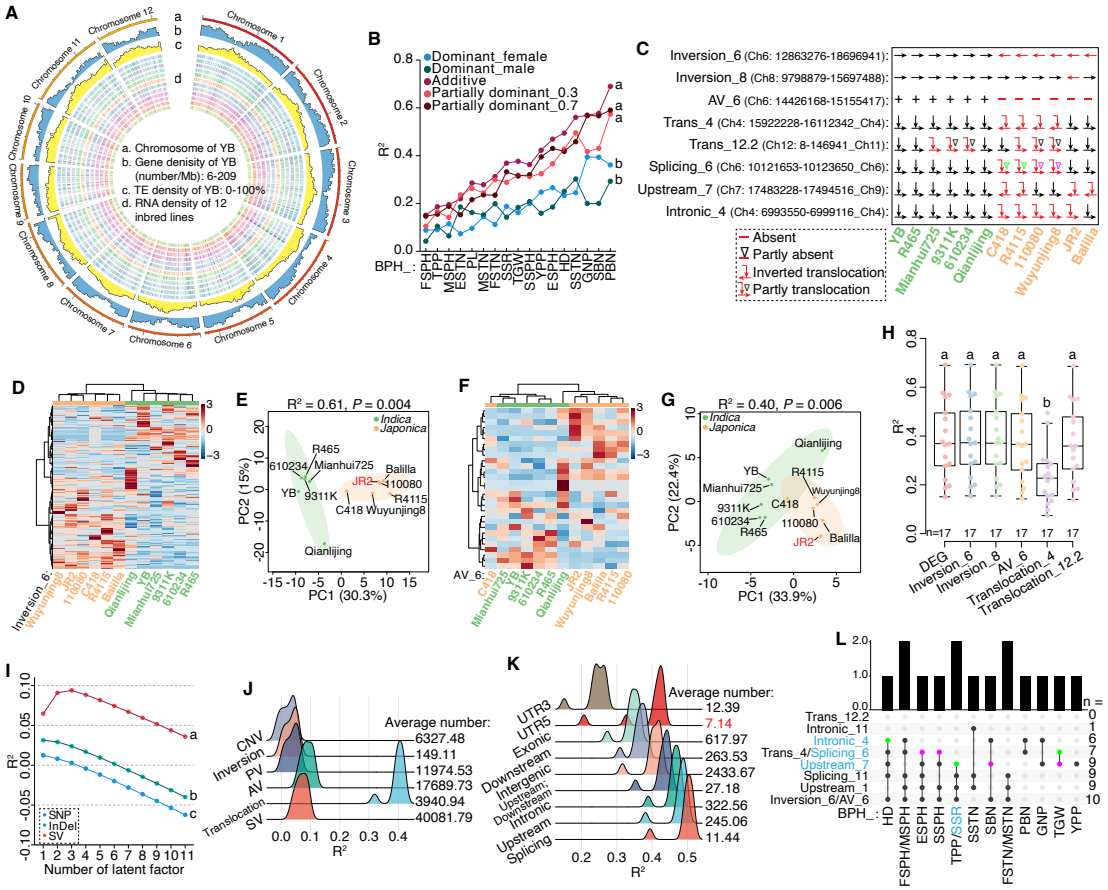

Figure 6. Relationships among genomic SVs, gene expression, and heterosis. **A)** Transcriptomic profiles of 12 inbred lines (ILs). Gene expression data were obtained from 15-day-old seedlings. **B)** Changes in R^2 (adjusted) for heterosis of 17 traits. Different inheritance patterns, including dominant (female and male), additive, and partially dominant (0.3 and 0.7) effects, of expressed genes were used as predictive variables in partial least square (PLS) analysis on better-parent heterosis (BPH) of 17 agronomic traits. The best regression model (with highest adjusted R^2) was selected as the final predictive model for heterosis of each trait. **C)** Genomic structures of two inversions, one absence variant (AV), and five translocations (trans) across 12 ILs. **D-G)** Heatmaps and PCA plots of 12 ILs based on expression levels of genes within inversion_6 and AV_6. *P*-values indicate PERMANOVA. **H)** Comparisons of R^2 (adjusted) for predicting BPH of 17 traits with differentially expressed genes (DEGs) of 12 ILs, genes in inversion_6, genes in inversion_8, genes in AV_6, genes in translocation_4, and genes in translocation_12.2. **I)** Changes in R^2 (adjusted) of PLS models with varied number of latent factors. The additive expression levels of DEGs from 12 ILs were used as variables to predict the number of parental genomic variants. The genomic variants include single nucleotide polymorphism (SNP), insertion and deletion (InDel), and structural variant (SV). **J)** R^2 (adjusted) for different types of SVs. **K)** R^2 (adjusted) for different types of translocations. **L)** Associations between SVs and heterosis. Genes linked to intronic_4, splicing_6, and upstream_7 were used to perform Spearman correlation analysis with heterosis. Their positive and negative correlations with heterosis are indicated in purple and green, respectively. Analysis of variance (ANOVA) with LSD in post-hoc test (*P*-value < 0.05) was performed in **B**, **H**, and **I**. Abbreviations: yield per plant = YPP, secondary branch number = SBN, grain number per panicle = GNP, tiller number per plant = TPP, seed setting rate = SSR, heading date = HD, thousand grain weight = TGW, seedling stage tiller number = SSTN, elongation stage tiller number = ESTN, flowering stage tiller number = FSTN, maturation stage tiller number = MSTN, seedling stage plant height = SSPH, elongation stage plant height = ESPH, flowering stage plant height = FSPH, maturation stage plant height = MSPH, panicle length = PL, and primary branch number = PBN.