Additive and partially dominant effects contribute to crop heterosis

Zhiwu Dan[1,2], Yunping Chen[1,2], Wei Zhou[1], Yanghong Xu[1], Junran Huang[1], Yi Chen[1], Jizhou Meng[1], Guoxin Yao[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. [2]These authors contributed equally: Zhiwu Dan, Yunping Chen. ✉e-mail: wenchaoh@whu.edu.cn

Abstract

The molecular mechanisms of heterosis remain debated due to discrepancies in factors such as species, phenotypes, hybrid populations, developmental stages or growth environments. Here, heterosis-associated genes and metabolites were screened from differential molecules in parents, and the main inheritance patterns of these molecules were additive and partially dominant effects, namely, at the middle parent levels and between the parental and middle parent levels. The additive and partially dominant effects contributed to heterosis of 17 rice agronomic traits, including grain yield and plant height across developmental stages, and to heterosis of F_1 hybrids from different populations or environments in both rice and maize. Furthermore, the two inheritance patterns were associated with parental genomic variants, particularly unique single-nucleotide polymorphisms (SNPs) and insertions/deletions (InDels), which were correlated with heterosis of agronomic traits. Our study systematically integrates parental genetic variation and highlights the predominant genetic effects of additive and partially dominant in crop heterosis.

Introduction

Heterosis describes the universal phenomenon of heterozygous F_1 hybrids outperforming their parents in a variety of phenotypes[1-3], including yield, growth rate, resistance to stress in plants, milk or meat production in animals, and production in yeast. In plants, heterosis was first scientifically reported by Charles Darwin[4]. The utilization of heterosis to increase crop yield and alleviate food shortages has been

31 treated as a landmark innovation in modern agriculture[5]. However, the molecular
32 mechanisms of heterosis for agronomic traits, especially complex traits and traits that
33 occur at different developmental stages or in different environments, remain mysterious.
34 The dominance (pseudo-overdominance)[6-10], overdominance[11,12], and epistasis
35 models[13-15], which emphasize complementation, heterozygosity, and interactions of loci,
36 respectively, have been proposed to interpret heterosis. The molecular mechanisms of
37 crop heterosis are still under debate because the conclusions of various studies differ
38 across species, phenotypes, populations, developmental stages or growth environments.
39 Four terms have been used to quantitatively describe the inheritance patterns of
40 hybrid phenotypes or molecules[16,17], including additive (middle parent values, MPVs),
41 dominant (parental values), partially dominant (between parental values and MPVs),
42 and overdominant (outside parental values) effects. The potential connection between
43 additive effect and heterosis has been inferred in several studies. The additive effect of
44 the component traits in a multiplicative interaction manner has been suggested to be the
45 genetic basis of heterosis for complex traits, including tomato and rice yield[18,19].
46 *SINGLE FLOWER TRUSS*, the expression level of which exhibits a near-additive effect,
47 drives strong yield heterosis in tomato F_1 hybrids[20]. The additive effect of proteins
48 involved in amino acid and carbohydrate metabolism or differentially expressed genes
49 in parents is associated with maize heterosis[21,22]. The additive effect of metabolite
50 concentrations explains variances in yield and yield-related traits in rice F_1 hybrids[23-26].
51 Moreover, partially dominant effect of genetic factors or component traits has also been
52 suggested to underlie heterosis[7,15,18,27].
53 Here, we investigated the molecular mechanisms of heterosis for seedling length
54 and 17 additional key agronomic traits in rice. Through the systematic identification of
55 heterosis-associated genes and metabolites, the proportions and contributions of the
56 four types of inheritance patterns in seedling length were analysed. Given the close
57 relationship between seedling length and plant height, the two main inheritance patterns,
58 namely, additive and partially dominant effects, were subsequently validated to
59 multiple agronomic traits of both rice and maize F_1 hybrids from different hybrid
60 populations under distinct environments. Furthermore, unique genomic variants in
61 parents that were correlated with heterosis of agronomic traits were associated with the
62 two main inheritance patterns.

63 # Results

64 **Heterosis-associated genes and metabolites of rice seedlings mainly originate from**
65 **differential molecules in parents**

66 The seedling length and weight of the rice cultivars Yuetai B (YB, *indica*), Balilla
67 (*japonica*), and their reciprocal F_1 hybrids were measured at 5, 10, and 15 days after
68 sowing (DAS) (Fig. 1a and Extended Data Fig. 1a). Better-parent heterosis (BPH) for
69 seedling length was observed in one of the reciprocal F_1 hybrids as early as 5 DAS, and
70 seedling length heterosis was observed in both F_1 hybrids at 15 DAS (Fig. 1a).
71 Moreover, the variance in seedling length was almost completely explained by
72 genotype, stage, and their interaction (Extended Data Fig. 1b), highlighting the
73 contributions of genotype and stage to seedling length.

74 Principal component analysis (PCA) of global gene expression levels showed that
75 the F_1 hybrids were located between both parents and closer to one parent, Balilla (Fig.
76 1b and Extended Data Fig. 1c). Differentially expressed genes (DEGs) were identified
77 for the two inbred lines (ILs) and F_1 hybrids, and the number of DEGs increased along
78 with development (Extended Data Figs. 1d-e). Unlike the pooled non-DEGs (NDEGs)
79 (Extended Data Fig. 1f), 66.6% of the sample variance in genotypes was explained by
80 principal component 1 (PC1), which was based on the DEGs (Extended Data Fig. 1g).
81 Furthermore, the F_1 hybrids deviated slightly from the mimicked "additive hybrids"
82 (calculated as MPVs), suggesting the contributions of additive and other genetic effects
83 to hybrid transcriptomics.

84 The greatest number of DEGs existed between the two parents, and most of the
85 DEGs (83.5% to 97.1%, average 92.4%) between the F_1 hybrids and the parents were
86 parental DEGs (Fig. 1c and Extended Data Figs. 1h-j). Moreover, the majority of the
87 significantly enriched Gene Ontology terms between the F_1 hybrids and the parents
88 were among the differentially expressed terms between Balilla and YB (Extended Data
89 Figs. 1k-p), indicating that the reconstruction of the hybrid transcriptomes was
90 predominantly determined by parental expression differences.

91 The 1,476 DEGs that overlapped for genotype, stage, and their interaction were
92 selected, and 275 DEGs that differed only between the two parents or between
93 reciprocal F_1 hybrids were excluded from the analysis (Extended Data Figs. 2a-g). The
94 remaining 952 upregulated and 249 downregulated DEGs in the F_1 hybrids were
95 associated with 17 terms, ten of which were also differentially expressed terms in the

96 parents (Extended Data Figs. 2h-i). Next, 626 heterosis-associated genes (HAGs)
97 related specifically to seedling length were then identified, 499 positively and 127
98 negatively correlated (Extended Data Figs. 2j-p), and these HAGs indicated that the F_1
99 hybrids were closer to "additive hybrids" (Fig. 1d). A total of 37 HAGs were enriched
100 in 12 terms (Fig. 1e), and these terms formed a complete network with the bridge of
101 heme binding (Fig. 1f), indicating that the genes involved in heterosis for seedling
102 length had tight connections.

103 Moreover, the PCA score plot based on untargeted metabolite profiles revealed that
104 the F_1 hybrids were located between the corresponding parents and presented a closer
105 relationship with YB (Fig. 1g and Supplementary Fig. 1a). A decreasing trend in the
106 number of differential metabolites (DMs) was clearly observed (Supplementary Figs.
107 1b-c). In contrast to the mixed distribution of the parents, F_1 hybrids, and "additive
108 hybrids" according to the pooled non-DM (NDM, Supplementary Fig. 1d), genotypic
109 differences were clearly observed according to the DMs (Supplementary Fig. 1e).
110 Although the percentages were lower than those of the transcriptomic data, most of the
111 metabolomic differences between F_1 hybrids and parents were due to differences in the
112 parents (Supplementary Figs. 1f-j, average 84.0%). Meanwhile, most of the
113 significantly enriched metabolic pathways between the F_1 hybrids and the parents also
114 differed between the two parents (Fig. 1h and Supplementary Figs. 1k-u). Thus, we
115 concluded that both the transcriptomic and metabolomic variation between F_1 hybrids
116 and parents were determined mainly by the molecular differences in corresponding
117 parents.

118 When the DMs were further screened for heterosis-associated metabolites (HAMs)
119 related to seedling length (Extended Data Figs. 3a-m), 645 HAMs were identified and
120 indicated that the F_1 hybrids were also closer to "additive hybrids" (Fig. 1i). Eight
121 HAMs annotated with the *mummichog* algorithm[28] were significantly enriched in three
122 pathways, and other chemically identified metabolites from these three pathways had
123 either positive or negative correlations with seeding length (Fig. 1j and Supplementary
124 Fig. 2a). Three of the six accurately annotated metabolites from the phenylpropanoid
125 biosynthesis had normalized metabolite levels and were correlated with the seedling
126 length of F_1 hybrids (Extended Data Figs. 3n-v), suggesting a connection between
127 phenylpropanoid biosynthesis and seedling length heterosis.

128 After the 626 HAGs and the 645 HAMs were combined for joint pathway analysis,

five pathways were found to be significantly enriched, among which the phenylpropanoid biosynthesis was most significant with eight metabolites and ten genes matched (Fig. 1k). Among the ten genes, nine genes were peroxidases, and seven genes were among the 37 HAGs (Supplementary Fig. 2b), confirming the combined contributions of genes and metabolites involved in phenylpropanoid biosynthesis to seedling length heterosis.

The integration of differentially expressed genes and metabolites in parents explains the variance in heterosis

Another nine pairs of reciprocal F_1 hybrids, which were divided into three groups—high, low, and no BPH for seedling length—were used to validate the HAGs and HAMs (Figs. 2a-b). The 626 HAGs failed to separate the three groups in both the PCA and the partial least squares-discriminant analysis (PLS-DA) (Supplementary Figs. 3a-d). The 645 HAMs did not classify the three groups according to PCA (Supplementary Fig. 3e), but they separated the high BPH group from the other two groups without overfitting according to PLS-DA (Supplementary Figs. 3f-h). Thus, the HAGs or HAMs for a single hybrid combination manifested predictive power for heterosis of other F_1 hybrids.

The HAGs and HAMs were subsequently integrated into a network (Network_51015), and the largest subnetwork (subnetwork_1) contained 2,138 edges, including 61.8% and 38.2% positive and negative correlations, respectively, with 211 HAGs and 196 HAMs as nodes (Fig. 2c). In contrast to the decreasing trend in the parents, the coefficients of module levels (average node levels) and seedling length remained greater than 0.95 in the hybrids (Supplementary Figs. 4a-c), indicating the robustness of the top three modules to heterosis for seedling length. The high and no BPH groups were not separated by either Module_1 or Module_3 in PCA but were separated by Module_2 (Supplementary Figs. 4d-f). Unexpectedly, the F_1 hybrids were divided into distinct groups in PLS-DA, and the predictabilities of the models varied, with that of Module_2 being as high as 0.97 (Supplementary Figs. 4g-i). Thus, the integration of HAGs and HAMs into network modules increased the predictability of heterosis for seedling length.

A total of 967 DEGs and 990 DMs were subsequently identified for six ILs of the ten pairs of reciprocal F_1 hybrids. The 967 DEGs roughly divided the three groups of F_1 hybrids according to PC1 and PC4, whereas the 990 DMs failed to separate the groups (Supplementary Figs. 5a-h). The 967 DEGs displayed better separations in PLS-

162 DA than in PCA, and the predictability was comparable to that of the 990 DMs
163 (Supplementary Figs. 5d,h). The combination of DEGs and DMs separated the three
164 groups according to PC1 and PC4, but the model was overfitted in PLS-DA
165 (Supplementary Figs. 5i-l). Next, another network (Network_HLN) integrating the 967
166 DEGs and 990 DMs was constructed and the largest subnetwork (subnetwork_1)
167 contained 868 edges consisting of 97 DEGs and 149 DMs (Fig. 2d). Although the top
168 three modules of subnetwork_1 from Network_HLN had fewer nodes than those of
169 Network_51015 did (Fig. 2e), the predictabilities of the top three modules of
170 Network_HLN were comparable to those of Network_51015 (Fig. 2f). And the variance
171 explained by Component 1 were larger in Network_HLN (Fig. 2g). The predictabilities
172 of the first modules in Network_51015 and Network_HLN were close (Extended Data
173 Figs. 4a-f). However, Network_HLN had 113 molecules, much fewer than the 150
174 molecules in Network_51015, demonstrating the advantage of informative molecules
175 from multiple ILs with different genetic backgrounds. The combination of Module_1
176 and Module_2 further increased the predictability to 0.94 (Extended Data Figs. 4g-h),
177 whereas the addition of Module_3 did not improve the predictability (Extended Data
178 Figs. 4i-j). Thus, the integration of parental transcriptomic and metabolomic differences
179 into network modules explained the variance in heterosis for seedling length.

180 **Most heterosis-associated molecules display additive and partially dominant**
181 **effects**

182 Heterosis-associated molecules specific to seedling length were further screened from
183 Module_1 and Module_2 of Network_HLN, and the 68 molecules (29 metabolites and
184 39 genes) with the greatest contribution separated the three groups with high
185 predictability (Fig. 2h-i, Q^2 = 0.90). The contributions of the molecules were
186 determined (Fig. 2j and Extended Data Fig. 4k), but the HAGs and HAMs did not
187 exhibit continuously changing patterns at the molecular level across groups. Moreover,
188 F_1 hybrids from the same group formed distinct clusters (Fig. 2k), displaying
189 heterogeneity at the molecular level. In addition, the hybrid profiles were intermediate
190 between the levels of the parents or close to those of one parent in the heatmap. And
191 the F_1 hybrids were located near the middle regions of corresponding parents on the
192 PCA score plot (Fig. 2l). These results highlighted the quantitative connections of
193 heterosis-associated molecules between parents and F_1 hybrids.

194 Additive, partially dominant, dominant, and overdominant effects were detected

195 by comparing the molecular levels of heterosis-associated molecules (Fig. 2m and
196 Extended Data Fig. 4l). Additive and partially dominant effects accounted for more than
197 70% of the molecular levels in most of the F_1 hybrids (Fig. 2n and Extended Data Figs.
198 4m-o). The predictability of the additive effect for heterosis of seedling length was 0.63
199 (Fig. 2o, $P < 5\times10^{-4}$). Two types of partially dominant (PD) effects, namely, PD_0.3
200 and PD_0.7, were identified, and their predictabilities were 0.37 ($P < 5\times10^{-4}$) and 0.29
201 ($P < 5\times10^{-4}$), respectively (Fig. 2p). The predictabilities of the dominant effects of the
202 female and male parents were 0.14 ($P = 0.40$) and 0.07 ($P = 0.015$), respectively (Fig.
203 2q). Importantly, the predictability of the additive effect was not significantly different
204 from that of the hybrid profiles, and the addition of additive, partially dominant, or
205 dominant effects to the hybrid profiles did not improve the predictability (Fig. 2q).
206 Furthermore, the sum percentages of additive and partially dominant effects had a
207 significant positive correlation with seedling length heterosis (Fig. 2r). The above
208 results demonstrated the predominant roles of additive and partially dominant effects
209 on seedling length heterosis.

210 **Additive and partially dominant effects contribute to the heterosis of 17 agronomic**
211 **traits in rice**

212 We then measured plant height for the above F_1 hybrids at four developmental stages
213 and detected significant positive correlations both between seedling length and plant
214 height and between seedling length heterosis and plant height heterosis (Extended Data
215 Figs. 5a-b), revealing close relationships of the same trait across developmental stages.

216 A hybrid population with a complete diallel cross design was constructed for 17
217 ILs (Figs. 3a-b and Extended Data Figs. 5c-d), which were representative of 223
218 accessions, including *indica*, intermediate types, and *japonica*. Seventeen agronomic
219 traits were measured, and 14 traits presented higher trait values in the 245 F_1 hybrids
220 than in their parents (Extended Data Figs. 5e-f). The average BPH values for only four
221 traits were less than zero (Fig. 3c), indicating increased fitness of the F_1 hybrids to ILs.
222 The values of the 17 traits had different correlation patterns between the ILs and F_1
223 hybrids, and heterosis for the 17 traits displayed a similar correlation pattern to that of
224 the corresponding ILs (Fig. 3d and Extended Data Figs. 5g-h).

225 After the metabolite profiles of the 17 ILs were obtained (Extended Data Fig. 6a),
226 different inheritance patterns of five chemically identified metabolites from the
227 phenylpropanoid biosynthesis displayed significant correlations with heterosis for plant

228 height (Fig. 3e and Extended Data Figs. 6b-f). 4-hydroxycinnamic acid, which is
229 associated with heterosis for plant height[23], was significantly correlated with plant
230 height heterosis and had decreased correlation coefficients across developmental stages
231 (Extended Data Fig. 6g). Importantly, additive and partially dominant effects were more
232 strongly correlated with plant height heterosis than the dominant effects (Extended Data
233 Fig. 6h), even after the inclusion of another 13 traits (Fig. 3f). Therefore, the
234 phenylpropanoid biosynthesis was closely related to heterosis and additive and partially
235 dominant effects might be the main inheritance patterns for heterosis of multiple
236 agronomic traits.

237 Next, the three types of inheritance patterns of the 1,306 differential metabolites
238 from the 17 ILs were used to predict heterosis for each trait by PLS. We found that the
239 additive effect had higher R^2 (adjusted) values than the dominant and partially dominant
240 effects (Fig. 3g and Extended Data Figs. 6i-x). Moreover, the greatest R^2 values of the
241 regression models for the 17 traits with additive effect were significantly greater than
242 those with dominant effects but comparable to those with partially dominant effects
243 (Fig. 3h). Therefore, we thought that additive and partially dominant effects were the
244 main inheritance patterns associated with the heterosis of 17 agronomic traits.

245 A test cross population was designed with the Honglian-type male sterile line
246 Yuetai A and 211 recombinant inbred lines (RILs) as the female and the male parents,
247 respectively (Extended Data Fig. 7a). The F_1 hybrids and corresponding parents
248 exhibited significant differences in three of the four traits (Extended Data Figs. 7b-e),
249 and the test crosses exhibited negative BPH only for heading date (Extended Data Fig.
250 7f). The metabolite profiles of 107 representative RILs and Yuetai A were collected for
251 predicting heterosis (Extended Data Figs. 7g-h), and the R^2 values were greater than
252 0.85 with five latent factors (Fig. 3i). Thus, the differential metabolites of representative
253 ILs were robust biomarkers for the heterosis of F_1 hybrids from recombinant inbred
254 lines.

255 Eight traits of the 17 ILs and 36 F_1 hybrids from the diallel cross population were
256 recollected, and significant differences between environments were observed for most
257 of the traits (Supplementary Fig. 6a). In contrast, heterosis for six of the eight traits
258 displayed no significant variation (Supplementary Fig. 6b). The greatest R^2 values of
259 the eight traits with additive effect were greater than those with dominant effects, and
260 were close to those with partially dominant effects (Fig. 3j and Supplementary Figs. 7a-

h). Hence, the additive and partially dominant effects contributed more than dominant effects to the heterosis of eight agronomic traits across environments.

Finally, we changed the growth conditions of the 17 ILs and obtained another set of metabolite profiles (Supplementary Figs. 8a-c). Consistent with the above results, additive and partially dominant effects of the 2,391 differential metabolites had higher R^2 values than did the dominant effects (Supplementary Figs. 8d-u). The results confirmed that additive and partially dominant effects were the main inheritance patterns underlying the heterosis of key agronomic traits.

The main inheritance patterns of maize heterosis are additive and partially dominant effects

To investigate the contributions of additive and partially dominant effects to the heterosis of agronomic traits in maize, the field fresh weight of two hybrid populations[29], which significantly differ under distinct growth environments (Extended Data Figs. 8a-b), and the metabolite profiles of the roots and leaves of 24 Dent lines, 25 Flint lines, and 326 F_1 hybrids (Dent × Flint) were used to perform PLS analysis (Extended Data Figs. 8c-j). For the top R^2 values, additive and partially dominant effects were comparable to that of the hybrid profiles and were greater than the dominant effects (Fig. 3k). Moreover, additive and partially dominant effects had similar R^2 values to those of the hybrid profiles and had higher values than did the dominant effects for grain dry matter yield and grain dry matter content[30] (Fig. 3l and Extended Data Figs. 8k-n).

Moreover, the grain yield per plant of another maize hybrid population[31], which exhibited strong yield heterosis, and the transcript levels of 84 DEGs (mainly core genes) in 12 ILs were obtained for analysis (Extended Data Figs. 8o-q). The additive effect had the highest R^2 values in PLS models for hybrid yield and yield heterosis (Fig. 3m and Extended Data Figs. 8r-s), and the R^2 values of the additive and partially dominant effects were greater than those of the dominant effects (Fig. 3n). We therefore thought that the main inheritance patterns underlying the heterosis of maize agronomic traits were additive and partially dominant effects.

Unique genomic variants in parents are associated with both the main inheritance patterns and heterosis

Genomic sequencing was performed on seedlings of the 17 rice ILs, and four types of genomic variants, including single-nucleotide polymorphisms (SNPs), insertions and

deletions (InDels), copy number variations (CNVs), and structural variations (SVs), were collected (Figs. 4a-c). Compared with the Nipponbare genome, the 17 ILs presented fewer genetic differences from *indica* to *japonica* (Extended Data Figs. 9a-p and Supplementary Figs. 9a-p). The 17 ILs had the lowest cross-validation errors with two groups and showed similar clustering patterns with different types of genomic variants (Figs. 4d-f and Supplementary Figs. 10a-i), which were almost consistent with those of the metabolite profiles (Fig. 4g). Thus, the genomic data revealed genetic diversity of the 17 ILs and genomic variants was reflected accordingly by differences in metabolite profiles.

The genomic variants between the two parents of each F_1 were divided into common and unique ones, and the number of parental unique variants had higher correlation coefficients with heterosis for plant height than with heterosis for the other traits (Figs. 4h and Extended Data Figs. 10a-b). The parental unique variants also exhibited more correlations with heterosis than the common ones did (Fig. 4i). Furthermore, unique SNPs and InDels exhibited more correlations with heterosis than the unique CNVs and SVs did (Fig. 4j). The parental unique SNPs and InDels were significantly correlated with heterosis for grain yield, but CNVs and SVs were not. With more parental unique SNPs or InDels, the heterosis for grain number increased, but that for the seed setting rate decreased (Fig. 4k and Extended Data Figs. 10c-e), revealing the barrier of reproductive isolation in the utilization of yield heterosis between *indica* and *japonica* varieties[32]. The opposing trends of heterosis for grain number and the seed setting rate caused the heterosis for grain yield to form a parabola (Fig. 4l and Extended Data Figs. 10f-h). These results indicated that the unique genomic variants of the parents were associated with heterosis for agronomic traits and that the heterosis for grain yield was genetically determined by the inverse relationship between its two component traits.

The three types of inheritance patterns of the top 50 contributing metabolites for heterosis of each trait were used to perform correlation analysis with the numbers of parental unique variants, and the metabolites with additive and partially dominant effects exhibited more correlations with heterosis than those with dominant effects (Extended Data Fig. 10i). Unexpectedly, the additive and partially dominant effects also exhibited more correlations with all four types of parental unique variants (Fig. 4m and Extended Data Figs. 10j-l). Thus, parental unique genomic variants were strongly

associated with heterosis-associated metabolites with additive and partially dominant effects. Finally, a model integrating parental genetic variation in genomics, transcriptomics, and metabolomics was proposed to show the contribution of additive effect to positive better-parent heterosis of a complex trait (Fig. 4n).

Discussion

The genetic mechanisms underlying heterosis have been debated for more than a century, and no consensus has been reached to date. The mysterious molecular mechanisms of heterosis may be due to the large number of heterosis-associated molecules, including genes[1,3,16,33], metabolites or other molecules. Heterosis-associated genes with strong functions[20] are rarely reported, and mutations in single heterosis-associated genes may be difficult to drive the expected changes in heterosis. The regulation of gene expression levels for one pathway or network module is worthy of further study, as heterosis-associated molecules are always interconnected.

Models of heterosis based solely on gene interactions may hinder the interpretation of heterosis for complex traits because interactions among component traits occur at the phenotypic level[18]. The heterosis of complex traits, e.g., grain yield, is a product of several component traits, e.g., grain number and grain weight, which are always negatively correlated. The component traits also include component traits, e.g., the number of branches and grain length. It is therefore necessary to decompose the components of a complex trait, and the investigation of heterosis for component traits may be fruitful. Moreover, non-differential molecules may also be involved in heterosis as they can be responsible for component traits without heterosis.

Additive and partially dominant effects are the main inheritance patterns in crop heterosis and they contribute to heterosis of different traits across hybrid populations, developmental stages, growth environments, detection platforms of omics data, and even species. The additive effect, namely, the average parental levels, has no significant difference in predictability from that of hybrid profiles, and the additive effect has been used to predict multiple agronomic traits of F_1 hybrids[23-26], which all contribute to the precise pairing of male and female parents in hybrid breeding. To build robust prediction models for crop heterosis, the collection of genetically divergent accessions, the genetic diversity of which can be assessed by untargeted metabolite profiles, should first be implemented. For heterosis of traits with low predictabilities, e.g., tiller number,

obtaining metabolite profiles from corresponding tissues or mixed tissues[22], other types of omics data[34-37], and phenotypic data from distinct environments[38] may explain more of the variance. Furthermore, the determination of the effects of parental genomic variants on changes in hybrid transcriptomics or metabolomics using pan- or spatial omics strategies will provide a more comprehensive landscape of the molecular mechanisms of crop heterosis.

Methods

Plant materials and field growth conditions

According to the *indica–japonica* indices of 223 inbred lines (ILs) evaluated by 33 pairs of InDel markers[39,40], 17 ILs, ranging from *indica* to intermediate types to *japonica*, were selected as parents of a hybrid population using a complete diallel cross design[26]. Moreover, 211 recombinant inbred lines (RILs, F_5 followings) derived from F_1 individuals of the 223 ILs were crossed with the Honglian-type male sterile line Yuetai A. Plants of the diallel and test cross populations were planted at the Hybrid Rice Experimental Base of Wuhan University in Ezhou (Hubei Province, China) in 2012 and 2015, respectively.

For the diallel cross population (245 F_1 hybrids), ten experimental plants were planted with 16.5 cm between plants and 26.4 cm between rows, and two plants of the sterile line, namely Yuetai A, were planted on two sides of each row. To obtain F_1 hybrids, the young panicles of the female parents that would flower in the next few days were immersed in warm water at 44 °C (for *indica* or near-*indica* ILs) or 46 °C (for *japonica* or near-*japonica* ILs) for 5 min to sterilize the pollen. For the hand-pollination procedure, three to four panicles were combined in a paper bag, and one to two bags from each female parent without pollination were used as controls. For female parents with a high self-fertilization rate, such as R465 and Qianlijing, the pistils in each spikelet were manually removed and pollinated by hand. Three biological replicates of each genotype were randomly distributed in the field. The middle plants of each row, ranging from three to ten plants, were selected for the collection of phenotypic data for 17 agronomic traits. Plant height, measured from the base of the plant to the highest part of the plant using meter rulers, was recorded from the seedling to mature stages. The four time points included 28 days after transplanting (8 July, 2012), 42 days after transplanting (22 July, 2012), the time of flowering, and the time

391 of maturation, which were defined as the seedling stage plant height, elongation stage
392 plant height, flowering stage plant height, and maturation stage plant height,
393 respectively. The tiller number per plant was also counted together with the plant height
394 at four different stages, which were defined as the seedling stage tiller number,
395 elongation stage tiller number, flowering stage tiller number, and maturation stage tiller
396 number. The heading date was the number of days between the heading time when a
397 plant from a biological replicate started flowering and the date of sowing (13 May,
398 2012). At maturity, all panicles from each plant were harvested and dried in the sunshine
399 for approximately one week and then dried in ovens at 42 °C for two days before yield
400 and yield-related traits were measured. Panicle length was measured from the panicle
401 neck to the top for all panicles from each plant. The primary branch number and
402 secondary branch number per panicle were counted manually for all panicles. Panicles
403 from each plant were counted and treated as the tiller number per plant. After manually
404 removing seeds (with or without grains) from panicles for each individual plant, empty
405 seeds and seeds with grains were mechanically separated and counted using seed
406 counters (PME-1, Shanke Equipment, Shanghai). The grain number per panicle was
407 calculated by dividing the total seed number (sum of empty seeds and seeds with grains)
408 of a plant by the tiller number per plant. The seed setting rate was calculated as the ratio
409 of the number of seeds with grains to the total number of seeds. The weight of 1,000
410 grains from a plant was defined as the thousand-grain weight, and the weight of the
411 seeds with grains per plant was defined as the grain yield per plant.
412 For the test cross population (211 F_1 hybrids), each biological replicate contained
413 ten plants (per row) spaced at 16.5 cm × 16.5 cm apart. Three biological replicates of
414 each genotype were placed in three consecutive rows and randomly planted in the field.
415 The phenotypic data of four agronomic traits, namely, heading date, thousand-grain
416 weight, maturation stage plant height, and yield per plant were collected. The heading
417 date was recorded for one biological replicate. The middle five plants per row were
418 selected for measurement of plant height at the maturation stage. Panicles from the five
419 plants were bagged and used to collect data on grain yield per plant and thousand-grain
420 weight. Moreover, eight agronomic traits, namely, heading date, thousand grain weight,
421 panicle length, seed setting rate, grain number per panicle, maturation stage plant height,
422 grain yield per plant, and tiller number per plant, of the 17 ILs and 36 F_1 hybrids from
423 the diallel cross population were recollected along with the test crosses. Parts of trait

values of the diallel and test cross populations have been reported in previous studies[19,23-26,32], and the complete datasets for all investigated traits are provided in source data.

To evaluate the degree of heterosis for each trait, better-parent heterosis was calculated using the equation: BPH_trait = (F_1 - P_{high})/P_{high}, where P_{high} represents the high trait value of both parents. In the calculation of better-parent heterosis for heading date (BPH_HD), parents with earlier heading dates were selected as P_{high}. The calculated values were multiplied by -1 to represent the heterosis for heading date.

Planting of rice seedlings for metabolomics analysis

Three batches of rice seedlings were planted for untargeted metabolite profiling analysis. The first batch in 2016 (Batch_2016) included 17 ILs, 107 RILs selected from the 211 RILs, and the male sterile line Yuetai A. As previously described[24], the germinated seeds were sown in soil with a spacing of 2 × 2 cm. The seedlings were grown at 30 °C under a 16 h light/8 h dark photoperiod in a light incubator. Two biological replicates were established for each genotype. On the 15th day, the aboveground parts, excluding approximately 0.5 cm of tissue close to the soil, of approximately ten plants were harvested and washed with ddH_2O three times. Two biological replicates of the 17 ILs were harvested as two experimental samples for each genotype. Two biological replicates of the 107 RILs and Yuetai A were pooled as one experimental sample for each genotype. The seedlings were first placed in 5 mL tubes (Eppendorf #0030119460) and snap frozen in liquid nitrogen.

The second batch in 2018 (Batch_2018) consisted of ten pairs of hybrid combinations. Seeds of 6 ILs and ten corresponding reciprocal F_1 hybrids were first dehulled and sterilized with 75% ethanol (1 min) and 0.15% $HgCl_2$ (10 min). The seeds were then transferred to sterilized half-strength Murashige and Skoog rooting medium (pH = 5.8) supplemented with phytogel. Ten seeds were sown on the surface of 30 mL of medium per glass tube (diameter: 3 cm and height: 20 cm). Three to six biological replicates were set up for each genotype or the same genotype at different time points. Seedlings were grown at 28 °C under 24-hour light in an artificial climate chamber (Zhu Laboratory at the College of Life Sciences, Wuhan University). The seedling length and weight of a pair of hybrid combination containing two ILs (Balilla and Yuetai B) and the corresponding reciprocal F_1 hybrids were recorded on 5, 10, and 15 days after sowing. Five seedlings from each glass tube were removed from the medium

457 and photographed for subsequent measurement of seedling length using ImageJ
458 (version 1.5.3). After excluding the roots, the seedling bases were washed three times
459 with ddH_2O and the water on the surface was removed with absorbent paper before
460 weighing. The five seedlings were snap frozen in liquid nitrogen, and photographs and
461 weights were taken and used for subsequent experiments. The remaining nine pairs of
462 hybrid combinations were harvested for seedling length measurements after 15 days of
463 growth. The sealing films on the glass tubes were removed on day 11 and 5 mL of
464 ddH_2O was added to each tube for normal growth.

465 The third batch in 2020 (Batch_2020) included the 17 ILs, 9311, and Nipponbare.
466 The germinated seeds were sown into 96-well PCR plates (bottomless), with one
467 genotype occupying half a plate. The plates were placed in plastic basins and Yoshida
468 nutrient solution (pH = 5.5, 4 L/basin) was provided for growth. Plants were grown at
469 30 ± 2 °C under a 16 h light/8 h dark photoperiod in a light culture room (State Key
470 Laboratory of Hybrid Rice, Wuhan University). The nutrient solution was replaced
471 every 3 days. On the 15th day, five plants were selected from the middle parts of each
472 row, and another ten plants were obtained from the other two rows; these plants were
473 treated as three biological replicates. The roots were discarded, and the seedling bases
474 were washed three times with ddH_2O. The seedling samples were then frozen in liquid
475 nitrogen. All the experimental samples were stored at -80 °C.

476 **RNA sequencing**

477 Seedlings from each biological replicate were ground into fine powder using mortar
478 and pestle with liquid nitrogen. After total RNA was extracted with TRIzol (Invitrogen),
479 the RNA was first assessed by electrophoresis in 1% agarose gels to check for
480 degradation and contamination states. The concentrations (Qubit2.0 Fluorometer) and
481 integrity (Agilent 2100 bioanalyzer) of each RNA sample were analyzed. DNA library
482 preparation (NEBNext Ultra RNA Library Prep Kit, NEB) with index codes and
483 sequencing experiments were performed on the sequencing platform of Novogene Co.,
484 Ltd (Beijing, China). Briefly, mRNA was first purified from 3 μg of total RNA using
485 magnetic beads attached to poly(T) oligo. The mRNA with poly(A) tails was randomly
486 fragmented (NEBNext First Strand Synthesis Reaction Buffer) and used to synthesize
487 the first strand of cDNA with the addition of hexamer primers and M-MuLV reverse
488 transcriptase. After the RNA was degraded with RNase H, the second strand of cDNA
489 was synthesized using DNA polymerase I with dNTPs. The strands were then converted

490 to blunt ends using the activities of T4 DNA polymerase and DNA polymerase I. The
491 products were subjected to adenylation, and NEBNext adaptors were added to the
492 products. Approximately 200 bp of cDNA was purified using AMPure XP beads
493 (Beckman Coulter) and treated with 3 μL of USER Enzyme (NEB) at 37 °C for 15
494 minutes and 95 °C for 5 minutes. PCR was then performed using Phusion High-Fidelity
495 DNA polymerase with universal and index primers. The reaction products were
496 repurified with AMPure XP beads, and the quality of the library was assessed using an
497 Agilent Bioanalyzer 2100. After clustering of the index-coded samples using the cBot
498 Cluster Generation System (TruSeq PE Cluster Kit v3-cBot-HS, Illumina), the libraries
499 were sequenced on the Illumina Hiseq PE150 platform, and 150 bp paired-end reads
500 were acquired with more than 6 G per sample.

501 The raw reads were first processed by removing adapters, poly N and low-quality
502 reads. The Q20, Q30, and GC contents were calculated for the clean reads. The clean
503 reads were then aligned to the Nipponbare reference genome
504 (Oryza_sativa_Ensembl_43) using Hisat2 (v2.0.5)[41] and the numbers of gene reads
505 were counted using featureCounts (v1.5.0-p3)[42].

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

507 Normalization and differential analysis of the transcriptomics data (42,341 genes) for
508 the hybrid combination of Balilla and YB at 5, 10, and 15 DAS were performed using
509 NetworkAnalyst 3.0[43] with the following parameters: specify organism, not specified;
510 data type, bulk RNA-seq data (counts); ID type, not specified; gene-level
511 summarization, sum; variance filter, 30; low abundance, 30; filter unannotated genes,
512 yes; normalization, Log2-counts per million; statistical method, DESeq2; Log2 fold
513 change, 2; and adjusted *p*-value, 0.0001. As the numbers of differentially expressed
514 genes between reciprocal F_1 hybrids of Balilla and YB at 5, 10, and 15 DAS were only
515 1, 13, and 29, respectively, the differences between reciprocal F_1 hybrids were not
516 included in the analysis. To analyse the differentially expressed genes of the six ILs,
517 the data matrix containing 42,341 genes was separately normalized (sum, none, and
518 autoscaling) with MetaboAnalyst 4.0[44] and analysed with ANOVA (adjusted *p*-value =
519 0.05, post-hoc analysis = Fisher's LSD). Principal component analysis (PCA) of the
520 transcriptomic and metabolomic data, which were normalized data from
521 NetworkAnalyst 3.0[43], was performed at ClustVis (https://biit.cs.ut.ee/clustvis)[45].

522 Two-way ANOVA of the 5,844 DEGs was performed using the "Time Series

523 Analysis" (time series and one factor) module of MetaboAnalyst 4.0[44] with the
524 following parameters: ANOVA type, type I; interaction, yes; adjusted *p*-value cut-off,
525 0.05; and correction for multiple testing, false discovery rate.

526 **Pathway enrichment analysis of the differentially expressed genes**

527 The ClueGO (version 2.5.8)[46] and CluePedia (version 1.5.8)[47] plugins of Cytoscape
528 (version 3.9.0)[48] were used for pathway enrichment analysis of the differentially
529 expressed genes. Functional analysis was selected as the analysis mode. The up- and
530 downregulated genes were represented by diamonds (shape 1) and triangles (shape 2),
531 respectively. The GO (ellipse), KEGG (hexagon), and WikiPathways (V) were selected
532 as ontologies/pathways. The GO term fusion was selected and only pathways with pV
533 ≤ 0.01 were shown. The GO tree interval was between 3 (minimum level) and 5
534 (maximum level). The minimum number and percentage of genes were set to 3 and 4%,
535 respectively, for both clusters. The genes accounting for more than 60% of a cluster
536 were defined as cluster-specific genes. The kappa score of network connectivity was
537 set to 0.4, and enrichment/depletion (two-sided hypergeometric test) with Benjamini–
538 Hochberg pV correction was set as the statistical option. GO term grouping was
539 selected in the group options, and the leading term was based on the highest
540 significance with fixed colours. The kappa score was selected for term grouping. The
541 initial group size, percentage of genes for group merging and percentage of terms for
542 group merging were set to 1, 50% and 50%, respectively.

543 **Untargeted metabolite profiling**

544 For the seedling samples in Batch_2016, liquid chromatography-mass spectrometry
545 (LC–MS) analysis procedures have been previously described[24,49]. In brief, all tissues
546 in each EP tube were first ground into homogenized powder by mortar and pestle with
547 the addition of liquid nitrogen. Eighty milligrams of powder was then transferred to 2
548 mL EP tubes and the tubes were vortexed for 1 min after the addition of 1 mL of
549 precooled extraction solution (methanol/acetonitrile/water, 2/2/1). A quality control
550 sample was prepared by mixing 10 μL of each experimental sample. The experimental
551 and quality control samples were then subjected to two rounds of ultrasonic treatments
552 (30 min each) and kept at -20 °C for 60 min. The samples were then centrifuged at
553 14,000 g for 15 min at 4 °C and the supernatants were transferred to clean tubes. After
554 drying the supernatants in a vacuum concentrator, 100 μL of acetonitrile
555 (acetonitrile/water, 1/1) was added to the tubes. The tubes were vortexed and

556 centrifuged at 14,000 g for 15 minutes at 4 °C. The supernatants were transferred to
557 sample tubes for LC–MS analysis on the platform of Shanghai Applied Protein
558 Technology Co., Ltd.

559 The experimental and quality control samples were separated on an Agilent 1290
560 Infinity LC system (Agilent Technologies) equipped with a reverse-phase T3 column
561 (ACQUITY UPLC HSS T3, 1.8 μm, 2.1 mm × 100 mm, Waters). The injection volume
562 was 2 μL per sample. The flow rate was 0.3 mL/min and the column temperature was
563 25 °C. Regents A and B in positive ion mode were 0.1% formic acid (in water) and 0.1%
564 formic acid (in acetonitrile), respectively. In addition, 0.5 mM ammonium fluoride
565 (Sigma–Aldrich) and acetonitrile (Merck) were used in negative ion mode. The gradient
566 elution steps were as follows: 0-1 min, 1% B; 1-8 min, linearly increased to 100% B;
567 8-10 min, held at 100% B; 10-10.1 min, reduced to 1% B; and 10.1-12 min, held at 1%
568 B. Experimental samples were analysed in a random order, and the quality control
569 sample was injected every nine experimental samples. To obtain MS profiles in both
570 positive and negative electron spray ion modes, the LC system was coupled to an
571 Agilent 6550 iFunnel QTOF mass spectrometer (Agilent Technologies). The MS data
572 were collected from 60 to 1,200 Da. The acquisition rate was four spectra per second,
573 and the cycle time was 250 microseconds. The temperatures of the drying gas and the
574 sheath gas were 250 °C and 400 °C, respectively. The flow rates of the drying gas and
575 the sheath gas were 16 L/min and 12 L/min, respectively. The nebulizer pressure was
576 set at 20 pounds per square inch (psi). The capillary, nozzle, and fragment voltages were
577 3,000, 0, and 175 V, respectively. Tandem mass spectrometry data were then acquired
578 on the quality control sample using a TripleTOF 6600 mass spectrometer (AB SCIEX).
579 The mass range was between 60 and 1,200 Da. To increase the data acquisition rate of
580 the quality control sample, the sample was injected twice and the mass range was
581 further divided into four sequential windows: 60-200, 190-400, 390-600, 590-1,200
582 (full range, 60-1,200) Da (for chemical standard library) and 200-600, 590-750, 740-
583 900, 890-1,200 (full range, 200-1,200) Da (for lipid MS/MS library). The mass
584 spectrometry accumulation time was 200 ms per spectrum, and the tandem mass spectra
585 were obtained in an information-dependent manner with high sensitivity. Ion source
586 gas 1, ion source gas 2, and the curtain gas were set at 40, 80, and 30 psi, respectively.
587 The source temperature was 650 °C. The ion spray voltage floating were +5,000 V and
588 -4,500 V, respectively. The declustering potential and collision energy were ± 60 V and

589 35 ± 15 eV, respectively. The number of candidate ions for monitoring per cycle was
590 ten and the isotope isolation width was 4 Da. The MS/MS accumulation time was set
591 to 50 ms.

592 The raw MS files (.d) of the experimental and quality control samples were
593 converted to mzXML format using ProteoWizard (version 3.0.6150)[50], and the
594 transformed files were uploaded to XCMS (version 1.46.0)[51] for feature detection,
595 retention time (RT) correction and feature alignment with previously reported
596 parameters[52]. The 17 ILs formed one group, and the 107 RILs plus Yuetai A formed
597 another group. The features with more than 50% nonzero values in at least one group
598 were retained for further analysis. The error thresholds for m/z and RT were set at 15
599 ppm and 20 s for the alignment of MS and MS/MS data, respectively. The MS/MS data
600 of the quality control sample were matched against both the local chemical standard
601 library[53] and the lipid MS/MS library[54], and metabolites were annotated with a mass
602 accuracy of less than 25 ppm. For the seedling samples in Batch_2018, the
603 metabolomics analyses were the same as those in Batch_2016, except that the four
604 sequential windows of the quality control sample were changed to three sequential
605 windows: 60-300, 290-600, and 590-1,200 (full range, 60-1,200) Da. The Metlin
606 library[55] was also used for feature annotation.

607 The tissues of each biological replicate of the seedling samples in Batch_2020 were
608 first ground into fine powder using a CryoMill (precooled for 2 min, Retsch) for 4 min
609 at 20 Hz. Next, 50 mg of powder was transferred to 2 mL EP tubes on ice. After adding
610 400 μL of methanol (with internal standards) to each tube and vortexing for 1 min, the
611 samples were ultrasonicated for 15 min and centrifuged at 14,000 g for 15 min at 4 °C.
612 The 200 μL supernatants were transferred to new EP tubes and dried in a vacuum
613 concentrator (Labconco). Finally, 100 μL of methanol (methanol/water, 1/4, Merck)
614 was added to the tubes to dissolve the powder for subsequent LC–MS profiling. A
615 quality control sample was prepared by pooling portions of the experimental samples
616 and subjected to the same treatments as those used for the experimental samples.
617 Methanol (methanol/water, 1/4) was used as the blank sample and was injected as the
618 first sample and every 20 samples thereafter. The quality control sample was injected
619 as the second and third samples and every 8 experimental samples thereafter. After all
620 the experimental samples were randomly analysed, the quality control sample was
621 injected at the end of the instrument run.

The experimental, quality control, and blank samples were analysed using a Vanquish UHPLC system coupled to a Q-Exactive plus mass spectrometer (Thermo Scientific) at Dalian ChemDataSolution Information Technology Co. Ltd[56]. The LC system was equipped with BEH C8 (1.7 μm, 2.1 mm × 100 mm, Waters) and HSS T3 (1.8 μm, 2.1 mm × 100 mm, Waters) columns in ESI positive and negative ion mode, respectively. The injection volume was 10 μL, the column temperature was 50 °C, and the flow rate was 0.35 mL/min in both ion modes. Water (with 0.1% formic acid) and acetonitrile (with 0.1% formic acid) were Regents A and B, respectively, in positive ion mode. The gradient elution steps were as follows: 0-1 min, 5% B; 1-24 min, linear increase to 100% B; 24.1-27.5 min, 100% B; and 27.6-30 min, decrease to 5% B. Meanwhile, water (with 6.5 mM ammonium bicarbonate) and 95% methanol (with 6.5 mM ammonium bicarbonate) were Regents A and B, respectively, in negative ion mode. The corresponding elution procedures were as follows: 0-1 min, 5% B; 1-18 min, linearly increased to 100% B; 18.1-22 min, 100% B; and 22.1-25 min, reduced to 5% B. A full scan was used to acquire the MS data and data-dependent acquisition was used to acquire the MS2 data. The ion spray voltages in positive and negative ion modes were + 3.8 kV and - 3.0 kV, respectively. The other parameters in both modes were set as follows: capillary temperature, 320 °C; aux gas heater temperature, 350 °C; sheath gas flow rate, 35 arb; aux gas flow rate, 8 arb; S-lens RF level, 50; mass range, 70-1,050 Da; full MS resolution, 70,000; MS/MS resolution, 17,500; Top N, 5; and stepped collision energy, 20, 40 eV. All metabolite profiles were acquired using Xcalibur 4.2 (Thermo Scientific). The raw data were processed by XCMS (version 3.10.1) to extract peaks with the following parameters in both ion modes: ppm = 20; fragments of MS2 = 50; snthresh = 3; prefilterK = 4; minpeakwidth = 5; method = obiwrap; time = second; mzwid = 0.015; and bw1= 15 and bw2 = 5. The maximum peak width was set to 30 and 25 in positive and negative mode, respectively. The prefilter values were 20,000 and 50,000 in positive and negative mode, respectively. SVR was chosen to correct the data. The 57 experimental samples, namely, 19 ILs with three biological replicates each, were divided into two groups. Metabolic features with zero values were removed if their percentage was greater than 20% in each group. Peaks with relative standard deviations greater than 50% in the QC samples were also removed and peak annotation was implemented using the one-map platform. The threshold for errors in RT was 10 s. The tolerances for errors in m/z of MS1 and MS2 for matching with the corresponding

655 MS were 0.005 and 0.01 Da, respectively. The lowest peak purity was set at 0.5. The
656 m/z deviation of MS2 was set to 0.005 for ion fusion. The reference noise fold and the
657 ratio of the fragment with the highest intensity were set to 1 and 0.01, respectively, for
658 filtering MS2. The reference noise of the unknown and standard MS2 to be removed
659 was 1 and 200, respectively. The minimum number of nearby fragments to merge peak
660 clusters was 2 and the error in m/z was 0.005. Three metabolite libraries, including
661 standards, KEGG, and integrated open-source databases, were used for peak annotation
662 at three MSI levels (levels 1-3). The forward scoring method was selected, and the
663 threshold of the score was set to 0.3 when searching the standards and KEGG databases.
664 The score threshold was increased to 0.4 when searching the integrated open-source
665 database. The combinations of RT-MS1-MS2, MS1-MS2, and RT-MS1 were used for
666 annotation. The RT threshold was 15 s. The m/z error thresholds for MS1 and MS2
667 were 0.01 and 0.05 Da, respectively. The maximum number of fragments to be
668 examined was 20. The top ten candidate metabolites were obtained, and the metabolite
669 with the highest score was selected for further analysis.

670 **Normalization and differential analyses of metabolomics data**

671 The normalization of the metabolomics data for the hybrid combination of Balilla and
672 YB at 5, 10, and 15 DAS followed the same parameters as those used for the
673 transcriptomics data. ANOVA was used to analyse differential metabolites for the six
674 ILs with a data matrix of 15,083 analytes. For ease of interpretation, the word
675 "metabolite" was used in this manuscript to represent analytes with or without
676 annotations. The "statistical analysis (metadata table)" module of MetaboAnalyst 5.0[57]
677 was used to perform two-way ANOVA on the 2,331 differential metabolites with
678 genotype and stage as metadata. One of the three biological replicates of each genotype
679 at each stage was used as a subject. Multiple factors were chosen for the study design.
680 Further data filtering, sample normalization, data transformation, and data scaling were
681 not performed. Type I ANOVA was used for two-way independent ANOVA, and two-
682 way repeated ANOVA was used to calculate interactions. The false discovery rate was
683 used to correct for multiple testing, and the adjusted *P* value cut-off was set at 0.05.

684 To normalize the metabolomics data of the 17 ILs, Yuetai A, and 107 RILs, the
685 intensities of the two biological replicates of the 17 ILs were first averaged and then
686 normalized using MetaboAnalyst 5.0[57] (sum, none, and autoscaling for sample
687 normalization, data transformation, and data scaling). Metabolites with RSD greater

688 than 25% in the QC samples were removed and the interquantile range was selected to
689 filter the features (40%). For the differential metabolite analysis of the 17 ILs, the
690 feature intensities of two biological replicates for each IL were used and ANOVA was
691 performed after filtering 40% of the features and data normalization (sum, none, and
692 autoscaling) using MetaboAnalyst 5.0[57]. For the normalization of the 19 ILs, sum, none,
693 and autoscaling were also selected. As three metabolic profiles of the QC sample were
694 not acquired, there were only seven metabolic profiles of the QC sample in negative
695 mode. There were ten metabolic profiles of the QC sample in positive mode. To merge
696 the metabolic profiles detected in both positive and negative modes, three metabolic
697 profiles of the QC sample, namely QC103.2, QC2, and QC63, were not merged. The
698 remaining seven QC samples were merged accordingly, namely, QC101 to QC301,
699 QC102 to QC302, QC103 to QC303, QC 104 to QC304, QC201 to QC401, QC203 to
700 QC402, and QC204 to QC403. For the metabolite levels of the 19 ILs in the regression
701 analysis, the obtained intensities of three biological replicates were first averaged and
702 then normalized.

703 **Determination of inheritance patterns for heterosis-associated molecules**

704 To determine the inheritance patterns for heterosis-associated molecules,
705 transcriptomic and metabolic levels were compared among F_1 hybrids, the
706 corresponding parents, and the average levels of the corresponding parents via ANOVA
707 using MetaboAnalyst 4.0[44]. Three biological replicates of reciprocal F_1 hybrids from
708 two ILs were combined into a hybrid genotype. An additive effect was assigned when
709 the molecular levels of the hybrid genotype were not significantly different from the
710 average levels of the corresponding parents but were significantly different from those
711 of both parents. A partially dominant effect was assigned when the hybrid levels were
712 significantly different from both the average parental levels and the molecular levels of
713 the parent on the same side. A dominant effect was when the hybrid levels were not
714 significantly different from those of the parent but were significantly different from the
715 average levels of the two parents. An overdominant effect was the pattern when the
716 hybrid levels were significantly higher or lower than those of the parent on the same
717 side. The patterns of hybrid levels other than the above four types were defined as
718 unclassified and marked as NA. To quantitatively describe the partially dominant effect
719 (PD), the degrees for the molecules with partially dominant effect of the six hybrid
720 combinations was calculated with the equation $PD = (F_1-P_2)/(P_1-P_2)$, where F_1, P_1, and

721 P_2 represent the molecular levels of the F_1 hybrids, parent 1, and parent 2, respectively.
722 Several molecules that had a partially dominant effect in multiple F_1 hybrids were
723 selected multiple times. The degree of the partially dominant effect was divided into
724 two parts, with 0.5 as the threshold. The averages of the two groups were calculated
725 and defined as two types of partially dominant effects (PD_0.3 and PD_0.7). Moreover,
726 the equations PD_0.3 = $P_m+0.3\times(P_f-P_m)$ and PD_0.7 = $P_m+0.7\times(P_f-P_m)$ were used to
727 represent partially dominant effects based on parental molecular levels.

728 **Pathway enrichment analysis of differential metabolites**

729 The metabolomics data of the hybrid combination of Balilla and YB at 15 DAS were
730 first used for pathway enrichment analysis. The table with normalized feature
731 intensities and retention times (in seconds) was uploaded to the "functional analysis"
732 module of MetaboAnalyst 5.0[57]. Both positive and negative modes were selected for
733 analysis. In positive mode, the mass tolerance was set to 5 ppm. The interquantile range
734 was selected for data filtering because the number of peaks was greater than 5,000. No
735 data were used for sample normalization, data transformation, or data scaling. Both
736 mummichog[28] (*P* value cut-off: default top 10% peaks, version 2.0) and GSEA were
737 used for enrichment analysis. As no pathway was significantly enriched for the
738 differential metabolites between Balilla×Yuetai B and its parent Balilla with the default
739 *P* value cut-off (0.01) in positive mode, the *P* value cut-off was increased to 0.05. The
740 pathway library used was *Oryza sativa japonica* (Japanese rice) [KEGG]. Metabolite
741 currency and adducts were set with the default sets, and pathways or metabolite sets
742 containing at least 3 entries were set. The parameters for the negative mode analysis
743 were the same as those for the positive mode analysis, except for that no data filtering
744 was performed and the *P* value cut-off was set to 0.05. Putatively annotated metabolites
745 and enriched pathways were obtained in each analysis. For the pathway analysis of the
746 645 DMs using the "Pathway Analysis" module of MetaboAnalyst 5.0[57], feature
747 annotation information was obtained from the results of differential metabolites
748 between Balilla and YB at 15 DAS. The parameters were set as follows: scatter plot,
749 visualization method; hypergeometric test, enrichment method; relative-betweenness
750 centrality, topology analysis; all compounds in the selected pathway library, reference
751 metabolome; and *Oryza sativa japonica* (Japanese rice) (KEGG), pathway library.
752 Three pathways were significantly enriched based on the annotation of eight peaks in
753 positive mode. Moreover, 20 metabolites (23 peaks from both ion modes), which were

754 annotated with local chemical standards, lipids, and Metlin libraries, from the three
755 significantly enriched pathways were also used for analysis. Correlations between the
756 metabolite levels of 26 peaks and seedling length were determined. One peak
757 (M147T118_POS) was annotated as 4-hydroxycinnamic acid in both the local standard
758 library and the result of the pathway enrichment analysis, and four peaks had no
759 concentration in the normalized metabolomics data.

760 **Integration of genes and metabolites into networks**

761 The normalized transcriptomics and metabolomics data were used to construct
762 networks via OmicsAnalyst[58]. No organism or ID type was specified in the annotation
763 procedure. Features with 50% missing values were excluded and the remaining missing
764 values were replaced by LoDs. The filter threshold for variance and abundance was set
765 to 0. In the comparison step, the genotype was used for grouping, and no data
766 transformation was performed. ANOVA was used for the differential analysis. The fold
767 change cut-off was set at 2.0, and the *P* value cut-off was set at 0.01. To make the data
768 suitable for inter-omics analysis, both the transcriptomics and metabolomics data
769 underwent normalization by sum, variance stabilizing normalization, quantile
770 normalization and range scaling. The statistically significant features were then used to
771 construct an interomics correlation network using the Spearman method. The
772 correlation threshold was set at 0.8, and both positive and negative correlations were
773 included. For Network_51015, 626 genes and 645 metabolites were used, and a total of
774 213 genes and 198 metabolites were statistically significant. The network contained
775 three subnetworks, and the largest subnetwork with 2,138 edges (subnetwork_1) had
776 211 genes and 196 metabolites as nodes. The other two subnetworks had two
777 interactions. For Network_HLN, 967 genes and 990 metabolites were used for analysis,
778 and 103 genes and 158 metabolites were retained in the network. Network_HLN
779 contained seven subnetworks, and the largest subnetwork with 868 edges
780 (subnetwork_1) contained 97 genes and 149 metabolites. The remaining six
781 subnetworks had nine edges. The walktrap algorithm was used for module detection.
782 Subnetwork_1 of Network_51015 contained seven modules. The module size and *P*
783 value of each module were 150 ($P = 2.34\times10^{-21}$, Module_1), 105 ($P = 1.01\times10^{-14}$,
784 Module_2), 80 ($P = 2.29\times10^{-12}$, Module_3), 21 ($P = 0.00303$, Module_4), 13 ($P = 0.452$,
785 Module_5), 7 ($P = 1$, Module_6), and 6 ($P = 0.318$, Module_7), respectively.
786 Subnetwork_1 of Network_HLN contained eight modules. The module size and *P*

787 value of each module were 112 (P = 6.6×10^{-35}, Module_1), 37 (P = 2.61×10^{-7},
788 Module_2), 35 (P = 1.39×10^{-6}, Module_3), 16 (P = 3.8×10^{-5}, Module_4), 11 (P = 0.111,
789 Module_5), 7 (P = 0.0592, Module_6), 6 (P = 0.0176, Module_7), and 6 (P = 0.0309,
790 Module_8), respectively. The subnetworks were imported into Cytoscape[48] for display
791 using the compound spring embedder layout. The node size was determined by the node
792 degree.

793 **Joint omics analysis of differentially expressed genes and metabolites**

794 To integrate the transcriptomics and metabolomics results, the MSU IDs of the 626
795 DEGs were first converted to gene symbols using the NCBI "Batch Entrez" tool
796 (https://www.ncbi.nlm.nih.gov/sites/batchentrez). Genes with gene symbols in the
797 *Oryza sativa japonica* group (340 genes) were uploaded to the "Joint pathway" analysis
798 module of MetaboAnalyst 5.0[57]. Moreover, the annotation information of the 645 peaks
799 was extracted from the results of the pathway enrichment analysis between Balilla and
800 YB at 15 DAS. The KEGG IDs for 29 compounds in positive mode and 44 compounds
801 in negative mode were uploaded as metabolomics data. "All pathways (integrated)"
802 was selected as the pathway database. The parameters for enrichment analysis,
803 topology measure, and integration method were set as hypergeometric test, degree
804 centrality, and combine queries, respectively.

805 **Analyses of phenotypic and transcriptomic/metabolic data for inbred lines and F_1**
806 **hybrids in maize**

807 The field fresh weight, grain dry matter yield, and grain dry matter content were
808 obtained from published data in two previous studies[29,30]. Metabolite profiles of 3.5-
809 day-old roots and leaves at the third leaf stage were also collected for 210 Dent and
810 Flint lines and 363 F_1 hybrids in the two studies[29,30]. The root and leaf metabolic
811 datasets contained levels of 165 and 81 metabolites, respectively. Since metabolic data
812 were not available for the parents of some of the F_1 hybrids, the field fresh weight of
813 326 F_1 hybrids, which were crossed between 24 Dent lines and 25 Flint lines, were used
814 for analysis. No further data normalization was performed on the metabolic data. As
815 reported in the study[29], the field fresh weight of the 326 F_1 hybrids was recorded in two
816 distinct field tests in 2010 (146 F_1 hybrids) and 2012 (180 F_1 hybrids). In the PLS
817 analysis, the number was set to 100 to ensure a sufficient number of latent factors, and
818 the number was defined for each regression analysis when the R^2 value (adjusted)
819 peaked. The R^2 values (adjusted) corresponding to different numbers of latent factors

820 and the greatest R^2 values for hybrid profiles and different inheritance patterns were
821 used for comparison analysis via ANOVA. Among the 326 F_1 hybrids, 291 F_1 hybrids
822 had recorded data on grain dry matter yield and grain dry matter content in
823 approximately seven agro-ecologically diverse environments[30]. The root metabolic data
824 were used for PLS analysis, and the R^2 values (adjusted) were used for comparison
825 analysis.

826 The grain yield per plant (in three environments) of 12 maize inbred lines and 66
827 F_1 hybrids from the inbred lines with a half-diallel design reported in a previous study[31]
828 was used for analysis. The expression levels of 27,608 genes, which were pan and core
829 genes in Zheng58, from the 12 inbred lines were also obtained from the study[31]. As
830 described in the study[31], each genotype of the 12 maize ILs had two sets (not biological
831 replicate) of RNA-seq data, with one set containing equally mixed RNAs from eight
832 tissues and the other from five tissues at different developmental stages. The expression
833 data were processed with MetaboAnalyst 5.0[57], with 40% set as the percentage for
834 filtering. No data transformation, normalization, or scaling was performed. In total, 84
835 DEGs were identified in the 12 inbred lines by ANOVA, with Fisher's LSD used in post
836 hoc analysis (*P* value cut-off = 0.05). The average values of the two sets of RNA
837 sequencing data were used to represent the levels of each inbred line. The 84 genes
838 were classified according to the previously reported results[31]. The number of latent
839 factors was set to 50 in the PLS analysis of better-parent heterosis for grain yield per
840 plant, middle parent heterosis for grain yield per plant, and grain yield per plant.

841 **Genome sequencing and analysis of genomic data**

842 Seedlings from the 17 ILs were ground into fine powder using liquid nitrogen. The
843 genomic DNA was extracted, and the quality and concentration were assessed by 1%
844 agarose gel electrophoresis and Qubit 3.0 fluorometer. The DNA libraries (with index
845 codes) were prepared and sequenced on the Illumina sequencing platform of Novogene
846 Co., Ltd (Beijing, China). In brief, genomic DNA was randomly fragmented into ~350
847 bp size via sonication (Covaris). The strands were then added with adapters and used
848 for PCR amplification. Products were purified with AMPure XP beads (Beckman
849 Coulter) and the concentrations were measured with Agilent 2100 Bioanalyzer.
850 Libraries were sequenced on the Illumina PE150 platform, and the raw data were
851 recorded in FASTQ format (> 15 G per sample). After the clean reads were obtained,
852 the sequences were aligned to the reference genome using the Burrows–Wheeler

853 Aligner (mem -t 4 -k 32 -M, version 0.7.8-r455)[59]. Dislodged duplications in the BAM
854 files were removed via SAMtools (rmdup, version 1.3.1)[60]. SNPs and InDels in each
855 sample were called with the parameters "-C 50 - mpileup -m 2 -F 0.002 -d 1000" and
856 further filtered to reduce errors (depth of variant > 4 and mapping quality > 20). CNVs
857 were detected by CNVnator (-call 100, version 0.3)[61], and SVs were detected by
858 BreakDancer (depth of variant > 2, version 1.4.4)[62]. Variant annotation was
859 implemented with ANNOVAR[63]. Genomic variants on chromosomes were visualized
860 by Circos (0.69-9)[64]. To analyse the population structures of the 17 ILs, GATK (version
861 4.2.0.0) was first used to filter SNPs (QD < 2, FS > 60, MQ < 40, MQRankSum < -
862 12.5, ReadPosRankSum < -8) and InDels (QD < 2, FS > 200, ReadPosRankSum < -
863 20). SNPs and InDels were further filtered with VCFtools (maximum missing value <
864 10%, minus allele: 0.05, minimum read depth > 4, multiple alleles removed, version
865 0.1.17) and analysed with Plink (version 1.90) and Admixture (version 1.3.0). When
866 comparing the SNPs and InDels between each pair of samples of the 17 ILs, the variants
867 were defined as common variants if they shared the same values for five parameters,
868 including CHROM (chromosome), POS (position of the reference), REF (reference
869 base), ALT (alternative base) and GT (genotype). Differential variants were those that
870 were unique to each IL. For CNVs and SVs, variants were defined as common variants
871 if they shared the same values of CHROM, POS1, POS2, TYPE and SIZE. The
872 numbers of common and unique variants were counted, and the sums of unique
873 genomic variants between each pair of ILs were used to represent the genetic distance.
874 Phylogenetic trees (neighbour-joining method) based on one or multiple types of
875 genomic variants in the 17 ILs were generated with MEGA (version 10.2.6)[65].

876 **Statistical analyses**

877 Analysis of variance using SPSS 20.0 was performed with the least significant
878 difference used in the post hoc test to compare the seedling length of the F_1 hybrids and
879 corresponding parents in Fig. 1a, the seedling length of the F_1 hybrids and
880 corresponding better parent in Fig. 2a, the predictabilities in Fig. 2q, the heterosis for
881 plant height and tiller number at four developmental stages in Fig. 3c, the R^2 values of
882 heterosis for yield per plant with different inheritance patterns in Fig. 3g, the greatest
883 R^2 values of heterosis for 17 traits with different inheritance patterns in Fig. 3h, the
884 greatest R^2 values of heterosis for eight traits with different inheritance patterns in Fig.
885 3j, the greatest R^2 values of heterosis for 17 traits with different inheritance patterns in

886 Supplementary Fig. 8u, the numbers (weighted) of correlations for four types of
887 genomic variants in Fig. 4m and Extended Data Fig. 10i, and correlation numbers
888 (weighted) of the three inheritance patterns in Extended Data Figs. 10j-l. An
889 independent samples *t* test was performed to compare the numbers (weighted) of
890 differentially expressed genes between F_1 hybrids and corresponding parents in
891 Extended Data Fig. 1c, percentages of differential molecules between F_1 hybrids and
892 corresponding parents in parental differences in Supplementary Fig. 1j, the node
893 numbers (weighted), predictability, and percentage of Component 1 of three modules
894 from Subnetwork_1 between Network_51015 and Network_HLN in Figs. 2e-g, the
895 numbers (weighted) of correlations between five metabolites from phenylpropanoid
896 biosynthesis and heterosis in Figs. 3f and Extended Data Fig. 6h, trait values and
897 heterosis for four traits between the diallel and test cross populations in Extended Data
898 Fig. 7b and Extended Data Fig. 7f, and the numbers (weighted) of correlations for four
899 types of genomic variants in Fig. 4i. Spearman rank correlation was performed between
900 the numbers of genetic variants and heterosis in Fig. 4h, Extended Data Figs. 10a-f,
901 Extended Data Fig. 10h, and Extended Data Fig. 10j-k. Pearson correlation analysis
902 (two-tailed) with significance level set at 0.05 was conducted between gene expression
903 (1,201 genes)/metabolite (1,249 metabolites) levels and seedling length to screen for
904 heterosis-associated molecules in Extended Data Fig. 2j and Extended Data Fig. 3h,
905 between seedling length and heterosis-associated metabolites from three enriched
906 metabolic pathways in Supplementary Fig. 2a, among better-parent heterosis for 17
907 traits in Fig. 3d, between three types of inheritance patterns of five metabolites and
908 heterosis in Fig. 3e, and between three types of inheritance patterns for the 50 top
909 contributing analytes and heterosis in Extended Data Fig. 10i. Partial least squares
910 regression between gene expression or metabolite levels and seedling length in
911 Extended Data Fig. 2j and Extended Data Fig. 3h, between parental differential
912 metabolites and heterosis in Figs. 3g-j, Extended Data Figs. 6i-x, and Supplementary
913 Figs. 7a-h were implemented with SPSS. The maximum number of latent factors was
914 initially set at 100 to extract enough latent factors. The regression model for each trait
915 was determined when the value of the adjusted R^2 was the highest. The means of
916 variable importance in the projection of all latent factors were used to estimate the
917 contributions of each variable, and threshold was set to 1 in Fig. 2h. Partial least
918 squares-discriminant analysis (PLS-DA) in Fig. 2i, Fig. 2o, and Fig. 2q was performed

with class order taken into account in MetaboAnalyst5.0[57]. The maximum number of components to be searched was 5 and tenfold cross validation was chosen. The separation distance (B/W) method was used for permutation analysis (2,000 times) to assess whether the model was overfitting. Q^2 was chosen as the predictability in PLS-DA models. Venn diagrams were drawn with the tool from http://bioinformatics.psb.ugent.be/webtools/Venn/. The phylogenetic tree of the 223 ILs shown in Fig. 3b was based on genotypes and the tree was constructed using MEGA[65] via the neighbour-joining method. The *indica* genotype, heterozygous genotype, and *japonica* genotype based on InDel markers were represented by 3, 2, and 1, respectively. A similarity matrix of the 223 ILs was then derived from MORPHEUS (https://software.broadinstitute.org/morpheus/) using Euclidean distance. For each box generated with BoxPlotR[66] (http://shiny.chemgrid.org/boxplotr/), the central line represents the median, the upper and lower edges indicate the 75th and 25th percentiles, respectively, and the whiskers are 1.5× the interquartile range away from the 75th and 25th percentiles.

Data availability

The genomics and transcriptomics data reported in this study have been deposited in the NCBI database under BioProject accession PRJNA1122958. The metabolomics data have been deposited in the MetaboLights database under MTBL742[24], MTBL1343, and MTBL1344. Source data are available with this paper.

Code availability

The software as described in the Methods section in the study is publicly available.

References

1. Birchler, J.A., Yao, H. & Chudalayandi, S. Unraveling the genetic basis of hybrid vigor. *Proc. Natl. Acad. Sci. USA* **103**, 12957-12958 (2006).
2. Shull, G.H. What is "heterosis"? *Genetics* **33**, 439-446 (1948).
3. Birchler, J.A., Yao, H., Chudalayandi, S., Vaiman, D. & Veitia, R.A. Heterosis. *Plant Cell* **22**, 2105-2112 (2010).
4. Darwin, C.R. *The effects of cross and self fertilization in the vegetable kingdom. John Murray, London, UK* (1876).
5. Hochholdinger, F. & Baldauf, J.A. Heterosis in plants. *Curr. Biol.* **28**, R1089-R1092 (2018).

6. Bruce, A.B. The Mendelian theory of heredity and the augmentation of vigor. *Science* **32**, 627-628 (1910).
7. Jones, D.F. Dominance of linked factors as a means of accounting for heterosis. *Proc Natl Acad Sci U S A* **3**, 310-312 (1917).
8. Li, X., Li, X., Fridman, E., Tesso, T.T. & Yu, J. Dissecting repulsion linkage in the dwarfing gene *Dw3* region for sorghum plant height provides insights into heterosis. *Proc. Natl. Acad. Sci. USA* **112**, 11823-11828 (2015).
9. Davenport, C.B. Degeneration, albinism and inbreeding. *Science* **28**, 454-455 (1908).
10. Xiao, Y. *et al.* The genetic mechanism of heterosis utilization in maize improvement. *Genome Biol.* **22**, 148 (2021).
11. East, E.M. Heterosis. *Genetics* **21**, 375-397 (1936).
12. Crow, J.F. Alternative hypotheses of hybrid vigor. *Genetics* **33**, 477–487 (1948).
13. Yu, S.B. *et al.* Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. USA* **94**, 9226-9231 (1997).
14. Li, Z., Pinson, S., Park, W., Paterson, A. & Stansel, J. Epistasis for three grain yield components in rice (*Oryza sativa* L.). *Genetics* **145**, 453-465 (1997).
15. Powers, L. An expansion of Jones's theory for the explanation of heterosis. *Am. Nat.* **78**, 275-280 (1944).
16. Schnable, P.S. & Springer, N.M. Progress toward understanding heterosis in crop plants. *Annu. Rev. Plant Biol.* **64**, 71-88 (2013).
17. Lisec, J. *et al.* Corn hybrids display lower metabolite variability and complex metabolite inheritance patterns. *Plant J.* **68**, 326-336 (2011).
18. Williams, W. Heterosis and the genetics of complex characters. *Nature* **184**, 527-530 (1959).
19. Dan, Z. *et al.* Hierarchical additive effects on heterosis in rice (*Oryza sativa* L.). *Front. Plant Sci.* **6**, 738 (2015).
20. Krieger, U., Lippman, Z.B. & Zamir, D. The flowering gene *SINGLE FLOWER TRUSS* drives heterosis for yield in tomato. *Nat. Genet.* **42**, 459-463 (2010).
21. Wang, Z., Xue, Z. & Wang, T. Differential analysis of proteomes and metabolomes reveals additively balanced networking for metabolism in maize heterosis. *J. Proteome Res.* **13**, 3987-4001 (2014).
22. Zhou, P., Hirsch, C.N., Briggs, S.P. & Springer, N.M. Dynamic patterns of gene expression additivity and regulatory variation throughout maize development. *Mol. Plant* **12**, 410-425

983 (2019).
984 23. Dan, Z. *et al.* The metabolomic landscape of rice heterosis highlights pathway biomarkers
985 for predicting complex phenotypes. *Plant Physiol.* **187**, 1011-1025 (2021).
986 24. Dan, Z., Chen, Y., Zhao, W., Wang, Q. & Huang, W. Metabolome-based prediction of yield
987 heterosis contributes to the breeding of elite rice. *Life Sci. Alliance* **3**, e201900551 (2020).
988 25. Dan, Z. *et al.* A metabolome-based core hybridisation strategy for the prediction of rice
989 grain weight across environments. *Plant Biotechnol. J.* **17**, 906-913 (2019).
990 26. Dan, Z. *et al.* Metabolic prediction of important agronomic traits in hybrid rice (*Oryza*
991 *sativa* L.). *Sci. Rep.* **6**, 21732 (2016).
992 27. Sun, Z. *et al.* Dissecting the genetic basis of heterosis in elite super-hybrid rice. *Plant*
993 *Physiol.* **192**, 307-325 (2023).
994 28. Li, S. *et al.* Predicting network activity from high throughput metabolomics. *PLoS Comput.*
995 *Biol.* **9**, e1003123 (2013).
996 29. de Abreu, E.L.F. *et al.* Metabolic robustness in young roots underpins a predictive model
997 of maize hybrid performance in the field. *Plant J.* **90**, 319-329 (2017).
998 30. Schrag, T. *et al.* Beyond genomic prediction: combining different types of *omics* data can
999 improve prediction of hybrid performance in maize. *Genetics* **208**, 1373-1385 (2018).
1000 31. Wang, B. *et al.* De novo genome assembly and analyses of 12 founder inbred lines
1001 provide insights into maize heterosis. *Nat. Genet.* **55**, 312-323 (2023).
1002 32. Dan, Z. *et al.* Balance between a higher degree of heterosis and increased reproductive
1003 isolation: a strategic design for breeding inter-subspecific hybrid rice. *PLoS ONE* **9**,
1004 e93122 (2014).
1005 33. Huang, X. *et al.* Genomic architecture of heterosis for yield traits in rice. *Nature* **537**, 629-
1006 633 (2016).
1007 34. Price, N.D. *et al.* A wellness study of 108 individuals using personal, dense, dynamic data
1008 clouds. *Nat. Biotechnol.* **35**, 747-756 (2017).
1009 35. Zhang, M. *et al.* Heterosis in root microbiota inhibits growth of soil-borne fungal pathogens
1010 in hybrid rice. *J. Integr. Plant Biol.* **65**, 1059-1076 (2023).
1011 36. Xu, Y. *et al.* Incorporation of parental phenotypic data into multi-omic models improves
1012 prediction of yield-related traits in hybrid rice. *Plant Biotechnol. J.* **19**, 261-272 (2021).
1013 37. Kakoulidou, I. *et al.* Mapping parental DMRs predictive of local and distal methylome
1014 remodeling in epigenetic F1 hybrids. *Life Sci. Alliance* **7**(2024).
1015 38. Liu, N., Du, Y., Warburton, M.L., Xiao, Y. & Yan, J. Phenotypic plasticity contributes to

1016 maize adaptation and heterosis. *Mol Biol Evol* **38**, 1262-1275 (2021).
1017 39. Lu, B., Cai, X. & Jin, X. Efficient *indica* and *japonica* rice identification based on the InDel
1018 molecular method:its implication in rice breeding and evolutionary research. *Progress in*
1019 *Nature Science* **19**, 1241-1252 (2009).
1020 40. Yao, G. & Huang, W. Identifying *indica-japonica* characteristic of rice breeding materials
1021 with *indica-japonica* differentiation InDel markers. *Hybrid Rice* **28**, 53-57 (2013).
1022 41. Kim, D., Langmead, B. & Salzberg, S.L. HISAT: a fast spliced aligner with low memory
1023 requirements. *Nat. Methods* **12**, 357-60 (2015).
1024 42. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for
1025 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-30 (2014).
1026 43. Zhou, G. *et al.* NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene
1027 expression profiling and meta-analysis. *Nucleic Acids Res.* **47**, W234-W241 (2019).
1028 44. Chong, J. *et al.* MetaboAnalyst 4.0: towards more transparent and integrative
1029 metabolomics analysis. *Nucleic Acids Res.* **46**, W486-W494 (2018).
1030 45. Metsalu, T. & Vilo, J. ClustVis: a web tool for visualizing clustering of multivariate data
1031 using Principal Component Analysis and heatmap. *Nucleic Acids Res.* **43**, W566-70
1032 (2015).
1033 46. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene
1034 ontology and pathway annotation networks. *Bioinformatics* **25**, 1091-3 (2009).
1035 47. Bindea, G., Galon, J. & Mlecnik, B. CluePedia Cytoscape plugin: pathway insights using
1036 integrated experimental and in silico data. *Bioinformatics* **29**, 661-3 (2013).
1037 48. Saito, R. *et al.* A travel guide to Cytoscape plugins. *Nat. Methods* **9**, 1069-76 (2012).
1038 49. Chen, Y., Dan, Z. & Li, S. *GROWTH REGULATING FACTOR 7*-mediated arbutin
1039 metabolism enhances rice salt tolerance. *Plant Cell* (2024).
1040 50. Chambers, M.C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat.*
1041 *Biotechnol.* **30**, 918-920 (2012).
1042 51. Gowda, H. *et al.* Interactive XCMS Online: simplifying advanced metabolomic data
1043 processing and subsequent statistical analyses. *Anal. Chem.* **86**, 6931-9 (2014).
1044 52. Jia, H. *et al.* Predicting the pathological response to neoadjuvant chemoradiation using
1045 untargeted metabolomics in locally advanced rectal cancer. *Radiother Oncol* **128**, 548-
1046 556 (2018).
1047 53. Shen, X. *et al.* Metabolic reaction network-based recursive metabolite annotation for
1048 untargeted metabolomics. *Nat. Commun.* **10**, 1516 (2019).

1049 54. Tu, J., Yin, Y., Xu, M., Wang, R. & Zhu, Z.J. Absolute quantitative lipidomics reveals
1050 lipidome-wide alterations in aging brain. *Metabolomics* **14**, 5 (2017).
1051 55. Domingo-Almenara, X. *et al.* XCMS-MRM and METLIN-MRM: a cloud library and public
1052 resource for targeted analysis of small molecules. *Nat. Methods* **15**, 681-684 (2018).
1053 56. Zhao, X. *et al.* Comprehensive Strategy to Construct In-House Database for Accurate and
1054 Batch Identification of Small Molecular Metabolites. *Anal. Chem.* **90**, 7635-7643 (2018).
1055 57. Pang, Z. *et al.* Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics
1056 integration and covariate adjustment of global metabolomics data. *Nat. Protoc.* (2022).
1057 58. Zhou, G., Ewald, J. & Xia, J. OmicsAnalyst: a comprehensive web-based platform for
1058 visual analytics of multi-omics data. *Nucleic Acids Res.* (2021).
1059 59. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
1060 transform. *Bioinformatics* **25**, 1754-60 (2009).
1061 60. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-
1062 9 (2009).
1063 61. Abyzov, A., Urban, A.E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover,
1064 genotype, and characterize typical and atypical CNVs from family and population genome
1065 sequencing. *Genome Res.* **21**, 974-84 (2011).
1066 62. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic
1067 structural variation. *Nat. Methods* **6**, 677-81 (2009).
1068 63. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants
1069 from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
1070 64. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome*
1071 *Res.* **19**, 1639-45 (2009).
1072 65. Stecher, G., Tamura, K. & Kumar, S. Molecular Evolutionary Genetics Analysis (MEGA)
1073 for macOS. *Mol Biol Evol* **37**, 1237-1239 (2020).
1074 66. Spitzer, M., Wildenhain, J., Rappsilber, J. & Tyers, M. BoxPlotR: a web tool for generation
1075 of box plots. *Nat. Methods* **11**, 121-2 (2014).

1076 Acknowledgements

1077 The authors thank members of the Zhu lab for assistance in the collection of field
1078 phenotypic data. The authors thank Professor Baorong Lu (Fudan University) for the
1079 design of the diallel cross population. The authors thank Mr. Caibin Zhu for the fund of
1080 rice research. The authors thank Fulin Luo for the planting of rice. This work is

supported by National Natural Science Foundation of China (31771746 to W.H., 31801439 to Z.D.), the China Postdoctoral Science Foundation (2018M632910, 2019M660186, and 2022T150500 to Z.D.), 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..

Author contributions

Z. D. conceived the project and performed most part of the hybridizations. W. H. collected rice inbred lines. Z. D. and Yunping Chen collected phenotypic data of rice seedlings, prepared seedling samples for transcriptomics and metabolomics experiments, and analysed the omics data. Z. D., W. Z. and G. Y. constructed the diallel cross population and recorded phenotypic data of the hybrid population in 2012. Z. D. and W. H. screened the recombinant inbred lines and constructed the test cross population. Z. D., Y. X., J. H. and J. M. collected phenotypic data of the hybrid population in 2015. G. Y. performed the analysis of 223 inbred lines with InDel markers. Yi Chen participated in the construction of phylogenetic tree for the 223 inbred lines and the preparation of several heatmaps. Z. D., Yunping Chen, and W. H. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information is available for this paper. Correspondence and requests for materials should be addressed to Wenchao Huang (wenchaoh@whu.edu.cn).

Extended figure legends

Extended Data Fig. 1 Transcriptomic differences in two inbred lines (Balilla and Yuetai B) and their reciprocal F_1 hybrids at three developmental stages.

Extended Data Fig. 2 The identification of heterosis-associated genes related to seedling length from the differentially expressed genes between F_1 hybrids and their corresponding parents.

Extended Data Fig. 3 The identification of heterosis-associated metabolites related to seedling length from differential metabolites between F_1 hybrids and their corresponding parents.

Extended Data Fig. 4 Inheritance patterns of heterosis-associated molecules for seedling length.

Extended Data Fig. 5 Phenotypes of seventeen agronomic traits of the diallel cross population.

Extended Data Fig. 6 Predictive models for heterosis of 16 traits based on three types of inheritance patterns of 1,306 differential metabolites.

Extended Data Fig. 7 Cross design and four traits of the test cross population.

Extended Data Fig. 8 Additive and partially dominant effects contribute to maize heterosis.

Extended Data Fig. 9 Four types of genomic variants of 16 ILs.

Extended Data Fig. 10 Correlations of genomic variants, heterosis, and different inheritance patterns.

Figures and figure legends

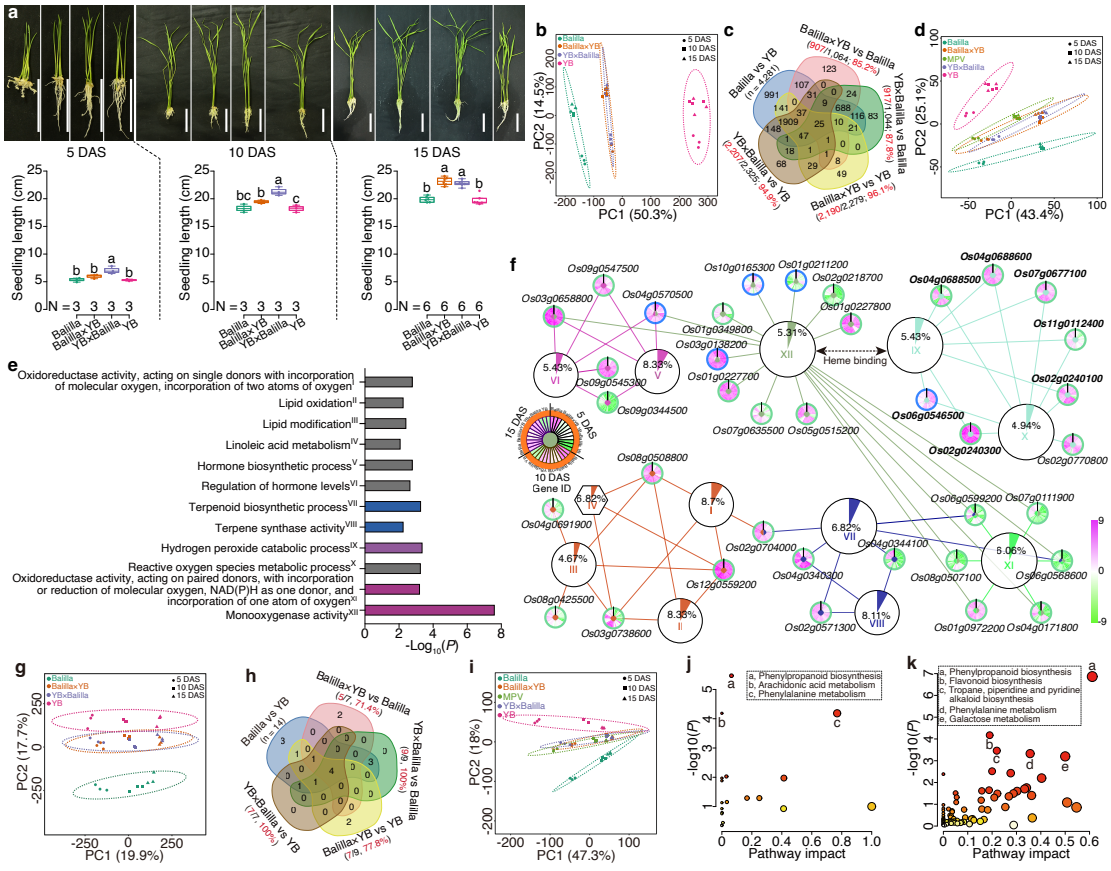

Fig. 1 Heterosis-associated genes and metabolites related to rice seedling length. a, Seedlings and seedling length of Balilla, Yuetai B (YB), and their reciprocal F_1 hybrids at 5, 10, and 15 days after sowing (DAS). The significance level was set at 0.05. Scale bar = 5 cm. **b**, PCA score plot of Balilla, YB, and their F_1 hybrids using the normalized expression levels of 22,194 genes. **c**, Venn diagram of differentially expressed genes (DEGs) at 15 DAS. The numbers in red represent the number of DEGs that are parental DEGs. **d**, PCA score plot of Balilla, YB, and their F_1 hybrids based on the expression levels of 626 heterosis-associated genes (HAGs). **e**, Gene Ontology analysis of 626 HAGs. The significantly enriched terms with colours are from parental differential terms. **f**, Network of 37 HAGs from 12 terms. The expression levels of each gene in the four genotypes at three stages are shown and the percentages of genes associated with a term are indicated. Genes with up- and downregulated expression levels in F_1 hybrids are marked with green and blue outer circles, respectively. **g,** PCA score plot of Balilla, YB, and their F_1 hybrids based on normalized metabolite levels of 10,522 metabolites. **h**, Venn diagram of differential metabolic pathways in positive ion mode. The numbers in red are the number of pathways among the parental differential pathways. **i**, PCA score plot based on 645 heterosis-associated metabolites (HAMs). **j**, Three significantly enriched pathways of HAMs in positive ion mode. **k**, Five significantly enriched pathways integrating heterosis-associated genes and metabolites.

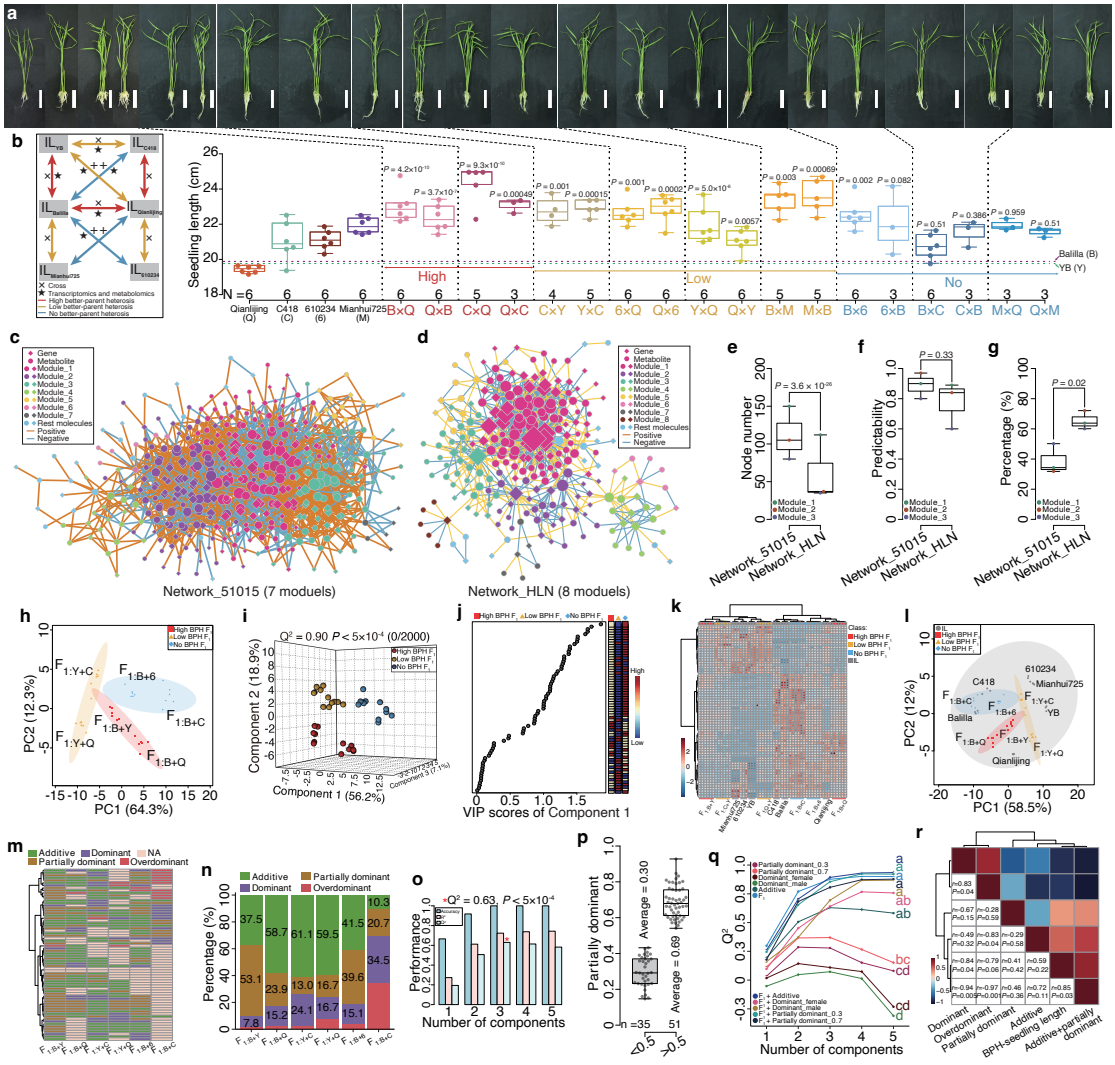

Fig. 2 Inheritance patterns of heterosis-associated genes and metabolites. a, Seedlings and seedling length of nine pairs of reciprocal F_1 hybrids. *P* values indicate the least significant difference in the post hoc test for the seedling length of the F_1 hybrid compared with that of its better parent. **b,** Cross design of six inbred lines (ILs). **c-d,** Subnetworks integrating heterosis-associated genes and metabolites. Subnetwork_1 of Network_51015 (**c**) and that of Network_HLN (**d**) contain 2,138 and 868 edges, respectively. The node size is determined by the node degree. **e-g,** Comparisons of Subnetwork_1 from Network_51015 and Network_HLN. Node number (**e**), predictability (**f**), and percentage of Component 1 (**g**) are compared for the top three modules. *P* values are for independent samples *t* test. **h-i,** Score plots of six pairs of reciprocal F_1 hybrids based on 68 heterosis-associated molecules in PCA (**h**) and PLS-DA (**i**). **j,** Scores of variable importance in projection of Component 1. **k-l,** Heatmap (**k**) and PCA score plot (**l**) based on the 68 heterosis-associated molecules. **m,** Distribution of four types of inheritance patterns. NA represents inheritance patterns outside the four types. **n,** Percentages of the four types of inheritance patterns. **o,** Performance of the additive effect. **p,** Determinations of two subtypes of partially dominant effects. **q,** Predictabilities (Q^2) of hybrid metabolite profiles (F_1), three types of inheritance patterns, and their combinations. **r,** Pearson correlations between better-parent heterosis (BPH) for seedling length and the percentages of inheritance patterns.

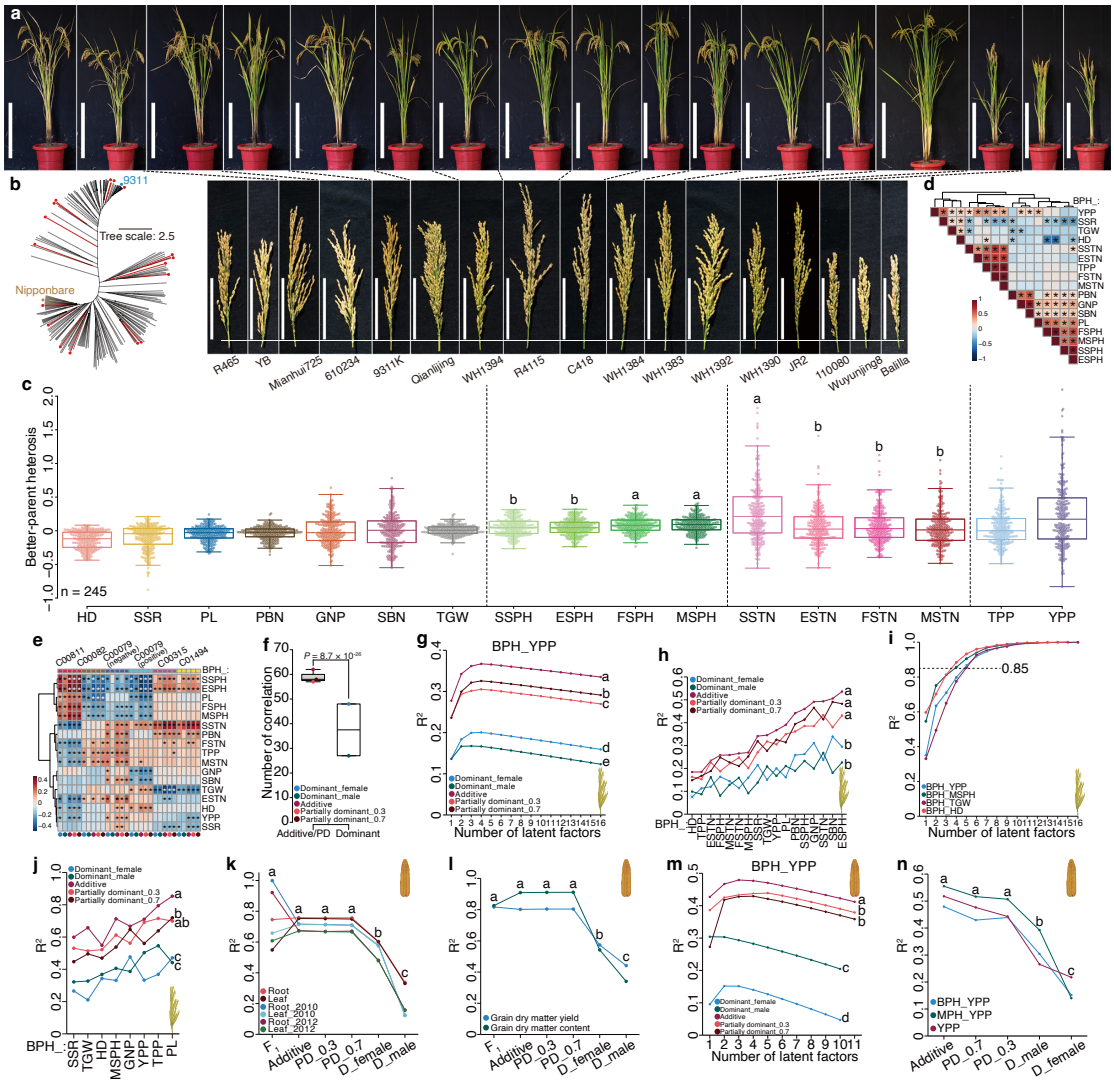

Fig. 3 Additive and partially dominant effects contribute to heterosis of agronomic traits in both rice and maize. a, Plants and panicles of 17 rice inbred lines (ILs). Scale bars represent 50 cm and 10 cm for plants and panicles, respectively. **b,** Neighbour-joining tree of 223 rice ILs. **c,** Better-parent heterosis (BPH) for 17 traits. **d,** Correlations of BPH for 17 traits. The asterisks indicate significant correlations ($P < 0.05$). **e,** Correlations between three inheritance patterns of five chemically identified metabolites and BPH. **f,** Correlation numbers of additive, partially dominant (PD), and dominant effects on 17 traits. *P* values indicate independent samples *t* tests. **g,** R^2 values of PLS-based predictive models for heterosis of grain yield per plant (YPP) with different inheritance patterns of 1,306 differential metabolites. **h,** Top R^2 for heterosis of 17 traits. **i,** R^2 values for heterosis of four traits. **j,** Top R^2 values for heterosis of eight traits. **k,** Top R^2 values for field fresh weight with hybrid profiles (F_1) and different inheritance patterns of metabolic data in maize. Metabolite profiles of root and leaf samples were used for two hybrid populations in 2010 and 2012. **l,** Top R^2 values for grain dry matter yield and grain dry matter content with root metabolic data in maize. **m,** R^2 for BPH_YPP with different inheritance patterns of 84 differentially expressed genes of 12 maize ILs. **n,** Top R^2 for BPH_YPP, MPH_YPP (middle parent heterosis), and YPP in maize. The R^2 are with adjusted values.

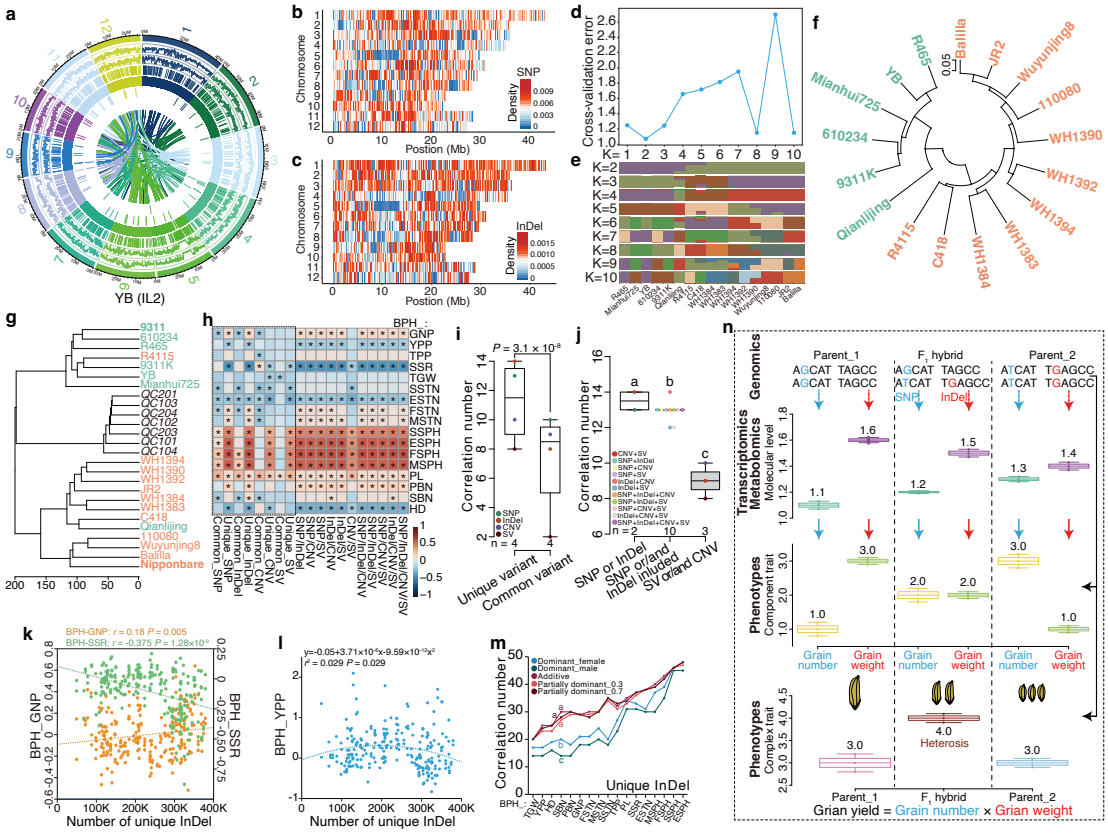

Fig. 4 Parental genomic variants are associated with additive and partially dominant effects. a, Genomic variants of Yuetai B. Chromosome position, SNP density, InDel density, CNV duplication position, CNV deletion position, SV insertion position, SV deletion position, and SV invertion position are listed from the outer to inner circles. **b-c,** Heatmaps of SNP (**b**) and InDel (**c**) density of Yuetai B. **d,** Cross-validation errors with K ranging from 1 to 10 based on SNPs. **e,** Population structure of 17 rice ILs based on SNPs. **f,** Neighbour-joining tree of 17 ILs based on SNPs. **g,** Dendrogram of 19 rice ILs based on normalized metabolite levels from 42,052 metabolites. Quality control = QC. **h,** Heatmap of correlations between parental genomic variants and heterosis of 17 traits. **i,** Correlation numbers of parental unique and common genomic variants. *P* value indicates independent samples *t* test. **j,** Correlation numbers of different types of parental unique genomic variants. **k,** Scatter plot between the numbers of parental unique InDels and heterosis for grain number per plant and seed setting rate. *P* value indicates Spearman rank correlation. **l,** Quadratic regression between the number of parental unique InDels and heterosis for grain yield per plant. **m,** Correlation numbers of parental unique InDels. **n,** Model for the contribution of additive effect to positive heterosis of a complex trait. The unique SNP and InDel in parents regulate grain number and grain weight, respectively. Meanwhile, these genomic variants are associated with transcriptomic and metabolomic additive effects in F_1 hybrids. The grain number and grain weight of F_1 hybrids are at the middle parent values because they are correlated with the levels of corresponding heterosis-associated molecules. Owing to the interaction between grain number and grain weight which determines grain yield, positive yield heterosis occurs when its two component traits have suitable values.